Optics and Renaissance Art

Charles M. Falco

C.M. Falco (✉)
College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA
e-mail: falco@email.arizona.edu

11.1 Introduction

An extensive visual investigation by the artist David Hockney [1] lead to the discovery of a variety of optical evidence in paintings as described in a number of technical papers [2–8]. This work demonstrated European artists began using optical devices as aids for creating their work early in the Renaissance well before the time of Galileo. These discoveries show that the incorporation of optical projections for producing certain features coincided with the dramatic increase in the realism of depictions at that time. Further, it showed that optics remained an important tool for artistic purposes continuing until today.

Our earliest evidence of the use of optical projections is in paintings of Jan van Eyck and Robert Campin in Flanders c1425, followed by artist including Bartholome Bermejo in Spain c1474, Hans Holbein in England c1530, and Caravaggio in Italy c1600, to name a few. Significantly, the optical principles of the camera obscura were described the eleventh century Arab scientist, philosopher, and mathematician, Abu Ali al-Hasan ibn al-Haytham, known in the West as Alhazen or Alhacen (b.965 Basra d.1039, Cairo). This is important for the present discussion because by the early thirteenth century al-Haytham's writings on optics had been translated into Latin and incorporated in the manuscripts on optics of Roger Bacon (c1265), Erasmus Witelo (c1275), and John Peckham (c1280).

Concurrent with the growing theoretical understanding optics were practical developments, such as the invention of spectacles in Italy around 1276. Pilgrims carried small convex mirrors into cathedrals to use as wide-angle optics to enable a much larger area of the scene to be visualized, showing how common the uses of optics had become by this time. As described below, evidence within paintings shows that at some point during this period someone realized replacing the small opening in a *camera obscura* with a lens resulted in a projected image that was both brighter and sharper. One lens from a pair of reading spectacles allows projection of images of the size, brightness, and sharpness necessary to be useful to artists, although with the optical "artifact" of having a finite depth of field (DOF). It is important to note that concave mirrors also project images, but with the advantage for an artist that they maintain the parity of a scene. For this reason it seems likely that, at least in the initial period, artists used them rather than refractive lenses.

The earliest visual depiction of lenses and concave mirrors of which I am aware are in Tomaso da Modena's 1352 paintings of "Hugh of Provence" and "Cardinal Nicholas of Rouen."[1] Either the spectacles or the magnifying glass in these paintings would have projected an image useful for an artist. His "St. Jerome" and "Isnardo of Vicenza" both show concave mirrors as well. This shows that the necessary optics to project images of the size and quality needed by artists were available 75 years before the time of Jan van Eyck.

The examples in what follows are selected from several well-known European artists. As will be shown, in each case features are shown in portions of their works that are based on optical projections.

---

1   These paintings are located in the Chapter House of the Seminario building of the Basilica San Nicolo in Treviso, Italy.

11.2 Analysis of Paintings

11.2.1 Jan van Eyck, *The Arnolfini Marriage*, 1434

One of the earliest examples we have found of a painting that exhibits a variety of evidence that the artist-based portions of it on optical projections is shown in ◘ Fig. 11.1. Several different types of optical analysis demonstrate the chandelier, enlarged in ◘ Fig. 11.2, is based on an optical projection.

The advantage of an optical projection of a real chandelier for an artist even of the skill of van Eyck is it would have allowed him to mark key points of the image. Even without tracing most of the image this would have enabled him to obtain the level of accuracy seen for this complex object that never had been previously achieved in any painting. The use of a lens results in an optical base for certain of the features, even though a skilled artist would not have needed to trace every detail in order to produce a work of art even as convincing as this one.

Since an optical projection only would be useful for certain features of any painting, and not for others, it is important to analyze appropriate aspects of the chandelier to determine whether or not they are based on optical projections. After establishing an optical base it would have been easier for van Eyck to "eyeball" many of the features [1]. As a result, paintings like the *Arnolfini Marriage* are collages consisting of both optical and non-optical elements, with even the optical elements containing eyeballed features as well [1]. Another important point is that all paintings of three-dimensional objects reduce those objects to two dimensions and, in doing so, lose some of the spatial information.

Elsewhere, based on the size of the candle flame, we estimated the magnification of the chandelier is 0.16 [6]. This means the outer diameter of the original chandelier was approximately 1 m which is consistent with the sizes of surviving chandeliers of that period. This magnification is small enough that the DOF for a lens falling within any reasonable range of focal lengths and diameters would be over 1 m. Because of this, van Eyck would have seen the entire depth of the real chandelier in the projected image without needing to refocus. Hence, if based on an optical projection the positions of the tops of each of the six candle holders should exhibit something close to perfect hexagonal symmetry after correcting for perspective. However, even if he had carefully traced a projected image there should be deviations from ideal symmetry due to the imperfections of any such large, hand-made object. If, instead, he had painted this complex object without the aid of a projection, and without the knowledge of analytical perspective that was only developed many decades later [9], larger deviations in the positions of these candle holders would be expected.

Marked with dots in ◘ Fig. 11.3 are the positions of the tops of each of the candle holders. The six-sided shape connecting them is an ideal hexagon that has been corrected for perspective. As can be seen, the agreement of the positions of the candle holders with the points of a perfect hexagon is remarkable. The maximum deviation of any of the candle holders from a perfect hexagon is only 7°, corresponding to the end of that half-meter-long arm being bent only 6.6 cm away from its "ideal" hexagonal position. Importantly, this analysis shows the arms are bent away from their "ideal" positions, but that none of them is either longer or shorter than the others. This is just what would be expected for a real chandelier. The deviations from perfect hexagonal symmetry are all on a circle, with the root-mean-square deviation only 4.1 cm. Although we shouldn't expect a hand-made fifteenth century chandelier to exhibit accuracy greater than this, some or all of the deviations could have resulted from slight bends during fabrication, transportation, hanging, or subsequent handling.

11

**Fig. 11.1** Jan van Eyck, *The Arnolfini Marriage*, 1434

Although the overall chandelier is three dimensional, the individual arms are two dimensional. We devised an analysis scheme based on this, as shown in ■ Fig. 11.4 [3, 6]. In this figure we individually corrected each of the six arms of the chandelier for perspective and overlaid them to reveal similarities and differences. Where a complete arm is not shown in the figure it is because it is partially obscured by arms in front of it. While the loss of spatial information when projecting a three-dimensional object into two dimensions introduces ambiguities, the scheme we used to analyze this chandelier avoids this limitation.

After transformation of the arms to a plan view of each the main arcs are identical to within 5 % in width and 1.5 % in length. That they are the same length is consistent with our independent analysis of the radial positions of the candle holders described above [6]. However, since it would have been easier for van Eyck to eyeball many aspects of this chandelier, rather than to trace the entire projected image, it is not surprising that there are variations in the positions of the decorative features attached to those arcs.

From this evidence and other that we published [1–3, 6] we can conclude with a high degree of confidence that van Eyck's chandelier is based on an optical projection of a real chandelier. Further, the small differences provide insight into the artistic choices van Eyck made to deviate from simply tracing the projection. However, the most important point is that the unprecedented realistic perspective of this complex object is a result of an optical projection that was made over a century earlier than previously thought possible [9].
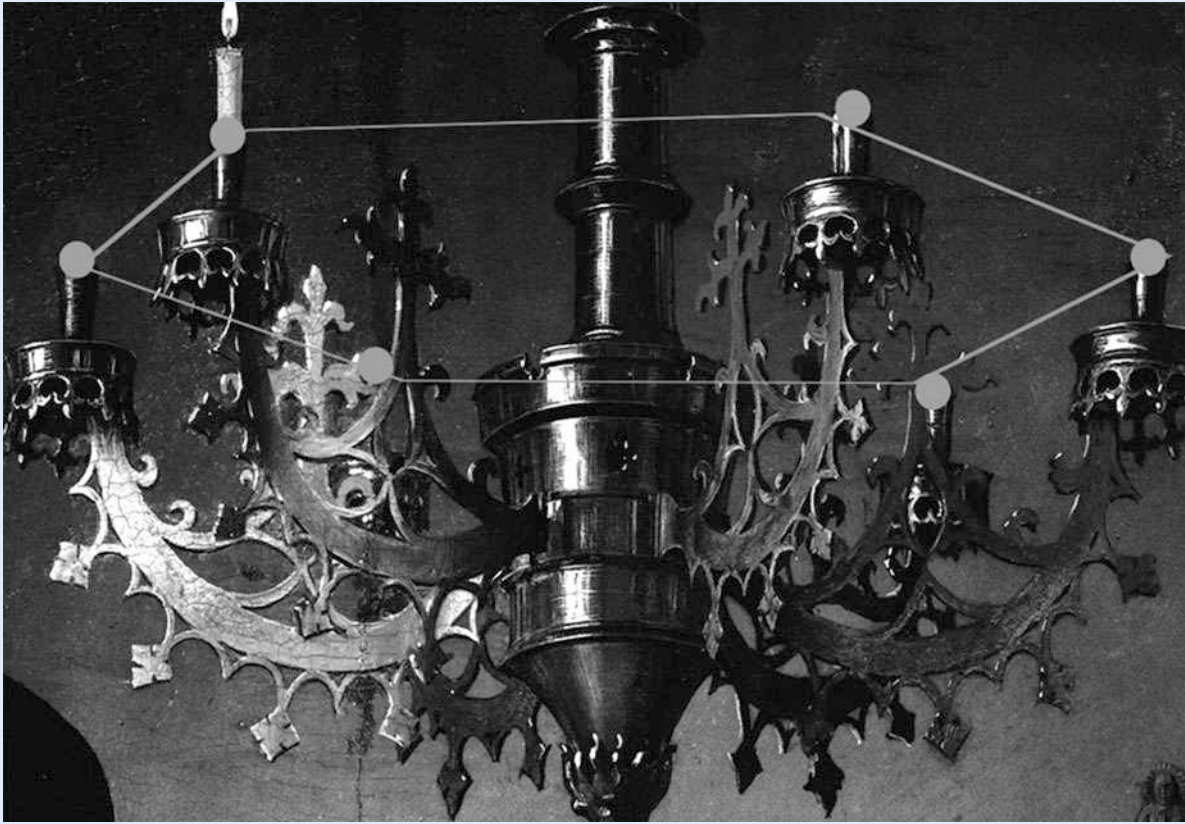
Fig. 11.3 *The Arnolfini Marriage* (detail). As can be seen, a perspective-corrected hexagon fits the positions of the tops of the candle holders to a remarkable accuracy, with small deviations from ideal symmetry consistent with a large, hand-made fifteenth century object

11.2.2 **Lorenzo Lotto, *Husband and Wife*, 1523–1524**

"Family Portrait" by Lorenzo Lotto (1523–1524) shown in Fig. 11.5 provides considerable quantitative information about the lens that optical evidence indicates Lotto used in creating this painting. Figure 11.6 is a detail from *Husband and Wife* showing an octagonal pattern on an oriental carpet that appears to go out of focus at some depth into the painting. Overlaid on this painting are three segments of a perspective-corrected octagon whose overall fit to the pattern is seen to be excellent, and whose quantitative details we calculate below.

As we have shown elsewhere [3, 6], based on the scale of the woman in the painting the magnification is approximately $M = 0.56$. Any optical projection at such a high magnification intrinsically has a relatively shallow DOF, the value of which depends on the focal length and diameter of the lens as well as the magnification. To change the distance of sharp focus requires physically moving the lens with respect to the subject and the image plane. To refocus an image on a region further into a scene from its original plane of focus requires moving the lens further away from the scene. This movement of the lens to refocus results in a small decrease in the magnification of the projected scene, as well as in a slight change in the vanishing points. Although such effects are fundamental characteristics of images projected by lenses, they are extremely unlikely to occur in a painting if an artist had instead laid out patterns using sighting devices or

Fig. 11.4 This figure contains the outlines of all six arms on the chandelier after correcting for perspective with the arms to the viewer's right flopped horizontally to overlay on the arms to the left. The main arc of all six arms is the same to within 1.5 % in length and 5 % in width. Variations are consistent with the decorative features on the main arc having been hand attached to the original chandelier as well as having been eyeballed when creating the painting

following geometrical rules first articulated in the fifteenth century [9]. Since we already have discussed several aspects of this painting elsewhere, here we summarize our previous analysis [3, 6–8].

The distance across the wife's shoulders in the painting, compared with measurements of real women, provides an internal length scale that lets us determine the magnification to be $M \approx 0.56$. This in turn allows us to determine the repeat distance of the triangular pattern on the actual carpet to be 3.63 cm. Since the first place where the image of the carpet changes character is approximately 4–5 triangular-repeats into the scene, we calculate the depth of field to be DOF $= 16 \pm 1.5$ cm. We now can use geometrical optics to extract quantitative information from this painting.

The focal length (FL) and magnification (M) are given by the following equations from geometrical optics: [3]

$$1/FL = 1/\left(d_{\text{lens Bsubject}}\right) \; + \; 1/\left(d_{\text{lens Bimage}}\right) \tag{11.1}$$

and

$$M \; = \; \left(d_{\text{lens Bimage}}\right)/\left(d_{\text{lens Bsubject}}\right) \tag{11.2}$$

Fig. 11.5 Lorenzo Lotto, *Husband and Wife*, c1523–c1524

As indicated by the overlays in ⬛ Fig. 11.6, there are three regions of this octagonal pattern. These regions are the result of Lotto having refocused twice as he exceeded the DOF of his lens. We label these Regions 1, 2, and 3, with Region 1 the closest to the front of the painting. Thus, for the first two Regions,

$$1/\text{FL} \; = \; 1/\left(d_{\text{lens B subject1}}\right) \; + \; 1/\left(d_{\text{lens B image1}}\right) \tag{11.3}$$

and

$$1/\text{FL} \; = \; 1/\left(d_{\text{lens B subject2}}\right) \; + \; 1/\left(d_{\text{lens B image2}}\right) \tag{11.4}$$

However, the measured DOF is 16 ± 1.5 cm, so for Region 2

$$d_{\text{lens B subject2}} \approx d_{\text{lens B subject1}} \; + \; 16 \text{ cm} \tag{11.5}$$

and thus

$$1/\text{FL} \; = \; 1/\left(d_{\text{lens B subject1}} + 16 \text{ cm}\right) \; + \; 1/\left(d_{\text{lens B image2}}\right) \tag{11.6}$$

Because Region 2 is further into the scene it is at a slightly lower magnification than is Region 1 so its DOF will be somewhat larger than 16 cm. We can calculate DOF_2 from
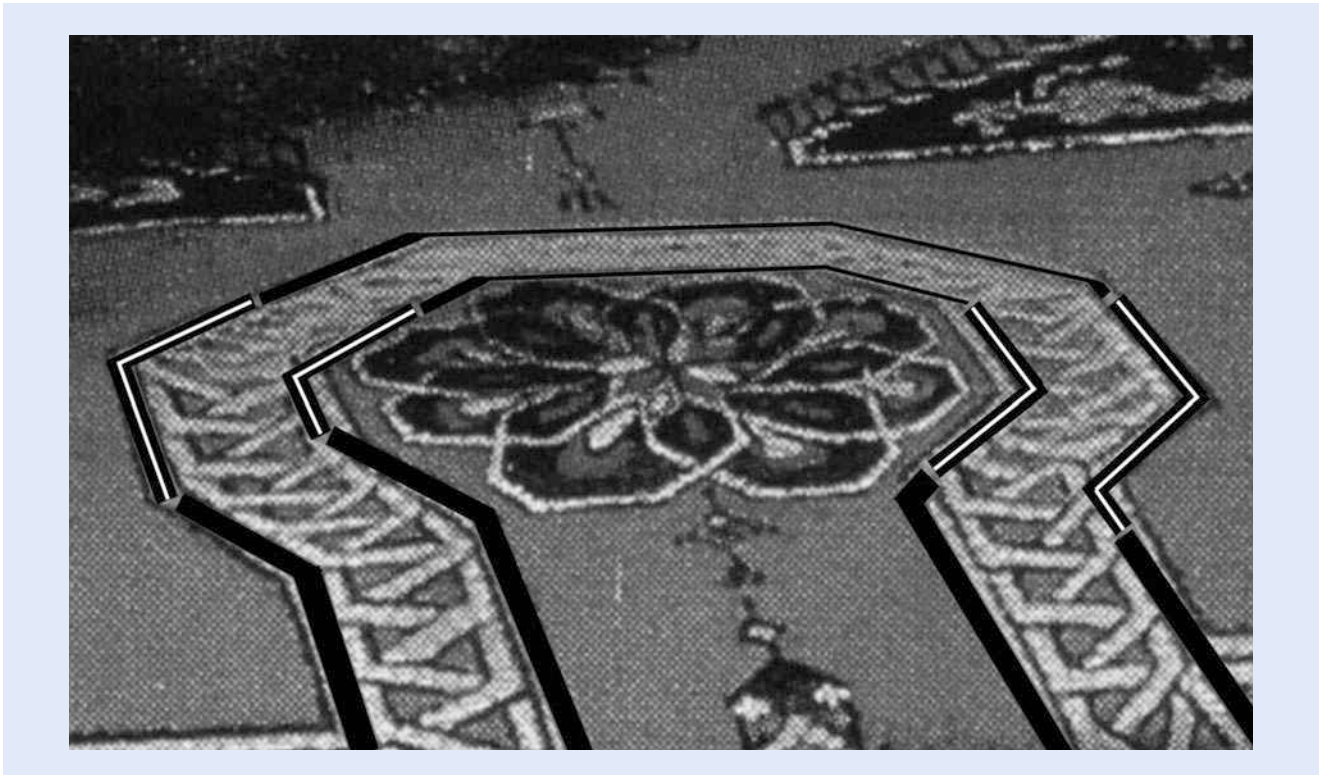
■ **Fig. 11.6** *Husband and Wife* (detail). The overlays are perspective-corrected sections of an octagonal pattern that we fit to the painting. As described in the text, the details of this portion of the painting are in excellent qualitative and quantitative agreement with the three-segment, perspective-corrected octagon that is predicted by the laws of geometrical optics for such a projected image

$$\text{DOF}_2 = 2\,C \times f\# \times (1 + M_2)/M_2{}^2 \qquad \textbf{(11.7)}$$

where C is the circle of confusion, $f\#$ is the lens diameter/focal length, and M_2 is the magnification of Region 2. Hence,

$$\text{DOF}_2 = \text{DOF}_1 \times (1 + M_2)/(1 + M_1) \times (M_1/M_2)^2 \qquad \textbf{(11.8)}$$

Region 3 of the pattern thus starts at a depth of 16 cm + DOF_2 into the scene, so

$$d_{\text{lensBsubject3}} = d_{\text{lensBsubject1}} + 16\text{ cm} + \text{DOF}_2 \qquad \textbf{(11.9)}$$

and

$$1/\text{FL} = 1/\big(d_{\text{lensBsubject1}} + 16\text{ cm} + \text{DOF}_2\big) + 1/\big(d_{\text{lensBimage3}}\big) \qquad \textbf{(11.10)}$$

The magnifications M of the three regions are given by:

$$0.56 = d_{\text{lensBimage1}}/d_{\text{lensBsubject1}} \qquad \textbf{(11.11)}$$

$$M_2 = d_{\text{lensBimage2}}/\big(d_{\text{lensBsubject1}} + 16\text{ cm}\big) \qquad \textbf{(11.12)}$$

$$M_3 = d_{\text{lensBimage3}}/\big(d_{\text{lensBsubject1}} + 16\text{ cm} + \text{DOF}_2\big) \qquad \textbf{(11.13)}$$

This analysis gives us seven Eqs. (11.3), (11.6), (11.8), (11.10), (11.11), (11.12), and (11.13) and eight unknowns: FL, $d_{\text{lensBsubject1}}$, $d_{\text{lensBimage1,2,3}}$, DOF_2, $M_{1,2}$. If we make a single assumption about any one of these unknowns we can then solve these equations uniquely for the other seven unknowns using simple algebra. Assuming that the distance from the lens to the carpet was at least 1.5 m, but not greater than 2.0 m (i.e., $d_{\text{lensBsubject1}} = 175 \pm 25$ cm) we find

$$\text{focal length} = 62.8 \pm 9.0 \text{ cm}$$
$$M_2 = 0.489 \pm 0.9$$
$$M_3 = 0.423 \pm 1.5$$

The magnification when moving from Region 1 to Region 2, as measured from our fit of a perspective-corrected octagon, decreases by 13.1 % from the original 0.56 of the painting, in excellent agreement with the -12.6 ± 1.5 % calculated from the above equations. Similarly, the measured magnification decreases by a further 13.3 % when going to Region 3, again in excellent agreement with the calculated value of -13.5 ± 1.6 %.

From Eq. (11.7),

$$f\# = \left[\text{DOF}_1 \times M_1{}^2\right] / \left[2\, C\, (1 + M_1)\right]$$

If we assume the simple lens available to Lotto resulted in a circle of confusion on the painting of 2 mm, we find $f\# \approx 22$, and hence a diameter of 2.9 ± 0.4 cm. As we have confirmed with our own experiments, a lens or concave mirror with these properties projects a quite useful image of a subject that is illuminated by daylight.

To summarize, using only the measured magnification of this painting (0.56, i.e., roughly half life size, as determined from the size of the wife), and making a reasonable assumption about the distance Lotto would have positioned his lens from the carpet (175 ± 25 cm), equations from geometrical optics uniquely determine both changes in magnification, -13.1 and -13.3 %, of the central octagonal pattern, as well as the focal length and diameter of the lens, 62.8 ± 9.0 cm and ~3 cm, respectively, used to project this image. The three sets of vanishing points exhibited by the octagonal pattern, as well as the depths into the painting where they occur, are a direct consequence of the use of a lens to project this portion of the painting. Other quantitative information extracted from this painting is discussed elsewhere [3, 6–8].

Recently we developed a portable high-resolution digital camera that allows us to acquire important information about paintings without needing to remove them from museums for detailed study [7]. Since infrared light penetrates many pigments further than does visible light it often can be used to reveal "underdrawings" or other features not apparent in the visible [10, 11]. ◘ Figure 11.7 is an infrared (IR) "reflectogram" of the Lotto painting captured in situ where it was located on the wall of the Hermitage State Museum in St. Petersburg. Although many features are revealed in this image, one immediate observation is we can see that Lotto used a different pigment for the woman's dark dress than he used for the man's jacket. This provides us with previously unknown information about the artist's working technique.

◘ Figure 11.8 shows the octagonal pattern of the table covering in greater detail. As can be seen by comparison with ◘ Fig. 11.5, the red and yellow pigments Lotto used are largely transparent in the IR so this image provides an uncluttered view of the black lines he used to create this feature on the painting.

Three distinct types of markings can be clearly seen for the lines making up the triangular pattern of the border of this feature. Well-defined lines are in the region nearest the front of the image, consistent with tracing a projected image. These "traced" lines abruptly change to tentative lines in the middle region, at just the

◘ **Fig. 11.7** Lorenzo Lotto, Husband and Wife, c1523–c1524. Infrared (IR) reflectogram

depth into the scene where our previous analysis showed the magnification was reduced by 12.6 ± 1.5 % due to having to refocus because of exceeding the depth-of-field. Because of this, Lotto faced significant difficulty to create a plausible match for this geometrical pattern after refocusing. His abrupt change to tentative lines reflects this difficulty. After re-establishing a plausible freehand sketch form of the geometrical pattern by the rear of this central region, the quality of the lines again abruptly changes to only short dashes in the region farthest into the scene, where our previous analysis shows the magnification was reduced by an additional 13.5 ± 1.6 % due to having to refocus a second time after again reaching the limit of the depth-of-field. These results from IR reflectography provide important insights into the actual working practices of an artist, revealing quite specific details about how he made use of projected images 75 years prior to the time of Galileo.

Our analysis of this painting found a change in the vanishing point that takes place part way back in the pattern in the border of the carpet to the right, quantitatively consistent with the change that is caused by the shift in position of a lens as it is refocused. ◘ Figure 11.9 shows the IR reflectogram of this portion of the painting. Overlaid to the left are seven units of a perfectly repeating structure that replicates the geometrical pattern of the border. As can be seen, after
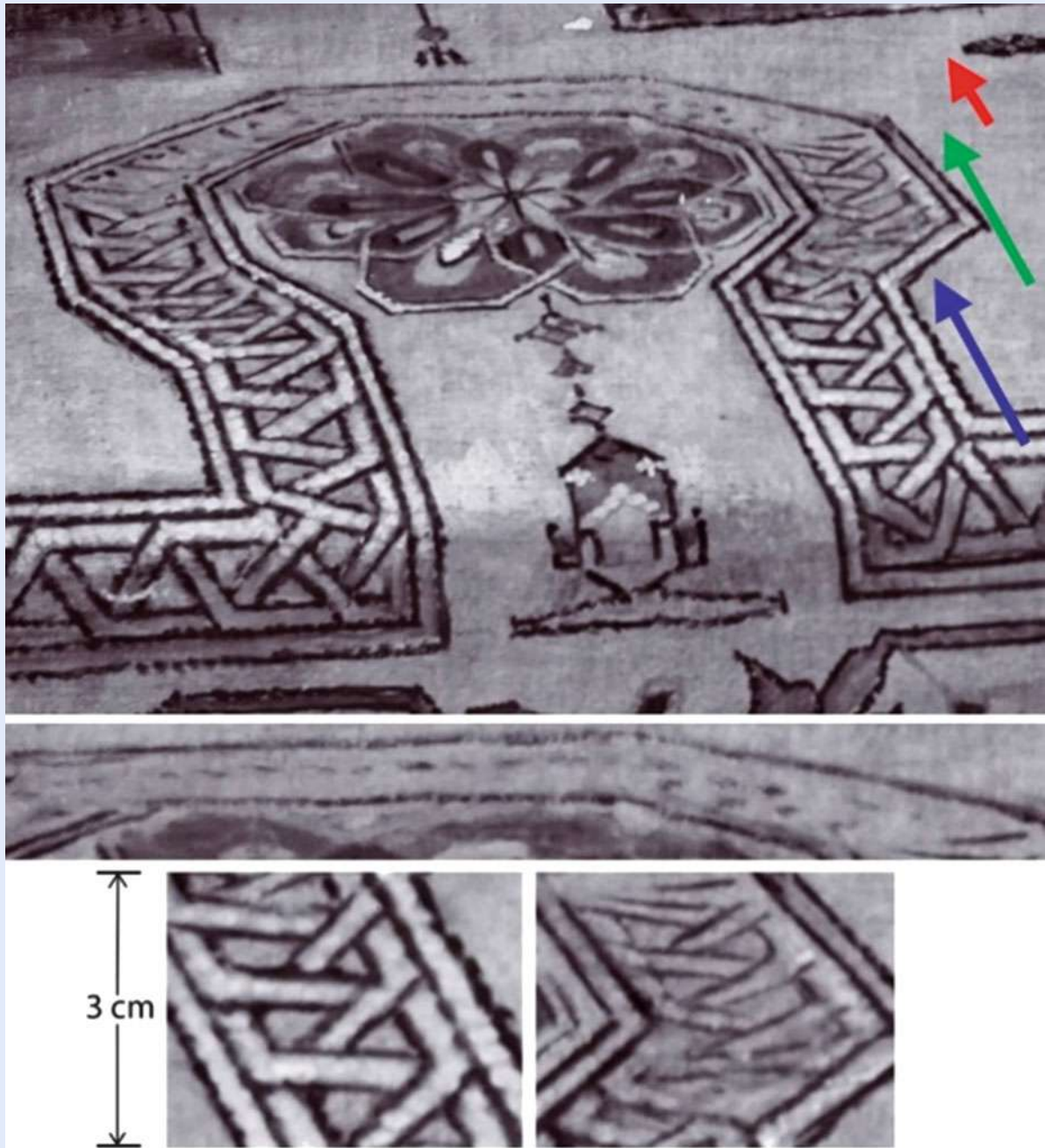
11

correcting for perspective, this structure is an excellent fit to the repeating pattern near the front of the carpet. The maximum deviation from a "perfect" fit is consistent with the degree of perfection found in the hand-made carpets of this type. Although an eighth unit of the structure does not fit at all, a small change in optical perspective makes the same repeating structure fit at the rear, again to within better than 2 mm. This change in perspective occurs at the same depth into the painting where our previous analysis found a shift in vanishing point, as
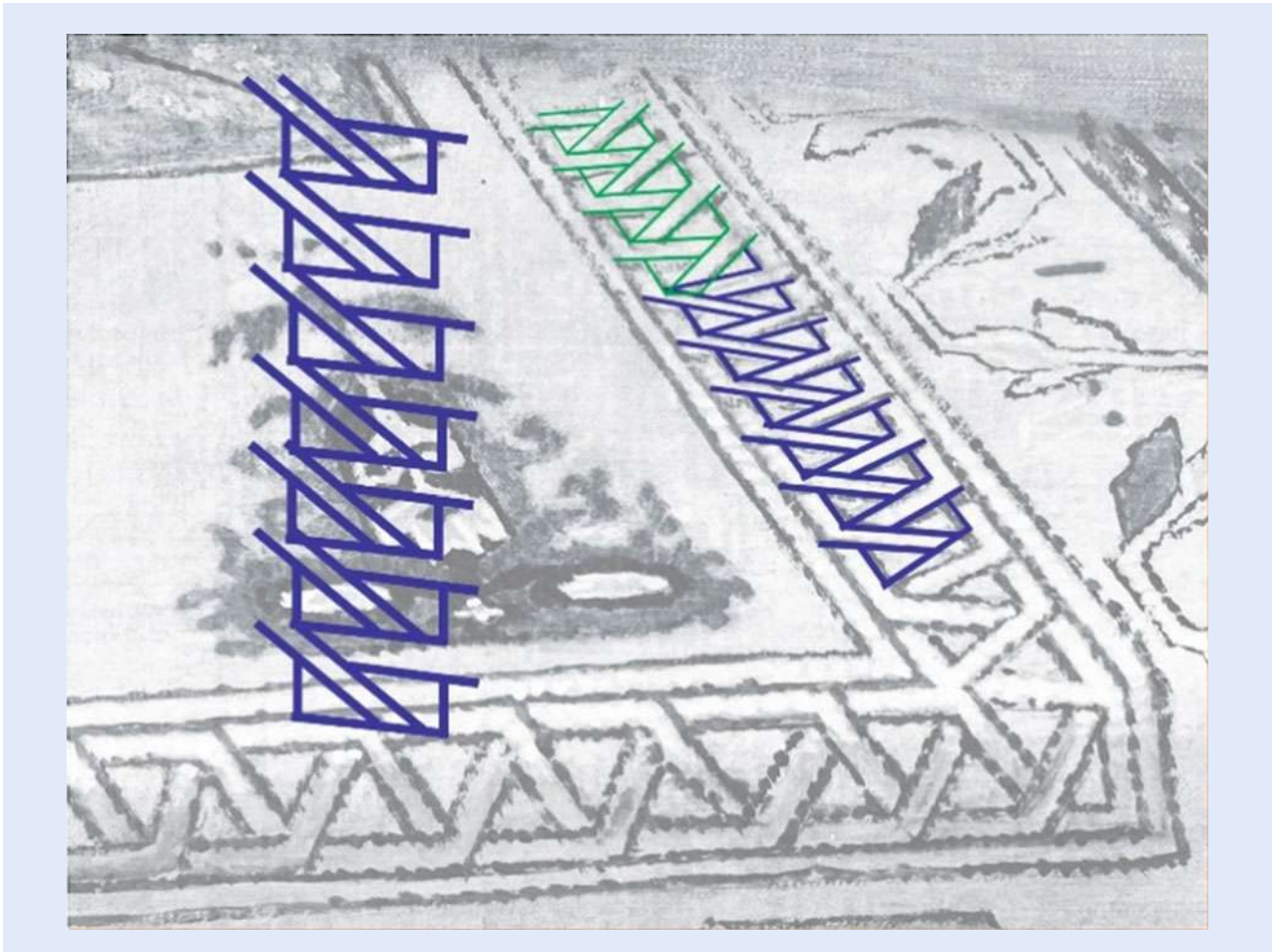
🔵 **Fig. 11.9** IR reflectogram of the border pattern of 🔵 Fig. 11.5. Overlay at left is seven segments of a repeating structure. When corrected for perspective, this is seen to be an excellent fit to the pattern at the front of the table covering. Changing the perspective, as happens when a lens is moved to refocus, gives an excellent fit to the pattern at the back. The maximum deviation of the perfect repeating structure from the pattern on the painting is 2 mm

happens when a lens is repositioned to focus further into a scene. Further, not only does the perspective change where a lens would have had to have been moved to refocus, the painting is missing a half-segment of the repeating pattern at this location. This is consistent with Lotto attempting to create a plausible match between two segments of a repeating structure after refocusing had caused the magnification and perspective to change. All of these detailed findings from IR reflectography are consistent with our other work showing this portion of the painting is based on the optical projection of an actual hand-made carpet [2, 3, 6, 7].

🔵 Figure 11.5 is the full image of this painting in the visible captured in situ using a standard digital camera with a 35 mm f/2 lens. This image reveals some of the difficulties with in situ image capture in a museum environment. The painting was illuminated by a combination of indirect sunlight from windows to the left, and overhead tungsten lights, each having its own color temperature. The shadows visible along the left and top borders were cast by the ornate frame in which the painting is mounted. The roughly equal darkness of these shadows indicates that the level of illumination from both types of sources was approximately equal. However, closer inspection shows that the illumination across the surface of the painting is not uniform. This can be most easily seen in the region of the man's

chest, which is too bright due to a partial specular reflection of one of the light sources that could not be eliminated by repositioning the camera within the constraints of the room.

◾ Figure 11.7 is an IR reflectogram of the full 96 × 116 cm painting, captured under the less than ideal lighting conditions described in the previous paragraph. Although many features are revealed by this IR reflectogram, one immediate observation is that Lotto used a different pigment for the woman's dress than he used for the man's jacket, providing us with previously unknown information about the artist's working technique.

Again, all of these detailed findings from IR reflectography are consistent with our earlier work that showed this portion of the painting is based on the optical projection of an actual hand-made carpet. I note that we have used fourteenth century optical technology (i.e., one lens of a pair of reading spectacles, as well as a metal concave mirror we fabricated following descriptions in texts of the time) to accurately reproduce all of the effects we have found in this carpet, as well as in all of the other paintings we have shown to contain elements based on optical projections, including projecting such patterns directly on a screen of the same shade of red used in this painting. Even on such a colored screen, the projected images are quite distinct and easy to trace.

11.2.3 Hans Holbein the Younger, *The French Ambassadors to the English Court*, 1532

A prominent feature of *The French Ambassadors to the English Court* by Hans Holbein is the anamorphic skull at the bottom of the 1532 painting. This feature is shown in ◾ Fig. 11.10. The way this appears to someone viewing it at a grazing angle is shown by linearly compressing it by 6× in ◾ Fig. 11.11 (Right), with a real skull for comparison in ◾ Fig. 11.11 (Left).[2] Very obvious differences include that the jaw of Holbein's skull is much longer than the real skull, the slope of the top of the skull is steeper, and the eye sockets and nose are much more pronounced as well as aimed more in the direction of the viewer.

To see if optical projections may account for the appearance of this skull in the painting, we used a concave mirror of focal length 41 cm to project the image of a real skull onto a screen at a grazing angle in order to produce an anamorphic image. ◾ Figure 11.11 (Left) is a photograph of the real skull taken from precisely the location of that concave mirror after the mirror had been removed from its holder. However, because of the limited depth of focus of the projected image on

---

2   The anamorphic skull is 106 cm long and 14.4 cm high. To visually compress its length to be the same as its height so that it appears approximately like ◾ Fig. 11.11 (Right) requires viewing the painting at a grazing angle of $\sin^{-1}(14.4/106) \approx 8°$. At this angle the far end of the anamorphic feature is over 100 cm further away from the viewer than is the near end, so that for reasonable viewing distances the magnification of the far end is significantly less than that of the near end. Also, since for any reasonable viewing distance the depth of the feature is greater than the depth of field of the eye, it requires the viewer to scan back and forth through the feature, with their eyes constantly refocusing when doing so, in order to "construct" a composite image in their mind that does indeed strongly resemble ◾ Fig. 11.11 (Right). Although our analysis shows that this anamorphic feature was constructed with the aid of optical projections, the multiple positions of the lens needed to generate it, coupled with the multiple movements and refocusing of the eye needed to view it, along with the mental compositing need to construct the final image of it in the brain, results in an underlying complexity to ◾ Fig. 11.11 (Right). For these reasons, because ◾ Fig. 11.11 (Right) was generated by a linear transformation, it only approximately reproduces what the feature looks like to the viewer when examining the painting from a grazing angle.
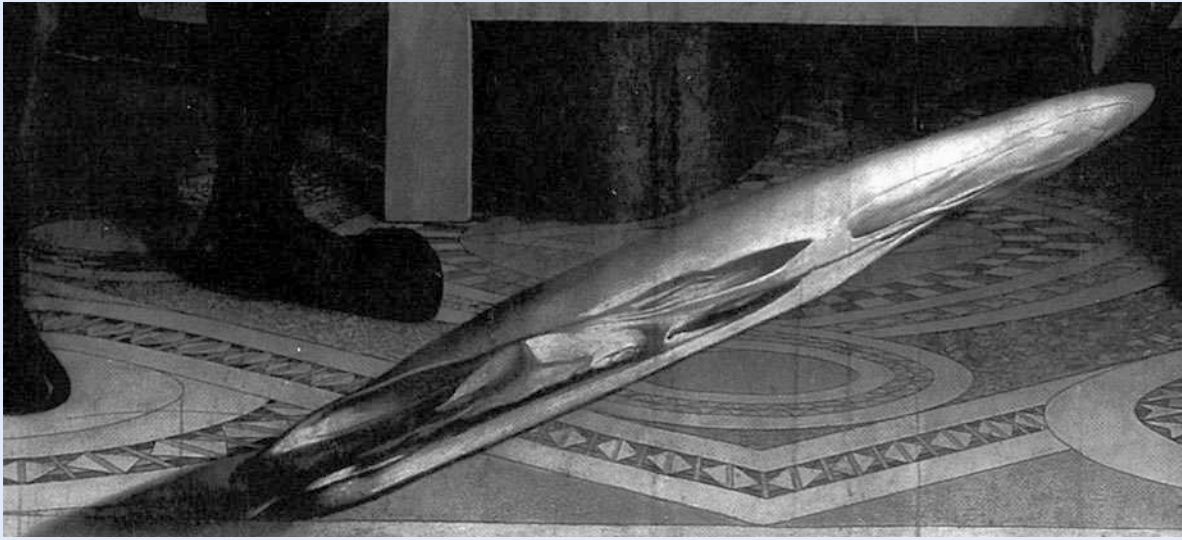
🔲 **Fig. 11.10** *The French Ambassadors to the English Court* (detail). This detail shows the unusual feature at the bottom of Holbein's painting. Viewed from a grazing angle to visually compress it, this feature appears as shown in 🔲 Fig. 11.11 (*Right*). Possibly not apparent in this small B&W reproduction is that this anamorphic skull does not occupy the same visual space as the rest of the painting
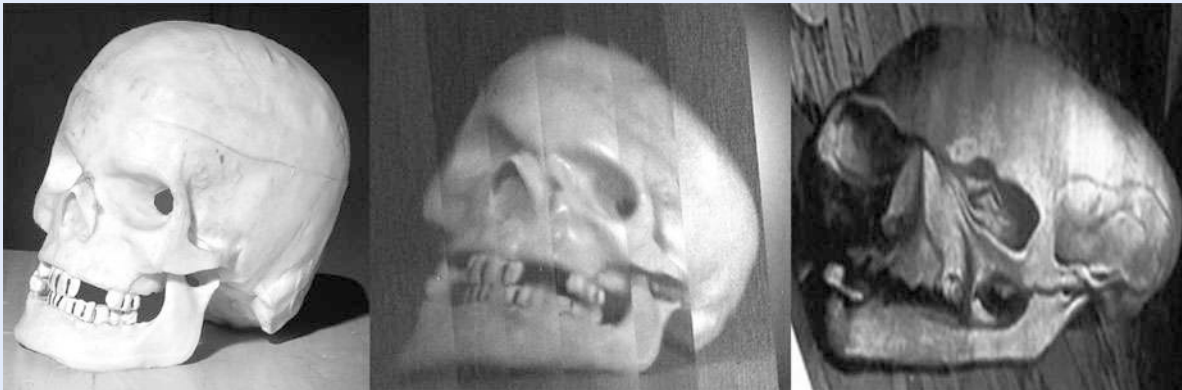
🔲 **Fig. 11.11** (*Left*) Photograph of a skull taken from the position of the concave mirror used to project its image onto a tilted screen to form an anamorphic image. (*Center*) Composite of the individual in-focus segments of the projected anamorphic image of the skull after linearly compressing it horizontally. (*Right*) Anamorphic skull in *The French Ambassadors* after linearly compressing it horizontally

the tilted screen, it was necessary to refocus the concave mirror a number of times in order to generate the composite anamorphic image that we have compressed linearly to produce 🔲 Fig. 11.7 (Center).

The segments of each of the in-focus images are visible in this composite. What is striking about 🔲 Fig. 11.11 (Center) is how well it reproduces the very unusual visual appearance of the linearly compressed skull from Holbein's painting. Although mathematical and graphical methods can be used to construct anamorphic images, the optics-produced composite of 🔲 Fig. 11.11 (Center) is far more complex than is obtained from any such construction. The magnification of each segment in the anamorphic photographic composite is linear in the vertical direction, but is proportional to 1/sin of the grazing angle in the horizontal. The overall composite of 🔲 Fig. 11.11 (Center) is thus the result of a nonlinear,

11

🔲 **Fig. 11.12** Anamorphic skull in *The French Ambassadors to the English Court*. For this figure we have rotated the feature in 🔲 Fig. 11.10 clockwise by 25° and then linearly compressed it by 6×. The height of the skull in this image compared to a real one gives a magnification $M = 0.71 \pm 0.5$. The lines indicate two regions where it can be seen Holbein duplicated features (notably, the two dark depressions just above the jaw, and the double-humped line midway up the skull). A discontinuity in the slope of the top of the skull is also visible at the left edge of the leftmost marked region

piecewise-segmented transformation. Although this complex transformation was naturally produced by the optical projection, it would be quite implausible to have resulted from any sort of a graphical or mathematical construction [9]. We conclude that the probability is extremely small that Holbein could have accidentally reproduced these complex features without having projected them with a lens.

🔲 Figure 11.12 shows 🔲 Fig. 11.11 (Right) at a larger scale. Marked on this figure are two regions where we observed that Holbein has duplicated features of the skull. Because the lens and canvas (or, less likely, the skull) has to be moved a number of times when piecing together an anamorphic image from segments projected at such a high magnification, it is very easy to accidentally duplicate a region, so its presence provides additional evidence that Holbein had to refocus a

lens. The duplicated segment corresponds to a region 3.0 ± 0.5 cm wide on a real skull. That same region corresponds to a width of 8.2 cm on the actual painting which gives us an approximate lower limit measure for the depth of focus. From the results of our experiments shown in ◘ Fig. 11.11 (Left) and (Center), that region of the skull is at an angle of $25° \pm 5°$ with respect to the perpendicular to the axis of the lens, so its depth into the scene is 1.3 ± 0.5 cm. Although a more accurate value for the depth of focus can be obtained by convoluting this measured DOF into the calculation, for our purposes here the approximate value 8.2 cm will suffice. Using this value, along with a circle of confusion of 2 mm and the measured $M = 0.71$, we calculate as a lower limit

$$f\# \geq \text{Depth of Focus} /[2C \times (M + 1)]$$
$$= 12.0$$

Because we have neglected the DOF in the calculation shown here, this value for the $f\#$ of Holbein's lens is somewhat smaller than the actual value, as well as represents a lower limit. However, this calculation is sufficient to show that the $f\#$ of Holbein's lens is consistent with the values we obtained for Lotto's and Campin's lenses (22 and 25.2, respectively).

11.2.4 Robert Campin, *The Annunciation Triptych (Merode Altarpiece)*, c1425–c1430

Robert Campin was a contemporary of Jan van Eyck and they are documented to have known each other. The center and right panels of Robert Campin's *Merode Triptych* of c1425$_B$28 contain the earliest evidence we have found to date of the use of direct optical projections. A detail of the right panel is shown at the lower left of ◘ Fig. 11.13. As we previously showed, this portion of the painting exhibits the same complex changes in perspective seen in Lorenzo Lotto's *Husband and Wife*, resulting from Campin also having refocused his lens twice [4, 6].

The upper right in ◘ Fig. 11.13 shows one of the two sets of slats (the set that is numbered on the lower inset), with each slat individually rotated to be vertical and expanded horizontally by a factor of $3.5\times$ to accentuate any deviations from being straight. Marked on the slats are the locations of "kinks" exhibited by each of them, with those kinks connected by lines. The positions of the lines connecting the kinks are shown on the inset at the lower left. Comparing with ◘ Fig. 11.2 of Reference 5 it can be seen that the slats are kinked at the same two depths into the painting where we previously showed, with a different type of analysis using different data, that Campin had to refocus due to the DOF of his lens. Geometrical constructions can be devised which exhibit kinks, but not in the overall configuration of this painting. The complex perspective exhibited by the latticework in this portion of the painting is a direct and inevitable outcome from the DOF of a lens, but would be extremely unlikely to have resulted from any geometrical construction, or from the use of a straightedge.

Using the height of the head in the full painting as a scale, the magnification of this portion of the painting is $M \approx 0.27$. If we assume a circle of confusion of 1 mm Eq. (11.7) yields $f\# = 25.2$. We can obtain an estimate for the focal length with the assumption the lens or concave mirror had a diameter of 3 cm, in which case the focal length $FL = f\# \times 3$ cm $= 76$ cm, which is quite reasonable.
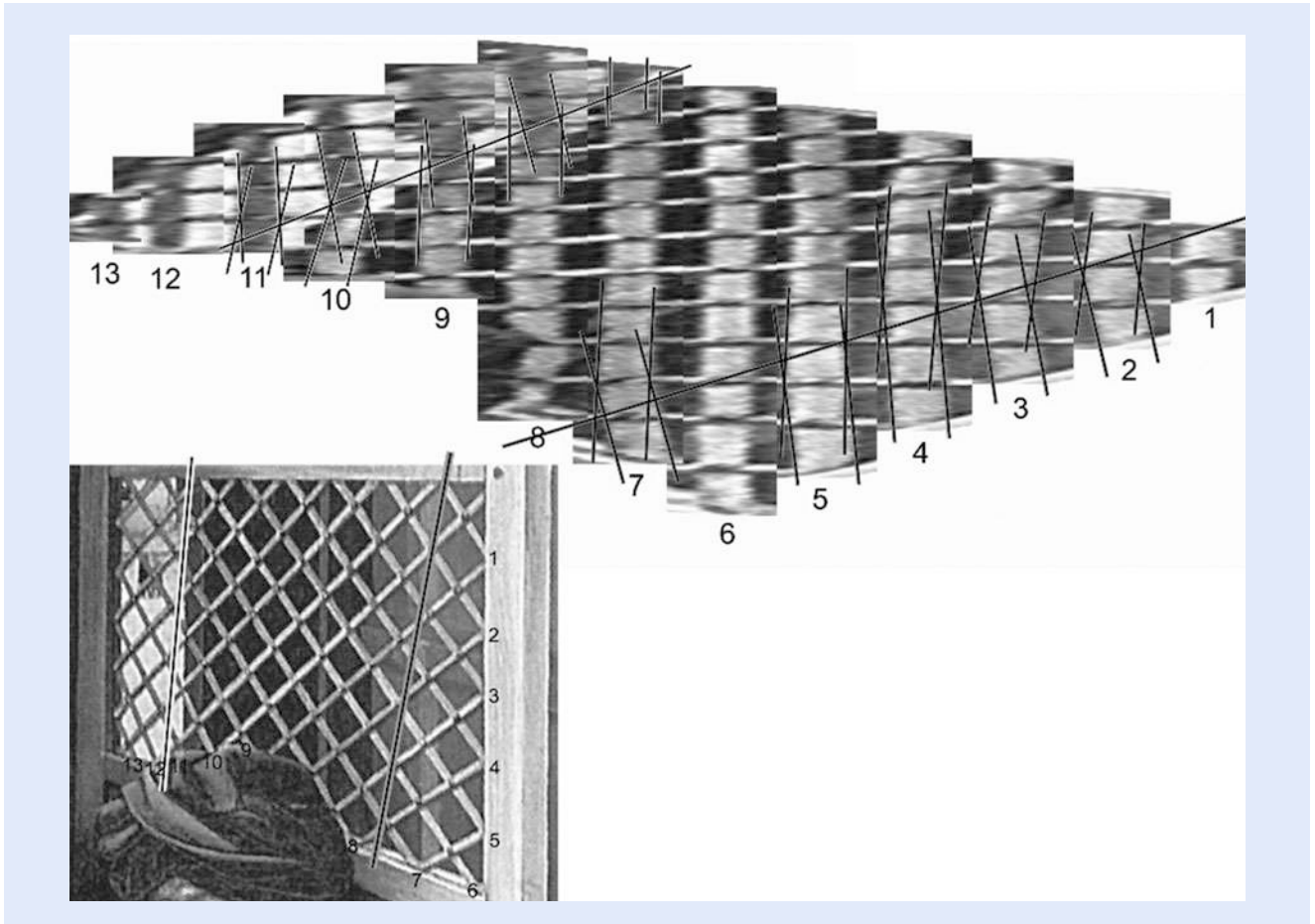
Fig. 11.13 (*Lower Left*) Detail of the *Merode Altarpiece* with one set of slats numbered. (*Upper Right*) Slats rotated to be vertical and expanded horizontally by 3.5×. We have connected the "kinks" that are apparent in the slats by lines, the locations of which are shown in the detail at the *lower left*

11.3   Conclusions

These discoveries demonstrate that highly influential European artists used optical projections as aids for producing some of their paintings early in the fifteenth century, at the dawn of the Renaissance, at least 150 years earlier than previously thought possible. In addition to van Eyck and Lotto we have also found optical evidence within works by well-known later artists including Bermejo (c1475), Holbein (c1500), Caravaggio (c1600), de la Tour (c1650), Chardin (c1750), and Ingres (c1825), demonstrating a continuum in the use of optics by artists, along with an evolution in the sophistication of that use. However, even for paintings where we have been able to extract unambiguous, quantitative evidence of the direct use of optical projections for producing certain of the features, it does not mean that these paintings are effectively photographs. Because the hand and mind of the artist are intimately involved in the creation process, understanding these images requires more than can be obtained from only applying the equations of geometrical optics. As to how information on optical projections came to these artists, evidence points to it having come via the Cairo-based scholar Ibn al Haytham [12].

11.4 Acknowledgments

References

1. Hockney D (2001) Secret knowledge: rediscovering the lost techniques of the old masters. Viking Studio, New York
2. Hockney D, Falco CM (2000) Optical insights into renaissance art. Opt Photonics News 11 (7):52–59
3. Hockney D, Falco CM (2003) Optics at the dawn of the renaissance. In: Technical digest of the optical society of America, 87th annual meeting, Optical Society of America, Washington DC
4. Hockney D, Falco CM (2004) The art of the science of renaissance painting. In: Proceedings of the symposium on 'effective presentation and interpretation in museums', National Gallery of Ireland, Dublin, Ireland, p 7
5. Hockney D, Falco CM (2005) Optical instruments and imaging: the use of optics by 15th century master painters. In: Proceedings of the SPIE photonics Asia, Bellingham, Washington, vol 5638, p 1
6. Hockney D, Falco CM (2005) Quantitative analysis of qualitative images. In: Proceeding of the IS&T-SPIE electronic imaging, SPIE, Bellingham, Washington, vol 5666, p 326
7. Falco CM (2009) High resolution digital camera for infrared reflectography. Rev Sci Instrum 80:071301–071309
8. Hockney D, Falco CM (2012) The science of optics: recent revelations about the history of art. In: Proceedings of the SPIE, Bellingham, Washington, vol 8480, p 84800A
9. Kemp M (1992) The science of art. Yale University Press, New Haven, CT
10. van Asperen de Boer JR (1968) Infrared reflectography: a method for the examination of paintings. Appl Opt 7(9):1711–1714
11. Faries M (2002) Techniques and applications, analytical capabilities of infrared reflectography: an art historian s perspective. In: Barbara B et al (eds) Scientific examination of art: modern techniques in conservation and analysis. National Academy of Sciences, Washington, DC
12. Falco CM, Weitz Allen A (2009) Ibn al-Haytham's contributions to optics, art, and visual literacy. In: Beckinsale M (ed) Painted optics symposium. Fondazione Giorgio Ronchi, Florence, Italy, p 115
13. Fendrich L (2002) Traces of artistry. Chron High Educ 53(36):B20

The Eye as an Optical Instrument

Pablo Artal

P. Artal (✉)
Laboratorio de Óptica, Universidad de Murcia, Campus de Espinardo (Edificio 34), 30100 Murcia, Spain
e-mail: pablo@um.es

12.1 Introduction

For us, humans, vision is probably the most precious of senses. Although our visual system is a remarkably sophisticated part of our brain, the process is initiated by a modest optical element: the eye. ◘ Figure 12.1 shows a schematic and simplified example of the visual process. The eye forms images of the visual world onto the retina. There, light is absorbed in the photoreceptors and the signal transmitted to the visual cortex for further processing. The eye, the first element in the system, is a simple optical instrument. It is composed of only two positive lenses, the cornea and the crystalline lens, that project images into the retina to initiate the visual process. In terms of optical design complexity and compared with artificial optical systems, often formed by many lenses, the eye is much simpler. However, despite this simplicity and the relative poor imaging capabilities, the eye is adapted to the requirements of the visual system. ◘ Figure 12.2 shows a schematic illustration of the eye as compared to a photographic objective that is composed of many single lenses.

Optical systems use transparent materials as glass or plastics with refractive index selected to bend the light rays to form images. In the case of the human visual system, our eyes have to form images of a large field of view for objects placed at different distances with high resolution at least at a central area of the retina. And these tasks have to be accomplished using living tissues.
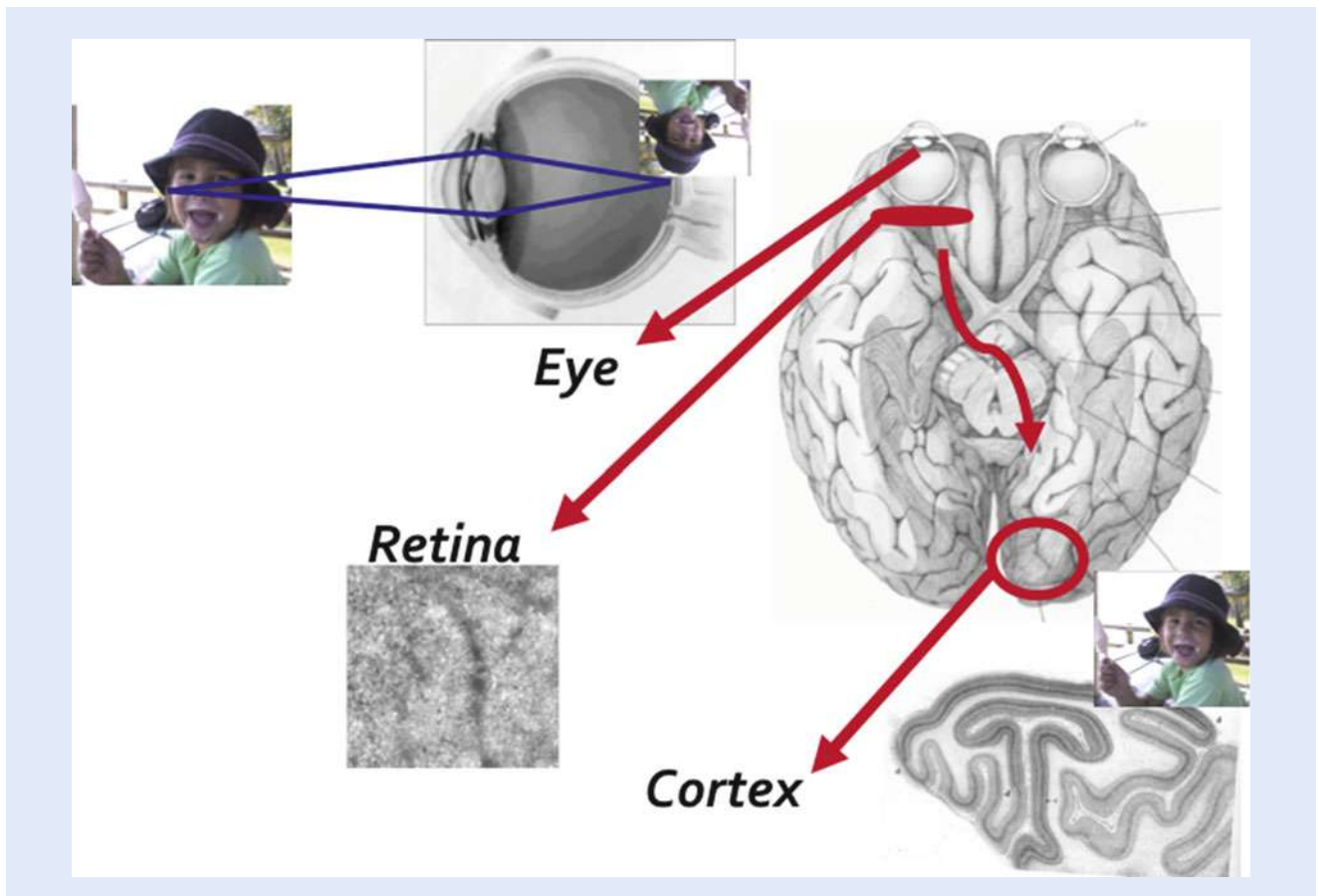
◘ **Fig. 12.1** Schematics of the visual system. The eye forms images of the world on the retina. There, optical images are sampled by the photoreceptors, converted into electrical signals and transferred to the visual cortex for further processing. The eye, although is the simplest part, since it is placed first in the visual cascade may impose fundamental limits
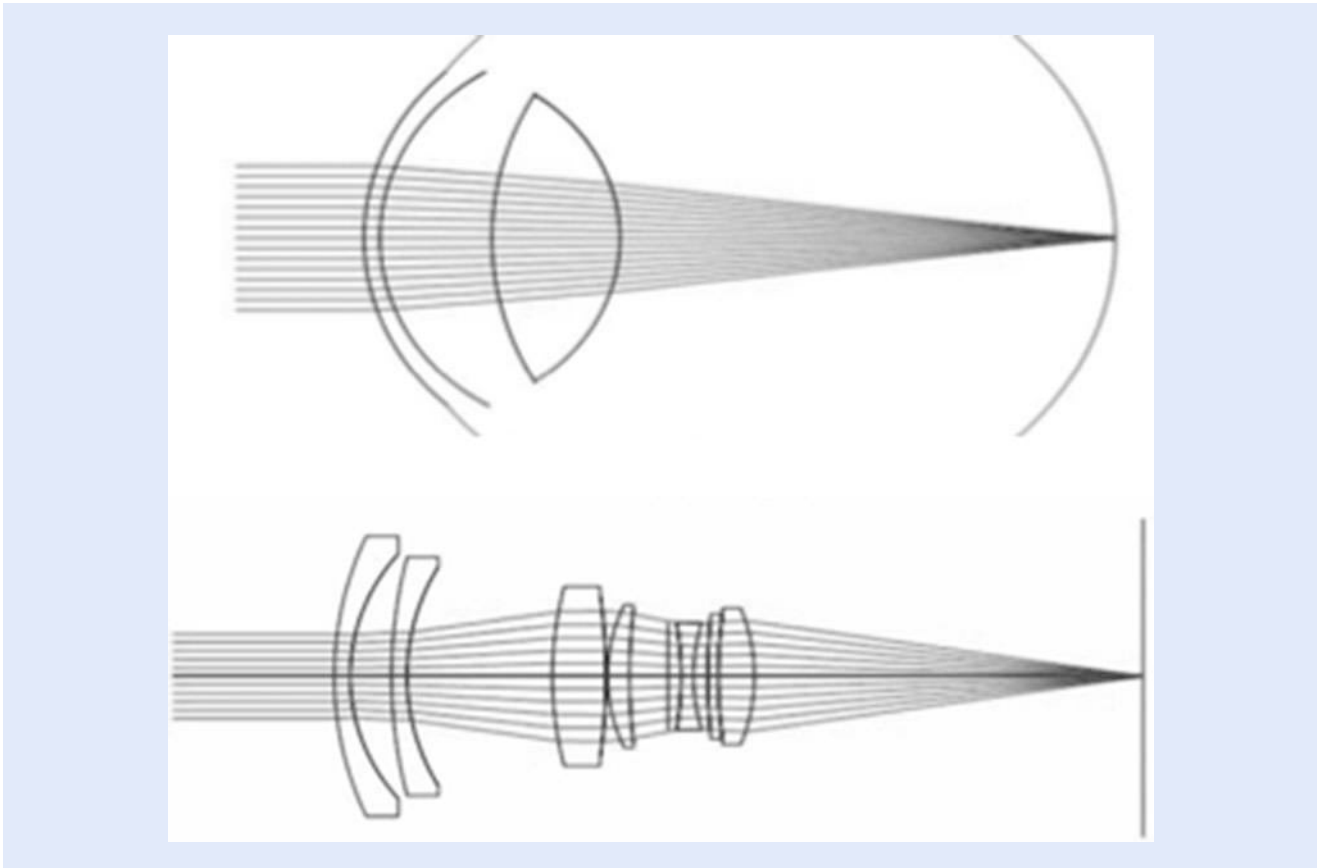
Fig. 12.2 The eye is only composed of two lenses, cornea and crystalline lens, as compared with artificial systems, such as camera objective that may have many single lenses

The eye as an optical instrument is extremely important because our vision is only good when the images formed on the retina are of sufficient quality. If the retinal images are too blurred, the visual system will not work properly. The opposite situation is not true since there are retinal and neural diseases that may impair vision even when the eye forms good quality retinal images.

The intrinsic nature of the light is somehow responsible for some of the characteristics of the eye. Or equivalently, the eye is adapted to transmit visible light and form images on the retina. The sensitivity of the retina is also optimized in the central part of visible spectrum and is similar to the solar emission spectrum. Light may be considered as a transverse electromagnetic wave. Monochromatic light waves have electric fields with sinusoidal oscillation perpendicular to the traveling path. Visible light has wavelengths ranging from approximately blue (400 nm) to red (700 nm), what is a small fraction of the electromagnetic spectrum. A simpler geometrical description of the light as rays pointing along the direction of wave propagation is often used to describe some of the image properties of the eye.

It is interesting to note that also the particle nature of light may have a role in vision under particular conditions. Absorption of light by matter only can be interpreted if the light is considered as a particle, called a photon. Photon absorptions occur in the photoreceptors following the rules of a random process, discontinuously in discrete quanta. Specifically, the light intensity reaching each photoreceptor only determines the probability of a photon being absorbed. This imposes another fundamental limit to vision related to the photon statistics. However, this is restricted to very low luminance conditions after dark adaptation.

12
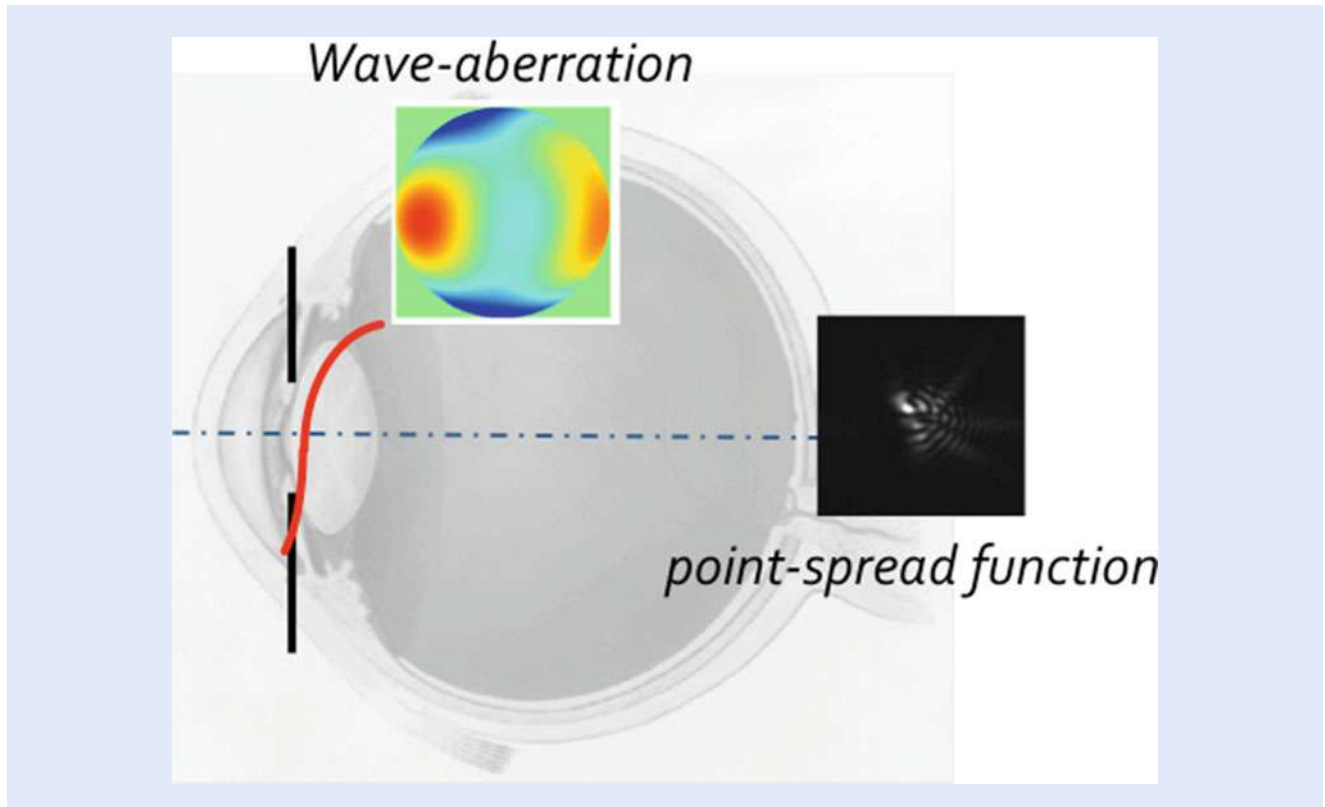
Fig. 12.3 The main functions describing the optical quality of the eye. The wave-aberration on the pupil plane represents the deviation in phase compared with a perfect optical system. The point-spread function (PSF) is the image of a point source formed in the retina. A good eye will form compact and symmetrical images, while in an aberrated eye the PSF will be more extended and asymmetric

Under most normal viewing conditions, the quality of the retinal images is governed by the wave-like nature of the light. The functions used to describe the quality of any optical instrument are showed in ◘ Fig. 12.3. The wave-aberration function is defined as the difference between the perfect (spherical) and the real wave-front for every point over the pupil. It is commonly represented as a two-dimensional map, where color level represents the amount of wave-aberration, expressed either in microns. The image of a point source is called the point-spread function, PSF. An eye without aberrations has a constant, or null, wave-aberration and forms a perfect retinal image of a point source that depends only on the pupil diameter. By performing a convolution operation, it is possible to predict the retinal images of any object. This can be easily understood as placing a weighted PSF onto each point of the geometrical image. Readers interested in more information on the nature of light and/or the functions describing image quality could read some general optics references [1–3].

12.2 The Anatomy of the Eye

The human eye can be described as a fluid-filled quasi-spherical structure. Anatomically consists essentially of three tissue layers: an outer fibrous layer (the sclera and cornea), an inner layer consisting largely of the retina, but including also parts of the ciliary body and iris, and an intermediate vascular layer made up of the choroid and portions of the ciliary body and iris. The eye in adult humans is approximately a sphere of around 24 mm in diameter. It is made up of a variety of cellular and non-cellular components derived from ectodermal and mesodermal

germinal sources. Externally it is covered by a resistant and flexible tissue called the sclera, except in the anterior part where the transparent cornea allows the light to pass into the eye. Internal to the sclera are two other layers: the choroid to provide nutrients and the retina, where the light is absorbed by the photoreceptors after image formation. The eye moves due to the action of six external muscles permitting fixation and the scanning of the visual environment. The light reaching the eye is first refracted by the cornea, a thin transparent layer free of blood vessels of about 12 mm in diameter and around 0.55 mm thickness in the central part. An aqueous tear film on the cornea assures that the first optical surface is smooth to provide the best image quality. After the cornea, the anterior chamber is filled with the aqueous humor, a water-like substance. The iris, two sets of muscles with a central hole whose size depends on its contraction, acts as a diaphragm with characteristic color depending on the amount and distribution of pigments. The aperture is the opening in the center of the iris, and limits the amount of light passing into the eye. The entrance pupil is the image of the iris through the cornea and the exit pupil the image of the aperture through the lens. The aperture size changes with the ambient light, from less than 2 mm in diameter in bright light to more than 8 mm in the dark. The pupil controls retinal illumination and limits the rays entering the eye affecting the retinal image quality. After the iris, the crystalline lens, in combination with the cornea, form the images on the retina. The crystalline lens is an active optical element. It changes its shape modifying its optical power. The lens is surrounded by an elastic capsule and attached by ligaments called zonules to the ciliary body. The action of the muscles in the ciliary body permits the lens to increases or decreases power.

The retina has a central area, the fovea, where photoreceptors are densely packed to provide the highest resolution. The eyes move continuously to fixate the desired details into the fovea. The peripheral parts of the retina render lower resolution but specialize in movement and object detection in the visual field. The typical field covered by the eye is quite large as compared with most artificial optical system, at least 160 × 130°.

The cornea is approximately a spherical section with an anterior radius of curvature of 7.8 mm, posterior radius of curvature of 6.5 mm, and refractive index of 1.3771. Since the largest difference in refractive index occurs from the air to the cornea (actually the tear film), this accounts for most of the refractive power of the eye, on average over 70 %. The lens is a biconvex lens with radii of curvature of 10.2 and −6.0 mm for the anterior and posterior surfaces. The internal structure of the lens is layered, which produces a non-homogeneous refractive index, higher in the center than in the periphery and with an equivalent value of 1.42. The refractive indexes of the aqueous and vitreous humors are 1.3374 and 1.336, respectively. More detailed information on different aspects of the eye's geometry and its optical properties can be found in references [4–6].

An average eye with these distances: 3.05, 4, and 16.6 mm for the anterior chamber, lens, and posterior chamber, respectively, will have a total axial length of 24.2 mm and will image objects placed far from the eye precisely in focus into the retina. This situation is called emmetropia. However, most eyes are affected by refractive errors since they do not have the adequate optical properties or the dimensions required for perfect focus. Refractive errors are classified as myopia, when the images of distant object are focused in front of the retina, and hypermetropia, when distant objects are focused behind the retina. In addition, the eye is not rotationally symmetric, being a common manifestation the presence of astigmatism: the retinal image of a point source consists of two perpendicular lines at different focal distances. ◘ Figure 12.4 shows an example of a myopic eye and the degradation founds in its retinal image.

The ocular media filter the wavelengths reaching the retina. There are a good matching between transmission and photoreceptor sensitivity. The cornea and the
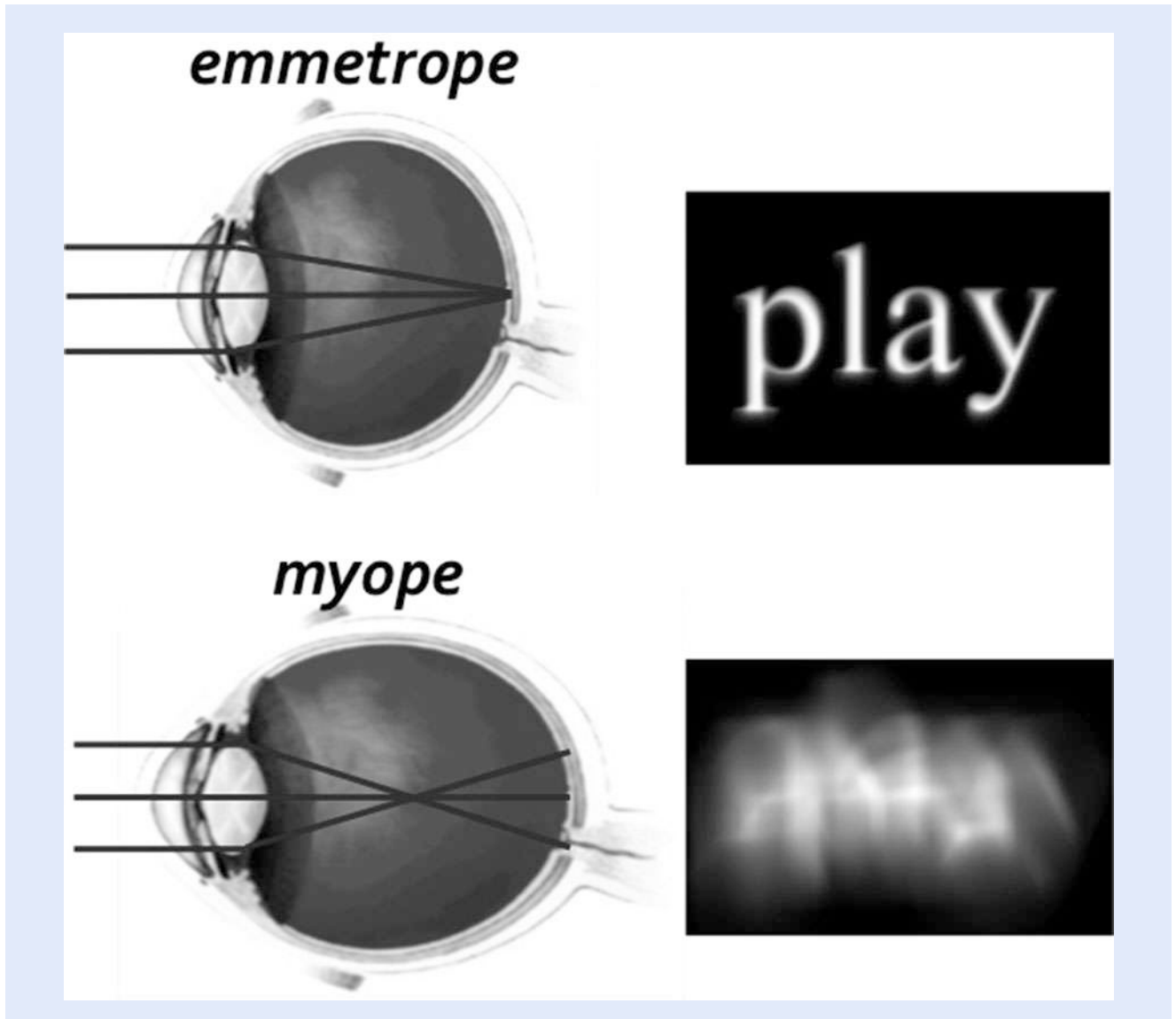
Fig. 12.4 Examples of an emmetrope and a myopic eye and the image formed in the retina of the word "play." In myopes, the image is formed behind the retina and the images are blurred

vitreous have bandwidths that exceed the visible spectrum, but the lens absorbs light in the short wavelength (blue) part of the spectrum. The retina has also pigments that filter the light reaching the photoreceptors. The main filter in the retina is the yellow macular pigment located within the macular region near the fovea. It has been suggested that the macular pigment may protect the retina from degenerative diseases and also improve vision by removing blue light.
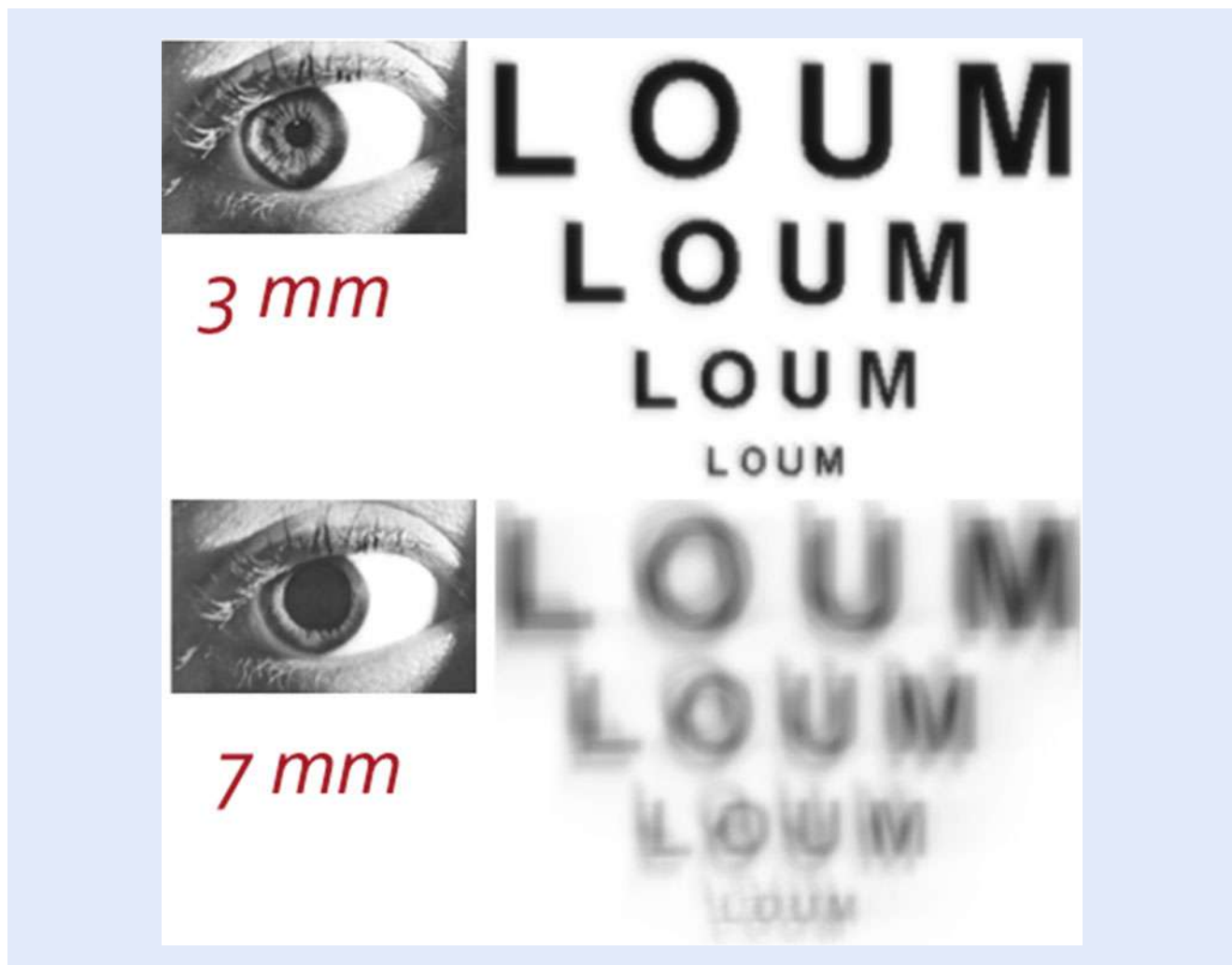
12.3   The Quality of the Retinal Image

Even when the eyes are at perfect focus, as in the case of an emmetrope, they do not produce completely perfect images. This means that the retinal image of a point source is not another point, but an extended distribution of light. Several factors are responsible for the degradation of the retinal images: diffraction of the light in

the eye's pupil, optical aberrations, and intraocular scattering. Diffraction blurs the images formed through instruments with a limited aperture due to the wave nature of the light. The effect of diffraction is usually small and only can be noticed with small pupils. The impact of the ocular aberrations in the eye's image quality is more significant for larger pupil diameters. The pupil of the eye varies diameter from around 2–8 mm in diameter. This corresponds approximately to an aperture range from f/8 to f/2, values which can be compared with the typical values in a camera objective. ◘ Figure 12.5 shows an example of realistic retinal images of letters for the same eye for small (3 mm) and a larger (7 mm) pupil. Note how aberration degrades the image for larger pupils.

The amount of aberrations for a normal eye with about 5 mm pupil diameter (f/4 aperture) is approximately equivalent to less than 0.25 D of defocus, a small error typically not corrected when dealing in the clinic with refractive errors.

The particular shapes of the eye's lenses, refractive index distribution, and particular geometry are responsible for the limited optical quality of the eye compared with artificial optical systems. A normal eye has at least six times lower quality than a good (diffraction-limited) artificial optical system. Each eye produces a peculiar retinal image depending on the optical aberrations present. This can be demonstrated by how a point source is projected in the retina. For example, the shape of stars would depend on our image quality. ◘ Figure 12.6

◘ **Fig. 12.5** Example of the effect of the pupil diameter in the retinal image quality. Aberrations affect more with larger pupils

12
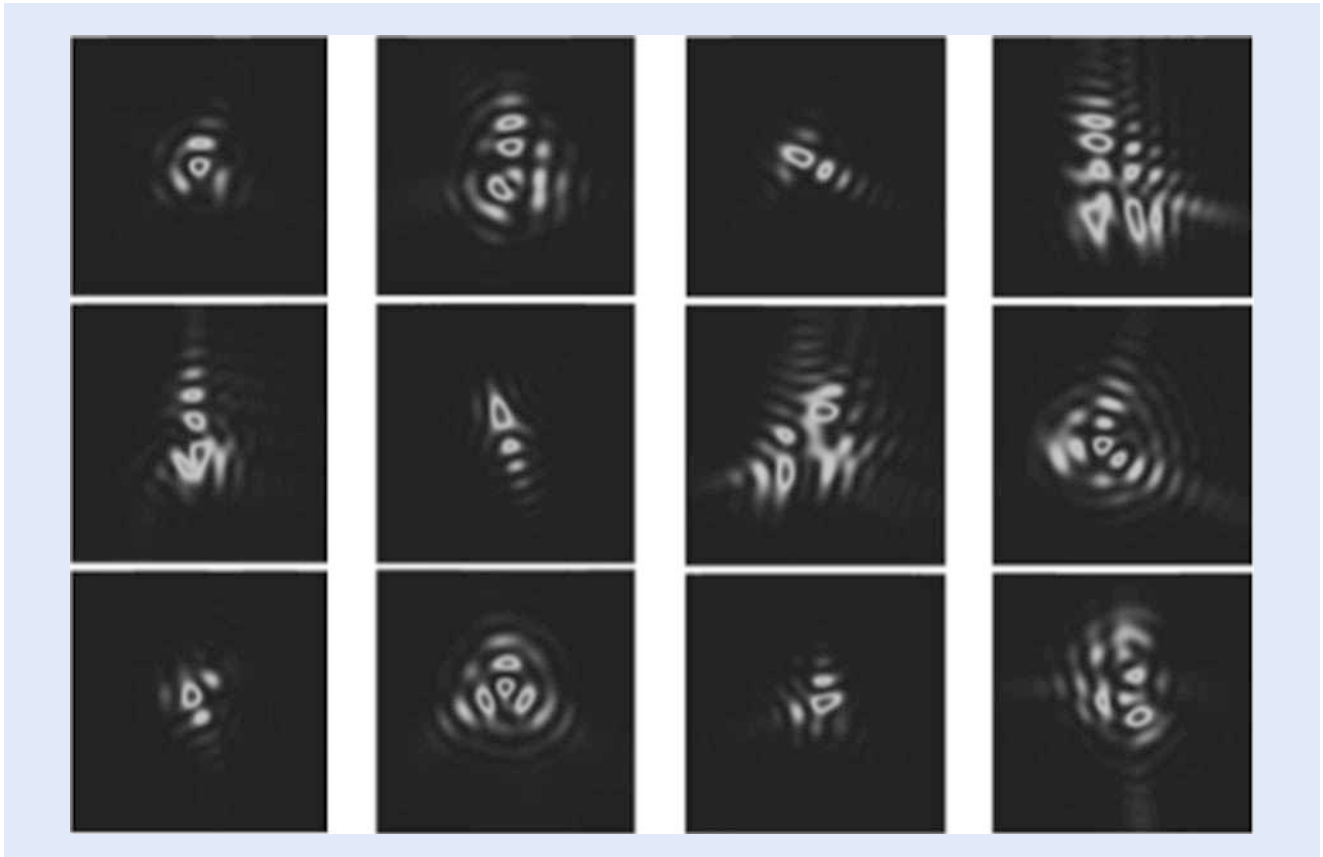
Fig. 12.6 Retinal PSFs measured in a group of normal young subjects. Each eye has particular optical characteristics that produce a unique PSF. It can be understood as if each one of us sees point objects (for instance, stars or distant lights) differently

shows PSFs for a group of normal eyes. This could be understood as how every subject sees an individual start. All are different in shape and size, so our experience of point objects is quite personal.

In addition, chromatic effects also contribute to reduce the retinal image quality since real scenes are usually polychromatic (in white light). This is due to the dependence of refractive index on wavelength that produces changes of the power of the eye with wavelength [7]. The chromatic difference in defocus for the eye from red to blue is large: around 2 diopters. This can be understood as if when you see simultaneously two letters, one red and one blue, and when the red is in perfect focus, the other would be defocus by nearly 2D in your retina. However, your perception of color images is not like that since the real impact of chromatic aberration is smaller than the equivalent of 2D defocus blur. The reason is that the visual system has mechanisms to minimize the impact. The relative larger filtering of blue light in the lens and the macular pigment, together with the spectral sensitivity of the retina, reduce the contribution of the most defocused bluish colors. ◘ Figure 12.7 shows an example of the appearance on the retina of a white letter a in a normal eye. It is important to note that due to retinal and neural factors, the actual impact of this chromatic blur is reduced and our perception less affected.

A question that attracted the interest of many scholars was how the cornea and lens contributed to the eye's optical quality. Early in the nineteenth century, Thomas Young neutralized the cornea by immersing his own eye in water and found that astigmatism persisted. This suggested that the crystalline lens itself have some degree of astigmatism. Recent experiments have also shown that the lens
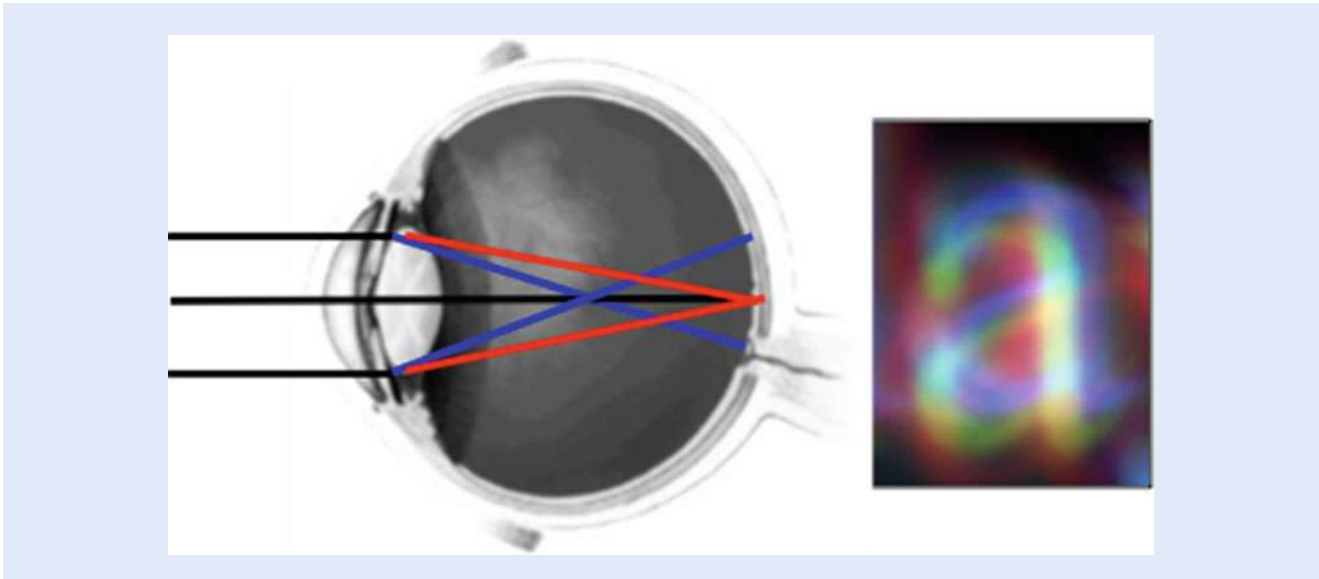
Fig. 12.7 Chromatic aberration in the eye. Due to the chromatic dispersion of the ocular media, if an eye is perfectly focused for red light, it will be myopic (up to 2 diopters) for blur light. This affects the quality of white light images, as the realistic simulation of an image of letter "a" shows (on the *right*)
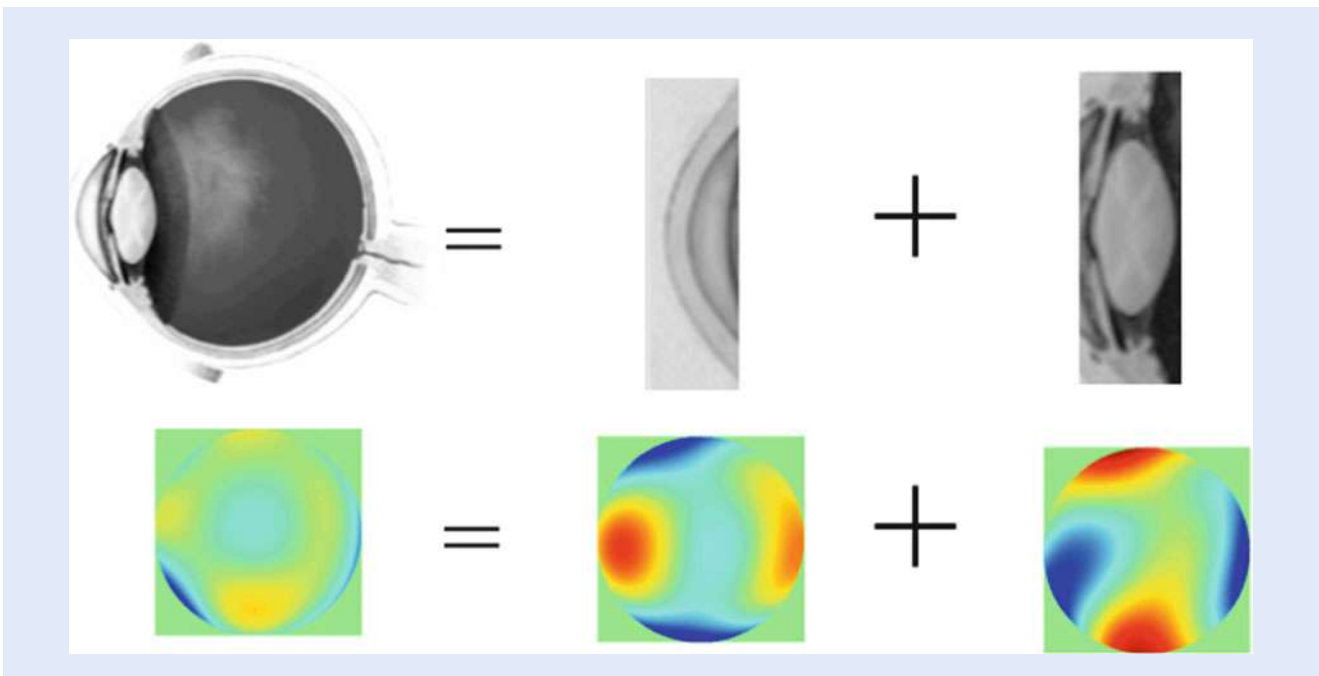
Fig. 12.8 The cornea and lens in young eyes have similar aberrations but with opposite signs rendering the whole eye with better quality than each component isolated. The eye behaves like an aplanatic optical system with partial correction of spherical aberration and coma

compensates not only for some moderate amounts of corneal astigmatism, but also spherical aberration and coma. Figure 12.8 shows as an example the aberrations for one author's eye for the anterior cornea, internal optics (mostly the lens), and the complete eye. The aberrations of the cornea and the lens are somehow opposite rendering an eye with improved optics. The eye as an optical system presents an aplanatic design of the eye, with partial correction of the spherical aberration and coma [8, 9]. This may help to maintain a rather stable optical quality independent of some alignment ocular variables. The reason for this compensation is found in
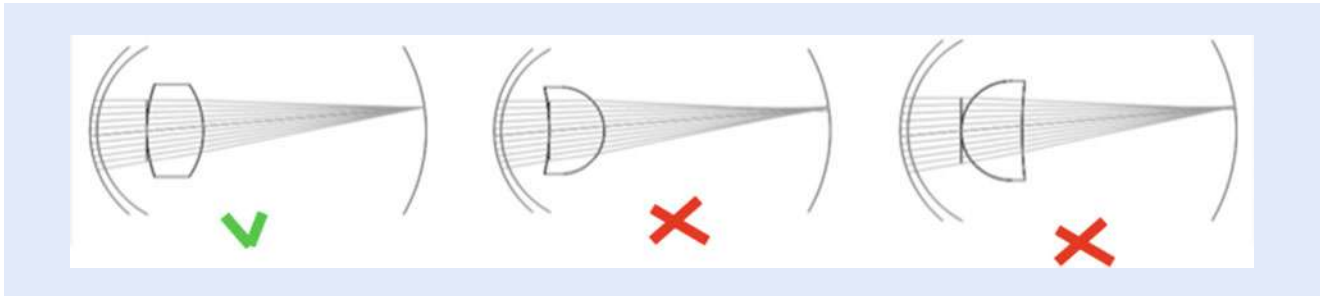
■ **Fig. 12.9** The particular biconvex shape of the lens is an optimized design. Other possible options (as those on the right of the figure) would produce eyes with the same power but with worst image quality

■ **Fig. 12.10** The center of the visual field (fovea) has the highest resolution and declines with retinal eccentricity. This figure shows a practical example. When fixating at the smallest letter at the left when placing this page at 30 cm distance, the larger letters should appear visible at the corresponding eccentricities. This shows the degradation in acuity with eccentricity

the particular shape of the cornea and lens. They have form factors (a relation between their curvature radii) of opposite sign. This means that their shape is optimized by evolution. ■ Figure 12.9 shows an example. Although the three schematic eyes have the same power, so could be considered as plausible solutions in a design, the optimum one is the biconvex lens that actually present in our eyes.

However, this optimized design is only present in younger eyes. During normal aging, the eye's aberrations tend to increases due to a partial disruption in this coupling between cornea and lens [10]. There is another compensatory mechanism: smaller pupil diameters in older eyes tend to compensate for this increase in aberrations.

12.4   Peripheral Optics

The central visual field (the fovea) has the highest spatial resolution; however, the periphery of the retina also plays a crucial role in our visual system. We use the peripheral parts of the visual field to detect objects of interest that we may bring to our fovea more detailed information. Ocular movements change fixation accordingly.

The optics of the eye has a different behavior when the images are formed eccentrically. The oblique incidence of light on the eye produces off-axis aberrations. The ability to discriminate small objects decreases severely with eccentricity. For example, while the normal resolution in the fovea is 1 min of arc, it will increase to 2.5, 5, and 10 at 10, 20, and 30° of eccentricity, respectively. ■ Figure 12.10 shows an example. When fixating to the smallest letter with the
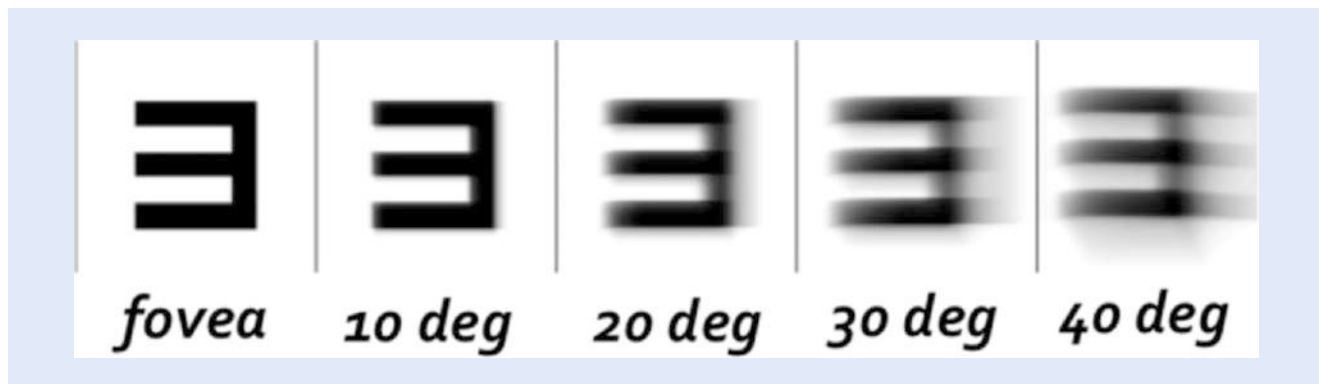
Fig. 12.11 Retinal images of a letter "E" in a normal eye for different angles at the retina, fovea, 10, 20, 30, and 40°. At larger eccentricities, the off-axis aberrations further degrade the quality of the retinal images

fovea placing the book, or the screen, at around 30 cm, the larger letters at the different eccentricity have the correct size to be still legible.

This resolution reduction is due to both optical and neural factors: the eccentric angular incidence induces optical aberrations, which lower the contrast of the retinal images, and the density of cones and ganglion cells also decline with eccentricity, resulting in sparse sampling of the image. In the fovea (central vision) the optics is in many cases the main limiting factor for vision, in the periphery vision is limited by neural factors. The optics is degraded for eccentric angles by distortion, field curvature, astigmatism, and coma [11]. Field curvature is a defocus for off-axis objects and implies that the best image is not formed on a plane but on a parabolic surface. In the eye, the screen is the retina, which has a spherical shape constitutes a curved image plane that in most cases compensates for field curvature. Astigmatism off-axis induces a significant optical degradation in the periphery.

Figure 12.11 shows examples of retinal images of a letter for different eccentricities. Despite the poor optics, visual acuity in the periphery cannot be improved with optical corrections. However, it is interesting to know that our peripheral optics is also optimized by the gradient index structure of the crystalline lens. This was demonstrated by comparing the peripheral image quality in the eyes of a group of patients with one eye implanted with an artificial intraocular lens and the fellow eye still with the natural pre-cataract lens. The eyes implanted had more astigmatism in the periphery than the normal eyes. This result suggests that the crystalline lens provides a beneficial effect also to partially compensate peripheral optics.

12.5 Conclusions

The eye is a simple and robust optical instrument that is fully adapted to serve our visual system. Although the optical quality is not as good in the eye as in the best artificial optical systems, it matches what is required by most of the visual capabilities. There are also a number of compensating mechanisms in the visual system that renders some of the potential optical limitations as invisible. For instance, the large potential deleterious effect of chromatic defocus is limited by proper color filters and the band-pass spectral sensitivity.

An interesting discussion in the last decades has been the possibility to correct for the aberrations of the eye using adaptive optics [12]. This is now technically possible in the laboratory and also, although partially, with correcting devices such
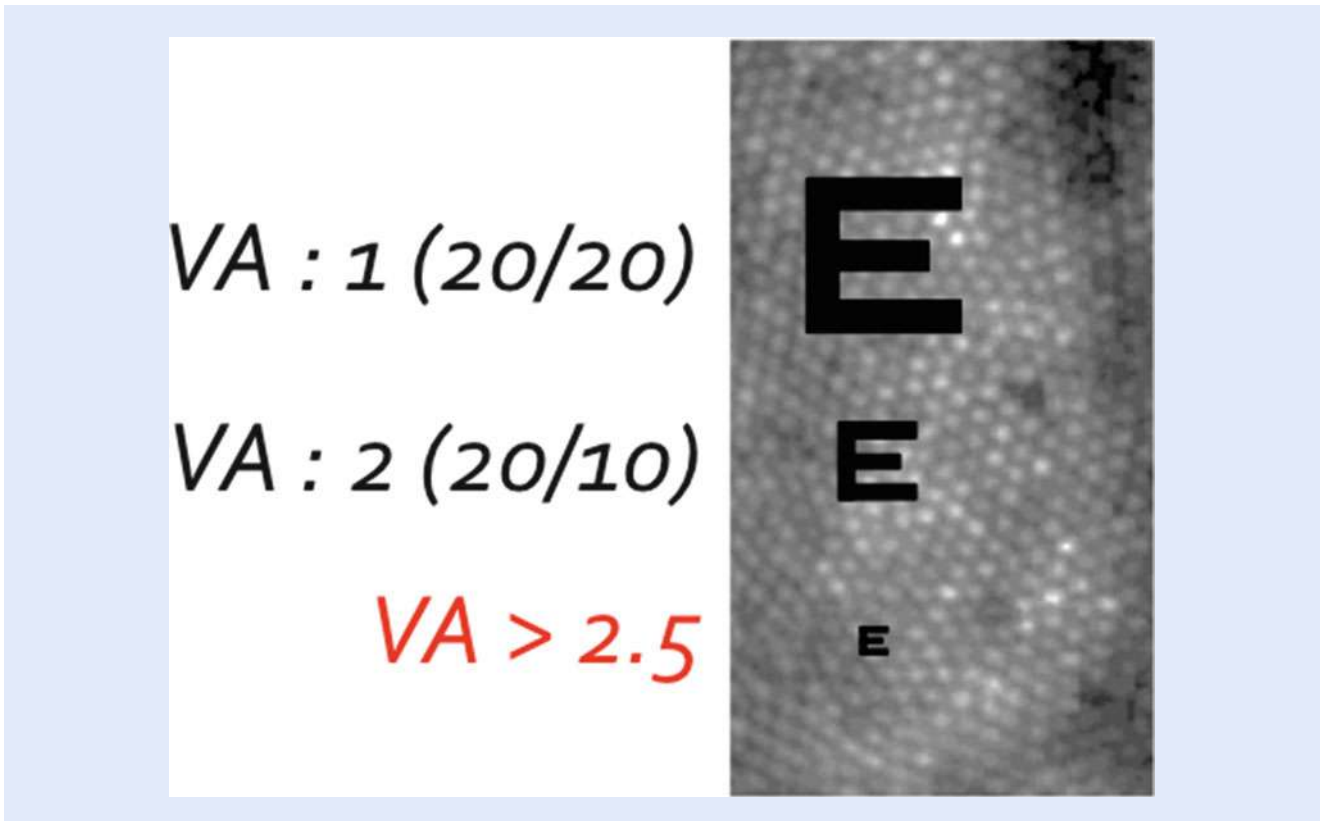
Fig. 12.12 The sampling of the images the retina by the photoreceptors imposes the second limit to vision after the optics. Images of letters corresponding to decimal visual acuity 1, 2, and larger than 2.5 are represented. At the limit, a very small letter even if well projected into the retina would cover one single receptor. Then, the details of the letter will not be detectable. This suggests that even if the optics were perfect, visual acuity would not be improved

as intraocular lenses. The correction of the eye's aberrations may improve vision in some subjects but there are fundamental limitations that cannot be surpassed. The first is the sampling of the retinal images by the photoreceptors. Even if sharp images are projected into the retina, the smallest letter to be perceived will require several photoreceptors across to be properly interpreted. ◙ Figure 12.12 shows this schematically at a correct scale. Images of letters smaller than those corresponding visual acuity (decimal) two will not be discriminated even if the letter is resolved by the eye's optics.

However, as was pointed out, the main cases for optical degradation are not higher order aberrations, but defocus and astigmatism. In that context, the manipulation of the eye's optics by different devices has been a successful technological development since the correction of defocus in the thirteenth century to the use of cylindrical lenses to correct astigmatism in the nineteenth century. Today, it is also possible to correct and induce also higher order aberrations in contact lenses, intraocular lenses, or laser refractive surgery procedures.

The future of correcting the eye's optics is both exciting and promising. And photonics and light technology will surely play a key role. The use of advanced optoelectronics would allow new prostheses to restore accommodation in the presbyopic eye. Two-photon interaction in the cornea by using femtosecond lasers may offer the possibility of changing the optical properties without the need to remove tissue, as is the case in the current ablation-based procedures. Optical technology is also fundamental in new diagnosis instrument. New swept-source optical coherence tomography allows full three-dimensional imaging in real time

of the eye in an unprecedented manner. And ophthalmoscopes equipped with adaptive optics obtain high-resolution images of the retinal structures in vivo. Optics and photonics are now, more than ever, at service to help our eyes to see well.

References

1. Pedrotti FL, Pedrotti LS (1993) Introduction to optics, 2nd edn. Prentice Hall, New Jersey
2. Goodman JW (2005) Introduction to fourier optics, 4th edn. McGraw-Hill, New York
3. Smith WJ (1990) Modern optical engineering, 2nd edn. McGraw-Hill, New York
4. Le Grand Y, El Hage SG (1980) Physiological optics. Springer, Berlin
5. Smith G, Atchison D (1997) The eye and visual optical instruments. Cambridge University Press, New York
6. Atchison DA, Smith G (2000) Optics of the human eye. Butterworth-Heinemann, Oxford
7. Bedford RE, Wyszecki G (1957) Axial chromatic aberration of the human eye. J Opt Soc Am 47:564–565
8. Artal P, Benito A, Tabernero J (2006) The human eye is an example of robust optical design. J Vis 6:1–7
9. Artal P, Tabernero J (2008) The eye's aplanatic answer. Nat Photonics 2:586–589
10. Artal P, Berrio E, Guirao A, Piers P (2002) Contribution of the cornea and internal surfaces to the change of ocular aberrations with age. J Opt Soc Am A 19:137–143
11. Jaeken B, Artal P (2012) Optical quality of emmetropic and myopic eyes in the periphery measured with high-angular resolution. Invest Ophthalmol Vis Sci 53:3405–3413
12. Fernández EJ, Iglesias I, Artal P (2001) Closed-loop adaptive optics in the human eye. Opt Lett 26:746–748

Optics in Medicine

Alexis Méndez

A. Méndez (✉)
MCH Engineering LLC, Alameda, CA 94501, USA
e-mail: alexis.mendez@mchengineering.com

13.1 Introduction

13.1.1 Why Optics in Medicine?

Unlike the present time, medical practitioners of the ancient world did not have the benefit of sophisticated instrumentation and diagnostic systems, such as X-rays, ultrasound machines, or CT scanners. Visual and manual auscultations were the tools of the day. Hence, since the early days of medicine, optics has been a useful and powerful technology to assist doctors and all forms of healthcare practitioners carry out examination and diagnosis of their patients. This is so because one of the fundamental aspects of medicine is observation and physical examination of the patience's general appearance. Hence, anything that can help "see" better the condition of a patient will be of aid. As such, optics, as the science that studies the behavior and manipulation of light and images, is an ideal tool to assist doctors gain better visual examination capabilities by providing improved illumination, magnification, access to small or internal body cavities, among others. But it is in reality light and its interaction with living tissues that is at the center of what makes optics in medicine possible. Light possesses energy and is capable of interacting with biological cells, tissues, and organs. Such interaction can be used to probe the state of such living matter for diagnostics and analytical purposes or, it could be used to induce changes on the same living systems and be exploited for therapeutic purposes. The science of light generation, manipulation, transmission, and measurement is known as *photonics*. The application of photonics technologies and principles to medicine and life sciences is known as *biophotonics*.

Nowadays, it is not only optics but also photonics that are used extensively in a myriad of medical applications, from diagnostics, to therapeutics, to surgical procedures. Hence, when we use the term medical optics, we are referring to biomedical optics and biophotonics as well. The interrelation between optics and light in medicine is ever present and it could be said that more significant advances in biophotonics are now due to the availability of more powerful, concentrated, and multi-spectral light sources which have been available only in the last 50 years. Historically, ambient light was the illumination source, which precluded performing exams late in the day or during certain hours in the winter time. Oil candles in the ancient world gave way to wax ones and alcohol burning lamps in the fifteenth through the nineteenth centuries until the development of electricity and the introduction of the electric lamp by Edison. Then, in the 1960s, with the development of semiconductor lasers, light emitting diodes (LEDs) and lasers, modern medical optics began to take shape and, coupled with the availability of optical fibers, a new generation of medical instruments and techniques began to be developed.

Fiber optics has been used in the medical industry even before their adoption and subsequent explosion as the technology of choice for long haul data communications [1]. The advantages of optical fibers have been recognized by the medical community long ago. Optical fibers are thin, flexible, dielectric (non-conductive), immune to electromagnetic interference, chemically inert, non-toxic, and of course, small in size. They can also be sterilized using standard medical sterilization techniques. Their major advantage lies in the fact that they are thin and flexible so they can be introduced into the body for both remotely sense, image and treat. Their initial and still most successful biological/biomedical application has been in the field of endoscopic imaging. Prior to the development of such devices, the only method of inspecting the interior of the body was through invasive surgery. Many patients owe their lives today to the existence of fiberoptic endoscopes. Optical fibers are not only useful for endoscopes, but can also be used to transmit light to tissue regions of interest either to illuminate the tissue so that it

> **◻ Table 13.1** Medical industry trends that promote the use of optical fibers

• Drives towards minimally invasive surgery (MIS)→Need for disposable probes and catheters • Miniaturization, Automation and Robotics→Need for instrumented catheters	
• Sensors compatible with MRI, CT, PET equipment as well as thermal ablative treatments involving RF or microwave radiation→Need for fiber sensors	
• Increased user of lasers→Need for fiber delivery devices • Increased use of optical imaging and scanning techniques→Need for fiber OCT probes	

can be inspected, or if much higher power laser light is used, to directly cut or ablate it. Hence, they are used extensively as laser-delivery probes, as well as imaging conduits in optical coherence tomography (OCT).

Optical fibers have revolutionized medicine in many ways and continue to do so thanks to the advent of new surgical trends, as summarized in ◻ Table 13.1. One such trend is the advent of minimally invasive surgery (MIS) where the trend now is to avoid cutting open patients and instead, perform small cuts and incisions through which a variety of different surgical instruments, such as catheters and probes, are inserted through these small opening, thus minimizing the postoperative pain and discomfort. Furthermore, there is today growing use of surgical robots where a surgeon operates them remotely using control arms to do a surgical procedure from the comfort of his office while the patient is at a remote hospital location. However, one of the issues with these types of systems is the fact that the surgeon loses the actual manual feedback and does not have sensitivity of the force needed to apply to a scalpel or other surgical tools. This is called *haptic* feedback. These "robotic surgeons" operate using very small tools and catheters, and in order to make sensing elements compatible with such slender instruments, fiber optics represent an ideal solution to provide shape, position, as well as force-sensing information to the remote surgeon's controls.

Fiber optic and photonic devices are also being exploited as sensing devices for patient monitoring during medical imaging and treatment using radiation devices such as MRI, CT, and PET type scan systems that involve the use of high-intensity electromagnetic fields, radiofrequencies, or microwave signals. Because the patient's risk of an electric shock conventional electronic monitoring devices and instrumentation cannot be used in these applications. Instead, patient monitoring is performed using optical fiber sensors.

Based on the above arguments it becomes evident the need for and benefits of optics (and photonics) in medicine. ◻ Table 13.2 summarizes the key general applications for optics in medicine. In general, it could be said that optics has been and will continue to be an enabling technology to further the development and advancement of medicine and the healthcare industry as a whole.

13

Table 13.2 Typical applications of optics in medicine

OPTICS

Medical Instruments	Sensors
Endoscopy	Imaging
Spectroscopy	Digital X-rays
Illumination	Laser Delivery
Microscopy	Surgical Instruments
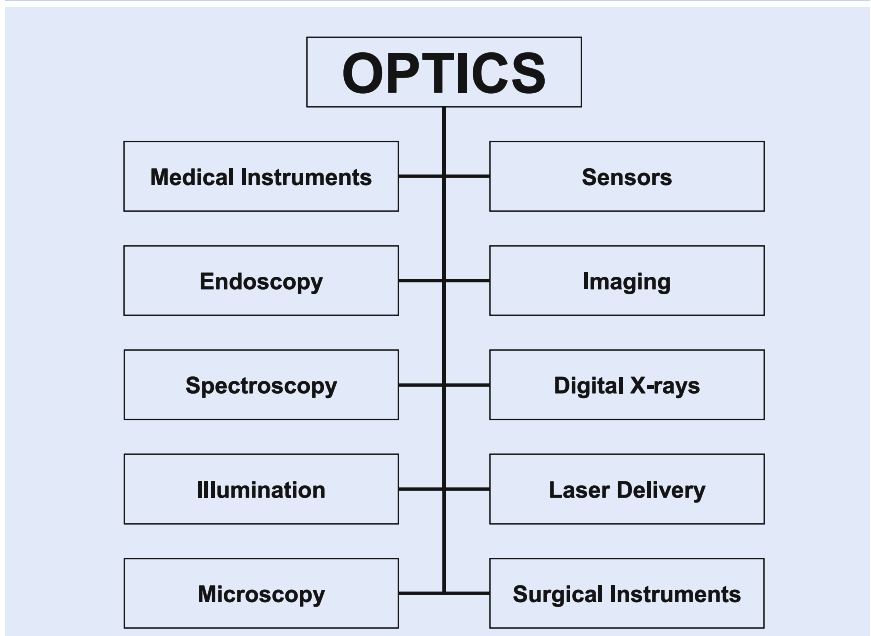
13.1.2 Global Healthcare Needs and Drivers

We all need medical care, from the day we are born, until the day we die. However, this need for medical care has now been affected and accentuated by a convergence of social, demographic, economic, environmental, and political global trends that have been developing over the last few decades. It is a world full of challenges that impact how to effective deliver healthcare in an effective, affordable, and sustainable fashion. On one hand, average lifestyles have changed drastically in the past century resulting in a more sedentary lifestyle with lack of exercise, poor diet, smoking, and excessive alcohol consumption that have resulted in a growing number of chronic diseases such as obesity, arteriosclerosis, diabetes, and cancer that have become leading causes of death and disability. On the other hand, the entire global population keeps growing. According to a recent United Nations Department of Economic and Social Affairs (DESA) report the world's population is estimated to be in excess of 7.3 billion and growing at ~1.1 % annual rate, and expected to reach 8.5 billion by 2030, 9.7 billion in 2050, and 11.2 billion in 2100 [2]. As illustrated in Fig. 13.1, the world population has experienced continuous growth since the end of the Great Famine and the Black Death back in 1350, when the total population stood at merely 370 million. Nowadays, total annual births are approximately 135 million/year, while deaths are around 56 million/year, but expected to increase to 80 million/year by 2040.

Add to this the fact that in certain parts of the world the population is aging, including the USA, Japan, and parts of Europe. Globally, the number of persons aged 65 or older is expected to reach to nearly 1.5 billion by 2050. An aging population puts additional demands on healthcare since older people are more vulnerable to illness and chronic diseases. Furthermore, life expectancy at birth has increased significantly. The UN DESA estimates a 6-year average gain in life expectancy among the poorest countries, from 56 years in 2000–2005 to 62 years in 2010–2015, which is roughly double the increase recorded for the rest of the world. Another key trend and global challenge is the expected shortage of medical doctors and physicians available to meet the healthcare needs of a growing world population.
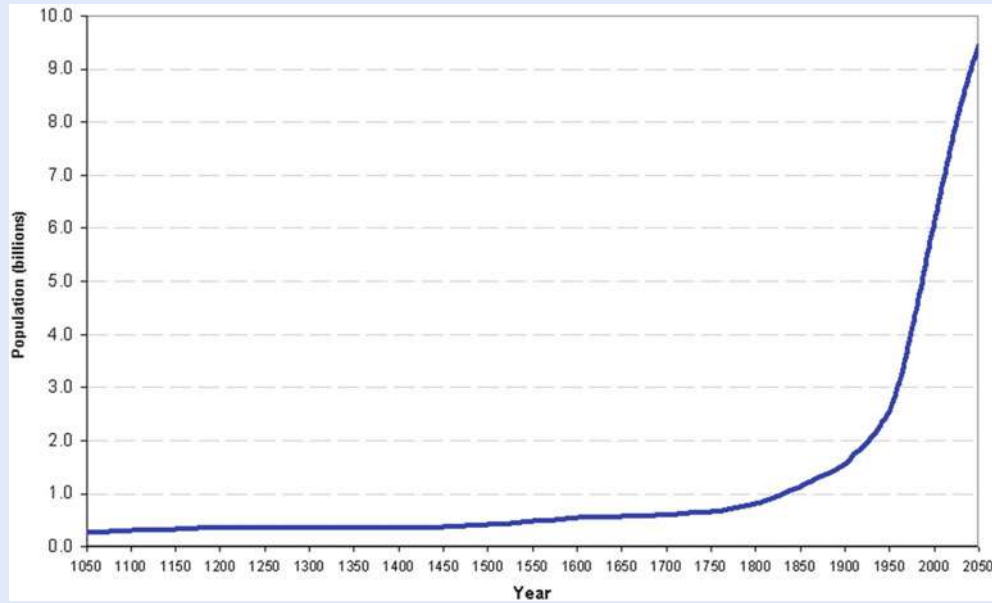
Fig. 13.1 Historical global population growth

A large global population requires more doctors, medical devices, medical supplies, clinics, hospitals, and overall healthcare infrastructure to address the needs of people needing immunizations, or getting sick or injured. Hence, there is and will continue to be an overall growth and expansion of the health care industry on a global basis, that continuous to demand more medical instruments and technical innovations that can facilitate and expedite medical examinations, while reducing costs. Historically, optics has been an enabling technology for the design and development of such medical devices and instruments.

Another relevant and converging present trend is how biomedical devices and instruments are so extremely pervasive across the healthcare industry today. We may not realize it, but whenever we get our blood pressure tested, monitor our blood sugar, or when a expectant mother is being monitored by her doctor, an instrument or sensing device is needed which, often times, is based on the use of an optical technique or based on the use of optical components. Couple this with the fact that in many parts of the underdeveloped world there is not enough doctors, hospitals, clinics, and instrumentation available to support local populations. Hence, it becomes critically important to develop simple, practical, effective, and inexpensive medical devices that can be used in rural and remote areas by non-professionals to examine and treat patients.

13.1.3 Historical Uses of Optics in Medicine

Mankind has always been fascinated with light and the miracle of vision, dating back to the first century when the Romans were investigating the use of glass and how viewing objects through it, made the objects appear larger. However, most of the significant developments of optics for medical diagnosis and therapy started occurring in the nineteenth century. Before that, the vast majority of the known published works on optics and medicine dealt mostly with the anatomy and physiology of the human eye. For instance, the Greek anatomist, Claudius Galen (130–201) provided early anatomical descriptions of the structure of the human

13
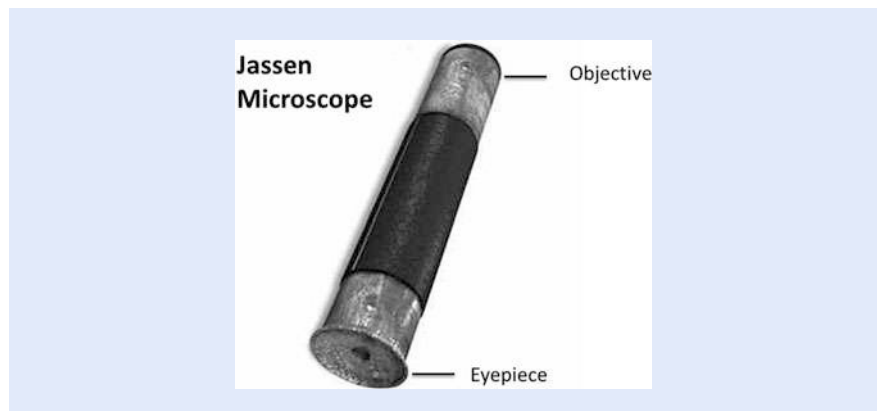
Fig. 13.2 Photograph of the Jansen compound microscope (c. 1595)

eye, describing the retina, iris, cornea, tear ducts, and other structures as well as defining for the first time the two eye fluids: the vitreous and aqueous humors. Subsequently, Arab scholars Yaqub ibn Ishaq al-Kindi (801–873) and Abu Zayd Hunayn ibn Ishaq alIbadi (808–873) provided a more comprehensive study of the eye in the ninth century in their *Ten Treatises on the Eye and the Book of the Questions of the Eye*. In the eleventh century Abu Ali al-Hasan ibn al-Haytham (965–1040)—known as Alhazen—also provided descriptions of the eye's anatomy in his Book of Optics (Kitab al-Manazir).

It is around this time that the so-called reading stones are being used as magnifying lenses to help read manuscripts. The English philosopher Robert Bacon (1214–1294) described in 1268 in his *Opus Majus* the mechanics of a glass instrument placed in front of his eyes. Then, in the thirteenth century, Salvino D'Armate from Italy made the first eye glass, providing the wearer with an element of magnification to one eye.

With the advent of the optical telescope optics took a significant step forward towards the development of one of the first early medical instruments—the microscope [3]. The compound microscope was developed around the late 1590s by Hans and Zacharias Janssen, a father and son team of Dutch spectacle makers, who experimented with lenses by placing them in series inside a tube and discovered that the object near the end of the tube appeared greatly enlarged (see Fig. 13.2).

A seminal optical medical instrument development came in 1804 when the German born physician Philipp Bozzini (1773–1809) developed and first publicized his so-called light conductor (*Lichtleiter*), which enabled the direct view into the living body [4]. The *lichtleiter* was an early form of endoscope which consisted of an open tube with a 45° mirror mounted at the proximal end with a hole in it. Illumination was provided by a burning alcohol and turpentine lamp was shone to a speculum mounted on the distal end and made to fit to the specific anatomy of the desired body opening to be inspected (see Fig. 13.3). In December 1806 Bozzini's light conductor was presented to the professors of the Josephinum, the "Medical-Surgical Joseph's Academy" in Vienna.

A period of significant activity and innovation in medical optics occurred from the mid-1800s through the early 1900s, when a variety of early medical instruments such as otoscopes, ophthalmoscopes, retinoscopes, and others, as well as improved illumination systems were developed. In 1851 German scientist and physician Hermann L. F. von Helmholtz (1821–1894) used a mirror with a tiny aperture (opening) to shine a beam of light into the inside of the eyeball [5]. Helmholtz found that looking through the lens into the back of the eye only produced a red reflection. To improve on the image quality, he used a condenser

Fig. 13.3 Bozzini's original light conductor with specula (c. 1806)

Fig. 13.4 Helmholtz ophthalmoscope (c. 1851)

lens that produced a 5× magnification (▪ Fig. 13.4). He called this combination of a mirror and condenser lens an *Augenspiegel* (eye mirror).

The term ophthalmoscope (eye-observer) did not come into common use until later. Helmholtz also invented the ophthalmometer, which was used to measure the curvature of the eye. In addition, Helmholtz studied color blindness and the speed of nervous impulses. He also wrote the classic Handbook of Physiological Optics.

In 1888 Prof. Reuss and Dr. Roth of Vienna used bent solid glass rods to illuminate body cavities for dentistry and surgery. This would be the earliest idea to

use a precursor of an optical fiber for medical applications. Decades later, in 1926, J. L. Baird of England and Clarence W. Hansell of the RCA Rocy Point Labs, propose independently of each other fiber optic bundles as imaging devices. A few years later, German medical student Heinrich Lamm assembles the first bundles of transparent optical fibers to carry the image from a filament lamp, but is denied a patent. Then in 1949, Danish researchers Holger M. Hansen and Abraham C. S. van Heel begin investigating image transmission using bundles of parallel glass fibers. Prof. Harold H. Hopkins from Imperial College in London begins work in 1952 to develop an endoscope based on bundles of glass fibers. University of Michigan Medical professor Basil Hirschowitz visits Imperial College in 1954 to discuss with Prof. Hopkins and graduate student, Narinder Kapany, about their ideas for imaging fiber bundles. Hirschowitz hires undergraduate student Larry Curtis to develop a fiber optic endoscope at the University of Michigan. Curtis fabricates the first clad optical fiber from a rod-in-tube glass drawing process. Prof. Hirschowitz tests first prototype fiber optic endoscope using clad fibers in February of 1957, and then introduces it to the American Gastroscopic Society in May of the same year.

The first solid-state laser was built in 1960 by Dr. T. H. Maiman at Hughes Aircraft Company. Within the year, Dr. Leon Goldman, chairman of the Department of Dermatology at the University of Cincinnati, began his research on the use of lasers for medical applications and later established a laser technology laboratory at the school's Medical Center. Dr. Goldman is known as the "father of laser medicine." He is also the founder of the American Society for Lasers in Medicine and Surgery [6]. However, the first medical treatment using a laser on a human patient was performed in December 1961 by Dr. Charles J. Campbell of the Institute of Ophthalmology at Columbia–Presbyterian Medical Center, who used a ruby laser that is used to destroy a retinal tumor. Since then, lasers have become an integral part of modern medicine [7].

During the 1980s and 1990s, extensive research was conducted to develop fiber-optic-based chemical and biological sensors for diverse medical applications [8].

OCT is a newer optical medical imaging technique, first introduced in the early 1990s, that uses light to capture micrometer resolution, three-dimensional images from within biological tissue based on low-coherence, and optical interferometry [9]. OCT is a technique that makes possible to take sub-surface images of tissues with micrometer resolution. It can be thought of as the optical equivalent of an ultrasound scanning system. This is an active area of medical research at the moment.

13.1.4 Future Trends

Optics and photonics, as mentioned earlier, are powerful, versatile, and enabling technologies for the development of present and future generations of medical devices, instruments, and techniques for diagnostic, therapy, and surgical applications.

Given the present R&D activity worldwide based on optical and photonic techniques it should be no surprise to expect a broader utilization of optically based solutions across the healthcare industry and medical profession. In the future, advances in the development of ever smaller and thinner medical probes and catheters should be expected, as well as broad utilization of OCT devices to become as common as ultrasound scanning devices are in today's society. There will also be a proliferation of laser-based treatments and therapies. Endoscopy, for its part, will continue to evolve and more sophisticated and smaller devices will be
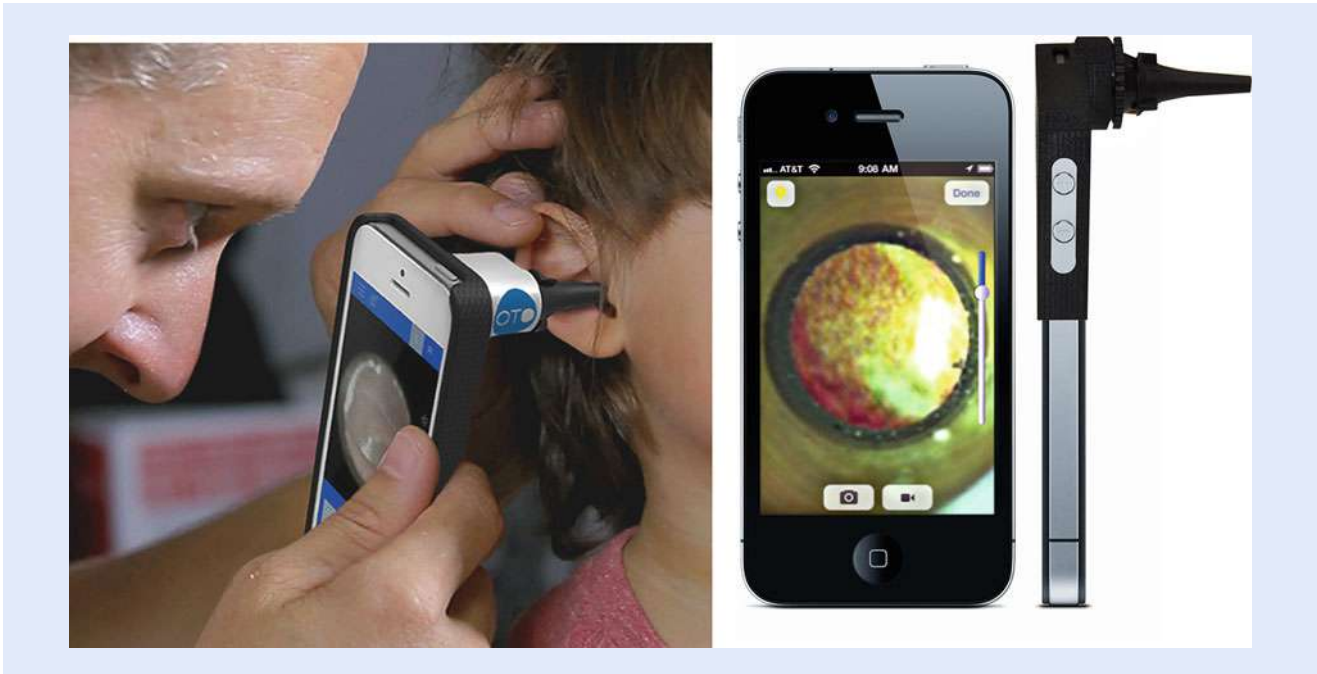
🔲 **Fig. 13.5** A smartphone otoscope

developed that will combine more functions (from the standard illumination and visualization) with direct tissue analysis and laser treatment. Optical imaging techniques will continue to advance along with digital X-rays to make non-invasive examination and diagnosis safe, fast and with greater resolution and pinpoint accuracy.

Other future capabilities brought on by optics will be in the form of the so-called lab-on-a-fiber or LOF for short [10], where optical fibers are combined with micro- and nano-sized functionalized materials that react to specific physical, chemical, or biological external effects and can thus serve as elements to build multi-function, multi-parameter sensing devices. Light would remotely excite the functionalized materials which are embedded in the fiber's coating material. These materials in turn will react to specific biological or chemical substances (*analytes*) and induce an optical signal change proportional to the given analyte concentration.

Some future innovations can already be witnessed today in the form of optical devices used in combination with smart portable cell phones [11, 12]. For example, several new companies have now developed accessories for attachment to smartphones, which turn them into electronic video equivalents of conventional medical examination instruments such as otoscopes (to view inside ears), ophthalmoscopes (to view the inside of eyes), or even simple microscopes. Such devices are passive, optical elements that couple images from the patient to the video lens onto the smartphone's digital camera transform it into a fully functioning, network-connected medical instrument, capable of sending images and video remotely to a consulting doctor. 🔲 Figure 13.5 depicts a cell phone otoscope in use, while 🔲 Fig. 13.6 depicts a smartphone version of an ophthalmoscope and a dermal loupe.

Another such smartphone innovation is the so-called CellScope developed by researchers at the University of California at Berkeley [13]. The CellScope is a microscope that attaches to a camera-equipped cell phone and produces two kinds
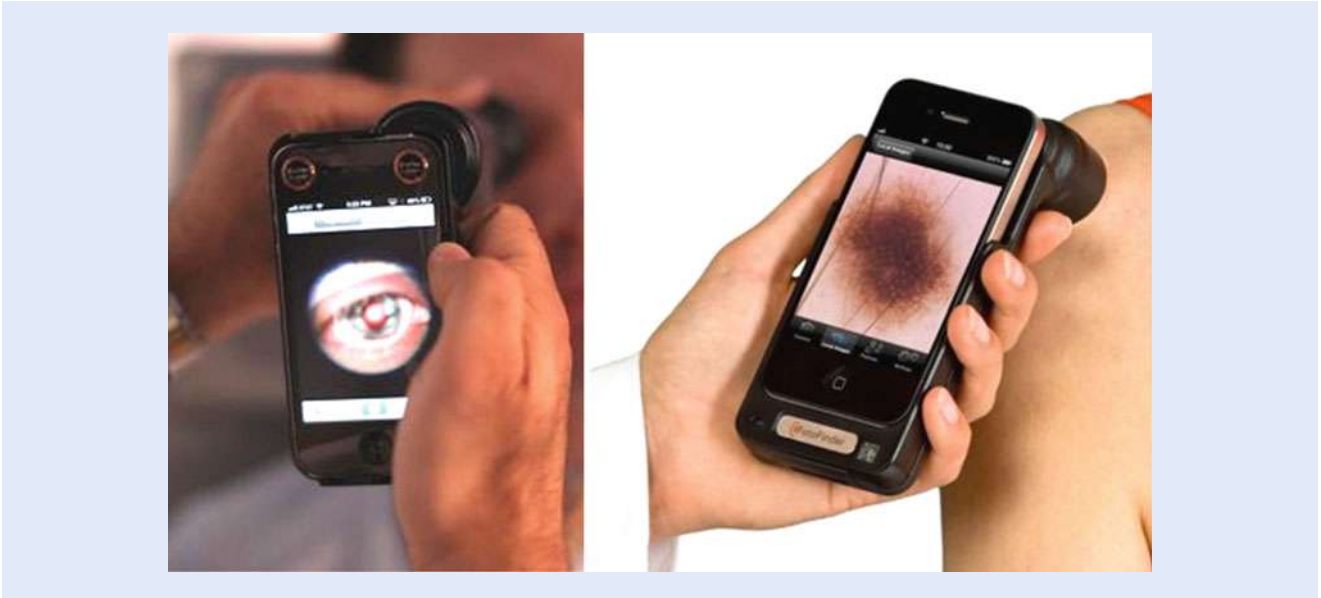
Fig. 13.6 Examples of an ophthalmoscope and a dermal loupe attached to a smartphone

of microscopy imaging: brightfield and fluorescence. The idea is that such device can then be used in the field (on remote locations or those where little medical infrastructure is available) and take snap magnified pictures of disease samples and transmit them to medical labs via mobile communication networks, and screen for hematologic and infectious diseases in areas that lack access to advanced analytical equipment.

13.2 Early and Traditional Medical Optical Instruments

As discussed earlier, optics has been used throughout the centuries as a technology to assist medical doctors perform examinations of patients. Many of the medical instruments in use today rely on optics and optical components to perform their intended function. In particular, there a set of very basic but very popular and common medical instruments that were developed in the nineteenth century and continue to be used in the medical profession of today. Among these optical instruments we have the *otoscope*, the *ophthalmoscope*, *retinoscope*, *laryngoscope*, and even basic devices such as the head mirror.

In general, many of the basic optical medical instruments have in common the goal to provide both a more direct illumination and optical magnification of the area under examination. Conceptually, these optical instruments are similar to a telescope or microscope, but their optical design is different. Typically, a medical instrument consists of a tubular structure fitted with an objective lens on the distal (patient) end, and an objective lens on the viewing (doctor) end, represented as (1) and (2) in ◘ Fig. 13.7.

This lens arrangement produces a magnification of the object under inspection on the objective side (distal end), which has a size Y, and is positioned a distance P from the entrance pupil of the objective lens. The visual magnification factor M_v is calculated as Eq. (13.1):

$$M_v = \theta' D / Y \tag{13.1}$$
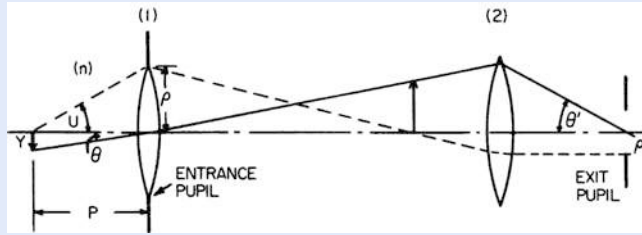
Fig. 13.7 Schematic of a basic medical optical instrument

Fig. 13.8 A medical head mirror and common placement on a doctor's head

where θ' is the angle of the light ray from the eyepiece, D is the viewing distance from the observer to the eyepiece. Hence, the magnification factor M is inversely proportional to the working distance P.

In the sections to follow, we shall describe the basic optical operating principles and uses of such devices. Our discussion of these devices is by no means exhaustive, but is intended to provide the reader with an overall idea on the utilization of optics in medicine and brief introduction on the subject of medical optical instruments [14].

13.2.1 Head Mirror

The most basic optical medical instrument is the so-called head mirror (see Fig. 13.8). A head mirror has historically been used by doctors since the eighteenth century for examination of the ear, nose, and throat. It consists of simple circular concave mirror—made of glass, plastic, or metal—with a small opening in the middle, and mounted on an articulating joint to a head strap made of leather or fabric. The mirror is positioned over the physician's eye of choice, with the concave mirror surface facing outwards and the hole directly over the physician's eye.

In use, the patient sits and faces the physician. A bright lamp is positioned adjacent to the patient's head, pointing towards the physician's face and hence towards the head mirror. The lamp's light gets concentrated by the curvature of the mirror and reflected off it towards the area of examination, and along the line

of sight of the doctor, thus providing shadow-free illumination. When used properly, the head mirror thus provides excellent shadow-free illumination.

A French obstetrician named Levert, who was fascinated with the intricacies of the larynx and dabbled with mirrors, is credited with conceiving the idea for the head mirror back in 1743. Today's head mirror has withstood the test of time and is still routinely used by ophthalmologists and otolaryngologists, particularly for examination and procedures involving the oral cavity.

13.2.2 Otoscope

An *otoscope* is a hand-held optical instrument with a small light and a funnel-shaped attachment called an ear speculum, which is used to examine the ear canal and eardrum (tympanic membrane). It is also called *auriscope*. The otoscope is one of the medical instruments most frequently used by primary care physicians [15]. Health care providers use otoscopes to screen for illness during regular check-ups and also to investigate ear symptoms. Ear specialists—such as otolaryngologists and otologists—use otoscopes to diagnose infections of the middle and outer ear (otitis media and otitis externa).

The design of a modern otoscope is very simple [16]. It consists of a handle and a head (Fig. 13.9). The handle is long and texture for easy gripping and contains batteries to power an integrated light. The head houses a magnifying lens on the eyepiece with a typical magnification of 8 diopters; a cone-shaped disposable plastic speculum at the distal end; and an integrated light source (either lamp bulb, LED, or fiber optic). The doctor inserts a disposable speculum into the otoscope, straightens the patient's ear canal by pulling on the ear, and inserts the otoscope to peer inside the ear canal. Some otoscope models (called pneumatic otoscopes) are provided with a manual bladder for pumping air through the speculum to test the mobility of the tympanic membrane.

The most commonly used otoscopes in emergency rooms and doctors' offices are monocular devices. They provide only a two-dimensional view of the ear canal. Another method of performing *otoscopy* (visualization of the ear) is use of a binocular microscope, in conjunction with a larger metal ear speculum, with the patient supine and the head tilted, which provides a much larger field of view and depth perception, thus affording a three-dimensional perception of the ear canal.

 Fig. 13.9 Otoscope for visual inspection inside the ear canal
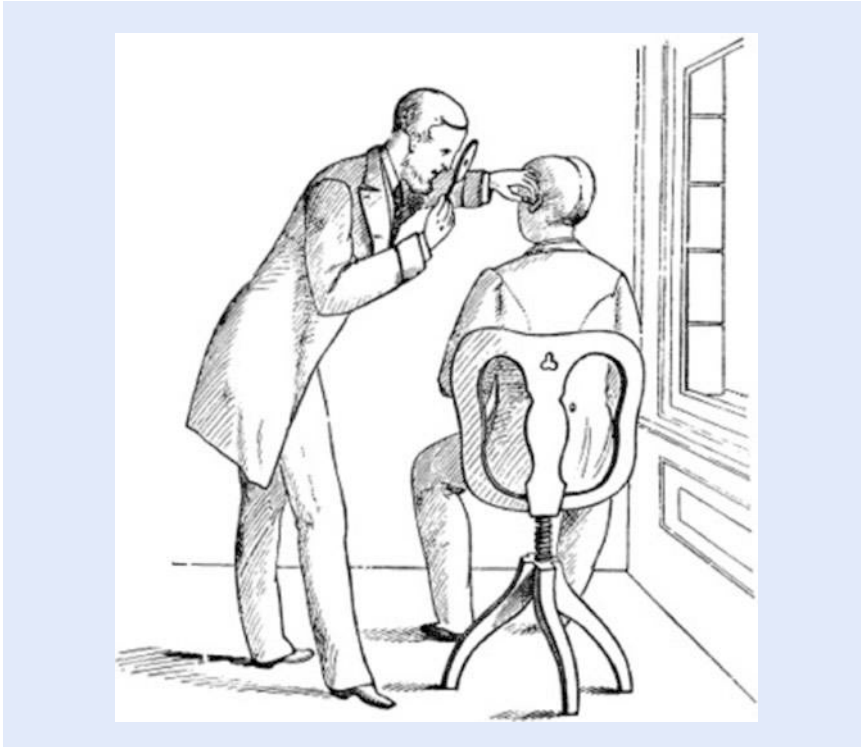
🔲 **Fig. 13.10** Ear examination in the nineteenth century

The microscope has up to 40× power magnification, which allows for more detailed viewing of the entire ear canal and eardrum.

The otoscope is a valuable tool beyond its primary role as an examination tool for detecting ear problems. It can also be used for transillumination, dermatologic inspection, examination of the eye, nose, and throat and as an overall handy light source.

13.2.2.1 History of the Otoscope

Early ear examinations were performed by direct observation of the ear canal during daylight. As a consequence, examinations were limited to times of the day and year when there was adequate bright daylight. Furthermore, a device was needed to gain more direct access to the ear canal and to keep it open and provide direct illumination inside. Hence, over the years, the use of a *speculum* (a conical shape device that can be safely inserted into the ear) was adopted. In 1363 Guy de Montpellier in France described the first aural and nasal specula [17]. However, some means or direct illumination was needed in order to perform more effective ear examinations. The next major requirement was for an adequate method of directing concentrated natural daylight into the depths of the ear canal, which was accomplished by using a perforated mirror mounted either on a handle or on the head, which shone light directly into the ear canal. This allowed the doctor to look down the center of the beam of light, thus eliminating shadow effects and parallax (difference in the apparent position of an object viewed along two different lines of sight) (🔲 Figs. 13.10 and 13.11).

Von Troltsch is generally credited with popularizing the use of a mirror in *otoscopy* after he showed it in 1855 at a meeting of the Union of German Physicians in Paris. He ultimately fastened the mirror to his forehead as is still currently practiced by some doctors. The size and focal length of the mirror was not standardized for some time. In an attempt to catch more light, used huge mirrors and only gradually was a diameter of 6–7 cm eventually adopted. A further

13
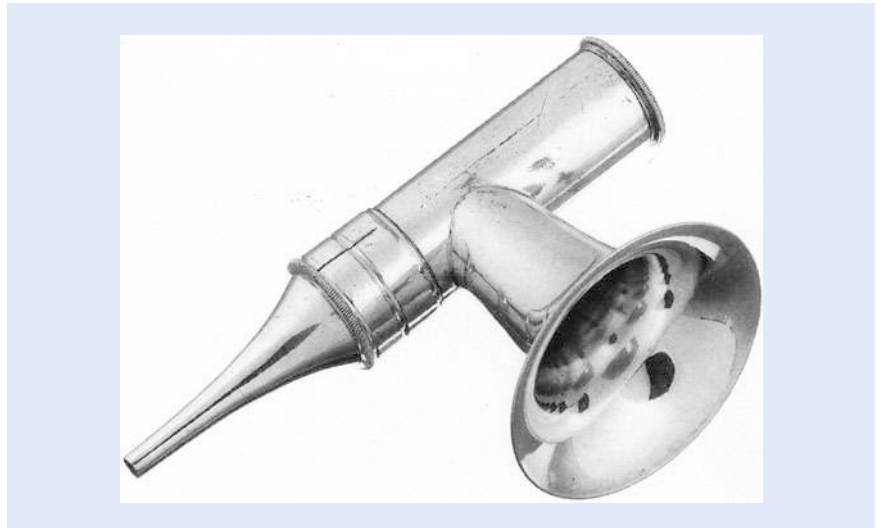
◘ **Fig. 13.11** Bunton's Auriscope (c. 1880). It can be observed the metal tip speculum, the rear objective lens for viewing, as well as the middle horn used to direct light from a candle or lamp

improvement to Von Troltsch's early auriscope is Brunton's device which was first described in an 1865 Lancet article. This auriscope combined mirror and speculum into a single instrument and worked on the principle of a periscope: light from a candle or lamp was concentrated by a funnel and then reflected by a plane mirror set at an angle of 45° into the ear canal. The mirror had a central perforation through which the doctor could view the ear. Brunton's auriscope was fitted with a magnifying lens for the observer and could also be sealed with plain glass at the illuminating end. These were the first otoscopes to be electrically illuminated.

13.2.3 Ophthalmoscope

An *ophthalmoscope* is an optical instrument for examining the interior of the eyeball and its back structures (called the *fundus*) through the pupil by injecting a light beam into the eye and looking at its back-reflection. An ophthalmoscope is also referred to as a *funduscope*. The fundus consists of blood vessels, the optic nerve, and a lining of nerve cells (the retina) which detects images transmitted through the cornea, a clear lens-like layer covering of the eye. Ophthalmoscopes are used by doctors to exam the interior of eyes and help diagnose any possible conditions or detect any problems or diseases of the retina and vitreous humor. For instance, a doctor would look for changes in the color the fundus, the size, and shape of retinal blood vessels, or any abnormalities in the *macula lutea* (the portion of the retina that receives and analyzes light only from the very center of the visual field). Typically, special eyedrops are used to dilate the pupils and allow a wider field of view inside the eyeball.

A modern ophthalmoscope (◘ Fig. 13.12) consists essentially of two systems: one for illumination and another for viewing. The illuminating system is comprised of light source (a halogen or tungsten bulb), a condenser lens system, a reflector (a prism, mirror, or metallic plate) to illuminate the interior of the eye with a central hole through which the eye is examined. The viewing system is made of a sight hole and a focusing system, usually a rotating wheel with lenses of different powers. The lenses are selected to allow clear visualization of the structures of the eye at any depth and compensate for the combined errors of refraction between patient and examiner.
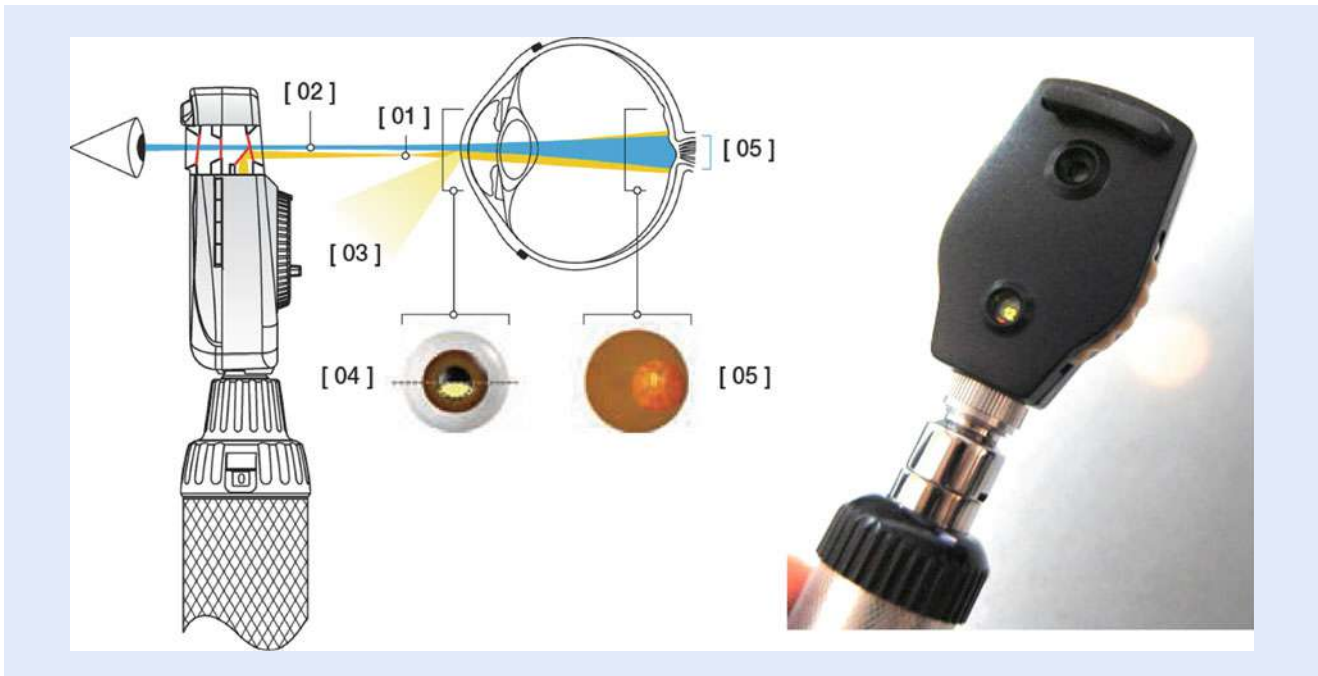
Fig. 13.12 Aspect of a modern ophthalmoscope. A light beam is projected into the eye (*1*). The medical examiner has a direct line of sight into the back of the eye (fundus) (*2*). Path of light reflected of the cornea and iris (*3*). Image observed at the pupil (*4*). Image observed at the back of the eye (*5*) (Images courtesy of Heine)
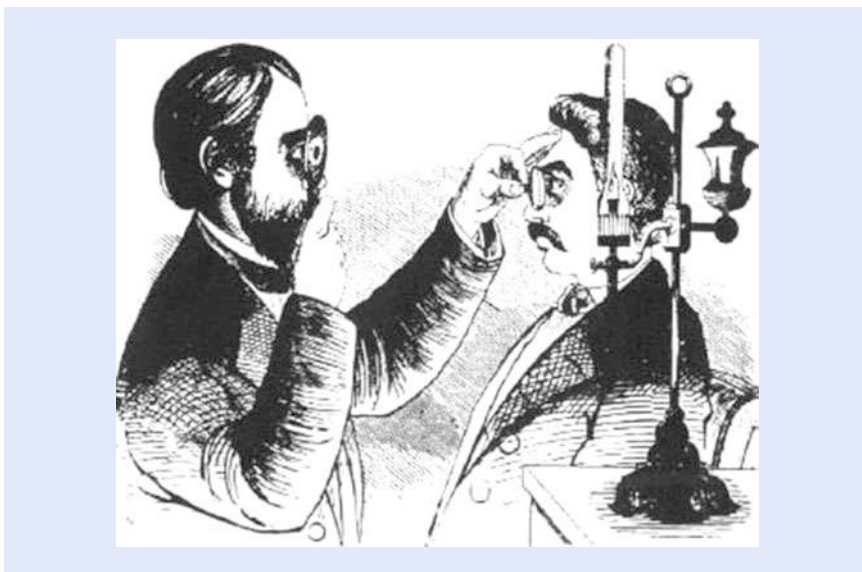
Fig. 13.13 Nineteenth century illustration of Helmholtz original ophthalmoscope

German physician Hermann von Helmholtz is credited with the invention of the ophthalmoscope back in 1851, which he based on an earlier version developed by Charles Babbage in 1847. Helmholtz original ophthalmoscope (see ▪ Fig. 13.13) was very basic (made or cardboard, glue, and microscope glass plates) but it allowed him to place the eye of the observer in the path of the rays of light entering and leaving the patient's eye, thus allowing the patient's retina to be seen. In 1915, Francis A. Welch and William Noah Allyn invented the world's
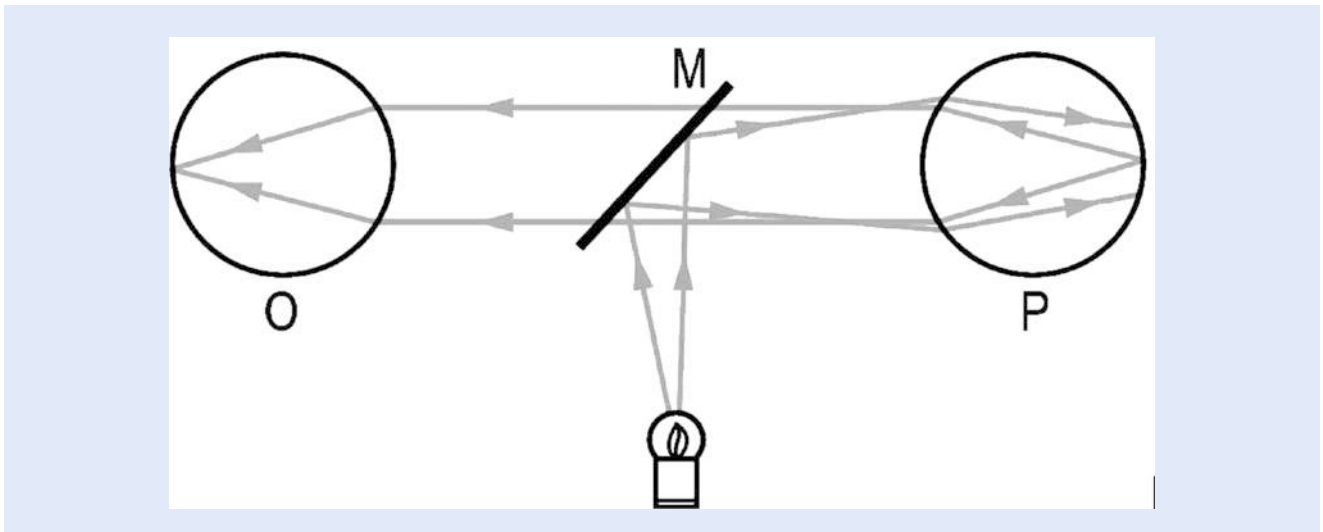
Fig. 13.14 Optical raytracing for a direct ophthalmoscope. Light from the illuminating source is reflected into the eye and then back-reflected by the fundus through a mirror (with either a hole through it or with partial reflectivity). O is the observer's eye, while P is the patient's eye; M, semi-silvered mirror. After [18]

first hand-held direct illuminating ophthalmoscope, and resulted in the formation of the Welch Allyn medical company—still in business today.

There are two types of ophthalmoscope: *direct* and *indirect*. A *direct ophthalmoscope* produces an upright (unreversed) image with 15× magnification. The direct ophthalmoscope is used to inspect the fundus of the eye, which is the back portion of the interior eyeball. Examination is best carried out in a darkened room. Macular degeneration and opacities of the lens can be seen through direct ophthalmoscopy. The instrument is held at close range to the patient's eye and the field of view is small (less than 10°) (◘ Fig. 13.14). The magnification M of a direct ophthalmoscope is equal to:

$$M = F_e/4 \tag{13.2}$$

where F_e is the power of the eye.

An *indirect ophthalmoscope* produces an inverted (reversed) image with a 2–5× magnification and formed. A small hand-held lens and either a slit lamp microscope or a light attached to a headband are used to form an image of the back of the eye in space, at approximately arm's length from the doctor. An indirect ophthalmoscope provides a stronger light source, a specially designed objective lens, and opportunity for stereoscopic inspection of the interior of the eyeball. It is invaluable for diagnosis and treatment of retinal tears, holes, and detachments.

This aerial image is usually produced by a strong positive lens ranging in power from +13 diopter to +30 diopter that is held in front of the patient's eye. The practitioner views this aerial image through a sight hole with a focusing lens to compensate for *ametropia* and accommodation. This instrument provides a large field of view (25–40°) and allows easier examination of the periphery of the retina. This instrument has been supplanted by the binocular indirect ophthalmoscope (◘ Fig. 13.15). The magnification of an indirect ophthalmoscope M is equal to:

$$M = F_e/F_c \tag{13.3}$$

where F_e and F_c are the powers of the eye and of the condensing lens, respectively.
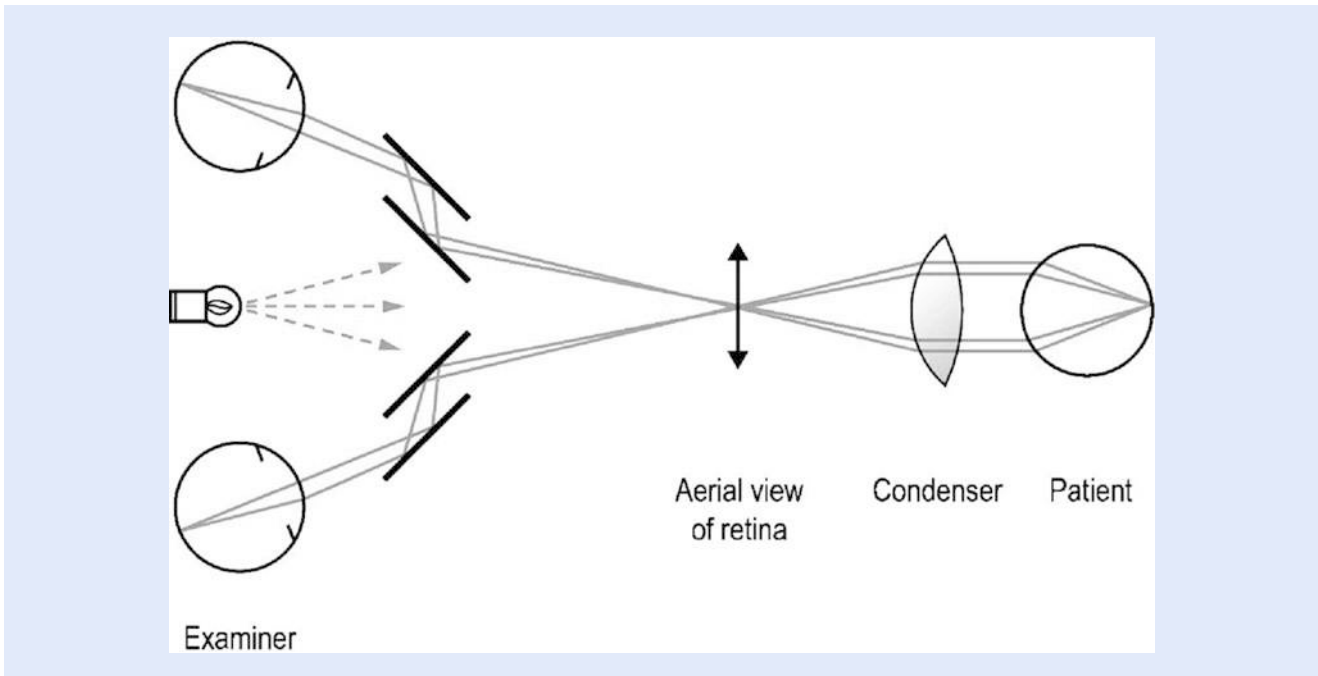
Fig. 13.15 Optical raytracing for a binocular indirect ophthalmoscope. The light source mounted on the doctor's head illuminates a hand-held condenser lens which forms an inverted stereoscopic image of the retina in free space (aerial image). After [18]
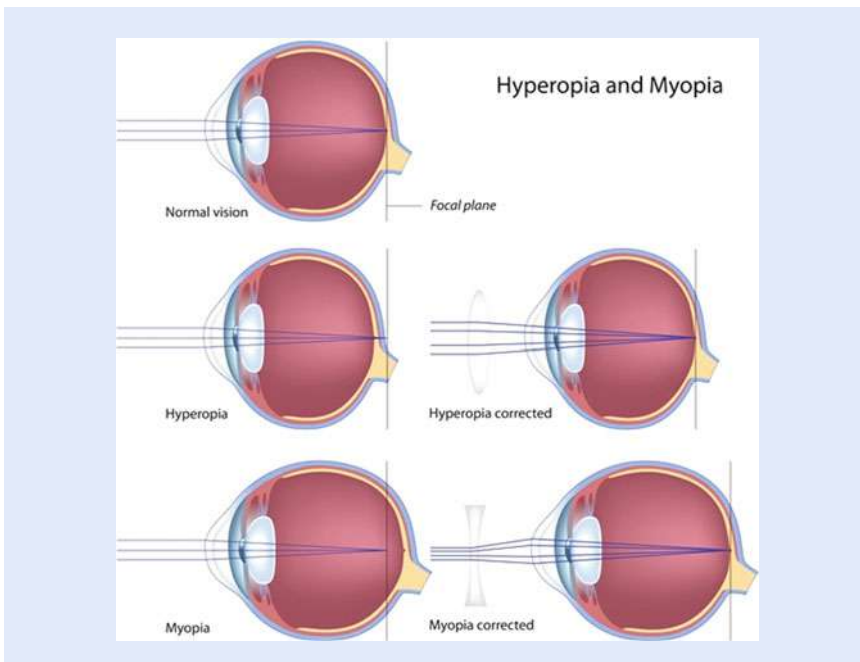
Fig. 13.16 Types of human vision and associated corrective optical lenses

13.2.4 Retinoscope

A *retinoscope* is an optical hand-held device used by optometrists to measure the optical refractive power of the eyes and whether corrective glasses might be needed and the associated prescription value. As shown in Fig. 13.16, a person can have normal vision (*emmetropia*), *myopia* (nearsightedness), *hyperopia*

13
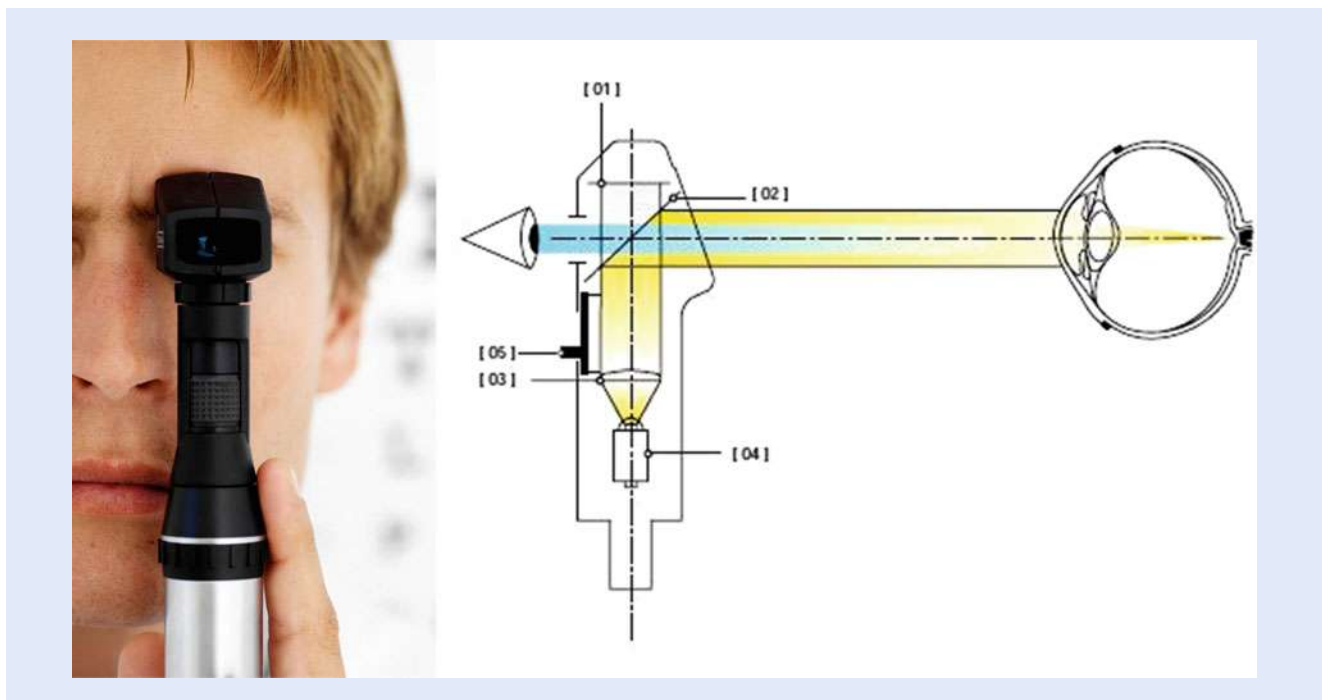
Fig. 13.17 A modern retinoscope. An integrated lamp or LED light source (*4*) shines light through a collimating lens (*3*) onto a partially reflective mirror (*2*), which directs the light to the eye. The back-reflected light from the fundus and the cornea is examined by the doctor through the eyepiece (*1*) and focus adjusted using the lens dial (*5*) (Image courtesy of Heine)

(farsightedness), or astigmatism. The retinoscope is used to illuminate the internal eye (while the patient is looking a far fixed object) and observe how the reflected light rays by the retina (called the *reflex*) align and move with respect to the light reflected directly off the pupil [19]. If the input light beam focuses in front of or behind the retina, there is a "refractive error" of the eye. A high degree of refractive power indicates that the light focus remains in front of the retina, in which case the eye displays myopia. Conversely, if the focal spot happens behind the retina, there is little refractive power and the eye has hyperopia. The error of refraction is then corrected by using a *phoropter*, which introduces a series of lenses of various optical strengths until the retinal reflex focuses at the right position on the retina.

The retinoscope consists of a light, a condensing lens, and a mirror (Fig. 13.17). The mirror is either semi-transparent or has a hole through which the practitioner can view the patient's eye. During the procedure, the retinoscope shines a beam of light through the pupil. Then, the optometrist moves the light vertically and horizontally across the patient's eye and observes how the light reflects off the retina (see pictures in Fig. 13.18). If the light reflex in the patient's pupil moves "with" or "against" motion. If the reflex moves in same direction, then the correction requires plus power (myopia) and motion against direction of the retinoscope, means negative power correction (hyperopia).

To determine the corrective refractive lens power needed, lenses of increasing refractive power are placed in front of the eye and the change in the direction and pattern of the reflex is observed. The optometrist keeps changing the lenses until reaching a lens power that provides adequate focusing on the retina, which manifests as alignment of the reflex with the streak light image outside of the pupil.
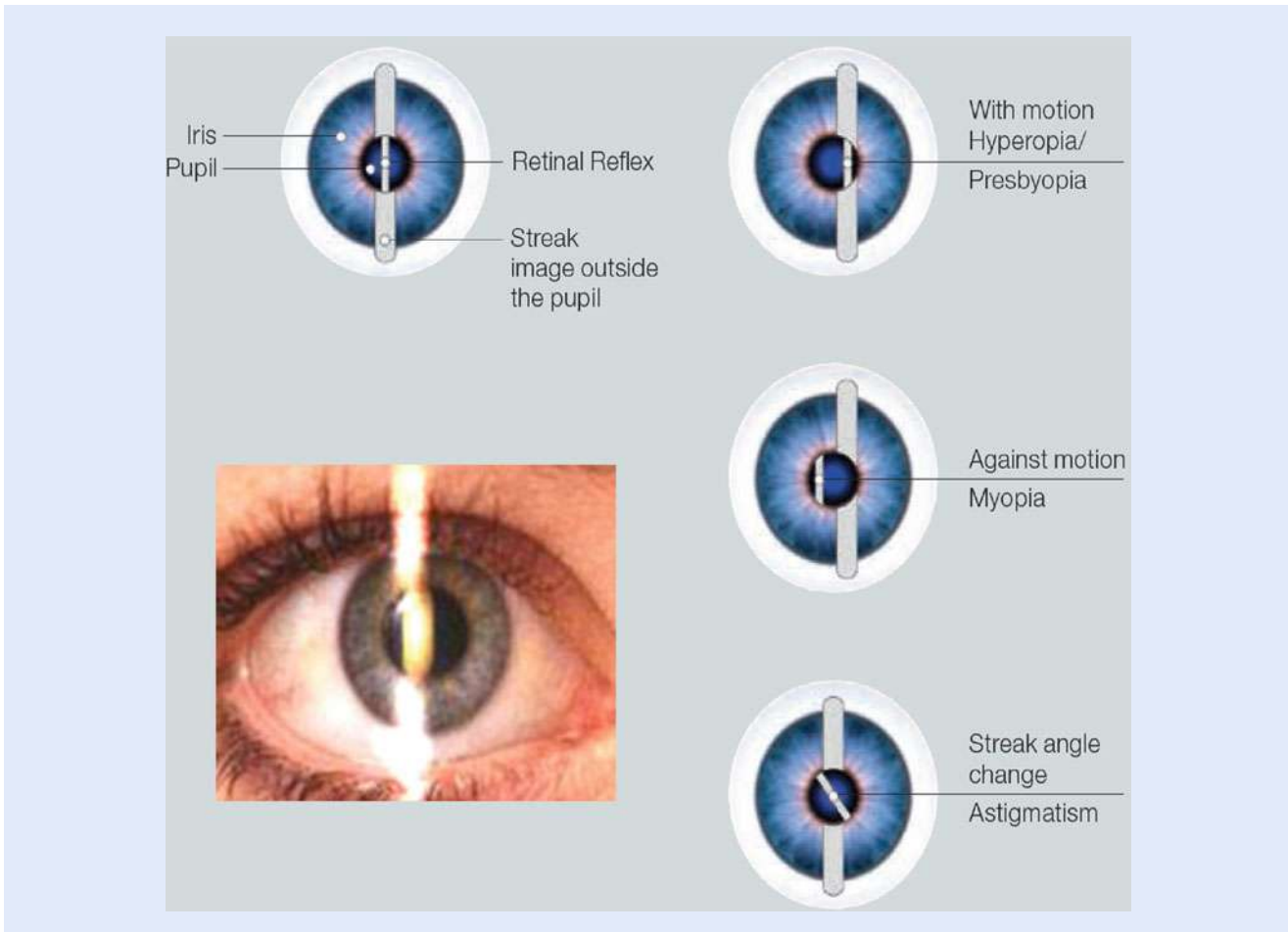
🔴 **Fig. 13.18** Aspect of reflex images from the human eye seen by a doctor using a retinoscope. If the reflex moves in same direction, then myopia is detected; if reflex is noted on motion against direction of the retinoscope, hyperopia is present. If the reflex line is oblique instead of vertical, then astigmatism is present. An aligned reflex means correct vision (*Source*: Heine)

13.2.5 Phoropter

A *phoropter* is an ophthalmic binocular refracting testing device, also called a *refractor*. It is commonly used by ophthalmologists, optometrists, and eye care professionals during an eye examination to determine the corrective power needed for prescription glasses. It is commonly used in combination with a retinoscope.

🔴 Figure 13.19 shows a photograph of phoropter which consists in double sets (one for each eye) of rotating discs containing convex and concave spherical and cylindrical lenses, occluders, pinholes, colored filters, polarizers, prisms, and other optical elements. The patient sits in front of the device and the lenses within a phoropter refract light in order to focus images on the patient's retina at the right spot to compensate for each individual eye refractive errors. The optical power of these lenses is measured in 0.25 diopter increments. By changing these lenses, the examiner is able to determine the spherical and cylindrical power, and cylindrical axis necessary to correct a person's refractive error. These instruments were first devised in the early to mid-1910s.

13
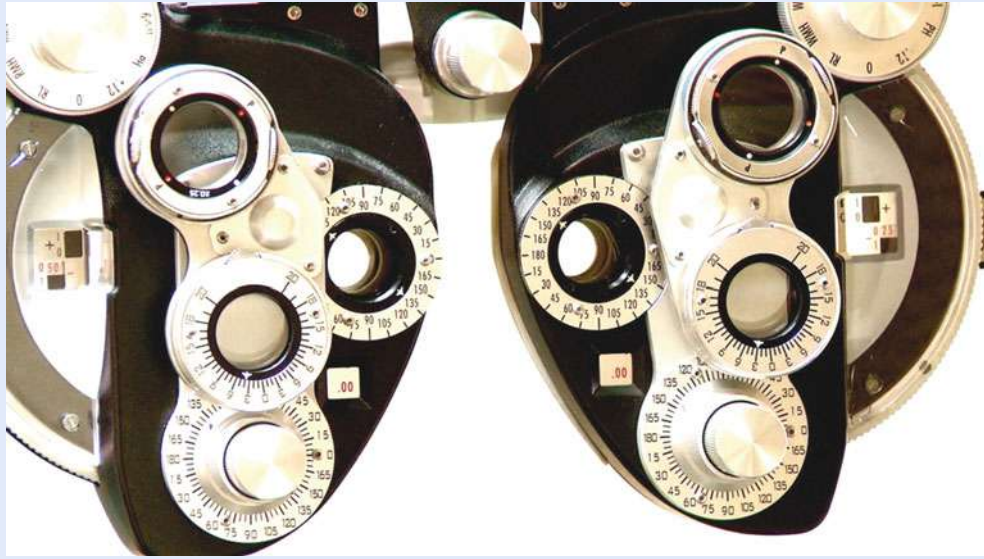
Fig. 13.19 A phoropter is commonly used by optometrists to determine the necessary corrective lens prescription

Fig. 13.20 A direct laryngoscope and its insertion into a human throat for examination

13.2.6 **Laryngoscope**

A *laryngoscope* is an optical instrument used for examining the interior of the larynx and structures around the throat. There are two types of laryngoscopes: *direct* and *indirect* laryngoscopes. A direct laryngoscope (Fig. 13.20) consists of a handle containing batteries, an integrated light source, and a set of interchangeable blades for easy reach and placement into a patient's throat. Besides being used for visualization of the glottis and vocal cords, a direct laryngoscope may also be used during surgical procedures to remove foreign objects in the throat, collect tissue samples (biopsy), remove polyps from the vocal cords, perform laser treatments and, very commonly, as a tool aid to facilitate tracheal intubation during general anesthesia or in cardiopulmonary resuscitation.

The blades in a laryngoscope help provide leverage to open wide the mouth and throat, as well as to keep the tongue in place and avoid a gag reflex. There are two basic styles of laryngoscope blades most commonly used: curved and straight. The Macintosh blade is the most widely used of the curved laryngoscope blades, while the Miller blade is the most popular style of straight blade. Blades come in different sizes, to accommodate different patients.

An *indirect* laryngoscope consists of a combination of a small mirror mounted at an angle on a long stem and a light source. The mirror is usually circular in form and made in various sizes, but is small enough to be placed in the throat behind the back of the tongue. The source of light is either a small bright lamp worn on the forehead of the observer, or a concave mirror, also worn on the forehead, for the purpose of concentrating light from some other source. Light is reflected to the back of the throat by the mirror and directed to illuminate up the interior of the larynx. The mirror also serves to reflect back to the doctor an image of the throat, to appreciate the structure of the glottis and vocal cords.

Some historians credit Benjamin Guy Babington (1794–1866), with the invention of the laryngoscope back in 1829 [20], who called his device the *glottiscope*. However, Manuel Garcia (1805–1906)—a Spanish tenor and singing maestro—experimented back in 1854 with a combination of throat mirror and light to observe the action of his own vocal cords and larynx when producing tones and sounds. His observations were published in the Royal Philosophical Magazine and Journal of Science in 1855 [21], and they constitute the first physiological records of the human voice as based upon observations in the living subject. For this, he is also recognized as the original inventor of the laryngoscope. ◘ Figure 13.21 shows a photograph and illustration of his original laryngoscope device.

Mirror-based laryngoscopy for the investigation of laryngeal pathology was pioneered back in 1858 by Johann Czermak, a professor of physiology at the University of Budapest. Czermak applied an external light source and a head-mounted mirror to improve visualization. During this period of time, a laryngoscopic examination was made as depicted in ◘ Fig. 13.22. The patient opens his mouth as widely as possible, protruding his tongue. The doctor, with a small napkin takes the protruded tongue between his thumb and forefinger and holds it in place, so as to enlarge opening of the mouth as much as possible. The laryngeal mirror is next inserted and dexterously positioned to the back of the mouth to direct the light from the external light source (mirror or lamp) into the back of the throat. An image of the lower throat is reflected back by the mirror for the doctor to view and assess the condition of the larynx.

All previous observations of the glottis and larynx had been performed under indirect vision (using mirrors) until 1895, when Alfred Kirstein (1863–1922) of Germany performed the first direct laryngoscopy in Berlin, using an *esophagoscope* he had modified for this purpose, calling device an *autoscope*, and the modern, direct laryngoscope was born [22].

13.3 Fiber Optic Medical Devices and Applications

The field of fiber optics has undergone a tremendous growth and advancement over the last 50 years. Initially conceived as a medium to carry light and images for medical endoscopic applications, optical fibers were later proposed in the mid-1960s as an adequate information-carrying medium for telecommunication applications. Ever since, optical fiber technology has been the subject of considerable research and development to the point that today light wave communication systems have become the preferred method to transmit vast amounts of data and information from one point to another.
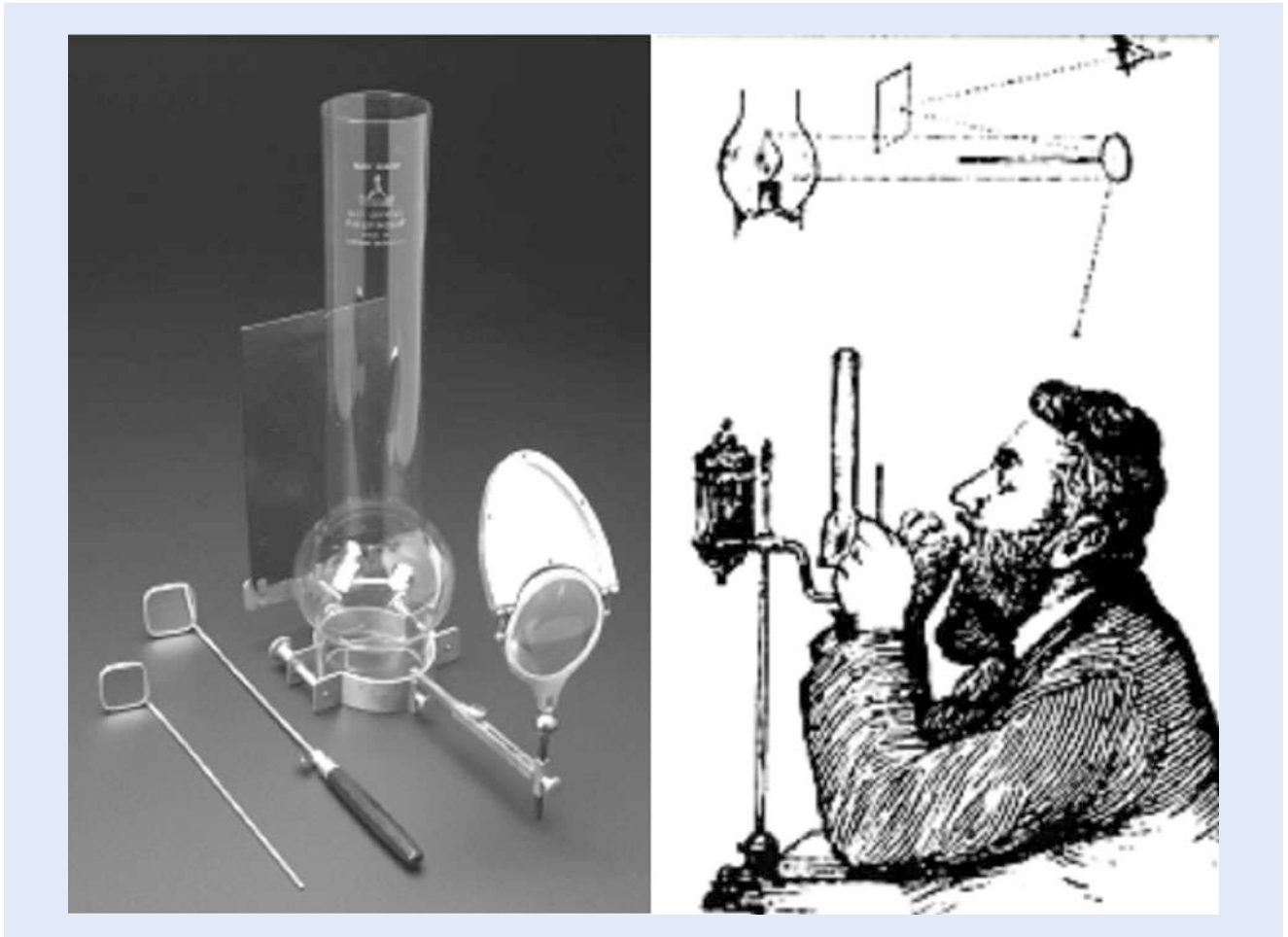
Fig. 13.21 Original indirect laryngoscope developed by Manuel Garcia to view the movement of his won vocal chords (c. 1870)

Given their EM immunity, intrinsic safety, small size and weight, autoclave compatibility and capability to perform multi-point and multi-parameter sensing remotely, optical fibers and fiberoptic-based devices are seeing increased acceptance and new uses for a variety of biomedical applications—from diverse endoscopes, to laser-delivery systems, to disposable blood gas sensors, and to intra-aortic probes. This section illustrates—through several application and product examples—some of the benefits and uses of biomedical fiber sensors, and what makes them such an attractive, flexible, reliable, and unique technology.

13.3.1 Optical Fiber Fundamentals

At the heart of this technology is the optical fiber itself. A hair-thin cylindrical filament made of glass (although sometimes are also made of polymers) that is able to guide light through itself by confining it within regions having different optical indices of refraction. A typical fiber structure is depicted in ◨ Fig. 13.23. The central portion—where most of the light travels—is called the core. Surrounding the core there is a region having a lower index of refraction, called the cladding. From a simple point of view, light trapped inside the core travels along the fiber by bouncing off the interfaces with the cladding, due to the effect of the total internal

Fig. 13.22 Nineteenth century illustration of a mirror-based laryngoscope examination of a patient's throat
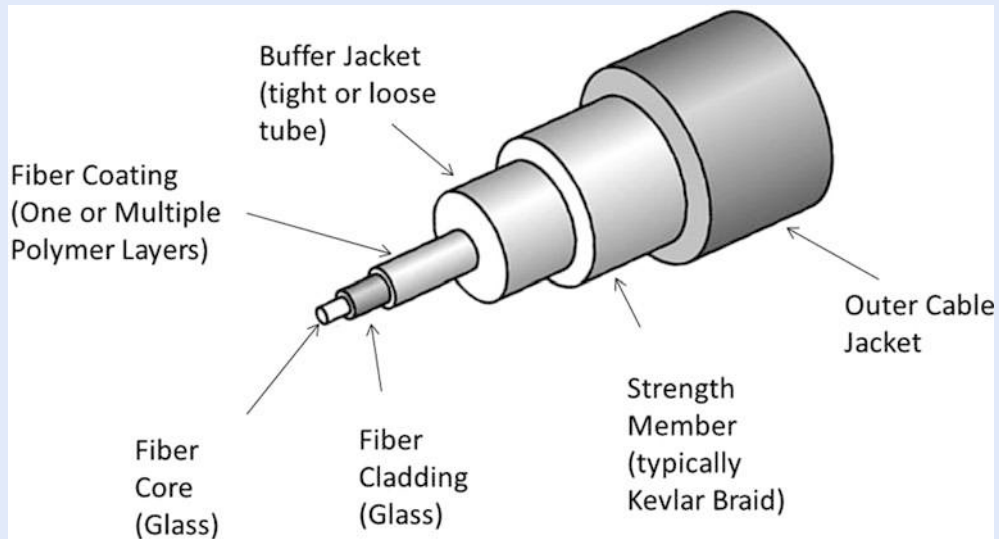
Fig. 13.23 Schematic of an optical fiber

13
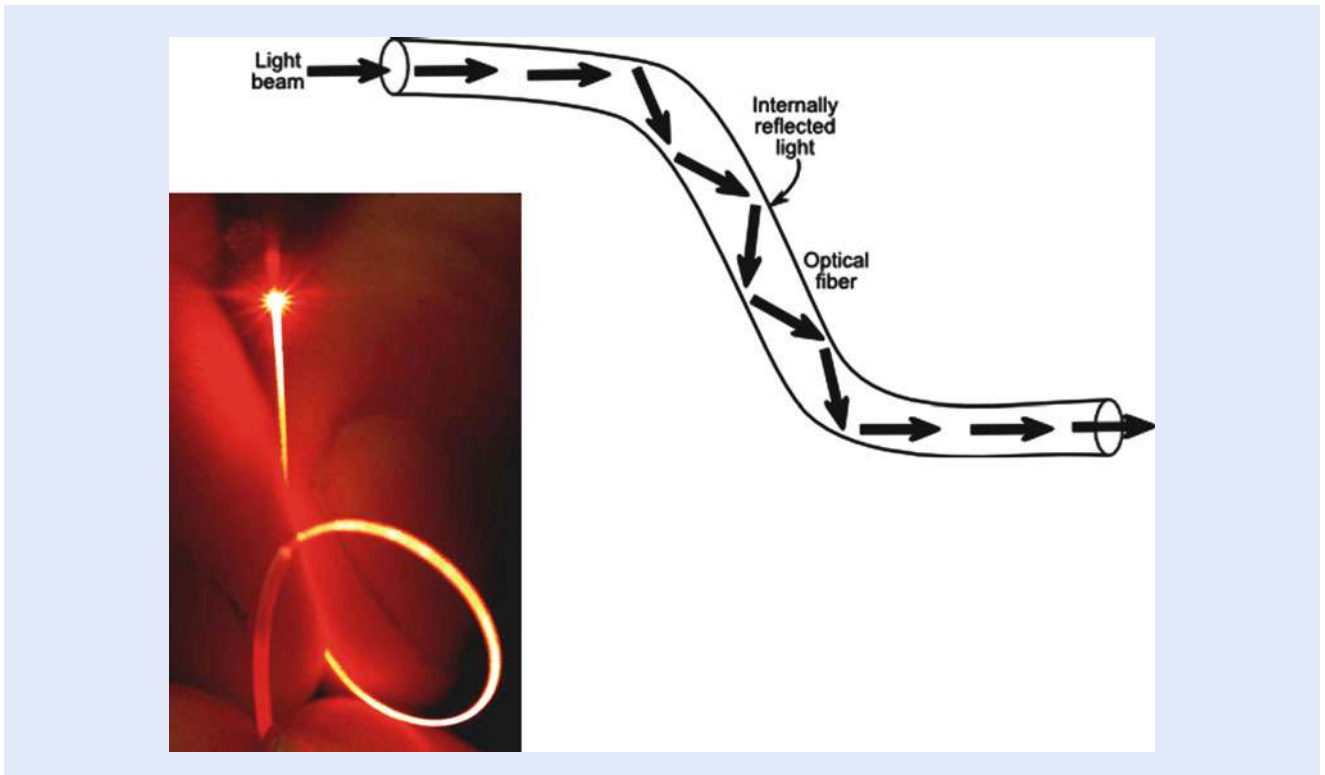
🔳 **Fig. 13.24** An optical fiber is able to guide light through the principle of total internal reflection. This allows the transmission of light energy (and signals) through any patch or shape taken by the optical fiber

reflection occurring at these boundaries (🔳 Fig. 13.24). In reality though, the optical energy propagates along the fiber in the form of waveguide modes that satisfy Maxwell's equations as well as the boundary conditions and the external perturbations present at the fiber.

Refraction occurs when light passes from one homogeneous isotropic medium to another; the light ray will be bent at the interface between the two media. The mathematical expression (Eq. (13.4)) that describes the refraction phenomena is known as Snell's law,

$$n_0 \sin \phi_0 = n_1 \sin \phi_1 \tag{13.4}$$

where n_o is the index of refraction of the medium in which the light is initially travelling, n_1 is the index of refraction of the second medium, Φ_o is the angle between the incident ray and the normal to the interface, and Φ_1 is the angle between the refracted ray and the normal to the interface.

🔳 Figure 13.25a shows the case of light passing from a high-index medium to a lower-index medium. Even though refraction is occurring, a certain portion of the incident ray is reflected. If the incident ray hits the boundary at ever-increasing angles, a value of $\phi_0 = \phi_c$ will be reached, at which no refraction will occur. The angle ϕ_c is called the critical angle. The refracted ray of light propagates along the interface, not penetrating into the lower-index medium, as shown in part 🔳 Fig. 13.25b. At that point, $\sin \phi_c$ equals to unity. For angles ϕ_0 greater than ϕ_c, the ray is entirely reflected at the interface, and no refraction takes place (see 🔳 Fig. 13.25c). This phenomenon is known as *total internal reflection*.

In 🔳 Fig. 13.26, a ray of light incident upon the end of the optical fiber at an angle θ will be refracted as it passes into the core. If the ray travels through the high-index medium at an angle greater than Φ_c it will reflect off of the cylinder
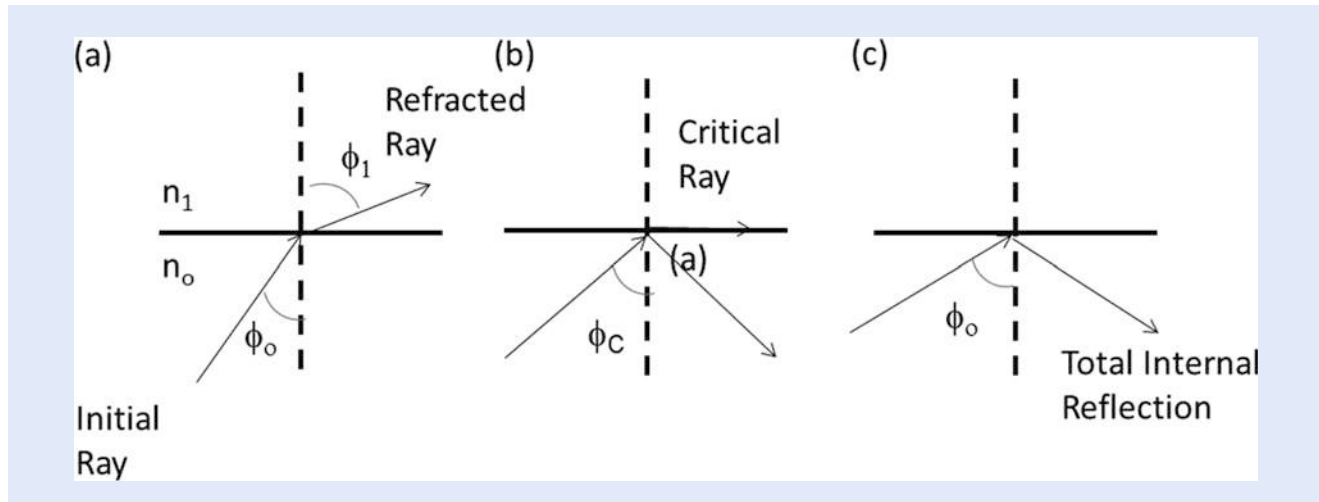
Fig. 13.25 Light reflection and refraction between two media with different indices of refraction
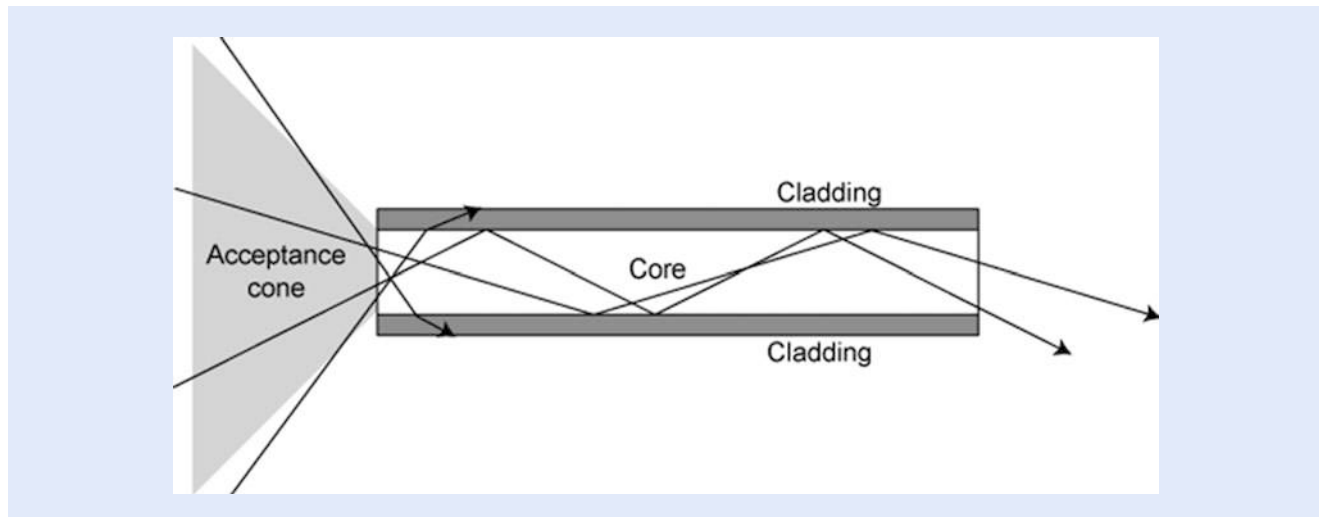
Fig. 13.26 Light ray propagation along the core of an optical fiber. At the entrance to the fiber, a conical region is defined by the so-called acceptance angle, which is the region in space where light can be effectively collected and coupled into the fiber for guiding

wall, will have multiple reflections, and will emerge at the other end of the optical fiber. For a circular fiber, considering only meridional rays, the entrance and exit angles are equal. Considering Snell's law for the optical fiber, core index n_0, cladding index n_1, and the surrounding media index n,

$$
\begin{aligned}
n \sin\theta &= n_0 \sin\theta_0 \\
&= n_0 \sin\left(\frac{\pi}{2} - \phi_c\right) \\
&= n_0 \left[1 - (n_1 - n_0)^2\right]^{1/2} \\
&= \left(n_0^2 - n_1^2\right)^{1/2} = \text{Numerical Aperture.}
\end{aligned}
\tag{13.5}
$$

The term $n\sin\theta$ is defined as the *numerical aperture* or NA for short. The NA is determined by the difference between the refractive index of the core and that of

13
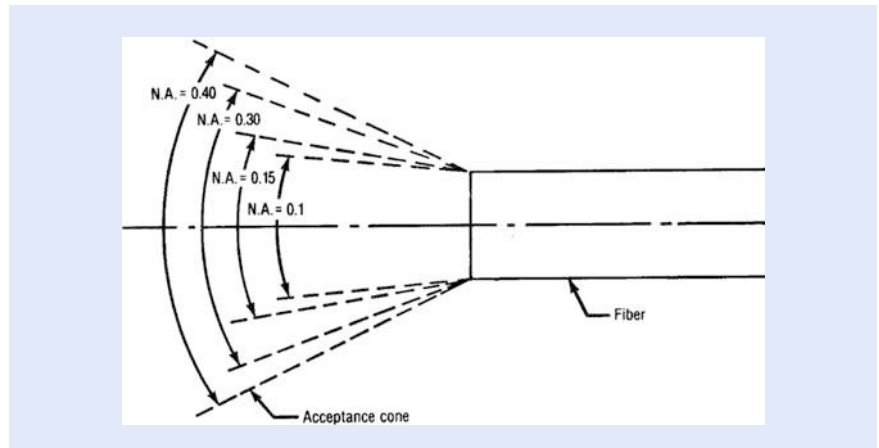
Fig. 13.27 The greater the NA of an optical fiber, the bigger the acceptance cone, and broader the angle of capture of light by the fiber

the cladding. It is a measure of the light-acceptance capability of the optical fiber. As the NA increases, so does the ability of the fiber to couple light into the fiber, as shown in Fig. 13.27. The larger NA allows the fiber to couple in light from more severe grazing angles. Coupling efficiency also increases as the fiber diameter increases, since the large fiber can capture more light. Therefore, the maximum light-collection efficiency occurs for large-diameter-core fibers and large-NA fibers.

13.3.2 Coherent and Incoherent Optical Fiber Bundles

In medicine, optical fibers have been considered for illuminating and imaging applications since the 1920s. Typically, a single glass optical fiber has a diameter ranging from 1 mm down to ~8 μm. However, a single optical fiber cannot transmit an image—only a bright light spot would be observed at its end. Hence, in order to carry a reasonable amount of light for illumination purposes, or to transmit and image, hundreds to thousands of optical fibers need to be assembled into bundles. Bundles of multiple single optical fibers of small diameter solid glass rods can thus be used to guide light or transmit images around bends and curved trajectories.

Glass optical fiber bundles are of two types: *incoherent* and *coherent*. An *incoherent* bundle consists of a collection of fibers randomly distributed in the bundle and is typically intended for illumination purposes only. In contrast, a *coherent* optical fiber bundle has an ordered array of fibers in which the relative position of each individual fiber at its input and output with respect to the bundle is maintained. That is to say, the position of individual fibers is at same locations over the cross section of both bundle ends as depicted in Fig. 13.28. In between the ends, the fibers need not have a fixed orientation and can move flexibly. Coherent bundles are used for conveying an image from one end to the other by the effect created by the grouping of the individual light conducted by each fiber which is perceived in the eye of the observer as a full image. To achieve better image quality and resolution, a large number of small diameter fibers are need for a given bundle diameter. Typically, fibers used in bundles have diameters on the order of 8–12 μm and their count can range from about 2000 up to 40,000 [23]. In the case of imaging bundles, larger diameter fibers are used of 30–50 μm in diameter.
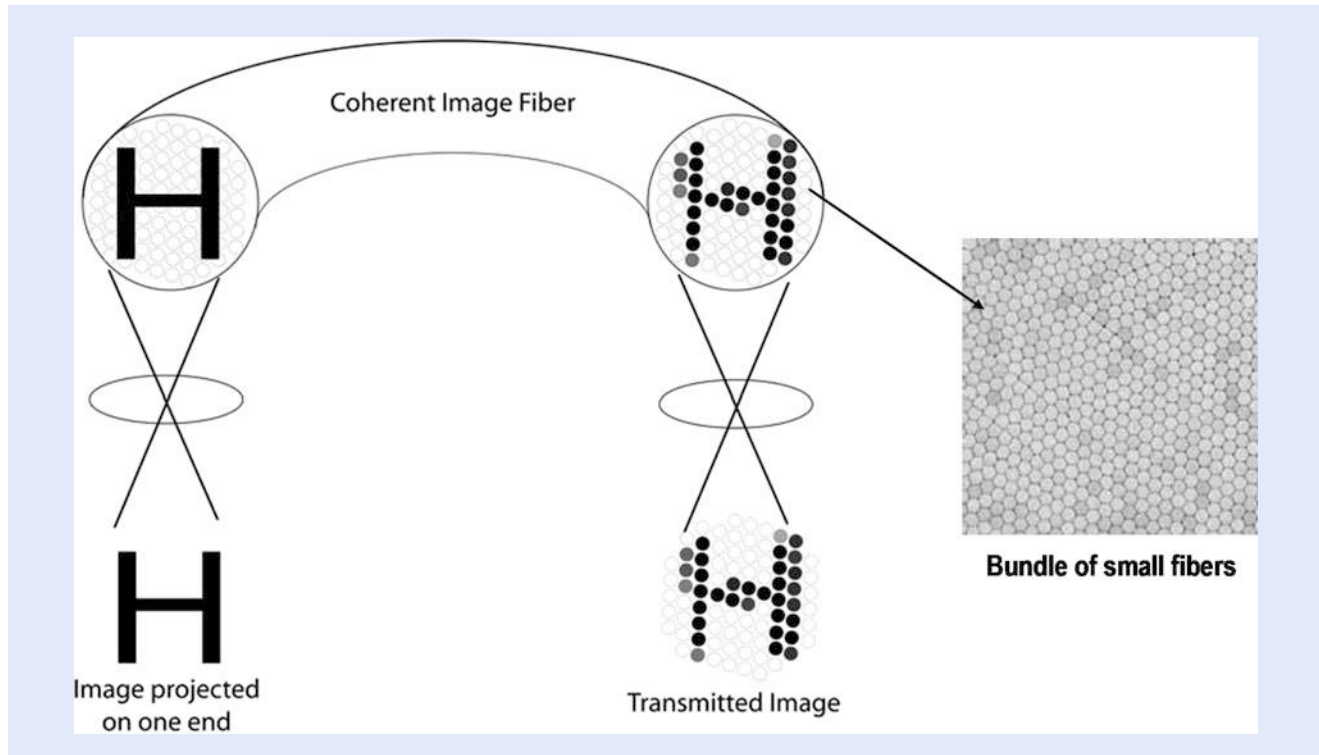
Fig. 13.28 A coherent optical fiber bundle. Images are accurately transmitted by preserving the relative position of the fibers at each end of the bundle. The bundle consists of a multitude of individual glass fibers of a small diameter (~12 μm) that create a lattice effect

Fabrication of illumination (non-coherent) and imaging (coherent) bundles is based on the same processes of drawing optical fibers or glass rods through heating furnaces and doing repeated draws of multi-stack sets, to achieve arrays with the desired quantity of fibers of the appropriate diameter. There are three common to fabrication methods for coherent bundles: fused image bundles, wound image bundles, and leached image bundles. ◘ Figure 13.29 illustrates the three steps needed to fabricate fused as well as leached image fiber bundles.

An individual fiber (or rod) is made by starting with a so-called perform made by the tube-in-rod technique where a single glass rod (which will become the fiber's core) is inserted into a tube made of glass with a lower refractive index (cladding). In the case of a leached bundle, an additional glass jacket made of a leachable glass is used. This glass perform is placed in an electric heating furnace that runs at a temperature close to the softening point of the glass. The heat causes the solid glass road to soften. Once soft, the glass is pulled down into a thin filament by a pulling mechanism. The final diameter of the filament is controlled by the ratio of the speeds between the advancing preform and the drawn fiber. Typically, the initial drawn fiber is more of a solid rod with a 2 mm diameter. In the next drawing stage, a multitude of mono fibers are stacked together and drawn in the furnace to produce a multi-fiber rod. The drawn filament from a multi-fiber preform consists of several 100 monofilament fibers. In the third stage, several multi-fiber rods are stacked together to perform the so-called multi-multi drawing process. The multi-multi stack assembly is fed through the furnace and drawn into a filament of rod of the desired diameter. Such filament will be composed of thousands of individual glass fibers. As shown in ◘ Fig. 13.30, imaging multi-fiber arrays of square, circular, or hexagonal shape and in different sizes can be fabricated with this process.
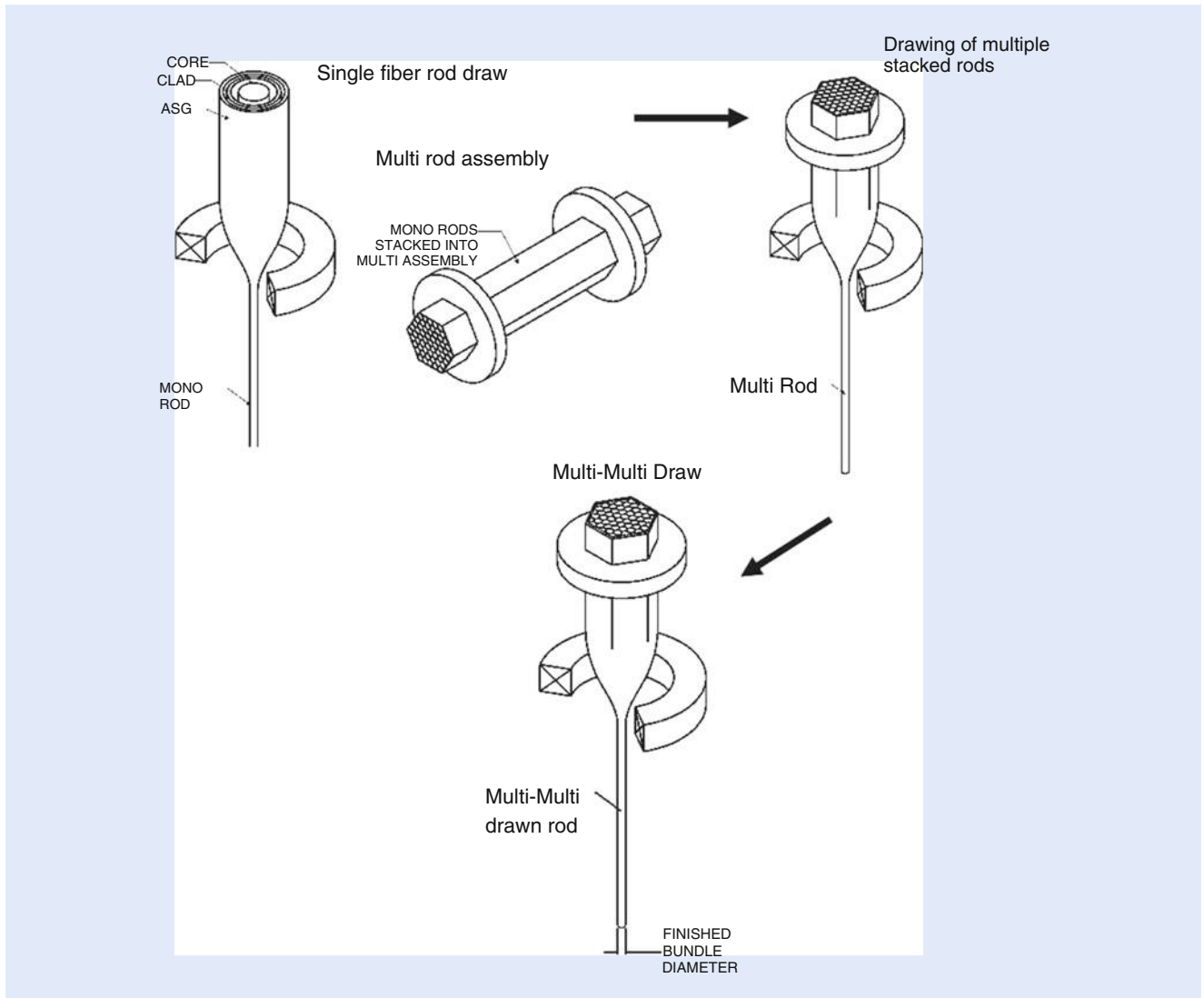
Fig. 13.29 Fabrication process to make fiber bundles

In the particular case of leached fiber bundles, each bundle end is properly secured and the entire bundle is soaked in an acid solution which will dissolve the leachable glass, allowing the fibers to move freely between the bundle ends.

Wound imaging bundles are made by winding a multi-fiber array as a single layer on a drum, and then stacking the desired number of layers manually in a laminating operation.

13.3.3 Illuminating Guides

Fiber optic illuminating guides are non-coherent and are used primarily to guide light to a desired point to provide illumination and enhance visual clarity. Imaging bundles are typically made of 30–50 μm diameter fibers, with NA values around 0.6. Most commonly, illuminating bundles are used as part of fiberscopes, endoscopes, and personal lights for surgeons. As seen in ▢ Fig. 13.31, when surgeons are operating on a patient, they need cool, bright light to help then see better tissues and organs—the closer the direct illumination to the operating field,
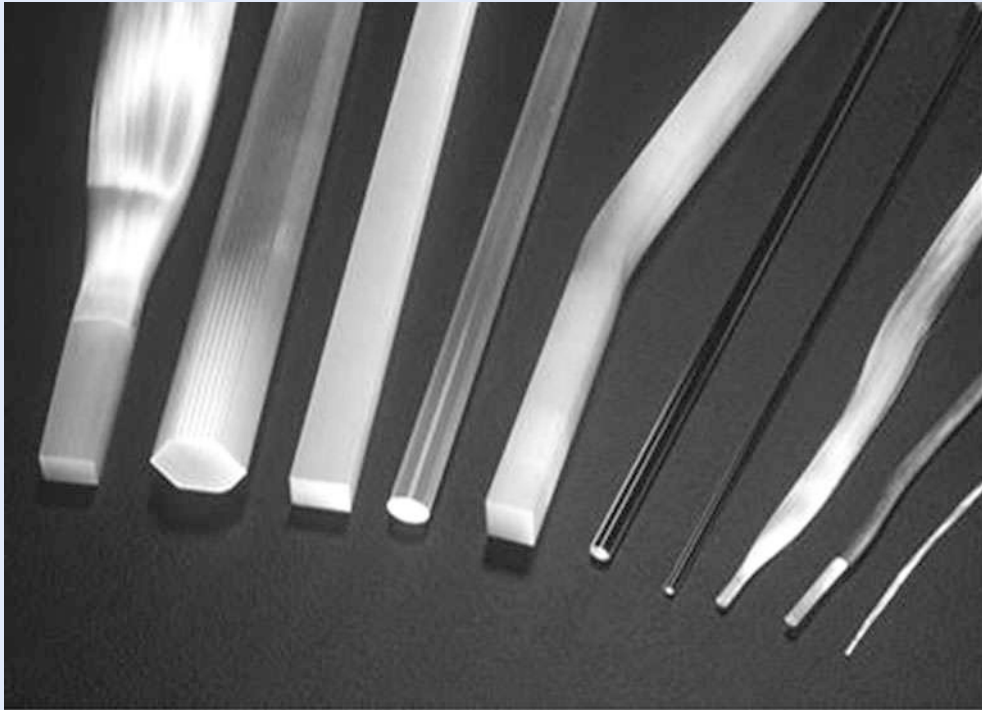
◨ **Fig. 13.30** Photograph of different styles and shapes of drawn coherent optical fiber bundles

the better. Rigid, light-guiding rods are also made (from single solid glass rods or from multi-core rods) for applications in dentistry and light therapy (see ◨ Fig. 13.32).

13.3.4 Fiberscopes and Endoscopes

An endoscope is an optical instrument used for direct visual inspection of hollow organs or body cavities. Typically, an endoscope is generally introduced through a natural opening in the body (◨ Fig. 13.33), but it may also be inserted through an incision. Instruments for viewing specific areas of the body include the broncho-scope, colonoscope, cystoscope, gastroscope, laparoscope, proctoscope, and several others. Although the design may vary according to the specific use, all endoscopes have similar construction and elements: an objective lens (distal end), illuminating fiber bundle, imaging coherent fiber bundle, fixed or articulating handle, and an eyepiece (proximal end). Accessories that might be used for diagnostic or thera-peutic purposes include irrigation channels, suction tips, tubes, and suction pump; forceps for removal of biopsy tissue or a foreign body; biopsy brushes; an electrode tip for cauterization; as well as a video camera, video monitors, and image recorder. Many modern endoscopes have also articulating ends, that are remotely controlled by the doctor using knobs on the handle that adjust pull wires inside the body of the endoscope. ◨ Figure 13.34 shows a modern, flexible, and fiber-optic endoscope.

Endoscopes can be rigid or flexible as depicted in ◨ Fig. 13.35. Modern endoscopes (both flexible and rigid) make use of fiber optic imaging bundles to achieve image transmission. However, earlier models relied on miniature flat or rod lenses to guide images from the objective end to the eyepiece as shown in ◨ Fig. 13.36. Hippocrates II (460–377 BC) reported using catheters and primitive

13

⬛ **Fig. 13.31** Fiber optic illuminators used in the operating room by surgeons

⬛ **Fig. 13.32** Solid fiber optic illuminating rods

forms of visualization tubes over two millennia ago. In the nineteenth century, endoscopy was very rudimentary and relied on the insertion of long, rigid metal tubes into body cavities. In 1910 Victor Elner used a gastroscope to view the stomach, while in 1912 the first semi-flexible gastroscope was developed. Then, Heinrich Lamm was the first person to transmit images through a bundle of optical
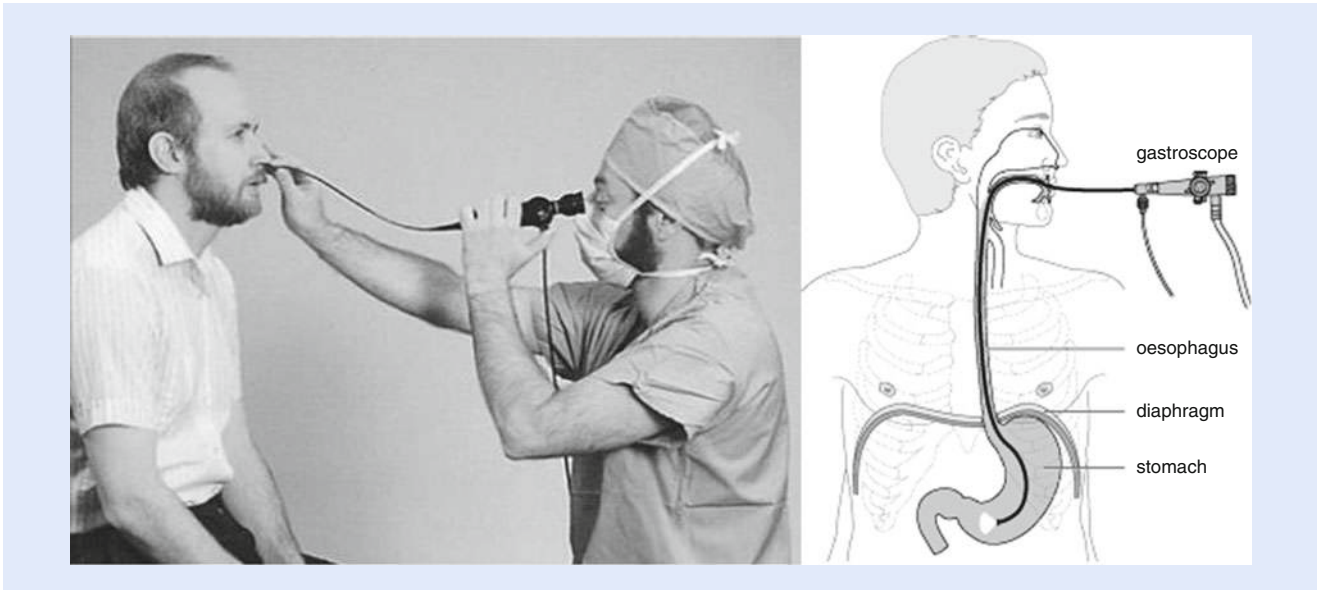
🔹 **Fig. 13.33** Endoscopes are commonly used by doctors to inspect patients' internal organs through body cavities, such as the nose or throat
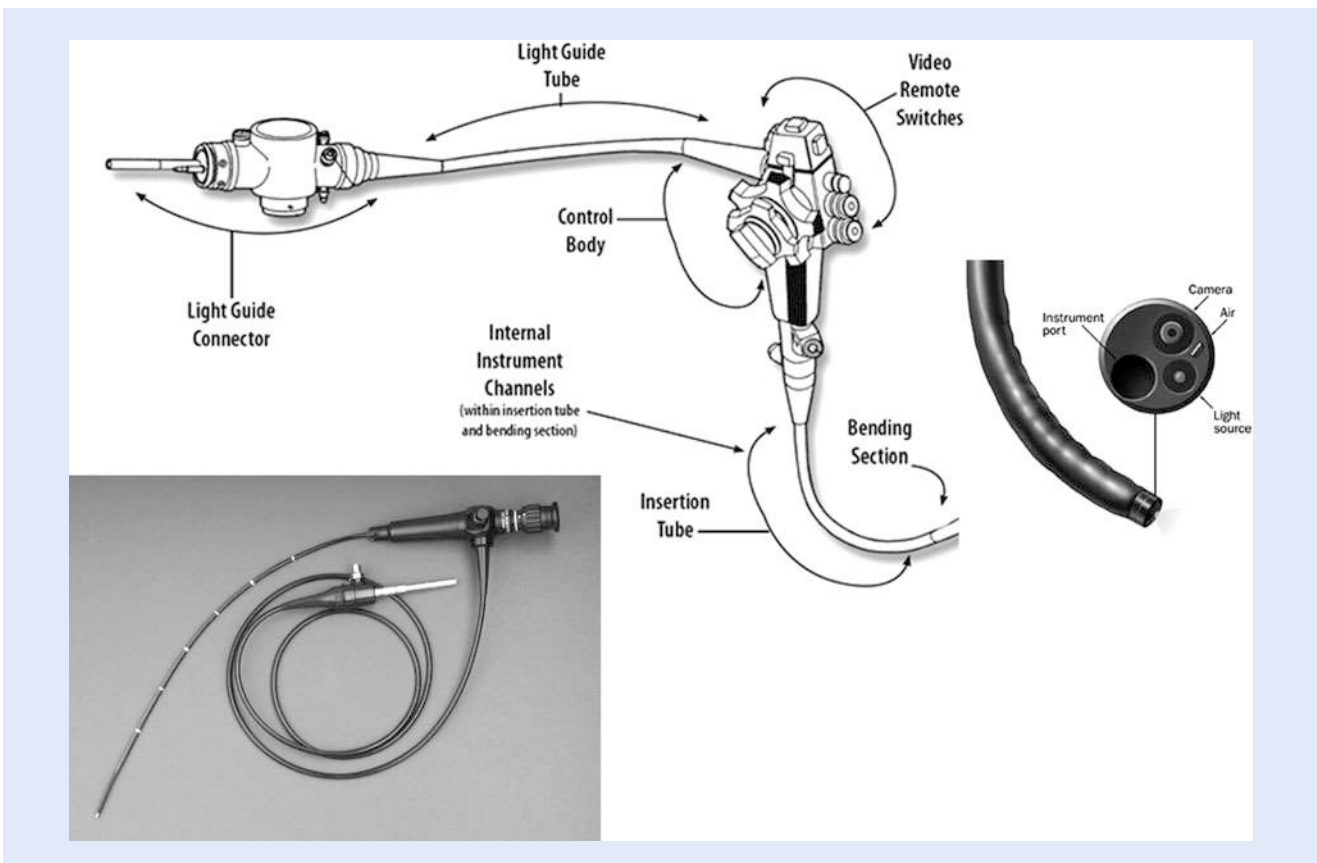
🔹 **Fig. 13.34** Aspect of a modern, flexible fiberscope fitted with articulating knobs, camera lens, and instrument port on the distal end

Fig. 13.35 Examples of a rigid (*upper*) and flexible (*lower*) endoscopes
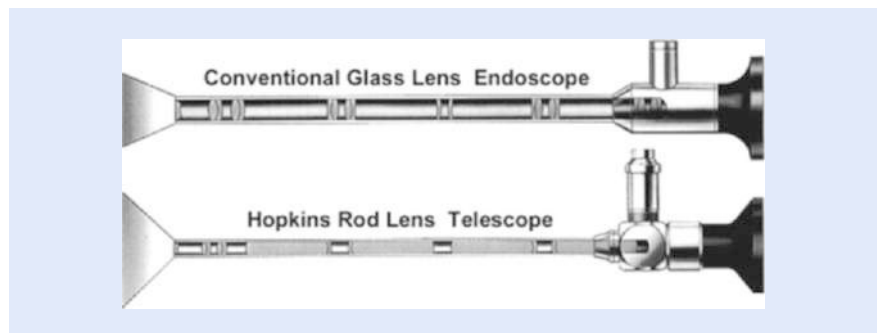
Fig. 13.36 Early designs of rigid endoscopes using rod or glass lenses for image relaying

fibers in 1930. In 1957, clad optical fibers were first proposed and developed by Lawrence Curtis as a graduate student at the University of Michigan, under the supervision of Dr. Basil Hirschowitz, who in 1957 demonstrated the first fiber optic endoscope [24]. From then on, the devices became known as *fiberscopes*. The fiberoptic endoscope has great flexibility, reaching previously inaccessible areas and has become the norm in medicine.

13.3.5 Fused Fiber Faceplates and Tapers for Digital X-rays

Another type of coherent imaging conduit is the fiber optic fused faceplate (FOFP). FOFPs are made as pre-arranged blocks of multiple pre-drawn multi-fiber glass rods (known as boules), which are then fused together under elevated heat and pressure to form a solid piece (Fig. 13.37). Typical individual fiber element sizes range from as small as 4 to 25 μm or larger. Thin plates are then sliced from the fused boule, ground and polished to the desired thickness—ranging from ~100 mm down to a practical limit of 50 μm. Typical shapes are round or rectangular. Depending on the intended application, the FOFP end faces can be coated with a specific spectral filtering, phosphorescent, or anti-reflective coating.

◘ **Fig. 13.37** A fused fiberoptic faceplate (*back*) and taper (*front*). Photo courtesy of Schott glass

Optically, an FOFP behaves as zero-thickness optical window transferring an image, fiber by fiber, from one face of the plate to the other. Image magnification or reduction can be achieved by tapering the cross section of the bulk plate during the manufacturing process. In this case, the boule is drawn down and a neck region is formed with an hour-glass shape piece. The piece is cut into two pieces, machined and the ends polished resulting in a fused fiber optic taper.

Faceplates and tapers also function as dielectric barrier and mechanical interface and are optically used as a two-dimensional image conduit for energy conversion, field-flattening, distortion correction, and contrast enhancement. They are typically used for imaging applications bonded to cathode ray tubes (CRT) and LCD displays, image intensifiers, charged coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) detectors, image plane transfer devices, X-ray digital detectors, among others.

In the medical area, fiber optic tapers and faceplates have found widespread use for both dental and medical digital radiography (such as mammography, fluoroscopy, intra-oral, panoramic, or cephalometric) where instead of using conventional film to obtain the X-ray images, an electronic photosensitive device such as a CCD or CMOS detector chip is used to convert the X-ray energy into electronic pixel signals via the use of an intermediate faceplate. Digital radiography offers high-resolution images while greatly reducing patient and sensor exposure to harmful X-rays by using low-dose X-ray sources. In addition, digital X-ray imaging speeds the availability of images for diagnostic, while also making the viewing, sharing, transmitting, and storing of X-ray patient data so much easy and compatible with modern electronic record systems. Furthermore, faceplates also provide a critical X-ray absorbing barrier between the X-ray emitter and the semiconductor detector device, prolonging their service life and reducing background noise.

As shown in ◘ Fig. 13.38, when an X-ray source emits radiation energy (that would pass through the patient) the transmitted energy impinges on a scintillator plate which converts the radiation rays into visible photons. The scintillating coating—e.g., cesium iodide (CSI) or gadolinium oxysulfide (Gadox) doped with Tl or Eu—is deposited directly on the large end of a fused fiber-optic taper. The light is then transferred and reduced through the taper and coupled to a digital CCD chip where a black and white image is formed which can then be viewed on a computer screen or monitor and readily archived as an electronic image file.

13

13.4 Conclusions

As discussed in this section, optics is a useful, practical, versatile, and powerful technology that, throughout history, has helped human kind perform visual examination, diagnostics, and therapeutics on both the sick and healthy. Optics technology and optical components are at the core in a variety of modern-day optical devices and instruments such as endoscopes, patient monitoring probes, and sensors, as well as in advanced robotic assisted surgery systems.

The harnessing power of light, and its interaction with living matter, is extremely useful and beneficial for a variety of medical purposes and treatments ranging from laser procedures for tattoo removal, to eye surgery to vessel and tissue ablation and coagulation, up to modern photodynamic therapy treatments. We have seen how the field of optics is in itself a subset of a more complex and interdisciplinary area of research known as Biophotonics.

New advancements in optics and photonics are driving the development of a new generation of imaging tools—such as optical coherence and photo-acoustic tomography—that can readily provide two and three-dimensional images of diverse human body tissues and organs.

Optics has, and will continue to be, an *enabling technology* for the advancement of medicine promoting unimaginable new devices, techniques, and applications to happen in the not too distant future.

References

1. Katzir A (1990) Selected papers on optical fibers in medicine, SPIE milestone series, vol MS 11. SPIE, Bellingham, WA
2. United Nations DESA (2015) World population prospects: the 2015 revision. United Nations, New York
3. Bradbury S (1967) The evolution of the microscope. Pergamon Press, Oxford
4. Bozzini P (1806) Lichtleiter, eine Erfindung zur Anschauung innerer Teile und Krankheiten. J Prakt Heilkunde 24:107–124
5. Von Helmholtz HLF (1856) Handbuch der physiologischen Optik. L. Voss, Leipzig
6. History of ASLMS. American Society for Laser Medicine and Surgery. ► http://www.aslms.org/aslms/history.shtml
7. Choy DS (1988) History of lasers in medicine. Thorac Cardiovasc Surg 36(Suppl 2):114–117
8. Mignani AG, Baldini F (1979) Fibre-optic sensors in health care. Phys Med Biol 42(5):967–979
9. Fujimoto JG et al (2002) Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. Neoplasia 2(1–2):9–25
10. Cusano A, Consales M, Crescitelli A, Ricciardi A (eds) (2015) Lab-on-fiber technology. Springer, Berlin
11. Lakshminarayanan V et al (2015) Smartphone science in eye care and medicine. Opt Photonics News 26:45–51, Optical Society of America
12. Maamari RN et al (2014) Novel telemedicine device for diagnosis of corneal abrasions and ulcers in resource-poor settings. JAMA Ophthalmol 132:894–895
13. Breslauer DN et al (2009) Mobile phone based clinical microscopy for global health applications. PLoS One 4, e6320
14. Hett JH (1969) Medical optical instruments. In: Kingslake R (ed) Applied optics and optical engineering, Vol. 5, optical instrument. Academic, New York
15. Devroey D et al (2000) Do general practitioners use what's in their doctor's bag? Scand J Prim Health Care 20:242–243
16. Kravetz RE (2002) The otoscope. Am J Gastroenterol 97:470
17. Feldmann H (1996) History of the ear speculum. Images from the history of otorhinolaryngology, highlighted by instruments from the collection of the German Medical History Museum in Ingolstadt. Laryngorhinootologie 75:311–318
18. Millodot M (2009) Dictionary of optometry and visual science, 7th edn. Butterworth-Heinemann, Edinburgh
19. Millodot M (1973) A centenary of retinoscopy. J Am Optom Assoc 44:1057–1059
20. Koltai PJ, Nixon RE (1989) The story of the laryngoscope. Ear Nose Throat J 68(7):494–502
21. García M (1855) Observations on the Human Voice. Proc R Soc Lond 7:399–410
22. Hirsch NP, Smith GB, Hirsch PO (1986) Alfred Kirstein Pioneer of direct laryngoscopy. Anaesthesia 41(1):42–45
23. Kawahara I, Ichikawa H (1987) Fiberoptic Instrument technology. In: Sivak MV (ed) Gastrointestinal endoscopy. W. B. Saunders, Philadelphia, pp 201–241
24. Hirschowitz B (1979) A personal history of the fiberscope. Gastroenterol 76;(4):864–869

Quantum Optics

Contents

Atom Optics in a Nutshell

Pierre Meystre

P. Meystre (✉)
Department of Physics and College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA
e-mail: pierre.meystre@optics.arizona.edu

14

14.1 Introduction

One of the most counter-intuitive aspects of quantum mechanics, the fundamental theory of nature that was developed starting in the early twentieth century, is the concept of wave-particle duality.

We are all familiar with the notions of waves and particles. We have observed water waves when throwing pebbles in ponds as children, and we have learned in high-school that sound and light consist of waves as well. What we may not have learned, though, is that light sometimes behaves as particles instead. Even more unsettling is the fact that atoms, and all massive particles for that matter, sometimes behave as waves.

Wave-particle duality is the central tenet of atom optics: since atoms, very much like light, behave sometimes as waves, sometimes as particles, it is possible in principle to do with them pretty much everything that can be done with light. One can build a broad variety of atom optical instruments such as mirrors, beam splitters, and lenses. One can develop techniques for imaging, microscopy, diffraction, interferometry, and more. One can even realize atom analogs of lasers.

This chapter sketches selected aspects of atom optics, a few of its recent developments, and some of its promise. We start with a brief historical overview of some of the milestones that have lead to our current understanding of atoms, from the Greek philosophers of antiquity to the development of quantum mechanics and the key experiments of the early twentieth century that confirmed the wave-particle duality of atoms and other massive particles. We then discuss how the wave nature of atoms becomes increasingly more evident as their temperature is decreased. For that reason it is oftentimes advantageous to work at extremely low temperatures, close to absolute zero, in applications such as atom microscopes and atom interferometers, or to build atom optics "analogs" of the laser, Bose–Einstein condensates. After outlining the basic ideas behind these devices we conclude with a brief overview of some current and future applications, with an emphasis on the role of atom optics in helping answer fundamental physics questions.

The bibliography is limited to a few milestone papers and is certainly not meant to be comprehensive. It also does not attempt to give proper credit to all research groups who have contributed significant advances to atom optics, sometimes within weeks of the work by research groups mentioned here. Due to their advanced technical content these papers will likely be of limited use to the casual reader beyond their historical interest. The excellent review [6] gives a comprehensive list of references through 2009. The elegant set of lecture notes of [18] also discusses some of the extraordinary promise of atom optics for tests of fundamental physics at a level appropriate for advanced graduate students and experts in the field.

14.2 Particles or Waves?

This short chapter is not the place to give a comprehensive review of our historical understanding of the nature of light. For our purpose it is sufficient to review a few of the key steps that resulted in that understanding. We then draw a similar sketch of the historical development of our understanding of atoms. This will set the stage for a discussion of the close parallels that have guided the development of atom optics.

The central idea that we will need to become somewhat comfortable with is the concept of "wave-particle duality," the co-existence of particle and wave properties in objects that we are used to think of as one or the other, but not both. From

everyday experience we are quite familiar with particles and waves, for instance, from watching the surf rolling on a beach, or tiny grains of sand on that same beach. What we need to grasp, though, is the rather counter-intuitive concept of both light and atoms behaving sometimes as particles, and sometimes as waves. *Why* this is the case is not a question that physics answers—it is a question perhaps best left to philosophers—but *how* this is the case is something that we now understand well. This is described beautifully and with extraordinary predictive power by modern quantum physics.

14.2.1 Light

Perhaps a good place to start is with the great Greek philosophers and mathematicians Pythagoras (c. 570–c. 495 BC), Plato (c. 428–c. 348 BC), and Euclid (c. 325–c. 265 BC). They thought that light consists of rays that travel in straight lines from the eye to the object, and that the sensation of sight is obtained when these rays touch the object, much like the sense of touch. Plato's student Aristotle (384–322 BC), though, had a different theory, considering instead that light travels in something like waves rather than rays. The understanding that light travels from the eye to the object remained largely unchallenged until it was finally disproved more than a thousand years later by Alhazen (965–1039), one of the earliest to write and describe optical theory. He studied in particular light and the nature of vision with the combined use of controlled experiments and mathematics.

Meanwhile the debate between the corpuscular and the wave nature of light already apparent in the conflicting views of the Pythagorean School and Aristotle continued unabated for centuries. Isaac Newton (1642–1726), who performed numerous experiments on light toward the end of the seventeenth century and whose extraordinary contributions include the understanding of the color spectrum and of the laws of refraction and reflection, argued that those effects could only be understood if light consisted of particles, because waves do not travel in a straight line. However, the corpuscular theory failed to explain the double-slit interference experiments carried out by Thomas Young (1773–1829)—we will return to these experiments at some length later on. It was replaced in the nineteenth century by Christiaan Huygens' (1629–1695) wave theory of light. Finally, James Clerk Maxwell (1831–1879) developed the equations that unify electricity and magnetism in a theory that describes light as waves of oscillating electric and magnetic fields. This is the culmination of the classical theory of light, and one of the greatest, if not the greatest achievement of nineteenth century physics. At that point, it appeared that light was indeed formed of waves, and the corpuscular theory seemed ruled out once and for all.

However things changed again at the beginning of the twentieth century in a way that revolutionized physics and profoundly transformed our understanding of nature. In 1900, Lord Kelvin gave a celebrated talk entitled "Nineteenth Century Clouds over the Dynamical Theory of Heat and Light" in which he stated with remarkable insight that [16]

» The beauty and clearness of the dynamical theory, which asserts heat and light to be modes of motion, is at present obscured by two clouds.

He went on to explain that the first of these two clouds was the inability to experimentally detect the "luminous ether"—the medium that was thought to be vibrating to create light waves; and the second was the so-called ultraviolet catastrophe of blackbody radiation—the fact that Maxwell's theory utterly failed to predict the amount of ultraviolet radiation emitted by objects as a function of

their temperature. As it turns out, these two clouds led to two earthshaking revolutions in physics: relativity theory and quantum mechanics.

Quantum mechanics reopened the centuries-old wave-particle debate, but it resolved it with a very unexpected and dramatic new answer: light behaves sometimes as waves, and sometimes as particles. In trying to understand how the radiation emitted by an object depends on its temperature, Max Planck (1858–1947) advanced the revolutionary idea that energy comes up in tiny discrete lumps, or quanta. With this ad hoc assumption, he was able to explain the experimental data that Maxwell's theory failed to explain. Following on that work, Albert Einstein (1879–1955) proposed that light also comes in small lumps of energy, now called photons. This allowed him to correctly characterize how electrons are emitted from surfaces of metal irradiated by light. (The other cloud mentioned by Lord Kelvin, the absence of a luminous ether, leads to Einstein's theory of relativity.)

What modern quantum theory teaches us is that light sometimes behaves as waves, as in the Young double-slit interference experiment, and sometimes as particles, as in the photoelectric effect. Although photons are massless they carry both energy and momentum that can be used to alter the motion of massive objects. We will talk about all this quite a bit more in this chapter, but first let's turn for a moment to what quantum theory has to say about atoms.

14.2.2 Atoms

As with light, a good place to start is again in ancient Greece. This is where Democritus (c. 460–370 BC) and Leucippus (fifth century BC) developed the theory of atomism, the idea of an ultimate particle and that everything was made out of indivisible "atoms." The first experiments that showed that matter does indeed consist of atoms are due to John Dalton (1776–1844). He recognized the existence of atoms of elements and that compounds were formed from the union of these atoms, and put forward a system of symbols to represent atoms of different elements. [The symbols currently used were developed by Jöns Jacob Berzelius (1779–1848).] In a major breakthrough, in 1897 J.J. Thompson (1856–1940) discovered the electron and advanced the so-called plum pudding model of the atom. In that description the volume of the atom was composed primarily of the more massive (thus larger) positive portion (the plum pudding), with the smaller electrons dispersed throughout the positive mass like raisins in a plum pudding to maintain charge neutrality.

Early in the twentieth century Ernest Rutherford (1871–1937) carried out a number of experiments that suggested that atoms consist instead of a tiny, positively charged nucleus, with electrons orbiting around it at relatively large distance, and discovered the existence of positively charged protons. In 1920 he further proposed the existence of the third atomic particle, the neutron, whose existence was experimentally confirmed quite a bit later, in 1932, by James Chadwick (1891–1974). Rutherford's experiments led to the development of the so-called Bohr–Sommerfeld model of the atom.

In a groundbreaking development and with extraordinary insight, Prince Louis-Victor de Broglie (1892–1987) then postulated that if light exists both as particle and wave then atoms, and all massive particles, should be the same (◘ Fig. 14.1). This was the key missing piece of the puzzle [9]. This property, now known as the *wave-particle duality*, guided Erwin Schrödinger (1887–1961), Werner Heisenberg (1901–1976), and others in developing quantum mechanics, the theory that has led to a wealth of extraordinary inventions from the internet to cell phones, from GPS to medical imaging, and to a myriad other developments that impact just about every aspect of modern life.

🔲 **Fig. 14.1** Prince Louis-Victor de Broglie, who came up with the idea that massive particles have a wave character

The first experiments confirming the wave nature of massive particles were carried out by Clinton J. Davisson (1881–1958) and Lester H. Germer (1896–1971), who in 1928 observed the diffraction of electrons by a crystal of Nickel [8]. The first experiments demonstrating the wave nature of atoms and molecules followed soon thereafter, in 1930, in Helium experiments performed by Immanuel Estermann (1900–1973) and Otto Stern (1888–1969) [10], thereby fully confirming the de Broglie hypothesis.

To briefly complete the story as we currently understand it, we now know that protons and neutrons are actually not elementary particles. They belong to a family of particles called baryons, made up of three elementary constituents called quarks (two "up" quarks and one "down" quark for the proton, and one "up" quark and two "down" quarks for the neutron) bound together by the nuclear force. Together with another family of particles called mesons, made up of two quarks, they form the hadrons family.[1] The electrons, by contrast, are believed to be true elementary particles and belong to a family called leptons. They interact with the atomic nuclei via the electromagnetic force, whose "force particle" is the photon. The Standard Model of elementary particle physics comprises two additional types of interactions: the weak interaction, responsible for radioactive decay and nuclear fission, and gravitation, which allows massive particles to attract one another in accordance with Einstein's theory of general relativity (🔲 Fig. 14.2).

However, to break atoms into their subatomic constituents requires very large energies, much larger than normally considered in atom optics experiments. For the purposes of this chapter it is therefore sufficient to consider atoms as essentially "elementary particles" that interact with each other via relatively weak electric and magnetic fields, most importantly for us with light fields.

14.2.3 Particles and Waves

By optical waves one usually means those electromagnetic waves that are visible to the human eye. They have very short wavelengths, of the order of a millionth of a meter or less, and very high frequencies, of the order of 100,000 billion of oscillations

---

1   In an exciting new development announced in summer 2015, experiments carried out at the CERN Large Hadron Collider near Geneva, Switzerland provided evidence for the existence on pentaquarks, a new type of hadrons consisting of five quarks.

◘ Fig. 14.2 Clinton J. Davisson and Lester H. Germer (Lucent Technologies Inc./Bell Labs, courtesy AIP Emilio Segrè Visual Archives)

per second.[2] Blue light consists of waves of higher frequency and shorter wavelength than green light, and green light consists of waves of higher frequency and shorter wavelength than red light. Past blue light and toward shorter wavelengths blue is followed by ultraviolet light, X-rays, and gamma rays. These waves are invisible to the human eye. On the other side of the spectrum and moving toward longer wavelengths, red is followed by infrared, microwaves, and radio waves, all also invisible to us. The wavelength of light is usually denoted by the Greek letter lambda, with symbol λ, and its frequency by the Greek letter "nu," written ν.

The particles of light are called photons. They are massless, and their energy E is proportional to their frequency. The proportionality constant is called Planck's constant, This is a fundamental constant that appears in the description of all quantum phenomena.[3] It traditionally denoted by the letter h, so that

$$E = h\nu. \tag{14.1}$$

Photons also carry a momentum, denoted by the letter p, which is inversely proportional to their wavelength,

---

2   We recall that waves are characterized by an amplitude, a wavelength, and a frequency: The wave amplitude is defined as half the vertical distance from a trough to a crest of the wave, the wavelength is the distance between two crests of the wave, and the frequency is the number of crests that an observer at rest sees passing in front of her eyes every second.

3   The value of the Planck constant is extremely small, $h = 6.62606957 \times 10^{-34}\ m^2\ kg/s$, hinting at the fact that quantum mechanics is especially important in the atomic and subatomic worlds.

$$p = h/\lambda. \tag{14.2}$$

In vacuum the product of the wavelength of light and its frequency is equal to the speed of light $c = 299, 792, 458$ m/s, $\lambda\nu = c$.

It might come as a surprise that a massless particle such as the photon carries momentum. We recall that the momentum of an object of mass M is the product of its mass times its velocity v, $p = Mv$.[4] Momentum is a very important quantity in physics: Newton's second law of motion $F = Ma$, where a is the acceleration—the change in velocity v—tells us that the force F required to change the velocity of the object is proportional to its mass (more precisely, that the change in momentum of the object is equal to the force acting on it). This is why it is harder to stop a freight train than a bicycle!

How, then, can a massless object carry momentum? To properly understand why this is the case requires invoking Einstein's special relativity theory. The basic idea is twofold: First, one needs to know that nothing can move faster than the speed of light c, and that the only particles that can move at that speed must be massless. This is because it would take an infinite amount of energy to bring any massive particle to that velocity. Second, the description of classical mechanics embodied in Newton's laws does not apply to particles moving at extremely high velocities, near the speed of light. Their motion must be described instead in the framework of the theory of relativity.[5] Unfortunately in that extreme regime of velocities our intuition tends to fail us, so we will simply take Einstein at his word and accept that photons do carry momentum, a property that has been confirmed in numerous experiments. Remarkably, the fact that light can modify the trajectory of a massive particle was already conjectured during the Renaissance, but without a sound theoretical basis, by none other than the great mathematician and astronomer Johannes Kepler (1571–1630) who observed that the tail of comets always points away from the sun and concluded [17] that

» The direct rays of the Sun strike upon it [the comet], penetrate its substance, draw away with them a portion of this matter, and issue thence to form the track of light we call the tail.

14.2.4 Atoms as Waves

We will soon come back to the photon momentum and its importance in atom optics. But before doing so we turn to the other actor in our story, the atoms, and sketch how they are described when they behave as matter waves, or de Broglie waves.

Very much like any other type of waves, they are characterized by a frequency and a wavelength, as first postulated by Louis de Broglie and then formalized in the framework of quantum theory. The de Broglie wavelength λ_{dB} of a non-relativistic massive particle of mass M is related to its momentum p by the equation

$$p = h/\lambda_{dB}, \tag{14.3}$$

---

4 Unfortunately the Greek letter ν used for frequencies and the roman letter v used for velocities look quite similar.

5 In the theory of special relativity the energy E of a particle is related to its momentum p by the equation $E^2 = p^2c^2 + M^2c^4$. For a massless particle, $M = 0$, this reduces simply to $E = pc$. For photons the energy is $E = h\nu$, and we have seen that in vacuum $\lambda\nu = c$, from which it follows that $p = h/\lambda$. As it turns out, the familiar definition of the momentum $p = Mv$ is only approximate. It holds for non-relativistic massive particles, that is, for particles moving much more slowly than the speed of light.

in complete analogy to the situation with photons. However, its kinetic energy takes the familiar form $E = Mv^2/2$ or, remembering that for atoms $p = Mv$ and with the relationship between p and λ_{dB},

$$E = h^2/(2M\lambda_{dB}^2). \tag{14.4}$$

So, while there are important similarities between light waves and matter waves, as evidenced by the relationships between momentum and wavelength of Eqs. (14.2) and (14.3), there are also important differences due to the fact that photons are massless objects while atoms have a mass. For optical waves, the energy is proportional to the momentum, $E = pc$, while for (non-relativistic) matter waves the energy is proportional to the square of the momentum, $E = p^2/2M$. This has important implications for atom optics.

Under everyday circumstances we experience atoms just as particles, not as waves. To understand why this is so let us estimate the size of the de Broglie wavelength. To do so, we need to figure out the momentum of an atom. Since the masses of the various atoms are known and can easily be found in a number of reference books or the internet all we need to do is determine their typical velocity. Let's imagine for a moment a box filled with some atomic gas, maybe Lithium or Sodium, at a temperature T. If it were possible to observe the individual atoms under a microscope, we would see that they move in random directions, going left or right or up or down or forward or backward, some faster, some more slowly, like little kids on a playground. Denoting the average velocity of all these atoms by the symbol $\langle v \rangle$, we would find that it is equal to 0, $\langle v \rangle = 0$: there are lots of atoms in the container, billions of billions of them, and just about as many of them are moving at a given velocity in one direction as in the opposite direction.

But if we took the square of all the individual velocities and averaged the result, call it $\langle v^2 \rangle$, we would find that it is different from zero. This is because the square of any number, be it positive or negative, is a positive number, and the average of a bunch of positive numbers is also positive. Importantly we would also discover that the higher the temperature T of the sample, the larger $\langle v^2 \rangle$, and that T is proportional to $\langle v^2 \rangle$. This is in fact precisely how temperature is defined: it is (in some units) the kinetic energy, or average energy of motion $M\langle v^2 \rangle/2$ of the atoms. The temperature at which all atoms cease to move is absolute zero, $T = 0$. It is impossible to cool anything below that temperature since the atoms cannot move more slowly than not moving at all![6]

Equation (14.3) teaches us that the de Broglie wavelength is inversely proportional to p—the smaller p, the larger λ_{dB}. Near absolute zero the atoms move extremely slowly. They have a very small momentum p, and hence a large de Broglie wavelength. At higher temperatures the atoms move faster, their momentum is larger, and their de Broglie wavelength is therefore smaller. This decrease is proportional to the square root of the temperature, or, in mathematical terms, λ_{dB} is proportional to $1/\sqrt{T}$. At room temperature, one finds that it is of the order of a tenth of a billionth of a meter, or a tenth of a nanometer, 10^{-10} m (this is 0.0000000001 m), a size comparable to the radius of an atom. This is why it is so difficult to observe the atoms as waves: their de Broglie wavelength is simply too small to be observable under normal circumstances.

A good strategy to investigate and exploit the wave nature of atoms is therefore to work at very low temperatures, where their de Broglie wavelength is more easily observable. For a typical atom cooled to a millionth of a degree above absolute zero

---

6   This is the classical physics view of things. The situation is somewhat more subtle in quantum mechanics, which teaches us that atoms still move a tiny bit at $T = 0$, but there is no need for us to worry about this here.

the de Broglie wavelength is of the order of a micron, a millionth of a meter. A millionth of a meter is still very small, so even at extremely low temperatures the wave nature of atoms is quite elusive—except that one micron also happens to be close to the wavelength of visible light. This is an important coincidence because as we know, to measure the size of any object we need to have an appropriate "measuring stick," not too big and not too small, just right. Because the wavelength of visible light turns out to be perfectly matched to the de Broglie wavelength of ultracold atoms, it can serve as that perfectly matched measuring stick. For this and several other reasons to which we will return the combination of visible light and ultracold atoms is a marriage made in heaven.

14.2.5 Cold Atoms and Molecules

Over the years optical and atomic scientists have developed exceedingly sophisticated methods to control the way light interacts with atoms. It is possible to exploit this know-how to prepare and manipulate atoms with extraordinary sophistication, and in particular to cool atoms to temperatures only a minute fraction of a degree above absolute zero.

At first sight, using light to cool atoms doesn't seem to make much sense: our intuition tells us that when we shine light on an object it becomes warmer, not colder. Therefore to use laser light to cool atoms requires one to be rather clever and to understand in detail the way they interact. For a simple qualitative discussion of the basic idea, though, it is sufficient to recall that atoms can both absorb photons (as happens, for instance, in your eyes) or emit them (for instance, in a light bulb.) We also need to keep in mind that all physical processes must satisfy two fundamental laws of nature: conservation of energy—one cannot create energy out of nothing; and conservation of momentum—a moving object will keep moving in a straight line at constant velocity unless one applies a force to it. This is Newton's first law. So, when a photon is absorbed by an atom both its energy and its momentum are transferred to the atom. In the reverse process, when an atom emits light, it loses the momentum h/λ that the photon carries away by changing its velocity.[7]

Remembering that the temperature of a gas is a measure of the energy of random motion of the atoms, this suggests a way by which light can be used to cool atoms: If we can somehow arrange for the moving atoms to predominantly absorb photons propagating *toward* them, then the momentum transferred to them by the photons will be opposite to their direction of motion. They will be pushed back and slowed down. The trick is of course that this needs to be done to all atoms, whether they move up or down, left to right or right to left, and backwards or forward. This is important: one needs to avoid as much as possible having atoms absorbing photons propagating in the same direction as they move, since this would accelerate them rather than slow them down. It turns out that this can actually be achieved by using six different light beams with just the right wavelengths. This mechanism, called Doppler cooling [14, 24], allows to cool atomic gases very significantly, down to roughly a thousandth of a degree above absolute zero. One can do even better by using more complex arrangements of light beams and by cleverly exploiting the internal structure of the atoms. Combining a variety of techniques it is now possible to cool atomic samples to within a

---

7   It turns out that the bulk of energy conservation is achieved via transitions of an atomic electron between different orbits around the nucleus, while the bulk of momentum conservation is normally achieved by changing the velocity of the atom.
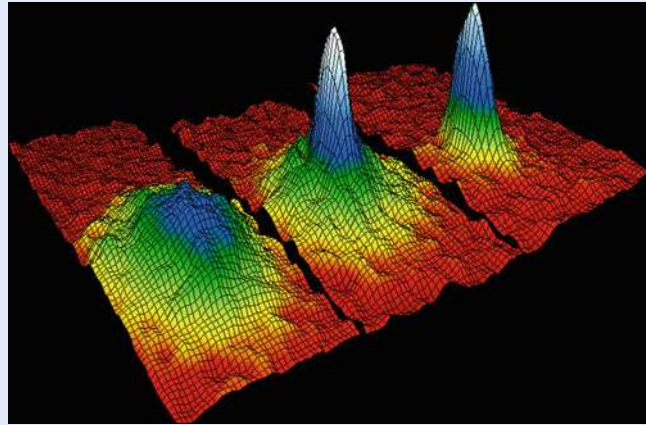
Fig. 14.3 The first atomic BEC ever created by the JILA group of E. Cornell and C. Wiemann in 1995 (Image credit Michael Matthews, JILA)

billionth of a degree above absolute zero or even colder, to a point where their de Broglie wavelength is of the order of a fraction of a millimeter !

A major experimental milestone resulting from the use of such cooling techniques was reached in 1995 (Fig. 14.3) by the groups of Carl Wieman and Eric Cornell at JILA [2], and soon thereafter by Wolfgang Ketterle and coworkers at MIT [7]. They succeeded in realizing atomic Bose–Einstein condensates, a tour-de-force for which they were awarded the 2001 Physics Nobel Prize. A Bose–Einstein condensate is a state of matter where all atoms "condense" into a single quantum object where they are all in the same state, some sort of a "super atom." Ideally the atoms form then a single macroscopic quantum wave, much like photons in a laser behave collectively as a single entity. This exotic object was predicted as early as 1924 by Albert Einstein, expanding on work by the Indian physicist Satyendra Nath Bose, but it is not until atomic samples could be cooled to the extraordinarily low temperatures now possible that it could be produced and observed in its almost pure form.

14.3 Atomic Microscope

We mentioned earlier that if we had at our disposal a microscope that could track individual atoms, we would be able to observe their random thermal motion. While this is not possible at room temperature at this time, the availability of ultracold atomic systems has now made such devices a reality at temperatures approaching absolute zero.

The key idea is that because ultracold atoms carry very little energy of motion it is possible to trap them in extremely shallow potentials, in particular in the periodic potentials that can be produced by standing optical waves—these are the waves produced by two light beams of the same wavelength propagating in opposite directions. Atoms can be trapped in the troughs of these waves in such a way that if they try to escape, then radiation pressure pushes them back down, somewhat like a ball always rolls down to the bottom of a slope. Using standing waves along two or three directions the landscape in which the atoms are trapped resembles an egg crate. It is called an optical lattice potential, or simply optical lattice, and atoms can be trapped at its local minima, as sketched in Fig. 14.4. In 2009 Marcus Greiner and his collaborators at Harvard University devised a microscope that successfully imaged individual atoms localized in such a tightly
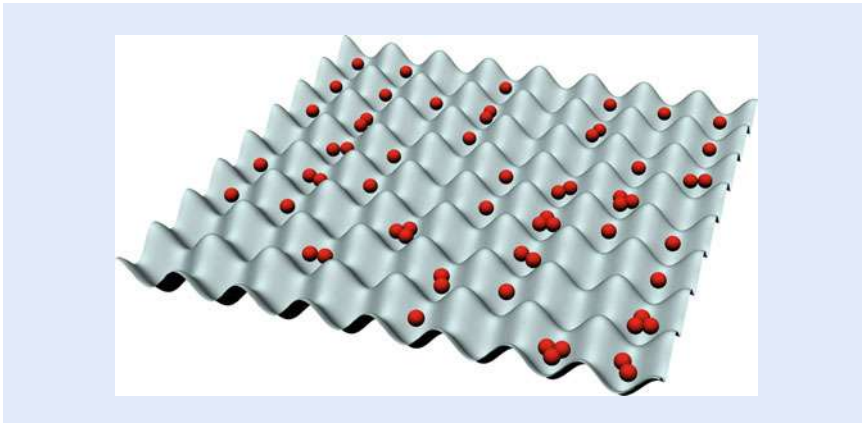
◻ **Fig. 14.4** Artist rendition of the way ultracold atoms can be trapped in an optical lattice (Image credit Andrew Daley, University of Strathclyde)
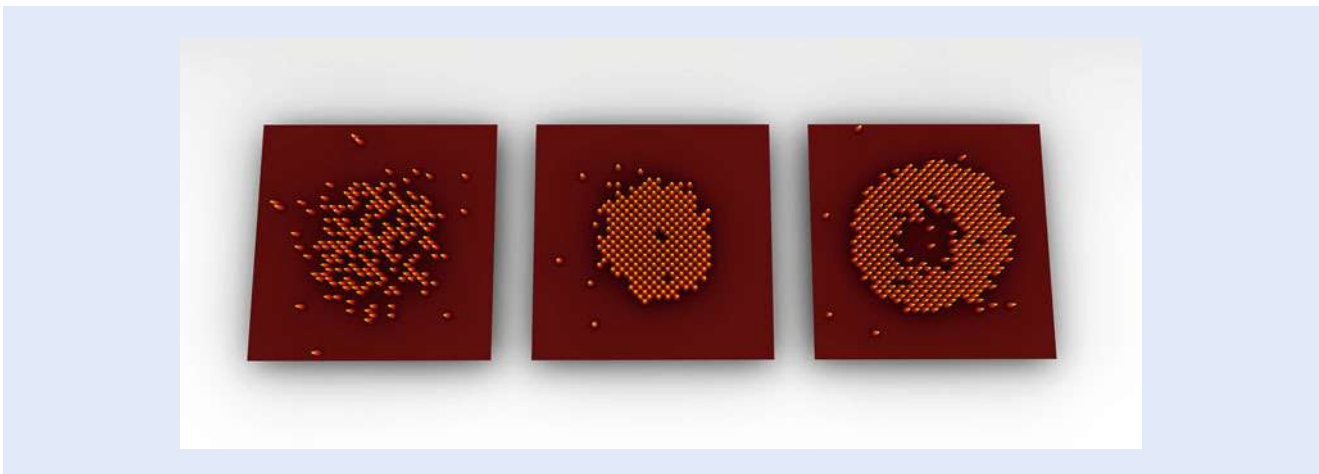
◻ **Fig. 14.5** Illustration of the way atoms can be manipulated and probed individually in an atom microscope. In this case, the system is modified to observe the transition from a Bose–Einstein condensate (*left*) to the so-called Mott insulator (© Immanuel Bloch, Max-Planck Institute for Quantum Optics)

spaced optical lattice [3]. This was soon followed by a second microscope (◻ Fig. 14.5) developed by Immanuel Bloch's group at the Max Planck Institute for Quantum Optics [23]. In these groundbreaking experiments the atoms trapped and individually imaged were bosons[8] but more recently that same technique has also been extended to fermions [5, 19].

Ultracold atoms trapped in optical lattices provide a powerful proving ground to study a number of effects in manybody physics, the situations dominated by the collective behavior of large ensembles of constituents. This is a broad and challenging area of research that is central to the understanding of many phenomena in fields ranging from condensed matter physics to nuclear physics. For example, the collective behavior of electrons in crystal structures is key in understanding their electrical and optical properties.

However experiments in solids can be challenging, in part because it is difficult to control the strength of inter-particle interactions. In contrast, a number of

---

8   Atoms come in two classes, bosons and fermions. Bosons are characterized by the fact that identical bosons can, and like to, occupy the same quantum state in unlimited number. In stark contrast, two identical fermions cannot be in the same quantum state.

powerful tools are available to control these interactions in ultracold atoms. And atom microscopes even permit to address and manipulate individual atoms in the system. For these reasons they provide a remarkable tool to simulate and investigate manybody effects in exquisitely controlled situations. They offer considerable promise to help understand a number of complex manybody phenomena. As noted by Martin Zwierlein [25], whose MIT group developed the first fermionic atom microscope [5],

» High-resolution imaging of more than 1,000 fermionic atoms simultaneously would enhance our understanding of the behavior of other fermions in nature, particularly the behavior of electrons. This knowledge may one day advance our understanding of high-temperature superconductors, which enable lossless energy transport, as well as quantum systems such as solid-state systems or nuclear matter.

14.4   Interferences

We can easily observe the interference of waves when we drop a pair of pebbles in a quiet pond. Each pebble is the source of a small wavelet that propagates away in regular circles, and when the two meet they interfere to produce a complex pattern of crests and troughs.

Similar interferences are also familiar in optics, most famously perhaps in the Young double-slit experiment mentioned earlier. In that case, an optical wave propagates from one side to the other of an absorbing screen through either one or two parallel slit openings (or even more simply one or two pinholes). In the case of a single slit, after it passes through the hole the light wave begins to spread much like the wave generated by a single pebble—the narrower the slit, the larger the angle of spread. With two slits the situation is then akin to what happens with the two pebbles: As they spread spatially the light beams originating at the two slits begin to overlap and interfere, much like the wavelets in the pond. This results in a pattern of alternating dark and bright regions, the analog of the crests and troughs. The more pure the color of the light, the higher the contrast between the bright and dark fringes. This interference phenomenon, perhaps the most direct demonstration of the wave nature of light, is what led Huygens to develop his wave theory of light.

Remarkably, interferences still occur if the light beam is so feeble that only one photon at a time flies past the screen, perhaps one every second, or one every minute, or even one per month! If one waits long enough for the successive photons to slowly build an image, say, on a photographic plate or a CCD camera, that image will still exhibit the same precise interference pattern as if the beam were intense and produced an image in the blink of an eye. The interference pattern builds up one photon at a time!

This should seriously bother you, because one would expect that each individual photon goes through either one slit or the other, but not both, and the situation should then be completely analogous of the one pebble case. *Obviously*, we should not observe interferences in that case. But this is not so: "Obvious" is obviously not a good characterization of what can happen in the quantum world.

14.4.1   Atom Interferences

The situation may seem even more bizarre with atoms, which we are used to think of as particles. But since they obey the same wave-particle duality as photons it is
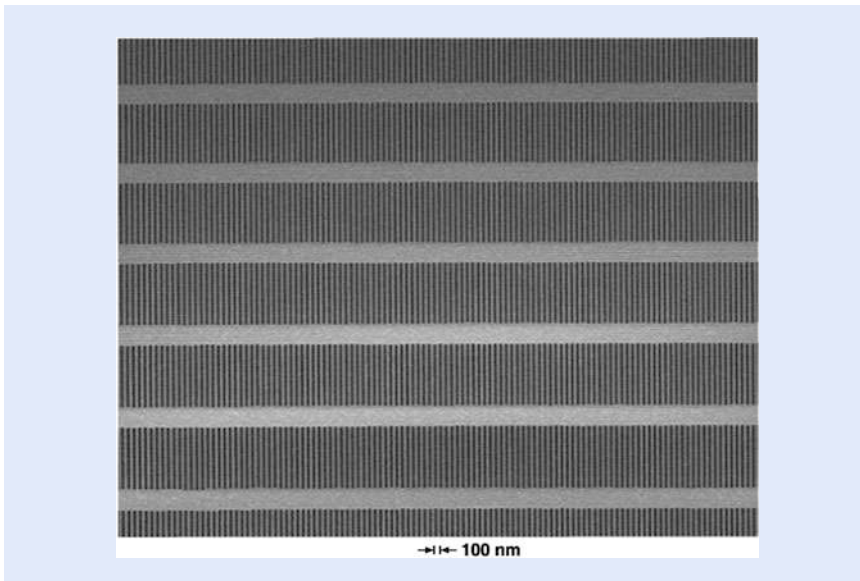
Fig. 14.6 Scanning electron micrograph of a free-standing 100 nm-period diffraction grating for atomic matter waves in a silicon nitride membrane of area 500 μm by 5 mm (courtesy David E. Pritchard, MIT)

possible to produce and observe the interference of matter waves as well. A simple way to do so is to mimic Young's optical double-slit experiment. Practical challenges are that for interferences to be easily observable the width of the slits must be much narrower than de Broglie wavelength, and also that the slit separation should typically be of the order of that wavelength. Modern nanotechnology has solved this problem and makes it possible to fabricate a variety of combinations of holes and slits through which matter waves can propagate. For example, one can pass a beam of atoms through a large array of parallel slits. This is an atom optics analog of the diffraction gratings widely used in optics. An important and useful property of such gratings is that the interferences of the individual wavelets result in different wavelengths (colors) exiting the grating at different angles. Likewise, a nanofabricated mechanical grating can redirect an atomic beam, or even a single atom, in a direction that depends on its energy and momentum. As a result, properly designed gratings can act as mirrors or as beam splitters for atoms (Fig. 14.6).

One can also use light instead of nanofabricated elements to achieve that goal. In the discussion of laser cooling we mentioned that if a photon is absorbed by an atom, then its momentum must be transferred to that atom because of momentum conservation. So, if a photon propagating from left to right is absorbed, then that atom must experience a small velocity kick in that same left to right direction. If, however, the photon propagates from right to left, the velocity kick to the atom will be from right to left as well. And if the atom interacts simultaneously with *two* light fields, one propagating to the left and the other to the right, then it suffers both a velocity kick to the left *and* a velocity kick to the right. As a result the atom "goes in both directions," or more precisely the atomic matter wave is split into two partial waves, one propagating to the left and the other to the right. Acting together, the two light beams act as an atomic beam splitter.

Much like the observation that optical interferences build up "one photon at a time," this is his very strange. Loosely speaking, the atom can move in two directions at the same time, and be in two places at the same time. In the classical world such a behavior would be impossible: The atom would go either to the left *or* to the right, but not to the left *and* to the right.

14

This counter-intuitive behavior is at the core of the double-slit experiment: The observed interferences can only be understood if the atom is described as a wave that propagates simultaneously through both slits, so as to produce the interfering partial waves, just like with the two pebbles. Yet, if the atom is a particle, then surely it must go through either one or the other slit, but not both, right? So, what is going on? Can we not place small detectors near the slits, and measure which of them the atom went through? The answer is that one can certainly do that, but if one makes this "which way" determination, then the interferences disappear! In other words, if we don't ask "which way" the atom went then it behaves as a wave and produces interferences, but if we measure which slit it went through then it behaves as a particle, with no interferences.[9] How can that be? The great physicist Richard Feynman put it beautifully when he wrote [11]

» Because atomic behavior is so unlike ordinary experience, it is very difficult to get used to, and it appears peculiar and mysterious to everyone – both to the novice and to the experienced physicist. Even the experts do not understand it the way they would like to, and it is perfectly reasonable that they should not, because all of direct, human experience and of human intuition applies to large objects. [. . .] We choose to examine a phenomenon which is impossible, absolutely impossible, to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the only mystery.'

14.4.2 Atom Interferometry

Optical interferometry is a remarkably powerful technique that uses the wave nature of light to measure small distances or displacements with extraordinary accuracy and sensitivity. As an example, the LIGO interferometers[10] built in Louisiana and the state of Washington to detect gravitational waves are able to measure length changes of one part in 10^{21} (this is one followed by 21 zeros!). Obviously, not all optical interferometers are that sensitive (or that expensive) but because of their remarkable properties, they are ubiquitous in R&D laboratories and industrial settings.

Optical interferometers come in many variations, but the basic idea is always pretty much the same. They rely on some combination of beam splitters and mirrors to divide a light beam into two or more partial beams that propagate in different environments where they are subjected to different forces and fields before being recombined to produce an interference pattern. For instance, one of the beams could go through an atomic vapor while the other propagates through vacuum, or one beam could have travelled a longer distance than the other; or perhaps one beam bounces off a moving mirror while the other is reflected by a mirror at rest. The key point is that the spatial and temporal features of the resulting interference pattern contain a great deal of information about the different environments that the partial beams propagated through (Fig. 14.7).

Especially since the invention of the laser, optical interferometers have found countless uses in fields as diverse as physics, astronomy, engineering, applied science, remote sensing, seismology, telecommunications, biology, medicine, and manufacturing, to list just a few examples. Applications range from the measurement of extraordinarily small distances to the precise determination of specific

---

9   The same is also true for the photons of the previous section, and for any quantum particle.

10   The LIGO acronym stands for Large Interferometer Gravitational Wave Observatory.
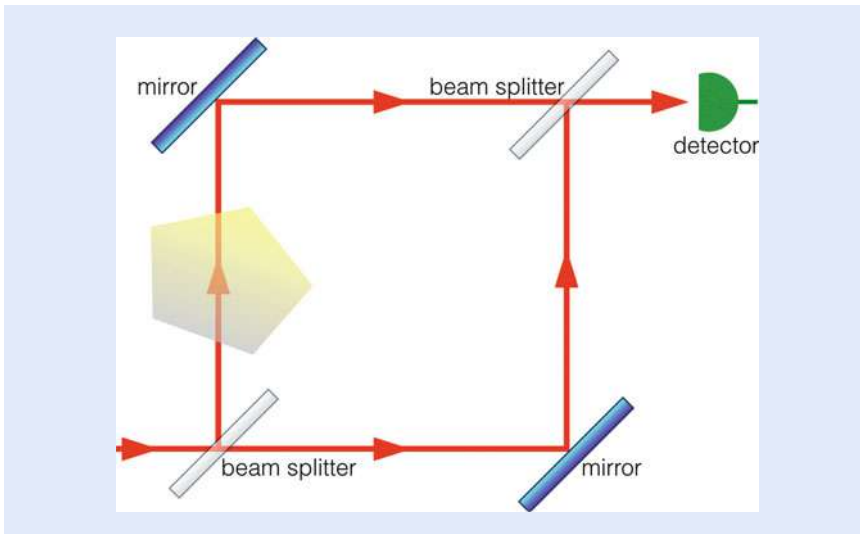
🔵 **Fig. 14.7** Schematics of an interferometer, with two beam splitters to first separate and then recombine the partial beams. One of the beams propagates through some kind of an environment that modifies that interference fringes observed at the detector

atomic or molecular properties, from navigation and guidance to tests of the fundamental laws of physics, from medical imaging to electronic chip fabrication, and much more.

Despite all these successes there are situations where there is considerable benefit in using matter-wave interferometry instead. This is because not surprisingly, massive particles are orders of magnitude more sensitive than photons when it comes to measuring accelerations. For example, one finds that everything else being equal, interferometric gyroscopes (called Sagnac interferometers) using atoms rather than photons have a sensitivity that is larger by the ratio of their rest energy Mc^2 to the energy $h\nu$ of a photon,[11] that is, by $Mc^2/h\nu$. For visible red light and a typical atom such a Cesium this is a factor of about 10^{10}, or 10 billions! This is why atom interferometers are so well adapted to the precise measurement of rotations and accelerations, and can also serve as sensors for other forces and fields [6].

Gravimeters are one example of a practical device that can benefit significantly from atom interferometry. They are important in oil and mineral exploration, where they rely on the fact that different types of rocks or liquids have different densities. They determine the local value of gravity by measuring the acceleration of a free falling mass. Atom interferometers permit in principle to significantly increase the precision of these measurements over other methods. Using small atomic samples as free masses, they operate by splitting their atomic matter waves into two partial waves of different velocities. After some time during which the atoms are free falling the velocities of the two partial waves are then interchanged by a matter wave "mirror." Finally they are recombined to produce an interference pattern from which one can infer the acceleration with high precision and accuracy. To take advantage of the fact that the precision increases with the free fall duration one sometimes uses "atomic fountains" to increase that time. In that case

---

11   The rest energy of a massive particle is its energy when it is not moving. For a particle at rest, $p = 0$, the relativistic energy equation $E^2 = p^2c^2 + M^2c^4$ of footnote 5 reduces to $E = Mc^2$, the equation for the rest energy of a massive particle famously associated with Einstein.

the atoms are launched upward before eventually turning around and falling back down toward the earth (for a brief history of atomic fountains see [15]).

14.4.3 Fundamental Studies

Because of their remarkable potential sensitivity atom interferometers are now a tool of choice not just in practical applications, but also in tests of the fundamental laws of physics, such as the Equivalence Principle. This is the fundamental principle which states that all objects fall with the same acceleration under the influence of gravity. It forms the foundational basis of Einstein's Theory of General Relativity. The best tests of the Equivalence Principle to date have shown that the accelerations of two falling objects differ by no more than one part in 10^{13}—this is one followed by 13 zeros [21]. A group led by Mark Kasevich at Stanford University aims for an improved test of this principle to one part in 10^{15} by dropping atoms of two different isotopes of rubidium[12] in a 10 m high drop tower [22] (◘ Fig. 14.8).
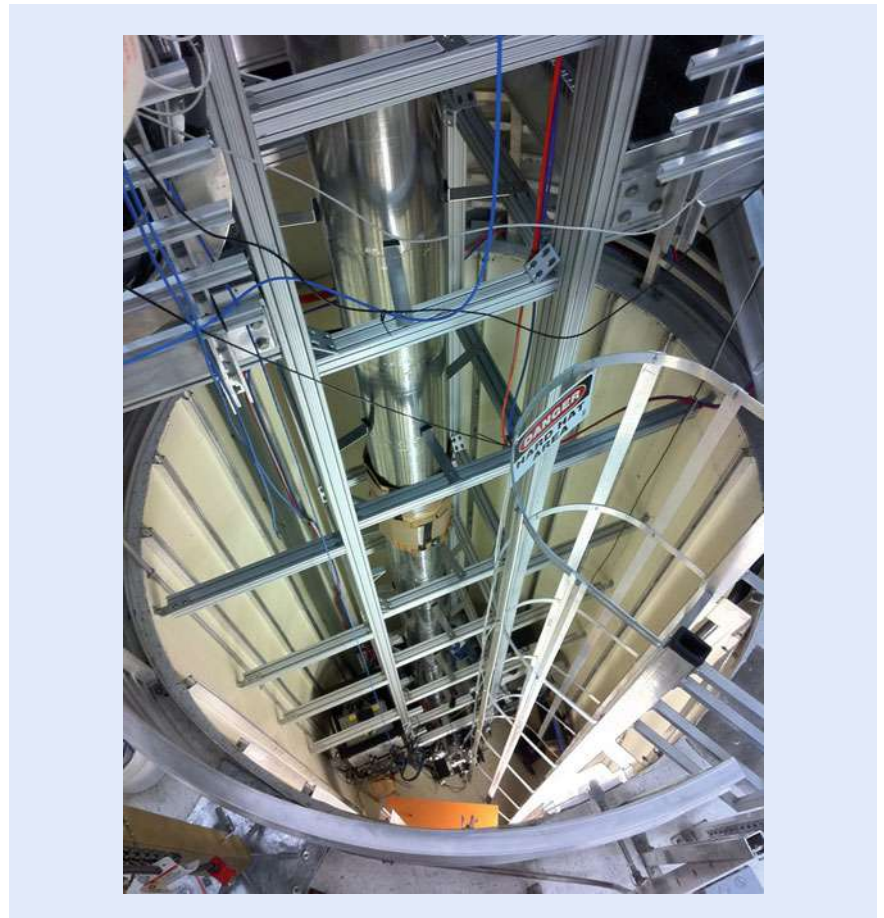
◘ **Fig. 14.8** Stanford 10 m tower used for tests of the equivalence principle (courtesy Mark Kasevich, Stanford University)

---

12   Isotopes are variations of a chemical element that all have the same number of protons and electrons, but differ by the number of neutrons in their nucleus, and hence have different masses.

In another example, atom interferometry offers great promise for the development of a new mass standard. Surprisingly perhaps, the kilogram is the last physical unit that is defined by an artifact, the International Prototype Kilogram. It is the mass of a block of platinum–iridium alloy stored in an environmentally controlled vault in the basement of the International Bureau of Weights and Measures in Sèvres, near Paris. In addition to being subject to damage, this standard presents the fundamental issue of not being based on a physical law.

One proposal is to define the unit of mass in terms of a frequency. This would be possible provided that one assigns to the Planck constant h a fundamental value, much like the speed of light c is assigned a fundamental value that allows to connect lengths to times.[13] The basic idea is that the momentum imparted on atoms by light is proportional to the frequency of that light, and inversely proportional to the mass of the atom, the proportionality factor being given by Planck's constant. If it were assigned a fundamental value, then one could connect masses to frequencies extraordinarily accurately. Future space-borne atom interferometers might then allow the measurement of atomic masses anywhere on Earth better than 1000 times more accurately than is presently the case.

A third example of a basic science application of atom interferometry is in gravitational wave detectors. Four centuries after Galileo (1564–1642) used telescopes to study and revolutionize our understanding of the Universe they remain our most powerful tool to learn about it, whether they detect radio waves, sub-millimeter waves, infrared radiation, visible light, ultraviolet radiation or X-rays. However, it is believed that extremely significant additional information would be provided by the detection and characterization of the gravitational waves produced by the motion of massive objects, in particular closely orbiting compact massive objects such as neutron stars or black holes binaries, merging supermassive black holes, collapsing supernovae, or pulsars. Gravitational waves might also provide information on the processes that took place in the early Universe, shortly after the Big Bang. However they interact only extremely weakly with matter, and so far they have remained elusive.[14]

It is expected that in the near future Advanced LIGO, the upgraded version of the LIGO gravitational wave antennas, will be sensitive enough to detect gravitational waves at the rate of maybe a few events per year. To further increase sensitivity and the frequency of observations, future systems will likely need to be space-based, one example being the proposed Laser Interferometer Space Antenna (LISA). Mark Kasevich and his coworkers at Stanford have proposed an alternative space-based hybrid approach (◘ Fig. 14.9) combining optical methods and atom interferometry [13]. Their proposal draws on the use of an optical method to measure the differential acceleration of two spatially separated, free falling atom interferometers whose mirrors and beam splitters are produced by light pulses sent back and forth between them through space. Comparing the matter-wave interference fringes in the two interferometers would provide a record of the effect of gravitational waves on the travel time a laser pulse linking the two atom interferometers. It is argued that using atoms instead of mirrors as test masses would reduce a number of systematic errors.

---

13   State-of-the-art atomic clocks can measure times to accuracies in excess of one part in 10^{17}, so that reducing the determination of a physical quantity to a measurement of time or frequency is particularly favorable.

14   Note added: This is no longer the case! A few months after this article was completed the LIGO Scientific Collaboration and the Virgo Collaboration reported the first direct observation of a gravitational wave signal, resulting from the collision of two massive black holes [1]. This historic breakthrough, 100 years after Einstein's 1916 prediction, opens the way to a new era in observational astronomy.

14
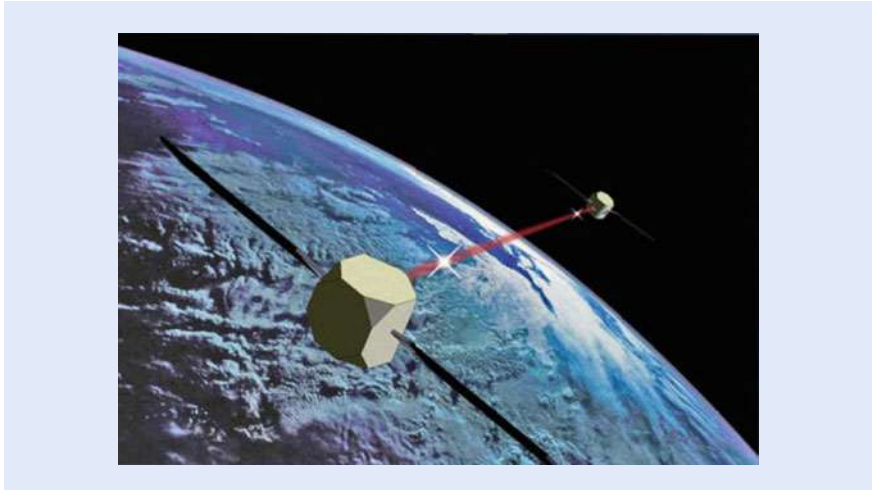
Fig. 14.9 A proposed space-borne atom interferometer gravitational wave detector. Two widely separated atom interferometers are controlled by light pulses sent back and forth between them. Passing gravitational waves will modify the travel time of the light pulses between the interferometers, resulting is a shift on their interference fringes that provide a record of the gravitational wave (courtesy Mark Kasevich, Stanford University)

14.4.4 BEC Atom Interferometers

In conventional optics it is often favorable to use lasers rather than conventional light sources, if only because they can have very high photon fluxes within a very narrow range of wavelengths. So one might ask whether the same is true in atom optics, and whether it would be advantageous to use an "atom laser," or Bose–Einstein condensate, rather than a regular beam of non-condensed ultracold atoms.

The simple answer is that this is typically not the case, because of an important difference between photons and atoms: Two light beams propagating in different directions in free space can cross without perturbing each other, because photons don't directly interact. In contrast, atoms do collide. Collisions are random events that result in uncontrolled changes in the interferences between the atom matter waves. This leads to additional detection noise that can significantly limit the sensitivity and accuracy of measurements. As collisions become more frequent when the atomic flux is increased this limits the applicability of Bose–Einstein condensates in atom interferometry.

This problem can, however, be circumvented to some extent by reducing the collisions between atoms. This can sometimes be achieved in ultracold atomic samples by using magnetic fields to control the collision rate. In the best cases it is possible to almost completely suppress collisions, leading to the potential for high precision interferometry with Bose–Einstein condensates. Alternatively, under appropriate conditions other quantum effects can be exploited to increase the sensitivity of the system and the precision of measurements, using, for example, the so-called squeezed states or number states of the matter waves.[15] An atom interferometer based on this principle was recently realized in the group of Jörg

---

15 Quantum objects are subject to fundamental random fluctuations called quantum noise—this is why even at $T = 0$ atoms are not completely still, see footnote 6. But it is sometimes possible to prepare atoms or photons in such a way that this noise is "squeezed away," or more precisely transferred to a place where it does not add imprecision to a specific type of measurement.

Schmiedmeyer at the University of Vienna [4], see ◘ Fig. 14.10. Another possible approach involves the use of fermionic atoms instead of bosons, although the interference contrast tends to be reduced in that case.

14.5 Outlook

It is widely accepted that quantum mechanics is the fundamental theory of nature. It has been and continues to be put to numerous, increasingly elaborate tests that it has so far passed with flying colors. Yet, in everyday life we don't observe the remarkable quantum effects that we can achieve with small ensembles of atoms or with photons under exquisitely controlled conditions. We cannot make a car be "in two places at the same time," or, in the famous example of Schrödinger, we cannot have a cat that is both alive and dead at the same time. Our everyday world seems to be most definitely governed by the laws of classical physics, not by quantum mechanics. This is extremely puzzling, because if the quantum mechanical description of nature is more fundamental than its classical description, then quantum mechanics should govern not just the microscopic world, but the macroscopic world as well.

Why and how macroscopic systems lose their quantum features and become essentially classical are challenging questions that are being addressed by a number of researchers, both theoretically and experimentally. On the theoretical side, proposed explanations range from relatively mundane mechanisms, such as increasingly fast decoherence resulting from the contact of objects of increasing size to their environment, to speculations about the role of gravity in washing out quantum features in massive objects.

On the experimental side, there are exciting efforts to observe quantum interferences in increasingly macroscopic objects, with the goal of improving our understanding of the physical mechanisms that wash out quantum features in objects of increasing complexity. For example, a group around Markus Arndt at the University of Vienna has succeeded in demonstrating the wave nature of large organic molecules, from the "buckeyball" C_{60} to the very large molecule TPPF152 $(C_{168}H_{94}F_{152}O_8N_4S_4,)$ which contains 430 atoms and has a thermal de Broglie wavelength of about one picometer, a millionth of a millionth of a meter [12]. It is hoped that eventually such experiments will help determine whether the quantum to classical transition is a practical and relatively mundane issue or a truly

fundamental occurrence. Is there a fundamental limit on the size of objects that can behave as de Broglie waves, or are the challenges only practical?

In an ambitious proposal, Oriol Romero-Isart, Markus Aspelmeyer, Ignacio Cirac, and coworkers have recently proposed a method to prepare and verify spatial quantum superpositions of a nanometer-sized object separated by distances comparable to its size [20]. It is hoped that such experiments will eventually be able to operate in a parameter regime where it will be possible to test various proposed mechanisms beyond quantum mechanics that have been advanced to explain the washing out of quantum properties in macroscopic objects. It will be exciting indeed to see these proposed experiments being realized and start answering questions that have surrounded quantum mechanics and its interpretation since its early days, nearly 100 years ago.

References

1. Abbott BP et al (2016) Observation of gravitational waves from a binary black hole merger. Phys Rev Lett 116:061102
2. Anderson MH, Ensher JR, Matthews, MR, Wieman, CE, Cornell, EA (1995) Observation of Bose-Einstein condensation in a dilute atomic vapor. Science 269:5221
3. Bakr, WS, Gillen, JI, Peng, A, Foelling, S, Greiner, M (2009) Quantum Gas microscope detecting single atoms in a Hubbard regime optical lattice. Nature 462:74
4. Berrada, T, van Frank, S, Bücker, R, Schumm, T, Schaff, J-F, Schmiedmayer, J (2013) Integrated Mach-Zehnder interferometer for Bose-Einstein condensates. Nat Commun 4:2077
5. Cheuk, LW, Nichols, MA, Okan, M, Gersdorf, T, Ramasesh, VV, Bakr, WS, Lompe, T, Zwierlein, MW (2015) Quantum-gas microscope for fermionic atoms. Phys Rev Lett 114:193001
6. Cronin, AD, Schmiedmayer, J, Pritchard, DE (2009) Optics and interferometry with atoms and molecules. Rev Mod Phys 81:1051 gives a comprehensive discussion of atom optics and interferometry, including a review of scientific advances and a broad range of application of atom interferometers
7. Davis, KB, Mewes, M-O, Andrews, MR, van Druten, NJ, Durfee, DS, Kurn, DM, Ketterle, W (1995) Bose-Einstein condensation in a gas of sodium atoms. Phys Rev Lett 75:3969
8. Davisson C, Germer, LH (1927) Diffraction of electrons by a crystal of nickel. Phys. Rev. 30:705
9. de Broglie L (1925) Recherche sur la Théorie des Quanta. PhD thesis, University of Paris

10. Estermann E, Stern O (1930) Beugung von Molekularstrahlen. Z Phys 61:95

11. Feynman RP (2011) In: Leighton RB, Sands, M (eds) Feynman lectures on physics, Chap. 1, vol. III. Free to read online edition available at ▶ http://www.feynmanlectures.caltech.edu/

12. Gerlich S, Eibenberger S, Tomandl M, Nimmrichter S, Hornberger K, Fagan PJ, Tüxen J, Mayor M, Arndt M (2011) Quantum interference of large organic molecules Nat Commun 2:263

13. Graham PW, Hogan JM, Kasevich MA, Rajendran S (2013) New method for gravitational wave detection with atomic sensors. Phys Rev Lett 110:171102

14. Hänsch TW, Shawlow AL (1975) Cooling of gases by laser radiation. Optics Commun 13:68

15. Kasevich MA, Riis E, Chu S (1989) Atomic fountains and clocks. Optics News 15:31

16. Kelvin L (1901) Royal Institution lecture. Nineteenth-century clouds over the dynamical theory of heat and light. Philosophical magazine, series 6, vol. 2, p 1

17. Kepler J, as quoted in *A Comet Called Halley*, by I. Ridpath, Cambridge University Press (1985), see ▶ http://www.ianridpath.com/halley/halley2.htm

18. Müller H (2012) Quantum mechanics, matter waves, and moving clocks. In: Proceedings of the international school of physics "Enrico Fermi". Vol. 188. Atom interferometry, pp 1339–418. IOS Press, available to read on ▶ http://arxiv.org/pdf/1312.6449.pdf

19. Parsons MF, Huber F, Mazurenko A, Chiu CS, Setiawan W, Wooley-Brown K, Blatt S, Greiner M (2015) Site-resolved imaging of fermionic ^{6}Li in an optical lattice. Phys Rev Lett 114:213002

20. Romero-Isart O, Pflanzer AC, Blaser F, Kaltenbaek R, Kiesel N, Aspelmeyer M, Cirac JI (2011) Large quantum superpositions and interference of massive nanometer-sized objects. Phys Rev Lett 107:020405

21. Schlamminger S, Choi KY, Wagner TA, Gundlach JH, Adelberger EG (2008) Test of the equivalence principle using a rotating torsion balance. Phys Rev Lett 100:041101

22. See the M. Kasevich group website at Stanford University. ▶ http://web.stanford.edu/group/kasevich/cgi-bin/wordpress/?pageid=11.

23. Sherson JF, Weitenberg C, Endres M, Cheneau M, Bloch I, Kuhr S (2010) Single-atom-resolved fluorescence imaging of an atomic Mott insulator. Nature 467:68

24. Wineland DJ, Dehmelt H (1975) Proposed $10^{14}\varDelta v < v$ laser fluorescence spectroscopy on Tl$^+$ mono-ion oscillator III. Bull. Am. Phys. Soc. 20:637

25. Zwierlein M, as quoted in ▶ http://www.sci-news.com/physics/science-microscope-fermionic-atoms-02799.html.

Slow, Stored and Stationary Light

Michael Fleischhauer and Gediminas Juzeliūnas

M. Fleischhauer (✉)
Department of Physics and research center OPTIMAS, University of Kaiserslautern, 67663 Kaiserslautern, Germany
e-mail: mfleisch@physik.uni-kl.de

G. Juzeliūnas
Institute of Theoretical Physics and Astronomy, Vilnius University, Saulėtekio 3, LT-10222 Vilnius, Lithuania
e-mail: gediminas.juzeliunas@tfai.vu.lt

15.1 Introduction

Since the experiments of Michelson and Morely and their brilliant explanation by Albert Einstein more than 100 years ago which have laid the foundation for the theory of relativity, we know that light propagates in empty space with the largest possible velocity. This speed of about 300,000 km/s is so fast that we can have a phone conversation around the globe without noticing that an electromagnetic signal has to be transmitted for every bit of information. When we look through a window or a prism of quartz we see that light gets refracted. Refraction is due to the fact that light propagates in a transparent medium at a slightly lower speed than allowed by the universal traffic laws of nature. This speed, called phase velocity depends on the color of light and the variation of the phase velocity in media, is what causes the beauty of a rainbow or the bright fan of colors produced by a prism. Yet the change of the velocity of light in water, in glass or even in diamond is small, it is typically less than a factor of 2. But what if this factor is 10^7, a ten with 6 extra zeros, i.e. 10,000,000? Such light can truly be called ultra slow. As opposed to propagation faster than the vacuum speed of light, this is not forbidden by Einstein's theory of relativity, but for a long time did not seem feasible. It did so until the late 1980s and early 1990s, when Steve Harris from Stanford University pointed out that an effect he termed electromagnetically induced transparency (EIT) [1, 2] can lead to a massive reduction of the effective speed of pulsed light [3]. When we talk about 'slow' light we talk about the speed of *pulses* of light, called group velocity, which needs to be distinguished from the phase velocity mentioned above.

Although a number of experiments have seen evidence of velocity reduction in EIT media, it took until 1999 [4–6] that slow light received a great deal of attention. In 1998 the group of Lene Hau at the Rowland Institute for Science together with Steve Harris managed to decelerate the propagation of light in an atomic gas to 17 m/s, i.e. almost 20 million times slower than in vacuum. The cover page of the journal Nature (◘ Fig. 15.1), where this experiment was

◘ **Fig. 15.1** *Slow light:* Cover page of the 18th February 1999 issue of the journal Nature illustrating an experiment on slow light by the group of Lene Hau at the Rowland Institute for Science. Using an ultracold gas of atoms the physicists managed to slow down a pulse of light to a velocity of 17 m/s (Reproduced with permission of the journal Nature)

published in 1999, illustrates the achievement showing that a trained cyclist could even outrace such a light pulse. This is of course only a figurative way to demonstrate how slow the light was compared to the usual. Actually there was no cyclist involved in the experiment. The light was propagating in a tiny cloud of ultracold atoms contained in a vacuum chamber over a very small distance, as one can see by closer inspection of the figure. This spectacular result then triggered a rapidly growing activity in the field leading to many fascinating applications.

So, what is slow light and what is it good for? How can we understand the physics of it and how can we practically make light go so slow? These are the questions we want to answer in the following using simple pictures, on the one hand, and supplementing them with a little bit of details, on the other hand, for those who want to go slightly deeper. Yet we will avoid math as much as possible and refer those who seek more detailed information to the specialized literature [7–11].

15.2 Slow Light, Stopped Light and Stationary Light: A Simple Picture

How can one slow down light to such extremely low velocities? Imagine a fast racing car (Fig. 15.2). If a heavy trailer is attached to the car, its engine has now also to pull the trailer. This slows down the car considerably. Something similar happens with light in a specially arranged atomic medium used in EIT experiments. Light is composed of photons—tiny particles which are very fast, so one can visualize them as fast racing cars. When entering the atomic medium, most of the photons are converted into a special kind of atomic excitations (which we here call spin excitations) which cannot move on their own, and thus behave like heavy trailers. The atomic excitations generated in this way are coupled to the small number of remaining photons which have to pull a vast number of immobile spin excitations while travelling in the medium. In this way, the propagation of the whole pulse of light is slowed down dramatically. The possibility to convert 'fast cars' into 'immobile trailers' is a small, but important difference to usual cars and trailers we encounter in real life. When the crawling light pulse reaches the end of the medium, the atomic excitations (trailers) are converted back to photons (fast cars), so the light exiting the medium becomes fast again.

Now imagine that the number of photons converted into atomic excitations (i.e. fast cars converted into trailers) can somehow be increased at will. This means there is an even lesser number of remaining cars to pull the whole bunch of trailers.
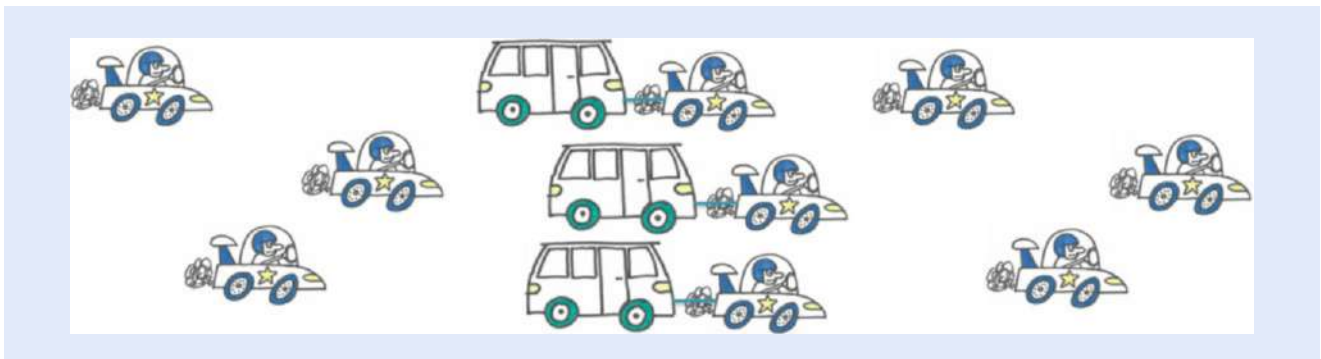
 Fig. 15.2 *A simple picture of slow light:* Imagine a bunch of racing cars that enter a parking lot where heavy trailers get attached to them. Since the racing cars have to pull the trailers, they get slowed down considerably. When they reach the end of the parking lot, the trailers get detached, and the cars can move on with their full speed. Slow light is almost like this, except that cars get partially converted into trailers at the entrance to the parking lot and converted back at the end

And now imagine further that the conversion between cars and trailers can be changed while the fast cars are going through the trailer park. What if all of them are converted and no racing car is left to pull? The pulse would stop! This is the essence of stopped, or more precisely stored light, theoretically predicted in [12] and soon after experimentally verified in [13, 14]. The important difference of this kind of light storing and using a black piece of paper, which just absorbs the light, is that here the information carried by the photons is still present in the medium, in our analogy in the form of heavy trailers. Thus in principle all information about the original photons stored in the atomic excitations (trailers) can be converted back into photons (fast cars) either completely or in part. When the slow-light pulse reaches the end of the medium, the atomic excitations can no longer be dragged along and are fully converted back into light. In this way the stored light pulse can be fully retrieved.

Light storage is of particular interest in information technology especially in quantum information science. Light is an ideal carrier of information be it classical information which we use in every-day life or be it quantum information which may encounter at some day in a quantum network. Yet in the second case it is rather difficult to store information without loosing the quantum character, referred to as quantum coherence. Here light storage is an extremely useful method to build what is called a quantum memory for light. In fact first proof-of-principle demonstrations of quantum memories for photons based on light storage have already been made in a number of labs [15, 16].

It is noteworthy that by storing a light pulse all its photons (i.e. all the racing cars) are converted into immobile atomic excitations (trailers). Yet there is another way to make photons immobile where the photons are still present in the medium. This is called stationary light. It is formed when two counter-propagating pulses of light are driving the *same* spin excitations of a properly prepared atomic medium [17–21]. This corresponds to having two types of racing cars, one going from the left to the right, and another one from the right to the left. Both types of cars are trying to pull the same immobile trailers in opposite directions, as illustrated in ◘ Fig. 15.3. The forces compensate, so the cars and the trailers remain at rest. More precisely stationary light behaves like massive quantum particles with zero average velocity. Note that in the quantum world physical quantities such as the particle velocity fluctuate and thus we need to talk about averages here.

One can also produce a situation where two counter-propagating pulses of light drive *different* spin excitations of the atomic medium. If the two types of spin excitations are coupled to each other in the right way, two-component slow light is formed which has a more complex structure resembling what is known in quantum physics as a particle with a spin degree of freedom [22–25]. This is like having two types of racing cars going in opposite directions, each pulling different types of
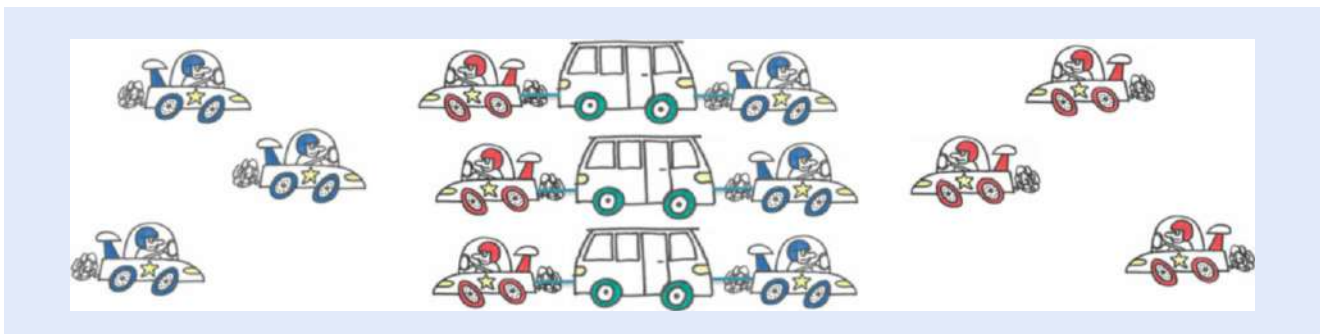
◘ **Fig. 15.3** *The principle of stationary light:* Imagine racing cars entering a parking lot with heavy trailers from opposite sides. When attaching trailers to the cars they are pulled in opposite directions with equal forces and thus don't move at all. In this way the racing cars can be brought to halt even without converting them completely into trailers as is the case for light storage
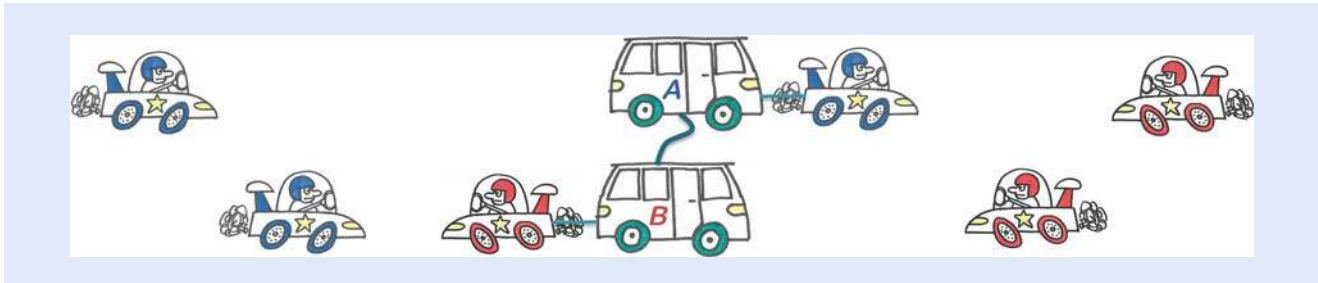
🔲 **Fig. 15.4** *Multi-component slow light:* When racing cares moving in opposite directions pull different types of trailers, both types of cars would slow down independently of each other. However, when coupling the trailers together in a proper way a situation is created that corresponds in physics to quantum particles with an internal degree of freedom
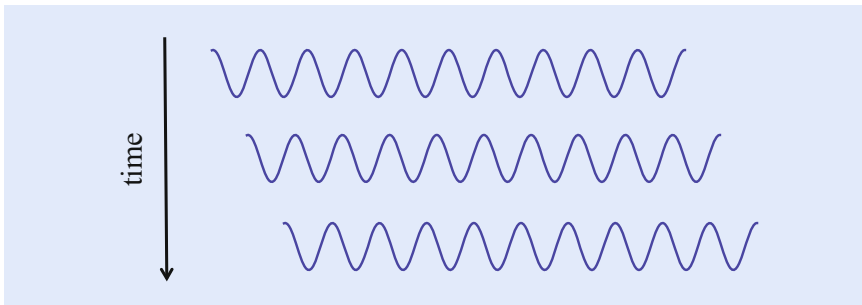
🔲 **Fig. 15.5** *Light waves:* Light are waves of the electric field oscillating in space with a certain period, the wavelength λ. The 'hills' and 'valleys' of the wave, i.e. the points of maximum and minimum wave amplitude propagate in space with phase velocity c, such that at a fixed point in space the electric field oscillates in time with frequency $\omega = 2\pi c/\lambda$

trailers, as shown in 🔲 Fig. 15.4. If the trailers were not coupled to each other, the two types of cars would slowly move in opposite directions pulling their respective trailers independently from each other. Yet if there is a coupling between the two sorts of trailers, the oppositely moving cars and trailers influence each other, making a more complex dynamics, resembling that of a relativistic quantum particle.

15.3 A Microscopic Picture of Light Propagation in a Medium

In order to understand the mechanism behind slow light we first have to talk about the microscopic physics of light propagation in a medium. In particular we will discuss what the physical origin of absorption and refraction is, two phenomena which we are familiar with in every-day life.

15.3.1 Absorption, Emission and Refraction

Light is nothing else than an electromagnetic wave build up from oscillating electric and magnetic fields. The color of light is determined by the oscillation frequency $\omega = 2\pi/T$, given by the inverse of the temporal period T of oscillations. The electric field of a plane wave propagating along say the x axis of some coordinate system has a sinusoidal form depicted in 🔲 Fig. 15.5. It is characterized by the frequency ω, and a corresponding wavelength λ, which is the spatial period of the wave. This can be written in the following form:
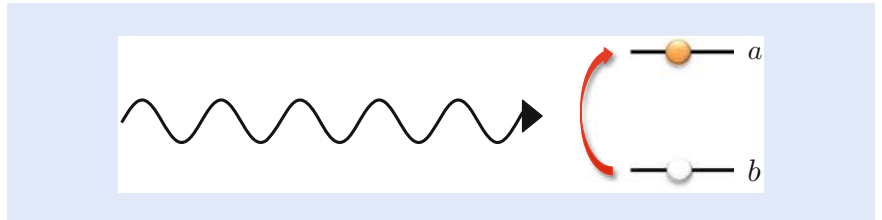
Fig. 15.6 *Absorption:* When an atom absorbs a photon it changes its quantum state from a low-energy state to a high-energy one

15

$$E = E_0 \sin\left(\omega t - 2\pi x/\lambda\right) = E_0 \sin\left(\phi\right), \tag{15.1}$$

where we have introduced the phase ϕ. The propagation velocity of such a wave can be found by asking: What is the position change Δx in a time Δt for a fixed value of the phase ϕ? One finds: $c = \Delta x/\Delta t = \omega\lambda/2\pi$, which is called the phase velocity.

Light carries energy, which, as figured out first by Max Planck in 1900, comes in quantized units. So a beam of light is composed of particles called photons. The amount of energy E contained in each of these photons is proportional to the oscillation frequency ω, i.e. it depends on the color, $E = \hbar\omega$, where the constant $\hbar$ entering here is the famous Planck constant. High-frequency photons, such as those of ultra-violet light or even X-rays, are very energetic, while low-frequency photons such as infrared light or microwaves, which we cannot see with our eyes, do contain much less energy per photon.

Matter, on the other hand, consists of atoms, which according to the laws of quantum mechanics have a number of states characterized by discrete energies. Very often it is sufficient to consider only two or three most relevant states. Atoms are also small quantum oscillators which can 'vibrate' at different frequencies corresponding to the energy differences between quantum states $\omega_{ab} = (E_a - E_b)/\hbar$. Many (but not all) of these 'vibration' modes are associated with an oscillating electric dipole. In this way an atom can absorb or emit radiation just like an antenna of a mobile phone. As we shall see later on, photons play the role of the fast racing cars described in the introductory section, whereas properly prepared atoms absorbing the photons play the role of the heavy trailers. When an atom absorbs a photon it changes its quantum state from the low-energy state to the high-energy state (see �’ Fig. 15.6) and vice versa if it emits a photon.

There are actually two types of emission of an excited atom. The most common is spontaneous emission, where a photon is emitted in a random direction leading to the loss of information on the state, the propagation direction and the polarization of the photon that excited the atom in the first place, see �’ Fig. 15.7a. The other one is stimulated emission which takes place in the presence of other identical photons and is pointed into the direction determined by these photons, see �’ Fig. 15.7b. In addition to spontaneous emission there are a number of other relaxation processes for excited states in atoms. As a consequence of these processes and due to spontaneous emission, excited atomic states decay with some rate γ. Thus when light shines on a cloud of atoms or atoms arranged in a crystal, it can be absorbed by exciting some of the atoms into high-energy states which subsequently decay. Clearly how much a medium absorbs depends on the density of atoms, which in a gas is much less than, e.g., in a solid.

Still, why is it that some solids like diamond are transparent to visible light and others like coal are pitch black? Both are just slightly different forms of carbon and their density does not differ significantly. The reason is simple: In order for a photon to be efficiently absorbed, its frequency has to be close to the frequency of the atomic oscillator, i.e. the frequency should correspond more or less to the
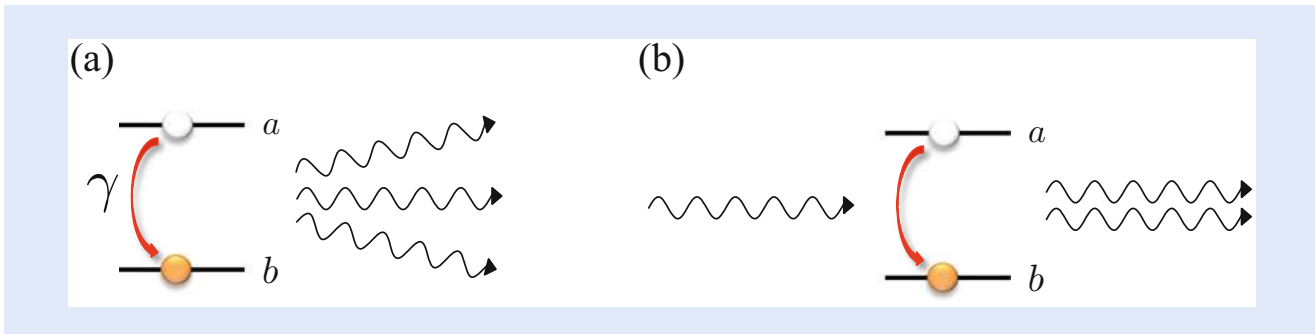
⬛ **Fig. 15.7** *Spontaneous and stimulated emission:* An excited atom can loose its excitation energy either by spontaneously emitting a photon in an arbitrary direction (**a**) or stimulated by an incoming photon (**b**) in which case the emitted photon has the same direction than the incident one. In both cases the atom changes its quantum state from the high-energy to the low-energy one

energy difference between some lower and higher state $\omega_{ab} = (E_a - E_b)/\hbar$. When this is the case, one talks about resonance. If the photon frequency is very different from any of the vibration frequencies of the atomic oscillator, i.e. if the light is off-resonant, not much can happen. It is like if you are trying to make a bridge vibrate by jumping up and down but are doing it at the wrong pace. Only a tiny bit of the photon energy is transferred to the atom, stored there for a very little moment and then is reemitted into the stream of photons. In this process the atom is actually not completely transferred from the lower-energy state to the higher-energy state, as in ⬛ Fig. 15.6, and the subsequent emission process is a bit different from the stimulated process shown in ⬛ Fig. 15.7b, but in essence it is like this. A word of caution is needed here: This picture of absorption is a bit of an oversimplification if applied to solids rather than to sparse atomic gases. The quantum states and energies in a solid are not the same than those of isolated atoms as they are affected by atom–atom interactions. Also even off-resonant transitions can eventually lead to sizable absorption if there are very many of them.

As we have mentioned before, waves are characterized by a wavelength λ, which gives the spatial period of a wave and is directly related to the frequency. In vacuum the relation between the two is $\lambda_0 = 2\pi c_0/\omega$. Here c_0 is the vacuum speed of light, i.e. the fastest velocity allowed by the laws of nature. In a medium this relation is changed, however. The short moment for which the photon is stored in the atom causes a delay. The effect of the very many, tiny delays at every atom in the medium makes light appear to propagate with a modified phase velocity

$$c(\omega) = c_0/n(\omega). \tag{15.2}$$

Here $n(\omega)$ is called the refractive index. In vacuum the refractive index is unity. The name 'refractive index' stems from the fact that it characterizes the refraction of light beams at an interface between say air and a piece of glass, as illustrated in ⬛ Fig. 15.8. Refraction comes about since along with the change of the phase velocity of a plane wave at frequency ω comes a change of the wavelength $\lambda = \lambda_0/n(\omega)$. This is because the frequency of the wave remains the same in the medium, giving $\omega = 2\pi c_0/\lambda_0 = 2\pi c/\lambda$.

The influence of a medium on the propagation of light is characterized by the susceptibility χ. In ⬛ Fig. 15.9 we have plotted both the absorption strength (red line) represented by the imaginary part of the susceptibility $Im[\chi] = \chi''(\omega)$ together with its real part $Re[\chi] = \chi'(\omega)$ (blue line) as function of the frequency in the vicinity of an atomic resonance frequency ω_{ab}. The latter χ' describes the deviation of the index of refraction from unity, $n = 1 + \chi'/2$. One recognizes that the absorption peaks on resonance and falls off quickly with increasing frequency mismatch $\Delta = \omega - \omega_{ab}$, called detuning. The refractive index has a bit more

15
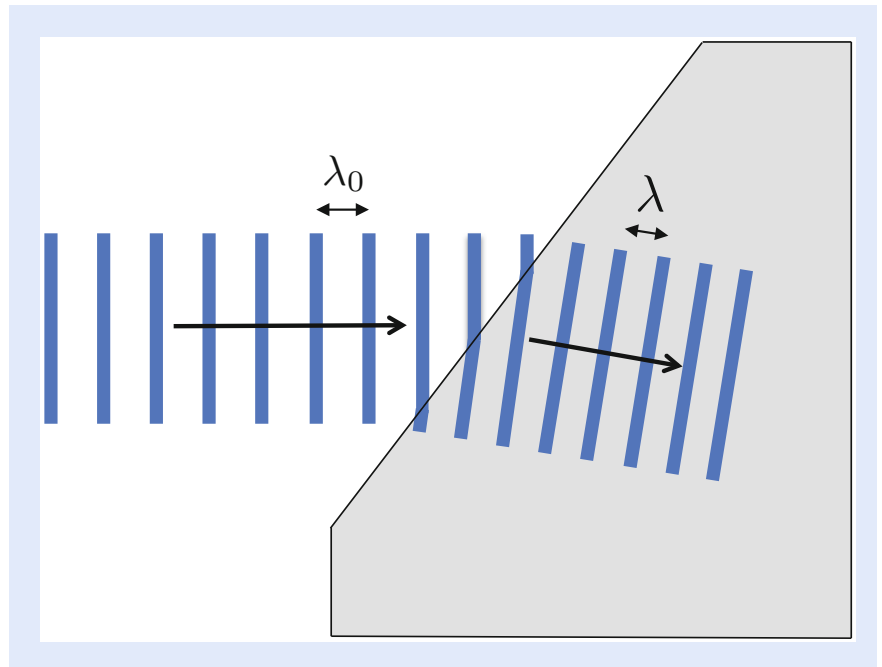
◻ **Fig. 15.8** *Refraction of a wave:* When a wave hits the surface of a medium with a different phase velocity, the wavelength has to change as the electric field oscillates in time always with the same frequency. This causes a change in the propagation direction of the wave
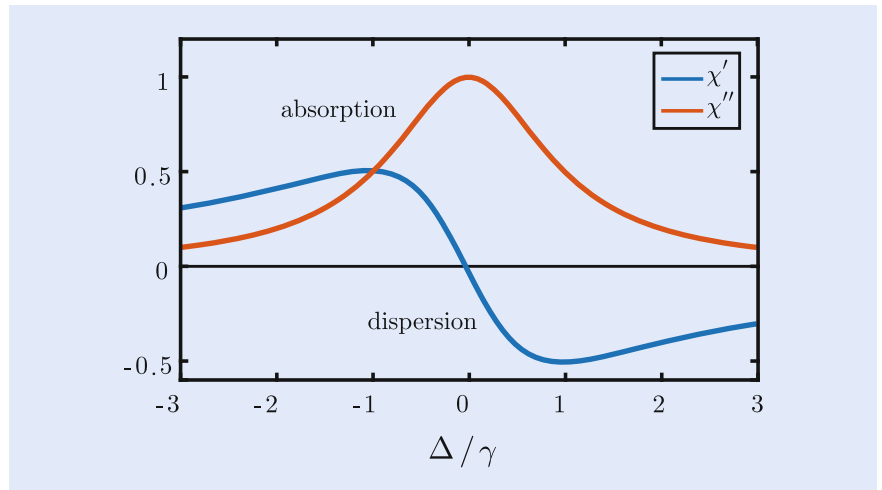
◻ **Fig. 15.9** *Absorption and dispersion of a two-level atom:* An atomic oscillator described by a two-level quantum systems leads to a strong absorption of light close to its resonance frequency. This is shown by the *red curve*, representing the imaginary part $\chi''(\omega)$ of the susceptibility as function of frequency ω. The refractive index $n(\omega) = 1 + \chi'(\omega)/2$, determined by the real part of the susceptibility $\chi'(\omega)$, is shown as the *blue curve*

complicated anti-symmetric shape. For frequencies above the resonance, $\omega > \omega_{ab}$, the medium leads to a reduction of the refractive index with respect to the background value, while below resonance, $\omega < \omega_{ab}$, the refractive index is enhanced. One notices the following from the figure: For large values of $|\Delta|$ the refractive index falls off much slower than the absorption, so for far off-resonant light only refractive effects of the medium matter. This is why even transparent media can still have a strong effect on the propagation of light. One of these effects
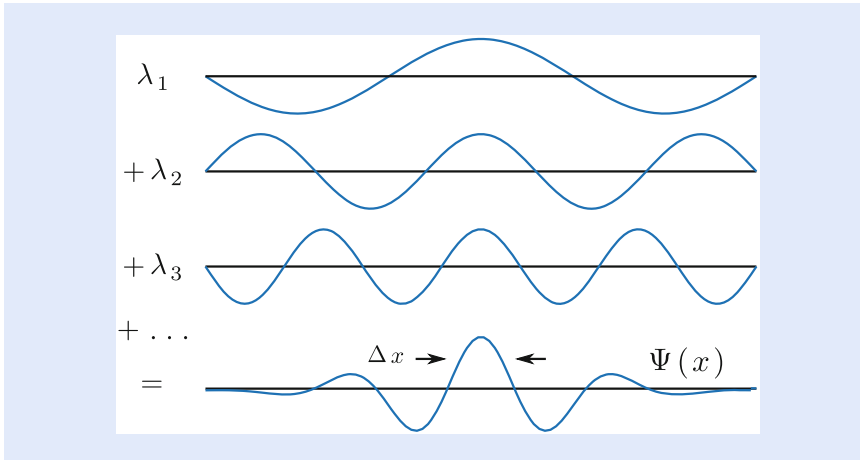
Fig. 15.10 *Wavepackets:* In order to create pulses of light with a finite spatial length, one needs to superimpose plane waves with slightly different wavelength in a proper way. In a medium the phase velocity of these components can differ. As a consequence the effective speed of the wavepacket is not given by the phase velocity but by the group velocity defined in Eq. (15.3)

is the refraction of a light beam at an interface between two media with different refractive indices. Another one is the modification of the propagation velocity of pulses, discussed in the following subsection.

15.3.2 Group Velocity

We have seen that the dependence of the refractive index on the frequency leads to different wavelength of light in a transparent medium as compared to free space. This dependence has another equally important effect, it determines the effective propagation speed of photon *wavepackets*. As illustrated in ▫ Fig. 15.10, one needs to superpose light waves with slightly different wavelength in order to create a wavepacket, i.e. a light pulse of finite length. In some sense we can envision photons as such wavepackets.

What is the propagation speed of such a wavepacket which consists of plane waves of different frequencies? In vacuum all frequency components propagate at the fundamental speed of light c_0, so wavepackets made of plane waves also propagate at this speed. But what about a medium, where each component has a different phase velocity $c(\omega) = c_0/n(\omega)$? It turns out that the slightly different phase velocities of each constituting plane wave cause the envelope of the pulse to move at the so-called group velocity v_{gr} which can be very different from the phase velocity $c = c_0/n(\omega)$. It is given by

$$v_{gr} = \frac{c_0}{n(\omega_0) + \left(\frac{\Delta n}{\Delta \omega}\omega_0\right)} \tag{15.3}$$

where ω_0 is the average frequency of the different components. The group velocity determines the effective speed of photons in a medium. When we talk about slow light, what we mean is light with a very small group velocity compared to c_0.

From Eq. (15.3) one recognizes that in addition to the refractive index itself, contained in the phase velocity $c = c_0/n(\omega)$, also the slope $\Delta n(\omega)/\Delta \omega$ enters at which the refractive index $n(\omega)$ changes by $\Delta n(\omega)$ when the frequency makes a small change $\Delta \omega$. As can be seen from ▫ Fig. 15.9 this slope is typically small far off resonance and the second term in the denominator of Eq. (15.3) is irrelevant. Thus in this frequency range the group velocity is essentially equal to the average

phase velocity. One also recognizes that on either side of the resonance, provided one is sufficiently far away from the resonance point, the slope of $n(\omega)$ is positive, which is called 'normal' dispersion. Here the group velocity is slightly smaller than the phase velocity. In order to see a dramatic difference between group and phase velocity one has to go closer to resonance. We immediately notice the problem with that: Whenever we are closer to resonance, the absorption of the medium becomes large and light gets quickly absorbed. In the following section we will explain how one can overcome this problem in an elegant way making use of an effect called EIT.

But before we proceed with this let's make a little side remark here: One notices that the situation is completely different in a very narrow frequency range around resonance: Here $\Delta n(\omega)/\Delta\omega$ is negative and large and the group velocity can become larger than the phase velocity. In principle it can even become larger than the vacuum speed of light c_0! But don't worry, this does not violate Einstein's principle of relativity as proven already by Arnold Sommerfeld [26]. One notices, for example, that in the same spectral region there is large absorption. As a consequence no signal can actually propagate faster than c_0.

15.4 Electromagnetically Induced Transparency

How can we get around the problem that strong effects on the group velocity of light seem to be always associated with large losses? The answer came from an effect known as EIT [2, 27, 28]. To understand what EIT is all about let us start with an analogy from mechanics [29]: Consider a mass m which can slide on a surface and is attached to a wall with a spring, as shown in ◘ Fig. 15.11a. This system forms an oscillator with frequency $\omega_0 = \sqrt{k/m}$, where k is the spring constant. Now assume that there is some friction, e.g. due to a rough surface on which the mass slides. If the oscillator is excited by a periodic force with frequency
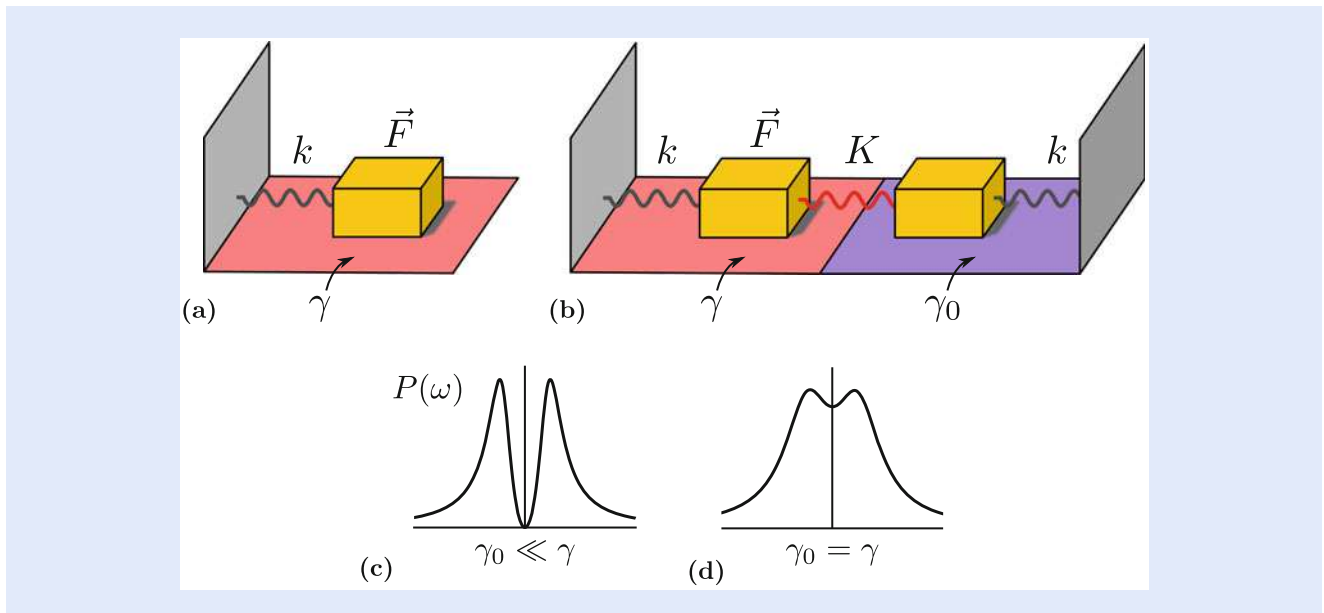
◘ **Fig. 15.11** *Coupled mechanical oscillators:* (**a**) A mechanical oscillator with resonance frequency $\omega_0 = \sqrt{k/m}$ driven by a periodic force **F** with frequency ω and subject to friction with energy loss rate γ generates a loss power spectrum similar to the absorption spectrum of a two-level system shown in Fig. 15.9. (**b**) If the mass is coupled to a second one with smaller friction (loss rate $\gamma_0 \ll \gamma$) a resonant periodic drive causes only the second mass to move and thus the power loss is dramatically reduced. (**c**) Loss power spectrum for $\gamma_0 \ll \gamma$. (**d**) If γ_0 equals γ, the total loss power spectrum is that of two independent absorption spectra slightly shifted in frequency (Adapted from [29])

ω close to the resonance frequency ω_0, energy is transferred to the oscillator and subsequently dissipated into heat due to the friction. The dissipated power $P(\omega)$ depends on the frequency mismatch between oscillator and drive frequency $\Delta = \omega - \omega_0$ and has a similar form as the absorption curve in ◘ Fig. 15.9.

Now suppose we couple this oscillator to another mass oscillating with the same frequency ω_0 using an additional spring with spring constant K (◘ Fig. 15.11b). Let us assume next that the second oscillator has little or no friction. If we now drive the first mass with a periodic force something interesting happens: Looking at ◘ Fig. 15.11c, where we have plotted the dissipated power again as function of frequency, one notices that if the driving frequency ω matches exactly the oscillator resonance frequency ω_0 little or no energy gets dissipated!

The reason is that the first mass, i.e. the one with friction, does not move at all. Only the second mass, the one with little or no friction, oscillates. It does this in such a way that it produces a force on the first mass exactly opposite to the external force F. The two forces compensate each other, and so the first mass stands still. One can say that the system of oscillators is driven into a dark mode, i.e. a mode without dissipation in which the lossy oscillator is not excited. Consequently the effect of friction is reduced considerably and no or little energy is dissipated.

The situation changes if the second mass also experiences a substantial friction. In particular, if the loss rates of both oscillators are the same, i.e. $\gamma_0 = \gamma$, the loss power spectrum is just the addition of two simple loss curves slightly shifted in frequency relative to each other, as shown in ◘ Fig. 15.11d. As long as γ_0 is not too large, there are two maxima corresponding to the two eigenfrequencies of the coupled oscillators. The splitting increases with $\sqrt{K}$, i.e. with the strength of the coupling. Most importantly if γ_0 vanishes or is very small, one can make the coupling very weak and still the dissipation essentially disappears when driving the first mass. This creates a situation where one can be close to resonance while there is almost no loss.

This principle can be translated to atomic oscillators. What is needed are two oscillators, one of them almost lossless, another one lossy, and the two oscillators need to be coupled by a 'spring'. This can be realized in a 3-level Λ-type system shown in ◘ Fig. 15.12. The atom-light coupling scheme is called Λ-type scheme because of the resemblance to the Greek letter Λ.

The first oscillator corresponds to the transition between the initially populated ground state g and the excited state e, as shown in ◘ Fig. 15.12a. This oscillator dissipates energy because of decay of the excited state e with rate γ, e.g. due to
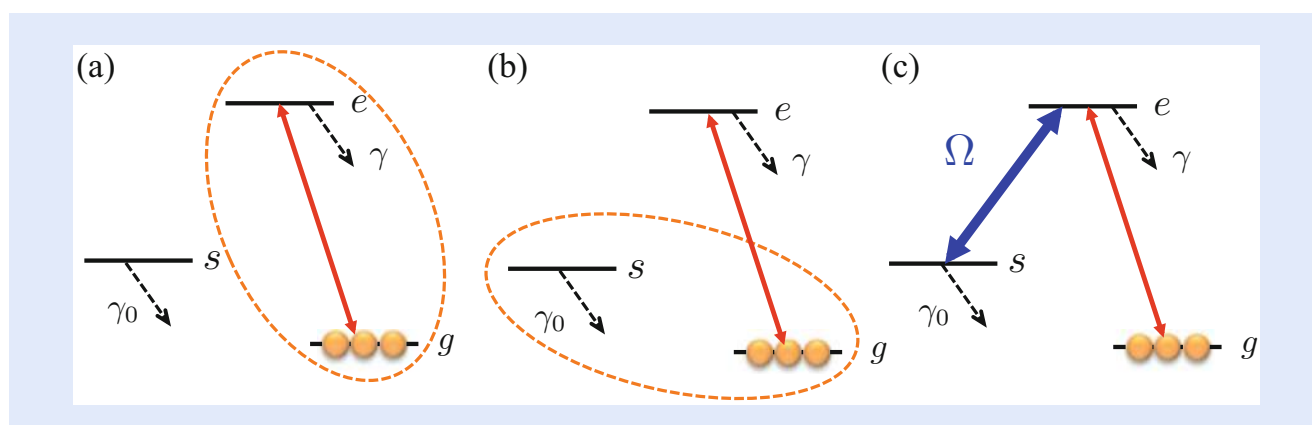
◘ **Fig. 15.12** *Principle of electromagnetically induced transparency:* (**a**) A lossy atomic oscillator consisting of the initially populated ground state g and an excited atomic state e is driven by a probe field (*red arrow*). (**b**) In a three-level Λ-type system there exists a second atomic oscillator between states g and s, which can be lossless or have very small losses, e.g. if s is a low-energy state. (**c**) Coupling the two oscillators by a control laser with a strength characterized by the Rabi frequency Ω produces a situation similar to that shown in ◘ Fig. 15.11b. Consequently the medium becomes (almost) transparent to the probe field
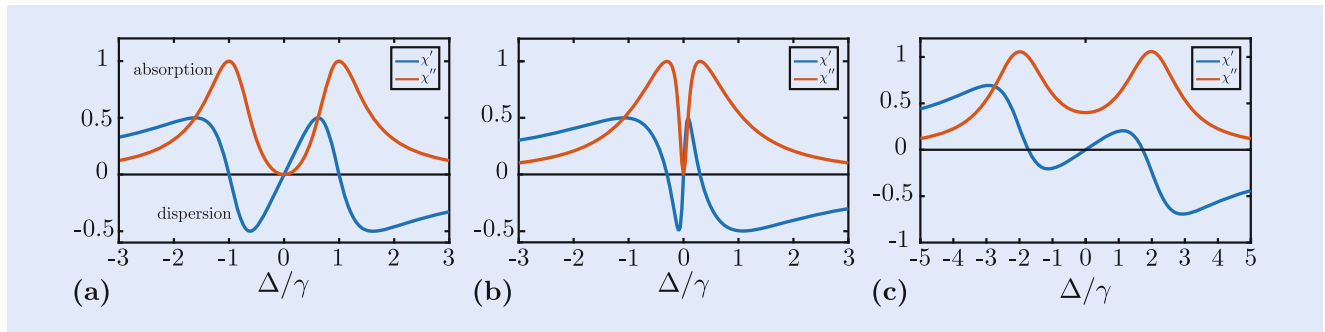
Fig. 15.13 *EIT versus two-level resonances:* (**a**) Real (χ') and imaginary (χ'') parts of the susceptibility of an EIT system characterizing the refraction and the absorption, respectively. Figure (**b**) shows the same with a smaller Rabi frequency of the drive field. For comparison we have shown in (**c**) the total susceptibility spectrum of two independent two-level systems with slightly shifted resonance frequencies. While χ', i.e. the index of refraction has a very similar shape in (**a**) and (**c**), there is an important difference in the absorption: In the EIT case it vanishes in between the two maxima, while for two two-level resonances it remains large

spontaneous emission. The oscillator is driven by the probe field (see Fig. 15.12a) corresponding to the external driving force in the mechanical picture from above. The ground state *g* together with another metastable ground state *s* forms the second oscillator (see Fig. 15.12b). The latter state *s* can be, e.g., a long-lived hyperfine spin state in the atomic ground state manifold, i.e. a low-energy state like *g*. Therefore the second oscillator is essentially lossless or has very small losses. Finally the role of the spring coupling the two oscillators is taken over by a coherent control laser field inducing transitions between the excited state *e* and state *s* (see Fig. 15.12c). The strength of this coupling is directly proportional to the amplitude of the electric field of the control laser, and the resulting splitting of the absorption peak (shown in Fig. 15.13) is denoted as Ω and is called Rabi frequency.

The absorption as a function of the probe field frequency ω relative to the resonance, expressed by the detuning $\Delta = \omega - \omega_0$ is shown in Fig. 15.13a, b as red lines. It consists of two absorption peaks like the spectrum of two coupled mechanical oscillators in Fig. 15.11c. Similar to the mechanical analog, the absorption shown in Fig. 15.13a, b vanishes exactly on resonance for $\gamma_0 = 0$, or is insignificant for small γ_0. This is quite remarkable since this means that despite the fact that one is very close to the resonance frequencies of the coupled system, the absorption is vanishingly small! Since a non-absorbing medium is transparent and since this effect is induced by the coupling of the two atomic oscillators by the drive laser, this phenomenon was called electromagnetically induced transparency or in short EIT.

The phenomenon of EIT has a widespread application in atomic and molecular physics and in optics. It can be used, for example, to make nonlinear optical processes much more efficient as it allows to operate close to atomic resonance without suffering from absorption. Some of the interesting applications will be discussed in detail in the following section.

15.5 Slow Light, Stored Light and Dark-State Polaritons

15.5.1 Slow Light

As we have discussed in Sect. 15.3 the absorption spectrum is associated with the imaginary part of the susceptibility. Figure 15.13a, b show the absorption spectrum of the atomic medium at an EIT resonance. The spectrum consists of two lines separated by an amount proportional to the strength of the driving field (Ω)

and in between these two peaks the absorption goes to zero. Also shown is the real part of the susceptibility as a function of frequency, which is called dispersion. In ◘ Fig. 15.13c we have plotted the absorption and dispersion spectra of two uncoupled oscillators with slightly different frequencies. We notice that the dispersion curves look qualitatively very similar in ◘ Fig. 15.13a, c. In particular the real part of the susceptibility, i.e. the refractive index, has a positive slope around $\Delta = 0$. In the case of two uncoupled two-level systems this just results from superposing the below-resonance tail corresponding to one oscillator with the above-resonance tail of the other. The most important difference between the case of two uncoupled resonances and EIT is that in the former case the absorption does not vanish in between the two resonances.

The dispersion curve has a remarkable feature right on resonance. It has a linear slope that can become very steep. In fact the closer the two absorption peaks are, the steeper is the dispersion curve. From Eq. (15.3) we notice that a steep slope of the index of refraction leads to a very large denominator in the expression for the group velocity. This means close to resonance the medium is, on the one hand, transparent due to EIT and at the same time the group velocity can be extremely small. This is the origin of ultra-slow light in EIT.

The value of the group velocity in an EIT medium is determined by the general equation (15.3) with the second term in the denominator being much larger than the first one, giving

$$v_{\mathrm{gr}} \approx \frac{c_0}{\omega_0 \frac{\Delta n}{\Delta \omega}} \quad \sim \quad \frac{\Omega^2}{\rho}, \tag{15.4}$$

where Ω is the Rabi frequency of the drive laser, and ρ is the density of atoms. By turning down the intensity of the drive laser, i.e. by reducing Ω, or alternatively by increasing the atom density ρ, one can reach very small values of the group velocity. This can also be seen from ◘ Fig. 15.13a, b: Reducing Ω the separation between the absorption maxima decreases making the dispersion curve steeper in the center and hence the group velocity smaller.

The first experiments measuring the group velocity reduction in EIT where done by Harris et al. [3] in an atomic vapor cell reaching $v_{\mathrm{gr}} = c_0/170$. The smallest group velocities achieved so far in experiments are obtained using very cold and dense clouds of atoms such as in a Bose Einstein Condensate and are on the order of 10 m/s, i.e. $v_{\mathrm{gr}} = c_0/30,000,000$ [4].

When a light pulse enters a medium with a small group velocity it will be transmitted if its central frequency is close enough to the resonance and if its spectral width, i.e. the spread of frequencies associated with any pulse of finite duration, is much less than the distance between the two peaks in the absorption spectrum shown in ◘ Fig. 15.13. The very steep slope of the refractive index has also a profound effect on the spatial shape of the pulse, as illustrated in ◘ Fig. 15.14. When the pulse just enters the medium its front end will propagate with the group velocity v_{gr}, while its back end still propagates with the vacuum speed of light. As a consequence the pulse will be dramatically compressed in length inside the medium. The compression ratio is given by

$$l/l_0 = v_{\mathrm{gr}}/c_0. \tag{15.5}$$

This resembles a situation where a number of vehicles moving fast on a highway suddenly approaches the beginning of an area with restricted speed. At this point the bunch of cars is compressed since when the first cars have already entered the area of restricted velocity, the ones at the back still drive at full speed. If the velocity of the vehicles is reduced by half, the distance between them becomes twice
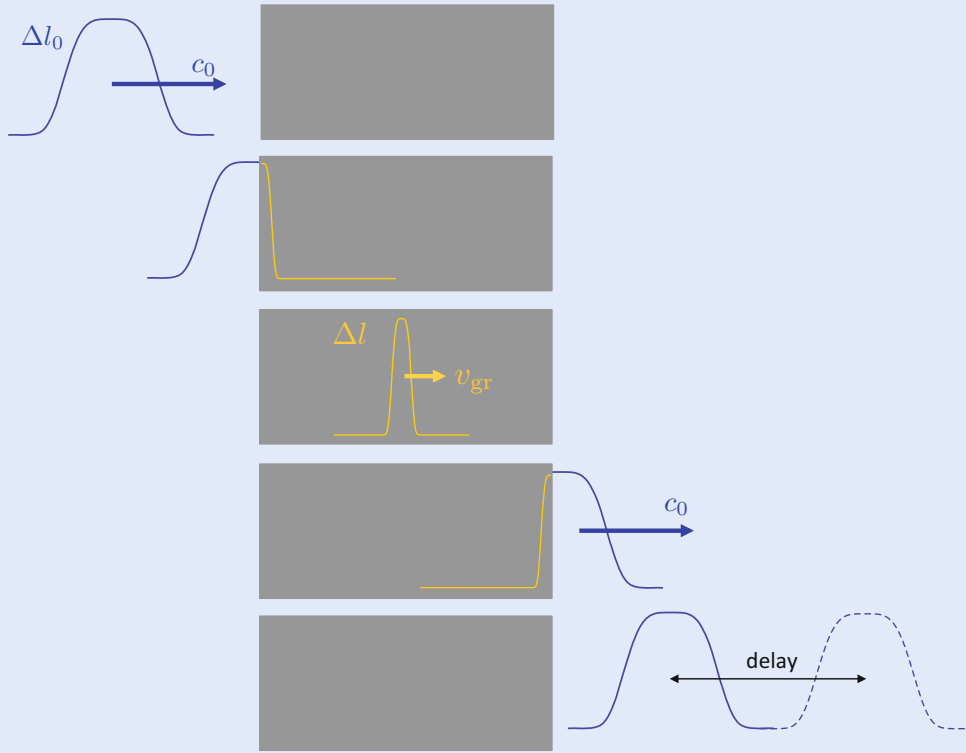
15

🔲 **Fig. 15.14** *Pulse compression:* When a light pulse enters a medium with a reduced group velocity it becomes spatially compressed by the ratio v_{gr}/c_0. When the front end is already in the medium it propagates with v_{gr}, while the back end still moves with the much larger speed c_0. This causes the pulse to shrink in space. The opposite is happening when the pulse leaves the medium

smaller, so the compression factor is 1/2. Since in the atomic media the light can be slowed down to such extremely small velocities as $v_{gr} \approx c_0/30,000,000 = 10$ m/s, the incoming pulse of fast light with original length l_0 of about 1 km will be compressed to a pulse of length l of about 30 μm (!). In this way even very long pulses of light can be made to fit into a small-sized material, such as an elongated (cigar shape) Bose Einstein Condensate of sodium atoms used in the 1999 experiment by the group of Hau [4]. When the pulse leaves the medium the opposite effect happens. The leading edge travels fast since it is in free space and the back end lags behind as it is still inside the medium. At the end the outgoing pulse has the same length as the incoming one, at least under ideal conditions. This is again like the spatial decompression of a bunch of cars when leaving the area of restricted speed on the highway.

15.5.2 Stopped Light and Quantum Memories for Photons

As can be seen from Eq. (15.4) the group velocity of slow light can be controlled by the strength of the coupling laser or the density of the medium. So what would happen if we turn the coupling laser off while the probe pulse propagates inside the EIT medium? The medium becomes immediately opaque for the probe light and thus we expect no probe field to survive. This is indeed the case. So does this mean the probe pulse is lost? Surprisingly this does not happen!

At the entrance of the medium most of the incoming photons are transferred to atomic excitations during the slowing down. In this process the pulse is also substantially compressed in space, so that it fits inside the medium. The atomic
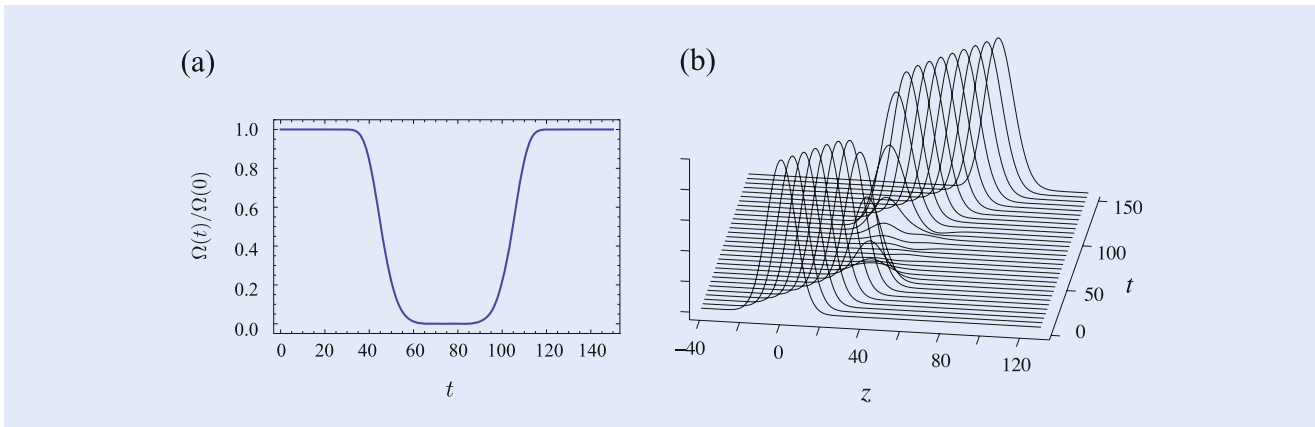
◘ Fig. 15.15 *Light storage and retrieval:* When the strength of the control field is switched off smoothly while the slow-light pulse is in the medium (**a**) the pulse stops but also all photons disappear (**b**). If the control field is, however, switched on again at a later time, the pulse miraculously reappears and continues to propagate as a slow-light pulse in the medium

excitations, carrying information about the incoming pulse, travel together with the remaining photons. If the control field is switched off while the compressed pulse is still inside the medium the light disappears, i.e. no probe light survives. But if the control field is switched on again at a later instant of time, the pulse miraculously reappears! This is shown in the numerical simulation of ◘ Fig. 15.15. The right-hand side shows the propagation of the compressed light pulse inside the medium when the control laser is switched off and on again as illustrated on the left-hand side. So obviously we have somehow managed to stop (or more specifically to store) the light pulse for a while and sent it off its way a while later.

This remarkable phenomenon of light stopping (storing) was theoretically predicted in 2000 [12] and experimentally demonstrated in 2001 by two groups at Harvard University [13] and the Roland Institute of Science [14]. ◘ Figure 15.16 is a reproduction of the data obtained in one of these experiments from [14]. In these experiments a storage time of up to half a millisecond was reached. In 2009 the group of Immanuel Bloch at the Max Planck Institute for Quantum Optics in Garching, Germany in collaboration with colleagues from Israel has increased the storage time to 240 ms using ultracold atoms in a Mott insulating state in a three-dimensional optical lattice [30]. In the so-called Mott insulating phase atoms are particularly protected from perturbations such as collisions and diffusion, which leads to the prolonged storage duration. The current record for storage times is 1 min [31]. It has been obtained in doped glasses, where impurity atoms behave almost like free atoms in a vapor with the advantage that they do not move as in the Mott insulating state discussed above, and the atomic density is higher than in a gas.

15.5.3 Slow-Light Polaritons

We have seen in ◘ Sect. 15.3 that the microscopic picture of light propagation in a transparent medium is that each atomic oscillator absorbs a tiny little bit of an incoming photon, stores it for a short moment and releases it again with a small time delay as electromagnetic energy. The amount of the time delay is determined by the ratio of group velocity and vacuum speed of light. Furthermore the reduction in the group velocity also leads to a spatial compression of a photon pulse at the entrance to the atomic medium, as discussed in the previous subsection. If a light pulse is spatially compressed without increasing its amplitude, this
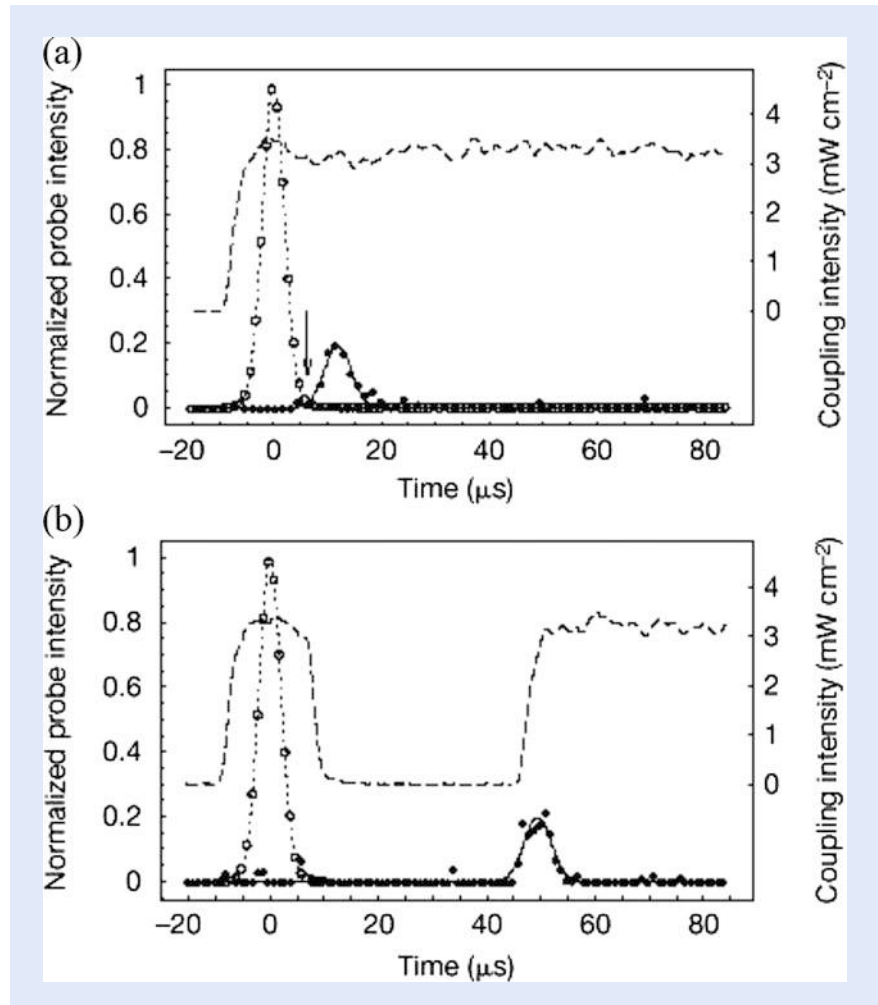
■ **Fig. 15.16** *Light storage experiment:* Reproduction from one of the first experiments on light storage [14] (with permission of the journal Nature). Shown are the control field (*dashed*), the input probe pulse (*open circles and dotted line*) as well as the output probe pulse (*full circles and full line*). The *top curve* shows the pulse delay when the control field is on all the time, the *lower curve* shows the storage of the probe pulse when the control field is switched off and subsequently on again after some time

means that its content of photon energy decreases, i.e. the total number of photons contained in the pulse must be reduced according to the spatial compression. Where do these photons go if the medium is not absorbing? The answer is: They are temporarily stored in the form of atomic spin excitations.

In an usual transparent medium, such as glass, the ratio between the number of atomic excitations and photons is fixed and is very tiny. In an EIT medium this ratio can be large and it can be dynamically modified by tuning the strength of the control laser or by changing the atomic density. The best way to describe this is not to think in terms of photons and atoms separately but in terms of a combined quasiparticle, called polariton, containing a contribution due to both a photon and an atomic spin excitation, i.e. the excitation of the atom from the initially populated atomic ground states g to another ground state s [12, 32, 33]. The polariton picture has been introduced in [12] to describe storing and releasing of slow light following an earlier single-mode treatment [32] used to describe Raman adiabatic passage between the atomic ground states which did not include pulse propagation. We can visualize this polariton as a vector with two components, the

$$\Psi = \cos\theta\,\mathcal{E} + \sin\theta\,\mathcal{S}$$
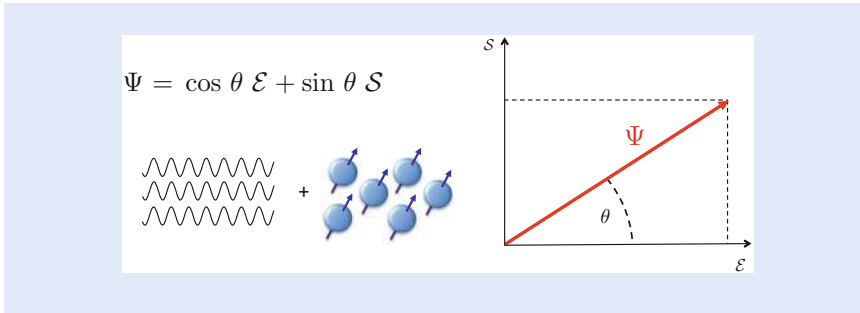
Fig. 15.17 *Slow-light polariton:* Slow and stored light can most easily be understood in terms of quasiparticles called dark-state polaritons introduced in [12]. They are a superposition of the electric field $\mathcal{E}$ of the probe pulse and an atomic spin excitation S, like a vector in a two-dimensional plane. The mixing angle θ depends on the strength of the control laser and the atom density and thus can be changed. The angle θ also determines the properties of the polaritons, such as their velocity, see Eq. (15.6)

electric field $\mathcal{E}$ and an atomic excitation S indicated in ▪ Fig. 15.17, where the mixing angle θ determines the ratio between the photonic and atomic components making up the polariton. Since the polariton is only partially a photon and only the photons move, the propagation speed is determined by the fraction of photons comprising the polariton:

$$v_{\mathrm{gr}}/c_0 = \cos^2\theta = \frac{\Omega^2}{\alpha\rho + \Omega^2}. \tag{15.6}$$

The group velocity v_{gr} is evidently less than that of pure photons. (Here α is some constant, which is not relevant for the present discussion.) When the probe pulse is outside the medium, where $\rho = 0$, it can be interpreted as a polariton with $\cos^2\theta = 1$ representing a pure photon without any atomic component. When it enters the medium, e.g. the cloud of ultracold atoms in the BEC experiment of Hau et al. [4], the density ρ increases smoothly in space. As a consequence the polariton turns smoothly into a mixed atomic-photonic excitation, with a large atomic component.

Since Ω, determining the group velocity in Eq. (15.6), is a tunable parameter, the composition of the slow-light polariton can be modified further while the pulse is propagating inside the medium. In the case of slow light, $\cos^2\theta$ is much less than unity already when the probe pulse has just entered the medium and most of the excitations which were originally photons propagate as an atomic excitation. By further reducing the strength of the control laser Ω from the initial value where $\cos^2\theta$ is finite (yet much smaller than unity) all the way to zero, the slow-light polariton looses its photon component altogether and reduces to a pure atomic excitation which does not move any more. By switching on the control laser again at a later time, $\cos^2\theta$ becomes finite again (yet much smaller than unity). The slow-light pulse resumes its motion inside the medium until reaching the end of the atomic cloud where it finally converts completely into a fast, purely photonic pulse. This explains the reappearance of the light pulse, when the control field is turned back on again. As shown in ▪ Fig. 15.18, illustrating the stopping and reacceleration of a slow-light pulse while inside the medium, the light storage and retrieval sequence becomes very clear in terms of the polariton picture. The polariton is there all the time. It only changes its character, first from fast light to slow light, and then to a frozen atomic spin excitation and finally back to a slow polariton and eventually to fast light again.
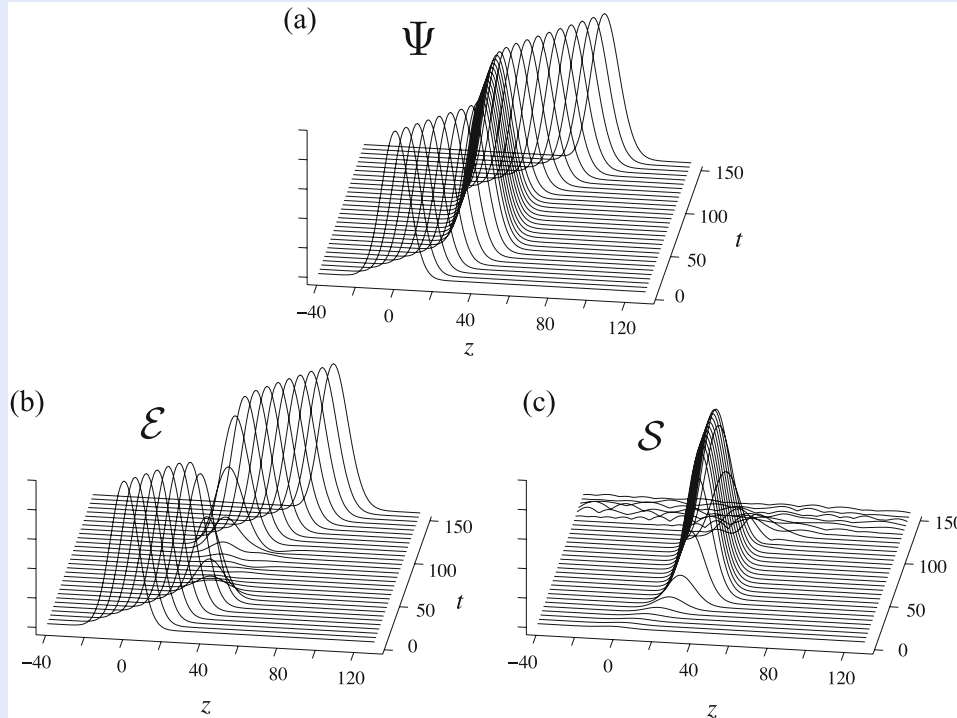
Fig. 15.18 *Polariton picture of light storage:* Light storage and retrieval, as shown in Fig. 15.15. This time also the propagation of the polariton (**a**) and the spin excitation (**c**) are shown. One recognizes that light storage is nothing else than a smooth conversion of the slow-light polariton from a polariton containing an electric-field component into a pure atomic excitation and back

15.6 Stationary Light

We have seen in the last section that a light pulse can be brought to a complete stop without loosing the information it contains by storing it in an atomic excitation. When the light pulse is at a halt, no photon is left in the medium anymore, so the polariton becomes entirely an atomic excitation. In the example of cars and trailers this corresponds to the case when all cars are converted into trailers. Thus there is no car left to pull and everything comes to a stop. There is, however, a way to keep the cars from driving without converting all of them to trailers. If two cars driving in opposite directions pull the same trailer, their forces can compensate and neither of the two can move forward. This is exactly what is happening in a situation called *stationary light*, which we will explain in the following.

A very interesting aspect of stationary light is that it mimics the behaviour of a massive quantum particle described by the Schrödinger equation for the amplitude of the stationary light polariton Ψ_{ss}:

$$ih \frac{d}{dt} \Psi_{ss} = -\frac{h^2}{2m^*} \frac{d^2}{dx^2} \Psi_{ss}. \tag{15.7}$$

Unlike photons in free space, which always propagate at the speed of light c, massive particles can stand still, or more precisely, as we are talking about quantum particles, can have a zero average velocity. Importantly the effective mass m^* of the stationary light polaritons is not a fixed quantity such as the mass of an electron or a proton, but is a tunable parameter. It can be changed by the strength of the control laser fields. This property makes stationary light an
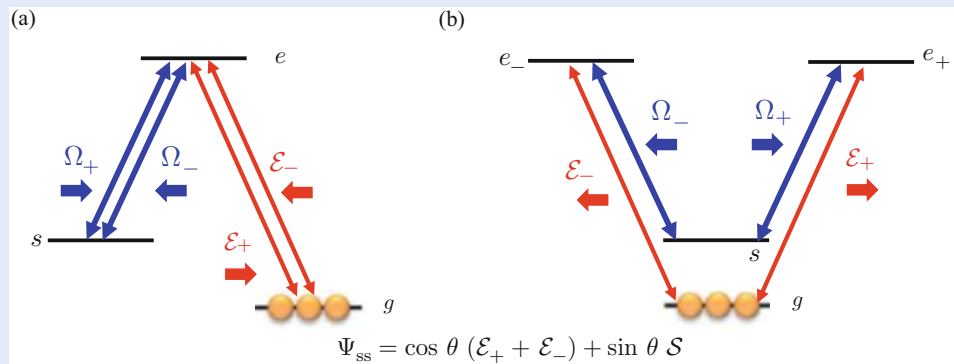
◘ Fig. 15.19 *Stationary light:* If two counter-propagating drive fields of equal strength couple a Λ system of atomic levels (**a**), two counter-propagating probe field components of equal strength are formed. These fields interfere with each other and form a stationary-wave pattern. The same happens if the two drive fields have orthogonal polarizations and couple to two different transitions in a four-level (double Λ) scheme (**b**). In this case two counter-propagating probe fields are generated which also have orthogonal polarizations, but which nevertheless form a stationary-wave pattern. In these ways an excitation wavepacket is created which does not move and has still a non-vanishing electric-field component

interesting model system for analyzing fundamental properties of massive quantum particles.

What is the physics behind stationary light? Suppose there are two (rather than one) control laser beams of equal strength $\Omega_+ = \Omega_- = \Omega$ and two (rather than one) probe fields $\mathcal{E}_\pm$ inducing transitions in a three-level Λ-system or a four-level system, as shown in ◘ Fig. 15.19.

The four-level system can be viewed as two Λ sub-systems, one for fields propagating in the forward direction (+), and another sub-system for the fields propagating in the backward direction (−). Since the two control fields have the same amplitude, so do the probe fields. In each of these Lambda systems the respective pairs of control and probe fields induce a transition from the ground atomic state g to the metastable state s (see ◘ Fig. 15.19). In such a situation the two counter-propagating probe beams drive the same atomic transition $g \rightarrow s$. Since the amplitudes of both probe fields are the same, each photon propagating forward has its counterpart, a photon propagating backward, and a stationary pattern of light is formed, frozen in the medium. This is as if two racing cars driving in opposite directions try to pull the same trailer but are not able to move it since their forces compensate (see ◘ Fig. 15.3).

For stationary light it is important that the counter-propagating probe fields are coupled to each other by the atomic medium. To see this let us draw an analogy with the string of a guitar: When a guitar player pulls the string at some place, *two* waves of equal frequency are created which propagate along the string in opposite directions. If two wavepackets of equal strength and opposite propagation directions are superimposed, a standing wave forms, but only for the short period of time for which they overlap. The two wavepackets would continue to propagate each in its own direction. Soon they would not overlap anymore and would be two spatially separated wavepackets. To prevent this another element is needed: At the points where the string is fixed to the body of the guitar, the wavepackets get reflected and the effect of this is a true standing wave that does not smear out. In a similar manner, one could produce a standing wave of light by confining the radiation in a resonator between parallel mirrors, so that the forward propagating light is permanently reflected to the backward propagating direction and vice versa. This principle is used, e.g., in a laser allowing the light to pass many times the lasing medium.

Stationary light also involves a permanent reflection of one component into the other but with no mirrors, and one can think of a kind of a mirrorless resonator. So what takes over the role of the fixing points of the guitar string or the mirrors reflecting the light? In fact we have here a whole periodic sequence of 'fixing' points, which causes reflection. In the case of a simple Λ-scheme, shown in ◘ Fig. 15.19a, the control lasers form a stationary *intensity* pattern that oscillates in space. Thus there is a periodic grating where the total control field intensity vanishes, which also means that in a periodic spatial pattern there are points without EIT for the probe light. This periodic array acts in a similar way as an absorption grating and reflects the forward and backward propagating components of the probe field. In the case of the four-level scheme, shown in ◘ Fig. 15.19b, the situation is somewhat different. Here the two control fields are not only propagating in opposite directions, but they also have opposite circular polarizations. Now, superimposing two light waves of equal intensity and with opposite circular polarization results in constant total field intensity with a linear polarization. Yet, since the two control beams propagate opposite to each other, the linear polarization rotates in space, forming a polarization grating. This polarization grating has the same effect as the intensity grating in the case of the simple Λ-scheme, it reflects forward- and backward propagating components into each other making the light stationary.

Stationary light has been first observed in 2003 by the group of Mikhail Lukin at Harvard University [17] using a Λ-type atom-light coupling which involves pairs of counter-propagating control (probe) beams with the same frequency, shown in ◘ Fig. 15.19a. One difficulty of these experiments is to make the 'non-moving light' visible. A trick used here is that the stationary light tends to excite also further off-resonant transitions to other excited states with a small probability. These excitations are then visible due to the spontaneous emission from these states.

Another form of stationary light, called bichromatic stationary light was observed in 2009 by the group of Ite Yu at the National Tsing Hua University in Taiwan [20] using a double Λ coupling scheme, as shown in ◘ Fig. 15.19b. Here the frequency (or color) of the two control fields and, respectively, the two probe fields were different, thus the name 'bichromatic'. Stationary light pulses maximize the interaction time and thus can provide a considerable interaction efficiency even at a single-photon level. Interaction of two stationary light pulses through the medium was experimentally demonstrated by the same group 3 years later [21].

15.7 Multi-Component Slow Light

We have seen that slow light can be turned into something that behaves like a massive quantum particle. It is known from quantum physics that certain particles can show up in different forms, i.e. they can have different internal states. Electrons, for example, possess two different spin states, spin-up and spin-down states. In a bit oversimplified picture the spin of a particle can be viewed as a tiny gyroscope resulting from rotation of the particle around its center. Such a rotation is often accompanied with a magnetic dipole, so an electron represents a little magnet pointing up or down depending on the spin state relative to the chosen axis. More exotic particles can have not only spin but also other internal degrees of freedom, such as isospin, colour or flavour. So an interesting question is: Can we give slow light internal properties such that it mimics massive quantum particles with, e.g., spin? The answer is yes, and this makes slow light an even more interesting object for quantum physicists. We note that in quantum mechanics the spin of, e.g., an electron is a relativistic effect, so slow light with spin can be used to investigate relativistic quantum physics.
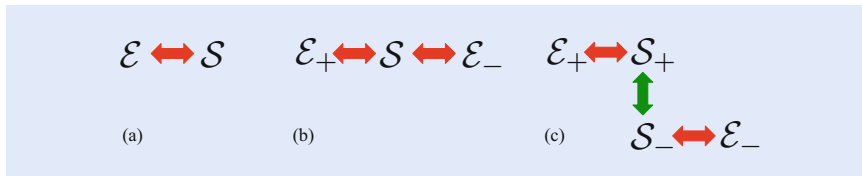
■ **Fig. 15.20** *Slow (a), stationary (b) and two-component (c) slow light:* In (**a**) a single probe field $\mathcal{E}$ is coupled to a single atomic coherence S. The radiation has to push the atomic coherence forwards and thus the light slows down. In (**b**) two counter-propagating probe beams $\mathcal{E}_\pm$ drive the same atomic coherence characterized by the amplitude S. One probe field pushes the atomic coherence forwards and the other backwards. The velocities of the probe photons compensate leading to stationary light. In (**c**) two counter-propagating probe beams $\mathcal{E}_\pm$ drive two different atomic coherences characterized by the amplitudes S_+ and S_-. If there is a coupling between these coherences indicated by the *green double arrow*, two-component stationary light is formed
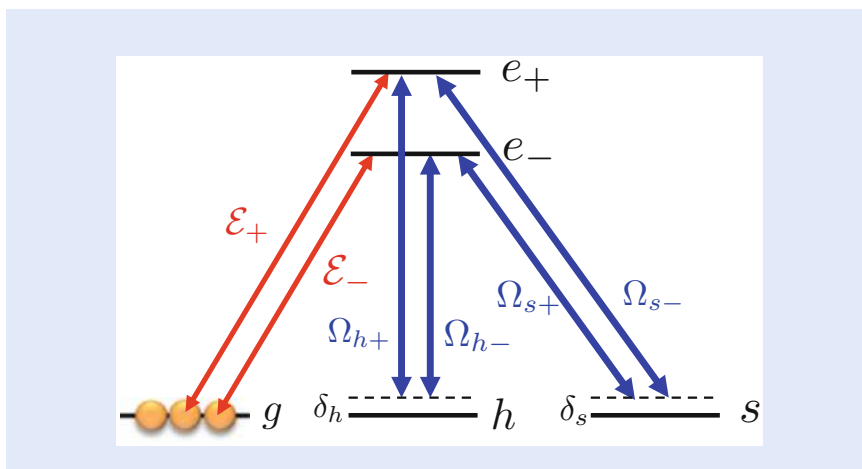
■ **Fig. 15.21** *Two-component slow light:* Atom-light coupling scheme of the double-tripod type for implementation of two-component stationary light adapted from [24]. The scheme involves three atomic ground states g, s and h coupled to two excited states $e_\pm$ by six fields: a pair of counter-propagating probe beams $\mathcal{E}_\pm$, as well as two pairs of counter-propagating control beams $\Omega_{s\pm}$ and $\Omega_{h\pm}$

Slow light as introduced in ■ Sect. 15.5 and the stationary light discussed in ■ Sect. 15.6 both involve only one spin component associated with a transition from the initially populated ground state g to one other ground state s, and described by the amplitude S, as illustrated in ■ Fig. 15.20a, b. This represents a single normal mode of oscillations of the coupled atom-light system (a single polariton) even though there are two counter-propagating probe fields, as in the case of the stationary light depicted in ■ Fig. 15.20b.
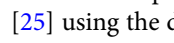
In order for stationary light to have two components, the counter-propagating probe fields $\mathcal{E}_\pm$ (together with a number of control beams) should drive *two different* spin coherences described by two amplitudes $S_\pm$. This is illustrated in ■ Fig. 15.20c. Two-component stationary light can be implemented using a tripod [23] or a double-tripod [24] atom-light coupling scheme, the latter shown in ■ Fig. 15.21. Here one has two pairs of counter-propagating control fields with Rabi frequencies $\Omega_{s\pm}$ and $\Omega_{h\pm}$ inducing the atomic transitions $s \to e_\pm$ and $h \to e_\pm$, respectively. Compared to the double Λ scheme used for stationary light (■ Fig. 15.19b) now there is an extra pair of counter-propagating control laser beams $\Omega_{h\pm}$, as well as an extra atomic ground state h. This leads to EIT for a pair of counter-propagating probe fields $\mathcal{E}_\pm$ inducing transitions (together

with the control fields) from the initially populated ground state g to two superpositions of the initially unpopulated atomic ground states s and h. Consequently the fields $\mathcal{E}_+$ and $\mathcal{E}_-$ drive different spin coherences characterized by the amplitudes S_+ and S_-.

If S_+ and S_- were not coupled to each other, the two probe beams would propagate in opposite directions slowly and independently from each other. The coupling emerges though a two-photon detuning $\delta = \delta_s = -\delta_h$ shown in ◘ Fig. 15.21. The corresponding two types of polaritons behave like particles with positive and negative effective masses, i.e. like electrons and positrons representing particles and antiparticles in the relativistic Dirac theory. Thus the two-component (spinor) slow-light polaritons Ψ obey an effective one-dimensional Dirac equation

$$i\hbar \frac{\partial}{\partial t}\Psi = \left(i\hbar v_{\mathrm{gr}}\sigma_z \frac{\partial}{\partial z} + m^* c^{*2}\sigma_y\right)\Psi, \qquad \Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix}, \qquad \textbf{(15.8)}$$

where σ_z and σ_y are the 2×2 Pauli matrices. For zero two-photon detuning δ, the two polaritons propagate in opposite directions with an effective speed $c^* = v_{\mathrm{gr}}$ given by the slow-light group velocity. A non-vanishing two-photon detuning introduces a coupling between the counter-propagating polaritons, providing a particle–antiparticle type dispersion with a variable mass $m^* = \hbar\delta/v_{\mathrm{gr}}^2$ [24]. An important feature of spinor slow light is that the relevant scales of velocity, energy and length, where relativistic effects start to matter, are very different from the values for say electrons. The effective 'vacuum speed of light' $c^* = v_{\mathrm{gr}}$ can now be a few meters per second instead of 300,000 km/s. The relativistic rest energy $m^* c^{*2} = \hbar\delta$ can be many orders of magnitude smaller than that for an electron, making it possible to observe particle–antiparticle pair generation processes in a conventional laser lab. Finally the relativistic length scale, called Compton length $\lambda_C^* = \hbar/m^* c^*$, is now large enough to be resolved in laboratory experiments as opposed to the value of 10^{-12} m for an electron. The possibility of a locally adjustable mass allows furthermore to observe a number of other interesting phenomena. For instance, if the mass m^* of the Dirac particle suddenly changes at a certain point in space from the value $+ |m|$ to $- |m|$, a localized, topological mid-gap (zero-energy) state is created. If m^* is a randomly varying function of space with a vanishing mean-value, there exist mid-gap states with unusual correlations [23, 34, 35].

Two-component slow light has been recently implemented in an experiment [25] using the double-tripod coupling scheme, like the one shown in ◘ Fig. 15.21 but with co-propagating rather than counter-propagating control and probe laser fields. Oscillations due to an effective interaction between the two components of the probe field have been observed revealing the two-component nature of the slow light. It was demonstrated that the double-tripod scheme enables precision measurements of frequency detunings. Furthermore a possible application of the double-tripod scheme as quantum memory/rotator for a two-colour qubit was experimentally demonstrated. This offers potential applications in quantum computation and quantum information processing.

15.8 Quo Vadis Slow Light?

Light is fascinating! Light has very many uses and modern life would be unthinkable without them. Thus there is plenty of reason for us to celebrate the Year of Light. We believe that the applications of slow light based on EIT and its generalizations, which we have discussed in this chapter of the book, are important additions to this list of reasons. We have seen that coupling light to atomic media, which are specially prepared by external laser fields, allows us to dramatically

modify the property of photons. We can change their effective propagation velocity, can store them or more precisely their information content with important applications for quantum information networks based on light, and we can turn them into massive quantum particles with tunable mass. Finally we can even use them to model relativistic quantum particles with spin.

It is interesting to note that EIT and slow light are not restricted to light in the optical frequency spectrum coupled to atoms. EIT can also be generated in other type of coupled oscillators, such as meta-materials build up of periodic arrays of small metallic antennas [36]. This allows to access the microwave part of the electromagnetic spectrum. On the other hand, the storage and release of light can also be carried out beyond atomic systems. Recently the conversion of light pulses into mechanical excitations of a silica optomechanical resonator and the subsequent retrieval of radiation using a method closely related to the EIT was experimentally demonstrated [37].

All phenomena we have discussed so far in this chapter address the single-particle properties of slow light, i.e. properties of *individual* photons. Yet it is also highly desirable to make photons interact with each other sufficiently strongly. Strong and controlled interactions between individual photons would, e.g., allow to implement quantum logic operations in the so-called quantum gates, the second important ingredient next to a quantum memory for photon-based quantum information technology. Interactions are also crucial for most applications of slow light to fundamental science. Several ideas have been put forward here to exploit the properties of slow light for implementing strong interactions. For example, the possibility offered by EIT to operate close to atomic resonances without suffering from absorption can be exploited to enhance nonlinear optical processes in atomic media [21, 38–44]. Another very promising direction is to combine EIT with the so-called Rydberg atoms. Here the atomic state *s* populated during the propagation and storage of light is not a hyperfine (spin) ground state of an atom, but rather a Rydberg state corresponding to a very high atomic level close to the ionization threshold. Such a state is metastable and has a very long lifetime. Atoms in Rydberg states exhibit very strong and long-range dipole–dipole interactions. This property is carried over to slow-light polaritons, whose spin component contains the Rydberg state, thus making these Rydberg polaritons strongly interacting [45]. The strongly nonlinear and nonlocal interaction between Rydberg polaritons has been observed in a number of recent experiments [46–50]. This opens many more fascinating applications in fundamental science and in quantum technology, and we anticipate a bright future for slow light.

15.9 Conclusions

In this chapter we have explained what slow light is and what it is good for, how to understand the physics of it and how one can practically make light go so slow. To answer these questions, we used simple pictures, on the one hand, and supplemented them with a little bit of details, on the other hand, for those who want to go slightly deeper into the field. Subsequently we discussed recent generalizations of slow light, such as stationary and spinor slow light which are interesting model system and can be used to understand more complex quantum systems. The chapter also presents important applications of the slow light in photon-based quantum information technology.

15

References

1. Boller KJ, Imamoğlu A, Harris SE (1991) Observation of electromagnetically induced transparency. Phys Rev Lett 66:2593–2596
2. Harris SE (1997) Electromagnetically induced transparency. Phys Today 50(7):36–42
3. Harris SE, Field JE, Kasapi A (1992) Dispersive properties of electromagnetically induced transparency. Phys Rev A 46:R29–R32
4. Hau LV, Harris SE, Dutton Z, Behroozi CH (1999) Light speed reduction to 17 metres per second in an ultracold atomic gas. Nature 397:594–598
5. Kash MM, Sautenkov V A, Zibrov AS, Hollberg L, Welch GR, Lukin MD, Rostovtsev Y, Fry ES, Scully MO (1999) Ultraslow group velocity and enhanced nonlinear optical effects in a coherently driven hot atomic gas. Phys Rev Lett 82:5229–5232
6. Budker D, Kimball DF, Rochester SM, Yashchuk VV (1999) Nonlinear magneto-optics and reduced group velocity of light in atomic vapor with slow ground state relaxation. Phys Rev Lett 83:1767–1770
7. Boyd RW, Gauthier DJ (2002) "Slow" and "fast" light. In: Wolf E (ed) Progress in optics, vol 43. Elsevier, Amsterdam, pp 497–530
8. Lukin MD (2003) Colloquium: trapping and manipulating photon states in atomic ensembles. Rev Mod Phys 75:457–472
9. Fleischhauer M, Imamoğlu A, Marangos JP (2005) Electromagnetically induced transparency: optics in coherent media. Rev Mod Phys 77:633–673
10. Milonni PW (2005) Fast light, slow light and left-handed light. Taylor and Francis, New York
11. Firstenberg O, Shuker M, Ron A, Davidson N (2013) Colloquium: coherent diffusion of polaritons in atomic media. Rev Mod Phys 85:941–960
12. Fleischhauer M, Lukin MD (2000) Dark-state polaritons in electromagnetically induced transparency. Phys Rev Lett 84:5094–5097
13. Phillips DF, Fleischhauer A, Mair A, Walsworth RL, Lukin MD (2001) Storage of light in atomic vapor. Phys Rev Lett 86:783–786
14. Liu C, Dutton Z, Behroozi CH, Hau LV (2001) Observation of coherent optical information storage in an atomic medium using halted light pulses. Nature 409:490–493
15. Chaneliére T, Matsukevich DN, Jenkins SD, Lan SY, Kennedy TAB, Kuzmich A (2005) Storage and retrieval of single photons transmitted between remote quantum memories. Nature 438:833–836
16. Eisaman MD, André A, Massou F, Fleischhauer M, Zibrov AS, Lukin MD (2005) Electromagnetically induced transparency with tunable single-photon pulses. Nature 438:837–841
17. Bajcsy M, Zibrov AS, Lukin MD (2003) Stationary pulses of light in an atomic medium. Nature 426:638–641
18. André A, Lukin MD (2002) Manipulating light pulses via dynamically controlled photonic band gap. Phys Rev Lett 89:143602
19. André A, Bajcsy M, Zibrov AS, Lukin MD (2005) Nonlinear optics with stationary pulses of light. Phys Rev Lett 94:063902
20. Lin YW, Liao WT, Peters T, Chou HC, Wang JS, Cho HW, Kuan PC, Yu IA (2009) Stationary light pulses in cold atomic media and without Bragg gratings. Phys Rev Lett 102:213601
21. Chen YH, Lee MJ, Hung W, Chen YC, Chen YF, Yu IA (2012) Demonstration of the interaction between two stopped light pulses. Phys Rev Lett 108:173603
22. Otterbach J, Unanyan RG, Fleischhauer M (2009) Confining stationary light: Dirac dynamics and Klein tunneling. Phys Rev Lett 102:063602

23. Unanyan RG, Otterbach J, Fleischhauer M, Ruseckas J, Kudriašov V, Juzeliūnas G (2010) Spinor slow-light and Dirac particles with variable mass. Phys Rev Lett 105:173603

24. Ruseckas J, Kudriašov V, Juzeliūnas G, Unanyan RG, Otterbach J, Fleischhauer M (2011) Photonic-band-gap properties for two-component slow light. Phys Rev A 83:063811

25. Lee MJ, Ruseckas J, Lee CY, Kudriašov V, Chang KF, Cho HW, Juzeliūnas G, Yu IA (2014) Experimental demonstration of spinor slow light. Nat Commun 5:5542

26. Brillouin L (1960) Wave propagation and group velocity. Academic Press, New York

27. Arimondo E (1996) Coherent population trapping in laser spectroscopy. In: Wolf E (ed) Progress in optics, vol 35. Elsevier, Amsterdam, pp 257–354

28. Scully MO, Zubairy, MS (1997) Quantum optics. Cambridge University Press, Cambridge

29. Garrido Alzar CE, Martinez MAG, Nussenzveig P (2002) Classical analog of electromagnetically induced transparency. Am J Phys 70:37–41

30. Schnorrberger U, Thompson JD, Trotzky S, Pugatch R, Davidson N, Kuhr S, Bloch I (2009) Electromagnetically induced transparency and light storage in an atomic Mott insulator. Phys Rev Lett 103:033003

31. Heinze G, Hubrich C, Halfmann T (2013) Stopped light and image storage by electromagnetically induced transparency up to the regime of one minute. Phys Rev Lett 111:033601

32. Mazets IE, Matisov BG (1996) Adiabatic Raman polariton in a Bose condensate. JETP Lett 64:515–519

33. Juzeliūnas G, Carmichael HJ (2002) Systematic formulation of slow polaritons in atomic gases. Phys Rev A 65:021601(R)

34. Balents L, Fisher MPA (1997) Delocalization transition via supersymmetry in one dimension. Phys Rev B 56:12970–12991

35. Shelton DG, Tsvelik AM (1998) Effective theory for midgap states in doped spin-ladder and spin-Peierls systems: Liouville quantum mechanics. Phys Rev B 57:14242–14246

36. Liu N, Langguth L, Weiss T, Kästel J, Fleischhauer M, Pfau T, Giessen H (2009) Plasmonic analogue of electromagnetically induced transparency at the Drude damping limit. Nat Mater 8:758–762

37. Fiore V, Yang Y, Kuzyk MC, Barbour R, Tian L, Wang H (2011) Storing optical information as a mechanical excitation in a silica optomechanical resonator. Phys Rev Lett 107:133601

38. Schmidt H, Imamoğlu A (1996) Giant Kerr nonlinearities obtained by electromagnetically induced transparency. Opt Lett 21:1936–1938

39. Harris SE, Yamamoto Y (1998) Photon switching by quantum interference. Phys Rev Lett 81:3611–3614

40. Lukin MD, Imamoğlu A (2000) Nonlinear optics and quantum entanglement of ultraslow single photons. Phys Rev Lett 84:1419–1422

41. Wang ZB, Marzlin KP, Sanders BC (2006) Large cross-phase modulation between slow copropagating weak pulses in 87Rb. Phys Rev Lett 97:063901

42. Shiau BW, Wu MC, Lin CC, Chen YC (2011) Low-light-level cross-phase modulation with double slow light pulses. Phys Rev Lett 106:193006

43. Venkataraman V, Saha K, Gaeta AL (2013) Phase modulation at the few-photon level for weak-nonlinearity-based quantum computing. Nat Photon 7:138–141

44. Chen W, Beck KM, Bücker R, Gullans M, Lukin MD, Tanji-Suzuki H, Vuletic V (2013) All-optical switch and transistor gated by one stored photon. Science 341:768–770

45. Gorshkov AV, Otterbach J, Fleischhauer M, Pohl T, Lukin MD (2011) Photon-photon interactions via Rydberg blockade. Phys Rev Lett 107:133602

46. Mohapatra AK, Jackson TR, Adams CS (2007) Coherent optical detection of highly excited Rydberg states using electromagnetically induced transparency. Phys Rev Lett 98:113003

47. Peyronel T, Firstenberg O, Liang Q, Hofferberth S, Gorshkov AV, Pohl T, Lukin MD, Vuletić V (2012) Quantum nonlinear optics with single photons enabled by strongly interacting atoms. Nature 488:57–60

48. Hofmann CS, Günter G, Schempp H, Robert-de-Saint-Vincent M, Gärttner M, Evers S, Whitlock J, Weidemüller M (2013) Sub-Poissonian statistics of Rydberg-interacting dark-state polaritons. Phys Rev Lett 110:203601

49. Firstenberg O, Peyronel T, Liang QY, Gorshkov AV, Lukin MD, Vuletić V (2013) Attractive photons in a quantum nonlinear medium. Nature 502:71–76

50. Maxwell D, Szwer DJ, Paredes-Barato D, Busche H, Pritchard JD, Gauguet A, Weatherill KJ, Jones PA, Adams CS (2013) Storage and control of optical photons using Rydberg polaritons. Phys Rev Lett 110:103001

Optical Tests of Foundations of Quantum Theory

Yanhua H. Shih

Y.H. Shih (✉)
Department of Physics, University of Maryland, Baltimore, MD 21250, USA
e-mail: shih@umbc.edu

16.1 Introduction

Since the beginning of Quantum Theory scientists questioned its very basic concepts, such as locality, reality, and complementarity, because they are so different from classical theory and from our everyday experience.

16.1.1 Locality

Einstein posed his students a question [1]: suppose a photon with energy $h\nu$ is created from a point source, such as an atomic transition; how big is the photon after propagating 1 year? This question seems easy to answer. Since the photon is created from a point source, it would propagate in the form of a spherical wave and its wavefront must be a sphere with a diameter of 2 lightyears after 1 year propagating. Einstein then asked again: suppose a point-like photon counting detector located on the surface of the big sphere is triggered by that photon, how long does it take for the energy on the other side of the big sphere to arrive at the detector? Two years? For a fast photodetector, it takes only a few picoseconds to produce a photoelectron by annihilating a photon with energy $h\nu$. Does this mean something has happened faster than the speed of light? Bohr provided a famous answer to this question: the "wavefunction collapses" instantaneously! Why does the wavefunction need to "collapse"? Bohr did not explain. In quantum theory, perhaps, the wavefunction does not need to "collapse." A wavefunction is defined as the probability amplitude for a particle to be observed at a space-time coordinate $(\mathbf{r}, t)$.

Quantum theory, however, does allow nonlocal interference. Assuming Einstein continued his question: if two photons are created simultaneously from the point source, and we set up a measurement with two different yet indistinguishable alternative ways for the photon pair to produce a joint photodetection event between two distant point-like photon counting detectors, what is the chance to observe a joint photodetection event at $(\mathbf{r}_1, t_1)$ and $(\mathbf{r}_2, t_2)$? According to quantum theory, the probability is the result of the linear superposition between the two probability amplitudes,

$$P(\mathbf{r}_1, t_1; \mathbf{r}_2, t_2) \quad = \left| \mathcal{A}_{\mathrm{I}}(\mathbf{r}_1, t_1; \mathbf{r}_2, t_2) + \mathcal{A}_{\mathrm{II}}(\mathbf{r}_1, t_1; \mathbf{r}_2, t_2) \right|^2 \qquad \textbf{(16.1)}$$

despite the distance between the two photodetection events, even if the two detectors are placed on the opposite sides of the big sphere.[1] How much time for this superposition to complete? Two years? Again, the two-photon interference must be completed within the "coincidence" time window which can be a few picoseconds. Furthermore, it is not necessary to use a hardware coincidence counter to count the coincidences. Two independent "event timers," which record the registration times of the two photodetection events of the two photon counting detectors, respectively, and PC software are able to calculate the joint photodetection probability. In some experiments, one detector-event timer package is placed on a satellite and the other one is placed in a ground laboratory. The recorded history of photodetection events are later brought together at the ground laboratory and analyzed by a PC. We found that the two-photon interferences are observable only when the time axis of the two event timers is correctly synchronized within the response time of the photodetectors which could be a few picoseconds.

---

1   This kind superposition has been named two-photon interference: a pair of photon interferes with the pair itself at distance.

16.1.2 **Reality**

Now, we ask a different kind of question: Does Einstein's photon have a defined momentum and position over the course of its propagation? On one hand, the photodetection event of a point-like photon counting detector tells us that the annihilated photon at $(\mathbf{r}, t)$ must carry momentum $\mathbf{p} = (\hbar\omega/c)\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit vector normal to the sphere; on the other hand, Einstein's spherical wavefunction means the momentum of that photon cannot be a constant vector when it is created at the point source of $(\mathbf{r}_0, t_0)$, otherwise it would not propagate to all 4π directions, yet, the uncertainty principle prevents a point source from producing a photon with $\Delta\mathbf{p} = 0$. Does it mean the photon has no momentum in the course of its propagation until its annihilation? To Einstein, the statement from Copenhagen "no phenomenon is a phenomenon until it is a registered phenomenon" [1] was unacceptable! Einstein believed a photon must be created and propagated with a defined momentum, the same as that observed from its annihilation. In Einstein's opinion, momentum and position must be physical realities accompany with a photon, otherwise, we may have to accept that some kinds of phenomena happen faster than the speed of light, such as "wavefunction collapse," or we may have to accept that a photon can be divided into parts, or that part of $h\nu$ is able to excite a photoelectron.

In 1935, Einstein, Podolsky, and Rosen (EPR) published an article to defend their opinion on physical reality [2]. In that article, EPR proposed a *gedankenexperiment* and introduced an entangled two-particle system based on the superposition of two-particle wavefunctions. The EPR system is composed of two distant interaction-free particles which are characterized by the following wavefunction:

$$\Psi(x_1, x_2) = \frac{1}{2\pi\hbar} \int dp_1 dp_2 \, \delta(p_1 + p_2) \, e^{ip_1(x_1-x_0)/\hbar} e^{ip_2 x_2/\hbar}$$
$$= \delta(x_1 - x_2 - x_0) \tag{16.2}$$

where $e^{ip_1(x_1-x_0)/\hbar}$ and $e^{ip_2 x_2/\hbar}$ are the eigenfunctions, with eigenvalues $p_1 = p$ and $p_2 = -p$, respectively, of the momentum operators $\hat{p}_1$ and $\hat{p}_2$ associated with particles 1 and 2; x_1 and x_2 are the coordinate variables to describe the positions of particles 1 and 2, respectively; and x_0 is a constant. The EPR state is very peculiar. Although there is no interaction between the two distant particles, the two-particle superposition cannot be factorized into a product of two individual superpositions of two particles. Quantum theory does not prevent such states.

What can we learn from the EPR state of Eq. (16.2)?

(1) In the coordinate representation, the wavefunction is a delta function: $\delta(x_1 - x_2 - x_0)$. The two particles are always separated in space with a constant value of $x_1 - x_2 = x_0$, although the coordinates x_1 and x_2 of the two particles are both unspecified.

(2) The delta wavefunction $\delta(x_1 - x_2 - x_0)$ is the result of the superposition of the plane wavefunctions of free particle one, $e^{ip_1(x_1-x_0)/\hbar}$, and free particle two, $e^{ip_2 x_2/\hbar}$, with a particular distribution $\delta(p_1 + p_2)$. It is $\delta(p_1 + p_2)$ that made the superposition special: although the momentum of particle one and particle two may take on any values, the delta function restricts the superposition with only these terms in which the total momentum of the system takes a constant value of zero.

Now, we transfer the wavefunction from coordinate representation to momentum representation:

$$\begin{aligned}
\Psi(p_1, p_2) &= \frac{1}{2\pi\hbar} \int dx_1 dx_2 \; \delta(x_1 - x_2 - x_0) \; e^{-ip_1(x_1-x_0)/\hbar} e^{-ip_2 x_2/\hbar} \\
&= \delta(p_1 + p_2).
\end{aligned} \tag{16.3}$$

What can we learn from the EPR state of Eq. (16.3)?

(1) In the momentum representation, the wavefunction is a delta function: $\delta(p_1 + p_2)$. The total momentum of the two-particle system takes a constant value of $p_1 + p_2 = 0$, although the momenta p_1 and p_2 are both unspecified.

(2) The delta wavefunction $\delta(p_1 + p_2)$ is the result of the superposition of the plane wavefunctions of free particle one, $e^{-ip_1(x_1-x_0)/\hbar}$, and free particle two, $e^{-ip_2 x_2/\hbar}$, with a particular distribution $\delta(x_1 - x_2 - x_0)$. It is $\delta(x_1 - x_2 - x_0)$ that made the superposition special: although the coordinates of particle one and particle two may take on any values, the delta function restricts the superposition with only these terms in which $x_1 - x_2$ is a constant value of x_0.

In an EPR system, *the value of the momentum (position) is not determined for either single subsystem. However, if one of the subsystems is measured to be at a certain momentum (position), the other one is determined to have a unique corresponding value despite the distance between them.* An idealized EPR state of a two-particle system is therefore characterized by $\Delta(p_1 + p_2) = 0$ and $\Delta(x_1 - x_2) = 0$ simultaneously, even if the momentum and position of each individual free particle are completely undefined, i.e., $\Delta p_j \sim \infty$ and $\Delta x_j \sim \infty$, $j = 1, 2$. In other words, each of the subsystems may have completely random values or all possible values of momentum and position in the course of their motion, but the correlations of the two subsystems are determined with certainty whenever a joint measurement is performed.[2]

According to EPR's criteria:

Locality - There is no action-at-a-distance;

Reality - If, without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of physical reality corresponding to this quantity;

Completeness - Every element of the physical reality must have a counterpart in the complete theory;

momentum and position must be physical realities associated with particle one and two. This led to the title of their 1935 article: "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" [2]

In early 1950s, Bohm simplified EPR's entangled two-particle state of continuous space-time variables to discrete spin variables [3]. Bohm suggested the singlet state of two spin 1/2 particles:

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left[|\uparrow\rangle_1 |\downarrow\rangle_2 - |\downarrow\rangle_1 |\uparrow\rangle_2 \right] \tag{16.4}$$

---

2   There have been arguments considering $\Delta(p_1 + p_2)\Delta(x_1 - x_2) = 0$ a violation of the uncertainty principle. This argument is false. It is easy to find that $p_1 + p_2$ and $x_1 - x_2$ are not conjugate variables. As we know, non-conjugate variables correspond to commuting operators in quantum mechanics, if the corresponding operators exist. To have $\Delta(p_1 + p_2) = 0$ and $\Delta(x_1 - x_2) = 0$ simultaneously, or to have $\Delta(p_1 + p_2)\Delta(x_1 - x_2) = 0$ is not a violation of the uncertainty principle.

where the kets $|\uparrow\rangle$ and $|\downarrow\rangle$ represent the states of spin "up" and spin "down," respectively, along an *arbitrary* direction. Again, for this state, *the spin of neither particle is determined; however, if one particle is measured to be spin up along a certain direction, the other one must be spin down along that direction, despite the distance between the two spin 1/2 particles*. Similar to the original EPR state, Eq. (16.4) is independent of the choice of the spin directions and the eigenstates of the associated non-commuting spin operators.

The most widely used entangled two-particle states might have been the "Bell states" (or EPR-Bohm-Bell states) [4]. Bell states are a set of polarization states for a pair of entangled photons. The four Bell states which form a complete orthonormal basis of two-photon states are usually represented as

$$|\Phi_{12}^{(\pm)}\rangle = \frac{1}{\sqrt{2}} [\, |\, 0_1 \ 0_2 \,\rangle \pm |\, 1_1 \ 1_2 \,\rangle],$$

$$|\Psi_{12}^{(\pm)}\rangle = \frac{1}{\sqrt{2}} [\, |\, 0_1 \ 1_2 \,\rangle \pm |\, 1_1 \ 0_2 \,\rangle]$$

(16.5)

where $|0\rangle$ and $|1\rangle$ represent two arbitrary orthogonal polarization bases, for example, $|0\rangle = |H\rangle$ (horizontal linear polarization) and $|1\rangle = |V\rangle$ (vertical linear polarization); or $|0\rangle = |R\rangle$ (right-hand circular polarization) and $|1\rangle = |L\rangle$ (left-hand circular polarization). We will have a detailed discussion of Bell states based on two types of experiments: (1) EPR-Bohm-Bell correlation measurement; and (2) Bell's inequality testing.

16.1.3 Complementarity

Wave-particle duality, which Feynman called the basic mystery of quantum mechanics [5], says that there is always a trade-off between the knowledge of the particle-like and wave-like behavior of a quantum system. In slightly different words, Bohr suggested a complementarity principle in 1927: one can never measure the precise position and momentum of a quantum simultaneously [6]. Since then, complementarity has often been superficially identified with the "wave-particle duality of matter." How quantum mechanics enforces complementarity may vary from one experimental situation to another (◘ Fig. 16.1).

In a single-photon Young's double-slit experiment, is the photon going to pass "both slits" like a wave or will it choose "which slit" to pass like a particle? This question has been asked since the early days of quantum mechanics [8]. Among most physicists, the common "understanding" is that the position-momentum uncertainty relation makes it impossible to determine which slit a photon or wavepacket passes through without at the same time disturbing the photon or wavepacket enough to destroy the interference pattern. However, it has been shown that under certain circumstances this common "understanding" may not be true. In 1982, Scully and Drühl showed that a "quantum eraser" may erase the which-path information [9]. The "random delayed choice quantum eraser" has been experimentally demonstrated with interesting results: the which-path information is truly erasable even after the annihilation of the quantum itself [10, 11].

Popper's thought experiment evaluated the same fundamental problem from a slightly different position [12]. Popper proposed a coincidence measurement on a pair of entangled particles. If the position of particle one is learned within Δy through the joint measurement of its twin, particle two, do we expect an uncertainty relation on particle one $\Delta y \Delta p_y \geq \hbar$? Namely, if we place an array of detectors at a distance at which particle one is restricted within Δy, do we expect a diffraction pattern with a minimum width that is determined by $\Delta y \Delta p_y \geq \hbar$?

Fig. 16.1 On one hand, a photon can never be divided into parts; on the other hand, we never lose interference at the single photon level. In fact, according to quantum theory, Young's double-slit interference is a single-photon phenomenon. In Diracts terms: "... photon ... only interferes with itself" [7]

Popper predicted a negative answer: particle one would not be diffracted unless a real slit of Δy is inserted. Similar to Einstein, Popper was a believer of realism. In his opinion, a particle must have defined momentum and position over the course of its propagation. We will review two experimental realizations of Popper's thought experiment. It is interesting to see that both experiments, one based on the measurement of entangled photon pairs, another based on the measurement of randomly paired photons in a thermal state, produced a similar result that $\Delta y \Delta p_y < \hbar$, agreeing with Popper's prediction [13, 14]. Is this result a violation of the uncertainty principle?

In the following, we will focus on three types of "optical tests of foundations of quantum theory": (1) EPR-Bohm-Bell correlation and Bell's inequality; (2) Quantum eraser; (3) Popper's experiment.

16.2 EPR-Bohm-Bell Correlation and Bell's Inequality

An important step to perform an optical test of the EPR-Bohm-Bell correlation and Bell's inequality is to prepare an entangled photon pairs in Bell state. Historically, the most popular entangled photon sources have been: (1) annihilation of positronium; (2) atomic cascade decay; (3) spontaneous parametric downconversion (SPDC). Both atomic cascade decay and SPDC have been experimentally tested. Most of the early EPR-Bohm-Bell experiments demonstrated in the 1970s and early 1980s used atomic cascade decay [15–20]. Since Alley and Shih introduced SPDC to the preparation of entangled states in the middle of 1980s [21], the signal-idler photon pair of SPDC has played an important role, especially in the tests of Bell's inequality. Using SPDC now-a-days one could easily observe a violation of Bell's inequality with hundreds of standard deviations [22]. The photon pair produced from SPDC has received an interesting name: *biphoton* [23].
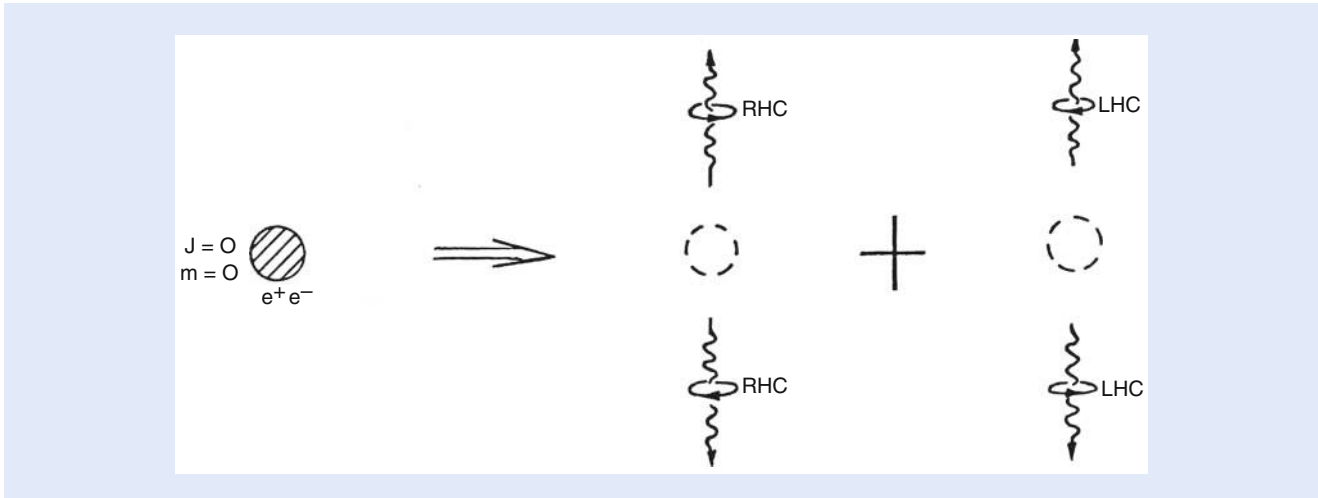
◘ **Fig. 16.2** Annihilation of Positronium. Due to the conservation of angular momentum, if photon 1 is right-hand circular (RHC) polarized, photon 2 must be right-hand circular polarized. If photon 1 is left-hand circular (LHC) polarized, then photon 2 has to be left-hand circular polarized

◘ Figure 16.2 schematically illustrates the pair-creation mechanism of positronium annihilation [5]. Initially, we have a positron and an electron in the spin-zero state with antiparallel spins. The positronium cannot exist very long: it disintegrates into two γ-ray photons within $\sim 10^{-10}$ s of its lifetime. The spin zero state is symmetric under all rotations. Therefore, the photon pair may disintegrate into any direction in space with equal probability. The conservation of linear momentum, however, guarantees that if one of the photons is observed in a certain direction, its twin must be found in the opposite direction (with finite uncertainty $\Delta(\mathbf{p}_1 + \mathbf{p}_1) \neq 0$). The conservation of angular momentum will decide the polarization state of the photon pair. As shown in ◘ Fig. 16.2, in order to keep spin-zero, if photon 1 is right-hand circular polarized (RHC), photon 2 must also be right-hand circular polarized. The same argument shows that if photon 1 is left-hand circular (LHC) polarized, then photon 2 has to be left-hand circular polarized too. Therefore, the positronium may decay into two RHC photons or two LHC photons with equal probability.

Furthermore, the law of parity conservation must be satisfied in the disintegration: the spin-zero ground state of positronium holds an odd parity. Thus, the state of the photon pair must keep its parity odd:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}[\,|R_1\rangle|R_2\rangle - |L_1\rangle|L_2\rangle\,], \qquad (16.6)$$

which is a non-factorizable pure state of a special superposition between the RHC and LHC states specified with a relative phase of π. Mathematically, "non-factorizable" means that the state cannot be written as a product state of photon 1 and photon 2. Physically, it means that photon 1 and photon 2 are not independent despite the distance between them. The two γ-ray photons are in an entangled polarization state, or spin state. The high energy γ-ray photon pair disintegrated from the annihilation of positronium is a good example to explore the physics of the EPR-Bohm state, however, the γ-ray photon pairs are difficult to handle experimentally: (1) There are no effective polarization analyzers available for the high energy γ-rays; (2) The uncertainty in momentum correlation, $\Delta(\mathbf{p}_1 + \mathbf{p}_2)$, has considerable large value, resulting in a "pair collection efficiency loophole" in Bell's inequality measurements, i.e., one may never have $\sim 100\,\%$ chance to "collect" a pair for joint photo-detection measurement [15]. Fortunately, the two-photon state of Eq. (16.6) is also observed in atomic cascade decay with

visible-ultraviolet wavelengths and we have plenty of high efficiency polarization analyzers available in that wavelengths. Thus, most of the early EPR-Bohm-Bell experiments demonstrated in the 1970s and early 1980s used two-photon source of atomic cascade decay [15]. These experiments, unfortunately, still experienced the difficulties in the momentum uncertainty. The "pair collection" efficiency is as low as that of the annihilation of positronium. It was in the middle of 1980s, Alley and Shih introduced the nonlinear optical spontaneous parametric down-conversion to the preparation of entangled states [21]. The entangled signal-idler photon pair can be easily prepared in visible-infrared wavelengths, and very importantly, the uncertainty in momentum correlation was improved significantly. The "pair collection efficiency loophole" was finally removed.

16.2.1 Biphoton and Bell State Preparation

The state of a signal-idler photon pair created in SPDC is a typical EPR state [24]. Roughly speaking, the process of SPDC involves sending a pump laser beam into a nonlinear material, such as a non-centrasymmetric crystal. Occasionally, the nonlinear interaction leads to the annihilation of a high frequency pump photon and the simultaneous creation of a pair of lower frequency signal-idler photons into an entangled two-photon state:

$$|\Psi\rangle = \Psi_0 \sum_{s,i} \delta\big(\omega_s + \omega_i - \omega_p\big)\delta\big(\mathbf{k}_s + \mathbf{k}_i - \mathbf{k}_p\big)a_s^\dagger(\mathbf{k}_s)\, a_i^\dagger(\mathbf{k}_i)\,|\,0\rangle \qquad \text{(16.7)}$$

where $\omega_j, \mathbf{k}_j(j = s,\, i,\, p)$ are the frequency and wavevector of the signal (s), idler (i), and pump (p), $a_s^\dagger$ and $a_i^\dagger$ are creation operators for the signal and the idler photon, respectively, and Ψ_0 is the normalization constant. We have assumed a CW monochromatic laser pump, i.e., ω_p and $\mathbf{k}_p$ are considered as constants. The two delta functions in Eq. (16.7) are technically named as phase matching condition:

$$\omega_p = \omega_s + \omega_i, \qquad\qquad \mathbf{k}_p = \mathbf{k}_s + \mathbf{k}_i. \qquad\qquad \text{(16.8)}$$

The names *signal* and *idler* are historical leftovers. The names probably came about due to the fact that in the early days of SPDC, most of the experiments were done with non-degenerate processes. One radiation was in the visible range (and thus easily detected, the signal), and the other was in IR range (usually not detected, the idler). We will see in the following discussions that the role of the idler is not any less than that of the signal. The SPDC process is referred to as type-I if the signal and idler photons have identical polarizations, and type-II if they have orthogonal polarizations. The process is said to be *degenerate* if the SPDC photon pair have the same free space wavelength (e.g., $\lambda_i = \lambda_s = 2\lambda_p$), and *nondegenerate* otherwise. In general, the pair exit the crystal *non-collinearly*, that is, propagate to different directions defined by the second equation in Eq. (16.8) and the Snell's law. Of course, the pair may also exit *collinearly*, in the same direction, together with the pump.

   The state of the signal-idler pair can be derived, quantum mechanically, by the first order perturbation theory with the help of the nonlinear interaction Hamiltonian. The SPDC interaction arises in a nonlinear crystal driven by a pump laser beam. The polarization, i.e., the dipole moment per unit volume, is given by

$$P_i = \chi_{i,j}^{(1)} E_j + \chi_{i,j,k}^{(2)} E_j E_k + \chi_{i,j,k,l}^{(3)} E_j E_k E_l + \ldots \qquad\qquad \text{(16.9)}$$

where $\chi^{(m)}$ is the mth order electrical susceptibility tensor. In SPDC, it is the second order nonlinear susceptibility $\chi^{(2)}$ that plays the role. The second order nonlinear interaction Hamiltonian can be written as

$$H = \varepsilon_0 \int\limits_V d\mathbf{r}\, \chi^{(2)}_{ijk}\, E_i E_j E_k \tag{16.10}$$

where the integral is taken over the interaction volume V.

It is convenient to use the Fourier representation for the electrical fields in Eq. (16.10):

$$\mathbf{E}(\mathbf{r},\, t) = \int d\mathbf{k}\, [\, \mathbf{E}^{(-)}(\mathbf{k}) e^{-i(\omega(\mathbf{k})t - \mathbf{k}\cdot\mathbf{r})} + \mathbf{E}^{(+)}(\mathbf{k}) e^{i(\omega(\mathbf{k})t - \mathbf{k}\cdot\mathbf{r})}\,]. \tag{16.11}$$

Substituting Eq. (16.11) into Eq. (16.10) and keeping only the terms of interest, we obtain the SPDC Hamiltonian in the interaction representation:

$$H_{\text{int}}(t) = \varepsilon_0 \int\limits_V d\mathbf{r} \int d\mathbf{k}_s\, d\mathbf{k}_i\, \chi^{(2)}_{lmn} E^{(+)}_{p\,l} e^{i(\omega_p t - \mathbf{k}_p \cdot \mathbf{r})}$$
$$E^{(-)}_{s\,m} e^{-i(\omega_s(\mathbf{k}_s)t - \mathbf{k}_s \cdot \mathbf{r})} E^{(-)}_{i\,n} e^{-i(\omega_i(\mathbf{k}_i)t - \mathbf{k}_i \cdot \mathbf{r})} + h.c., \tag{16.12}$$

where $h.c.$ stands for Hermitian conjugate. To simplify the calculation, we have also assumed the pump field to be plane and monochromatic with wave vector $\mathbf{k}_p$ and frequency ω_p.

It is easily noticeable that in Eq. (16.12), the volume integration can be done for some simplified cases. At this point, we assume that V is infinitely large. Later, we will see that the finite size of V in longitudinal and/or transversal directions may have to be taken into account. For an infinite volume V, the interaction Hamiltonian Eq. (16.12) is written as

$$H_{\text{int}}(t) = \varepsilon_0 \int d\mathbf{k}_s\, d\mathbf{k}_i\, \chi^{(2)}_{lmn}\, E^{(+)}_{p\,l} E^{(-)}_{s\,m} E^{(-)}_{i\,n}$$
$$\times \delta(\mathbf{k}_p - \mathbf{k}_s - \mathbf{k}_i) e^{i(\omega_p - \omega_s(\mathbf{k}_s) - \omega_i(\mathbf{k}_i))t} + h.c. \tag{16.13}$$

It is reasonable to consider the pump field classical, which is usually a laser beam, and quantize the signal and idler fields, which are both in single-photon level:

$$E^{(-)}(\mathbf{k}) = i\sqrt{\frac{2\pi\hbar\omega}{V}} a^\dagger(\mathbf{k}),$$
$$E^{(+)}(\mathbf{k}) = i\sqrt{\frac{2\pi\hbar\omega}{V}} a(\mathbf{k}), \tag{16.14}$$

where $a^\dagger(\mathbf{k})$ and $a(\mathbf{k})$ are photon creation and annihilation operators, respectively. The state of the emitted photon pair can be calculated by applying the first order perturbation

$$|\Psi\rangle = -\frac{i}{\hbar} \int dt\, H_{int}(t)\, |0\rangle. \tag{16.15}$$

By using vacuum $|0\rangle$ for the initial state in Eq. (16.15), we assume that there is no input radiation in any signal and idler modes, that is, we have a spontaneous parametric down conversion (SPDC) process.

Further assuming an infinite interaction time, evaluating the time integral in Eq. (16.15) and omitting altogether the constants and slow (square root) functions of ω, we obtain the *entangled* two-photon state of Eq. (16.7) in the form of integral:
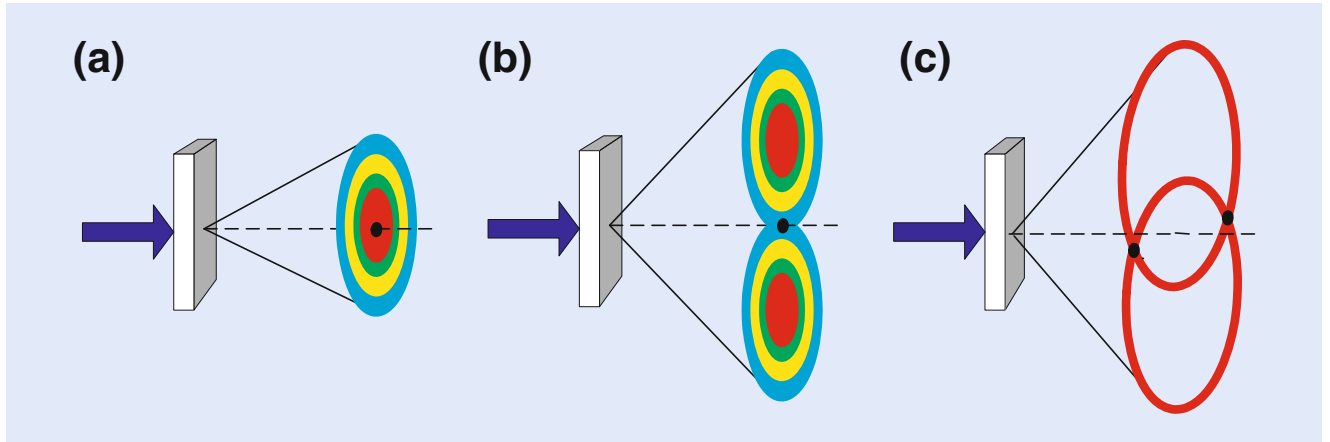
Fig. 16.3 Three widely used SPDC. (**a**) Type-I SPDC. (**b**) Collinear degenerate type-II SPDC. Two rings overlap at one region. (**c**) Non-collinear degenerate type-II SPDC. For clarity, only two degenerate rings, one for *e*-polarization and the other for *o*-polarization, are shown

$$|\Psi\rangle = \Psi_0 \int d\mathbf{k}_s d\mathbf{k}_i \delta[\omega_p - \omega_s(\mathbf{k}_s) - \omega_i(\mathbf{k}_i)]$$
$$\times \, \delta(\mathbf{k}_p - \mathbf{k}_s - \mathbf{k}_i) a_s^\dagger(\mathbf{k}_s) a_i^\dagger(\mathbf{k}_i)|0\rangle \qquad \textbf{(16.16)}$$

where Ψ_0 is a normalization constant which has absorbed all omitted constants.

The way of achieving phase matching, i.e., the way of achieving the delta functions in Eq. (16.16) basically determines how the signal-idler pair "looks." For example, in a negative uniaxial crystal, one can use a linearly polarized pump laser beam as an extraordinary ray of the crystal to generate a signal-idler pair both polarized as the ordinary rays of the crystal, which is defined as type-I phase matching. One can alternatively generate a signal-idler pair with one ordinary polarized and another extraordinary polarized, which is defined as type II phase matching. ▣ Figure 16.3 shows three examples of SPDC two-photon source. All three schemes have been widely used for different experimental purposes. Technical details can be found from textbooks and research references in nonlinear optics.

The two-photon state in the forms of Eq. (16.7) or Eq. (16.16) is a pure state, which describes the behavior of a signal-idler photon pair mathematically. Does the signal or the idler photon in the EPR state of Eq. (16.7) or Eq. (16.16) have a defined energy and momentum regardless of whether we measure it or not? Quantum mechanics answers: No! However, if one of the subsystems is measured with a certain energy and momentum, the other one is determined with certainty, despite the distance between them.

In the above calculation of the two-photon state we have approximated an infinite large volume of nonlinear interaction. For a finite volume of nonlinear interaction, we may write the state of the signal-idler photon pair in a more general form:

$$|\Psi\rangle = \int dk_s \, dk_i \, F(\mathbf{k}_s, \mathbf{k}_i) \, a_i^\dagger(\mathbf{k}_s) \, a_s^\dagger(\mathbf{k}_i)|0\rangle \qquad \textbf{(16.17)}$$

where

$$F(\mathbf{k}_s, \mathbf{k}_i) = \varepsilon \, \delta(\omega_p - \omega_s - \omega_i) \, f(\Delta_z L) \, h_{tr}(\vec{\kappa}_1 + \vec{\kappa}_2)$$

$$f(\Delta_z L) = \int_L dz \, e^{-i(k_p - k_{sz} - k_{iz})z}$$

$$h_{tr}(\vec{\kappa}_1 + \vec{\kappa}_2) = \int_A d\vec{\rho} \, \tilde{h}_{tr}(\vec{\rho}) \, e^{-i(\vec{\kappa}_s + \vec{\kappa}_i) \cdot \vec{\rho}}$$

$$\Delta_z = k_p - k_{sz} - k_{iz}$$

(16.18)

where ε is named as parametric gain, ε is proportional to the second order electric susceptibility $\chi^{(2)}$, and is usually treated as a constant; L is the length of the nonlinear interaction; the integral in $\vec{\kappa}$ is evaluated over the cross section A of the nonlinear material illuminated by the pump, $\vec{\rho}$ is the transverse coordinate vector, $\vec{\kappa}_j$ (with $j = s, i$) is the transverse wavevector of the signal and idler, and $f(|\vec{\rho}|)$ is the transverse profile of the pump, which can be treated as a Gaussion in most of the experimental conditions. The functions $f(\Delta_z L)$ and $h_{tr}(\vec{\kappa}_1 + \vec{\kappa}_2)$ can be approximated as δ-functions for an infinitely long ($L \sim \infty$) and wide ($A \sim \infty$) nonlinear interaction region. The reason we have chosen the form of Eq. (16.18) is to separate the "longitudinal" and the "transverse" correlations. We will show that $\delta(\omega_p - \omega_s - \omega_i)$ and $f(\Delta_z L)$ together can be rewritten as a function of $\omega_s - \omega_i$. To simplify the mathematics, we assume near co-linearly SPDC. In this situation, $|\vec{\kappa}_{s,i}| \ll |\mathbf{k}_{s,i}|$.

Basically, function $f(\Delta_z L)$ determines the "longitudinal" space-time correlation. Finding the solution of the integral is straightforward:

$$f(\Delta_z L) = \int_0^L dz \, e^{-i(k_p - k_{sz} - k_{iz})z} = e^{-i\Delta_z L/2} \, \text{sinc}(\Delta_z L/2).$$

(16.19)

where sinc $(x) = \sin(x)/x$.

Now, we consider $f(\Delta_z L)$ with $\delta(\omega_p - \omega_s - \omega_i)$ together, and taking advantage of the δ-function in frequencies by introducing a detuning frequency ν to evaluate function $f(\Delta_z L)$:

$$\omega_s = \omega_s^0 + \nu$$
$$\omega_i = \omega_i^0 - \nu$$
$$\omega_p = \omega_s + \omega_i = \omega_s^0 + \omega_i^0.$$

(16.20)

The dispersion relation $k(\omega)$ allows us to express the wave numbers through the detuning frequency ν:

$$k_s \approx k(\omega_s^0) + \nu \left. \frac{dk}{d\omega} \right|_{\omega_s^0} = k(\omega_s^0) + \frac{\nu}{u_s},$$

$$k_i \approx k(\omega_i^0) - \nu \left. \frac{dk}{d\omega} \right|_{\omega_i^0} = k(\omega_i^0) - \frac{\nu}{u_i}$$

(16.21)

where u_s and u_i are group velocities for the signal and the idler, respectively. Now, we connect Δ_z with the detuning frequency ν:

$$\Delta_z = k_p - k_{sz} - k_{iz}$$

$$= k_p - \sqrt{(k_s)^2 - (\vec{\kappa}_s)^2} - \sqrt{(k_i)^2 - (\vec{\kappa}_i)^2}$$

$$\cong k_p - k_s - k_i + \frac{(\vec{\kappa}_s)^2}{2k_s} + \frac{(\vec{\kappa}_i)^2}{2k_i} \tag{16.22}$$

$$\cong k_p - k(\omega_s^0) - k(\omega_i^0) + \frac{\nu}{u_s} - \frac{\nu}{u_i} + \frac{(\vec{\kappa}_s)^2}{2k_s} + \frac{(\vec{\kappa}_i)^2}{2k_i}$$

$$\cong D\nu$$

where $D \equiv 1/u_s - 1/u_i$. We have also applied $k_p - k(\omega_s^0) - k(\omega_i^0) = 0$ and $|\vec{\kappa}_{s,i}| \ll |\mathbf{k}_{s,i}|$. The "longitudinal" wavevector correlation function is rewritten as a function of the detuning frequency ν: $f(\Delta_z L) \cong f(\nu DL)$. In addition to the above approximations, we have inexplicitly assumed the angular independence of the wavevector $k = n(\theta)\omega/c$. For type II SPDC, the refraction index of the extraordinary-ray depends on the angle between the wavevector and the optical axis and an additional term appears in the expansion. Making the approximation valid, we have restricted our calculation to near-collinear process. Thus, for a good approximation, in the near-collinear experimental setup:

$$\Delta_z L \cong \nu DL = (\omega_s - \omega_i)DL/2. \tag{16.23}$$

Type-I degenerate SPDC is a special case. Due to the fact that $u_s = u_i$, and hence, $D = 0$, the expansion of $k(\omega)$ should be carried out up to the second order. Instead of (16.23), we have

$$\Delta_z L \cong -\nu^2 D'L = -(\omega_s - \omega_i)^2 D'L/4 \tag{16.24}$$

where

$$D' \equiv \frac{d}{d\omega}\left(\frac{1}{u}\right)\Big|_{\omega^0}.$$

The two-photon state of the signal-idler pair is then approximated as

$$|\Psi\rangle = \int d\nu \, d\vec{\kappa}_s \, d\vec{\kappa}_i f(\nu) \, h_{tr}(\vec{\kappa}_s + \vec{\kappa}_i) \, a_s^\dagger(\omega_s^0 + \nu, \vec{\kappa}_s) \, a_i^\dagger(\omega_i^0 - \nu, \vec{\kappa}_i)|0\rangle \tag{16.25}$$

where the normalization constant has been absorbed into $f(\nu)$.

SPDC has been one of the most convenient two-photon sources for the preparation of Bell state. Although Bell state is for polarization (or spin), the space-time part of the state cannot be ignored. One important "preparation" is to make the two biphoton wavepackets, corresponding to the first and the second terms in the Bell state, completely "overlap" in space-time, or indistinguishable for the joint detection event. This is especially important for type-II SPDC.

A very interesting situation for type-II SPDC is that of "noncollinear phase matching." The signal-idler pair are emitted from an SPDC crystal, such as BBO, cut in type-II phase matching, into two cones, one ordinarily polarized, the other extraordinarily polarized, see ◻ Fig. 16.4. Along the intersection, where the cones overlap, two pinholes numbered 1 and 2 are used for defining the direction of the **k** vectors of the signal-idler pair. It is very reasonable to consider the polarization state of the signal-idler pair as
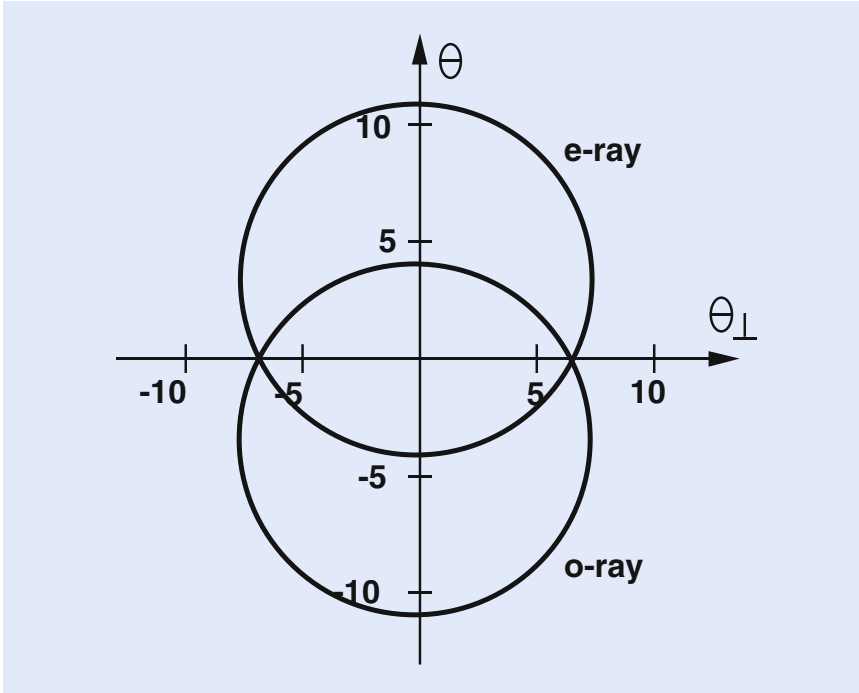
⬛ **Fig. 16.4** Type-II noncollinear phase matching: a cross section view of the degenerate 702.2 nm cones. The 351.1 nm pump beam is in the center. The numbers along the axes are in degrees

$$|\Psi\rangle = \frac{1}{\sqrt{2}} \left(|o_1 e_2\rangle + |e_1 o_2\rangle \right) = |\Psi^{(+)}\rangle \tag{16.26}$$

where o_j and e_j, $j = 1, 2$, are ordinarily and extraordinarily polarization, respectively. It seems straightforward to realize an EPR-Bohm-Bell measurement by simply setting up a polarization analyzer in series with a photon counting detector behind pinholes 1 and 2, respectively, and to expect observe the polarization correlation. This is, however, *incorrect*! One can never observe the EPR-Bohm-Bell polarization correlation unless a "compensator" is applied [4]. The "compensator" is a piece of birefringent material. For example, one may place another piece of nonlinear crystal behind the SPDC. It could be the same type of crystal as that of the SPDC, with the same cutting angle, except having half the length and a 90° rotation with respect to that of the SPDC crystal.

What is the role of the "compensator"? There have been naive explanations about the compensator. One suggestion was that the problem comes from the longitudinal "walk-off" of the type-II SPDC. For example, if one uses a type II BBO, which is a negative uni-axis crystal, the extraordinary-ray propagates faster than the ordinary-ray inside the BBO. *Suppose* the $o - e \leftrightarrow e - o$ pair is generated in the *middle* of the crystal, the e-polarization will trigger the detector earlier than the o-polarization by a time $\Delta t = (n_o - n_e)L/2c$. This implies that D_2 would be fired first in $|o_1 e_2\rangle$ term; but D_1 would be fired first in $|e_1 o_2\rangle$ term. If Δt is greater than the coherence length of the signal-idler field, one would be able to distinguish which amplitude gave rise to the "click-click" coincidence event. One may compensate the "walk-off" by introducing an additional piece of birefringent material, like the compensator we have suggested above, to delay the e-ray relative to the o-ray by the same amount of time, Δt. If, however, the signal-idler pair is generated in the *front face* or the *back face* of the SPDC, the delay time would be very
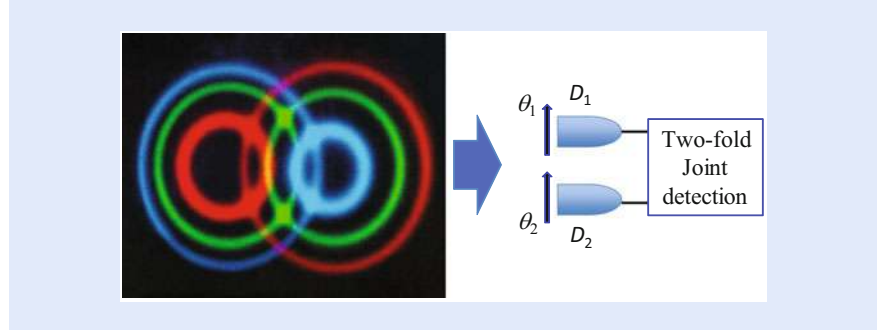
Fig. 16.5 Schematic setup of a Bell correlation measurement. The orthogonally polarized signal-idler photon pair is created from type-II SPDC. The X-direction (Y-direction) is defined by the ordinary polarization (extraordinary polarization) of the nonlinear crystal

different: $\Delta t = (n_o - n_e)L/c$ for the *front face* and $\Delta t = 0$ for the *back face*. One can never satisfy all the pairs which are generated at different places along the SPDC crystal. Nevertheless, since SPDC is a *coherent* process, the signal-idler pair is generated in such a way that it is impossible to know the birthplace of the pair. So, how is the delay time Δt determined?

▶ Figure 16.5 schematically illustrates a Bell correlation measurement in which an orthogonally polarized signal-idler photon pair of type-II SPDC is annihilated at $(\mathbf{r}_1, t_1)$ and $(\mathbf{r}_2, t_2)$ jointly by two point-like photon counting detectors D_1 and D_2 with two polarization analyzers oriented at θ_1 and θ_2, respectively.

The coincidence counting rate of D_1 and D_2 measures the probability for a pair of photons to produce a joint photodetection event at D_1 and D_2. In this setup, the pair has two different yet indistinguishable ways to produce a coincidence count: (1) the X-polarized photon passes θ_1 triggering D_1, the Y-polarized photon passes θ_2 triggering D_2; (2) the Y-polarized photon passes θ_1 triggering D_1, the X-polarized photon passes θ_2 triggering D_2. If the above two alternatives are indistinguishable, quantum theory requires a superposition of the two probability amplitudes which results in an EPR-Bohm-Bell correlation:

$$R_c(\theta_1, \theta_2) \propto |\mathcal{A}_{\mathrm{I}}(\theta_1, \theta_2) + \mathcal{A}_{\mathrm{II}}(\theta_1, \theta_2)|^2 = \sin^2(\theta_1 + \theta_2). \tag{16.27}$$

To calculate the joint detection counting rate, we follow the Glauber formula [25]:

$$\begin{aligned} R_c &\propto \langle \Psi | E^{(-)}(\mathbf{r}_1, t_1) E^{(-)}(\mathbf{r}_2, t_2) E^{(+)}(\mathbf{r}_2, t_2) E^{(+)}(\mathbf{r}_1, t_1) | \Psi \rangle \\ &= \left| \langle 0 | E^{(+)}(\mathbf{r}_2, t_2) E^{(+)}(\mathbf{r}_1, t_1) | \Psi \rangle \right|^2. \end{aligned} \tag{16.28}$$

Adopting our earlier result, we may rewrite the state of the type-II signal-idler pair in the following form:

$$|\Psi\rangle = \int d\mathbf{k}_o d\mathbf{k}_e \delta(\omega_o + \omega_e - \omega_p) \Phi(\Delta_k L) \, \hat{\mathbf{o}} \, a_o^\dagger(\omega(\mathbf{k}_o)) \hat{\mathbf{e}} \, a_e^\dagger(\omega(\mathbf{k}_e)) \, |0\rangle \tag{16.29}$$

where $\hat{\mathbf{o}}$ and $\hat{\mathbf{e}}$ are unit vectors along the o-ray and the e-ray polarization direction of the SPDC crystal, and $\Delta_k = k_o + k_e - k_p$. In $\Phi(\Delta_k L)$, the finite length of the nonlinear crystal has been taken into account. Suppose the polarizers of the detectors D_1 and D_2 are set at angles θ_1 and θ_2, relative to the polarization direction of the o-ray of the SPDC crystal, respectively, the field operators can be written as

$$E_j^{(+)}(t_j, r_j) = \int d\omega \, \hat{\theta}_j \, a(\omega) e^{-i[\omega t_j - k(\omega) r_j]}$$
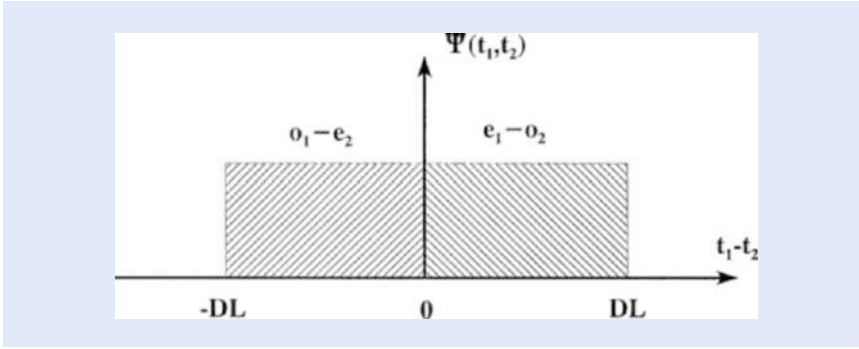
Fig. 16.6 Without "compensator," the two dimensional wavepackets of $\Psi(\tau_1^o, \tau_2^e)$ and $\Psi(\tau_1^e, \tau_2^o)$ do not overlap along $\tau_1 - \tau_2$ axis

where $j = 1, 2$, $\hat{\theta}_j$ is the unit vector along the orientation of the ith polarization analyzer. Substitute the field operator into Eq. (16.28),

$$
\begin{aligned}
R_c &\propto |(\hat{\theta}_1 \cdot \hat{\mathbf{o}})(\hat{\theta}_2 \cdot \hat{\mathbf{e}}) \, \Psi(\tau_1^o, \tau_2^e) + (\hat{\theta}_1 \cdot \hat{\mathbf{e}})(\hat{\theta}_2 \cdot \hat{\mathbf{o}}) \, \Psi(\tau_1^e, \tau_2^o)|^2 \\
&= |\mathcal{A}_1(\theta_1, \theta_2) + \mathcal{A}_2(\theta_1, \theta_2)|^2 \\
&= \cos^2\theta_1 \, \sin^2\theta_2 \, + \, \sin^2\theta_1 \, \cos^2\theta_2 \\
&\quad + \cos\theta_1 \, \sin\theta_2 \, \sin\theta_1 \, \cos\theta_2 \, \Psi^*(\tau_1^o, \tau_2^e) \, \Psi(\tau_1^e, \tau_2^o)
\end{aligned}
\tag{16.30}
$$

where $\Psi(\tau_1^o, \tau_2^e)$ and $\Psi(\tau_1^e, \tau_2^o)$ are the effective two-photon wavefunctions, namely the biphoton wavepackets, and $\tau_j^o = t_j - r_j/u_o$, $\tau_j^e = t_j - r_j/u_e$. The third term of Eq. (16.30) determines the degree of two-photon coherence. Considering degenerate CW laser pumped SPDC, the biphoton wavepacket can be simplified as

$$
\Psi(\tau_1, \tau_2) = \Psi_0 \, e^{-i\omega_p(\tau_1+\tau_2)/2} \, \mathcal{F}_{\tau_-}\{f(\Omega)\}.
$$

The coefficient of $\cos\theta_1 \, \sin\theta_2 \, \sin\theta_1 \, \cos\theta_2$ in the third term of Eq. (16.30) is thus

$$
e^{-i\omega_p(\Delta\tau_1-\Delta\tau_2)/2} \, \mathcal{F}_{\tau_1^o-\tau_2^e} \, \{f(\Omega)\} \otimes \mathcal{F}_{\tau_1^e-\tau_2^o} \, \{f(\Omega)\}.
$$

where $F_\tau - \{f(\Omega)\}$ labels a Fourier transform.

Therefore, two important factors will determine the result of the polarization correlation measurement: (1) the phase of $e^{-i\omega_p(\Delta\tau_1-\Delta\tau_2)/2}$; and (2) the overlapping between the biphoton wavepackets $\Psi^*(\tau_1^o, \tau_2^e)$ and $\Psi(\tau_1^e, \tau_2^o)$, i.e., the chances for both $\Psi^*(\tau_1^o, \tau_2^e)$ and $\Psi(\tau_1^e, \tau_2^o)$ take nonzero values simultaneously at $\tau_1^o - \tau_2^e$.

Examining the two wavepackets associated with the $o_1 - e_2$ and $e_1 - o_2$ terms, we found the two dimensional biphoton wavepackets of type II SPDC do not overlap, due to the *asymmetrical* rectangular function of $\pi(\tau_1 - \tau_2)$ as indicated in ▪ Fig. 16.6. In order to make the two wavepackets overlap, we may either (1) move both wavepackets a distance of $DL/2$ (case I) or (2) move one of the wavepackets a distance of DL (case II). The use of "compensator" is for this purpose. After compensating the two asymmetrical function of $\pi(\tau_1 - \tau_2)$, we need to further manipulate the phase of $e^{-i\omega_p(\Delta\tau_1-\Delta\tau_2)/2}$ to finalize the desired Bell states. This can be done by means of a retardation plate to introduce phase delay of 2π (+1) or π (−1) between the o-ray and the e-ray in either arm 1 or arm 2. The EPR-Bohm-Bell polarization correlation $R_c \propto \sin^2(\theta_1 \pm \theta_2)$ is expected only when the above two conditions are satisfied. We can simplify the polarization state of the signal-idler photon pair in the form of Bell states $|\Psi^{(\pm)}\rangle$ in this situation only.

In recent years, special attention has been paid to femtosecond laser pulse pumped SPDC due to its attractive applications in quantum information processing and communication. The biphoton wavepacket looks very different

16
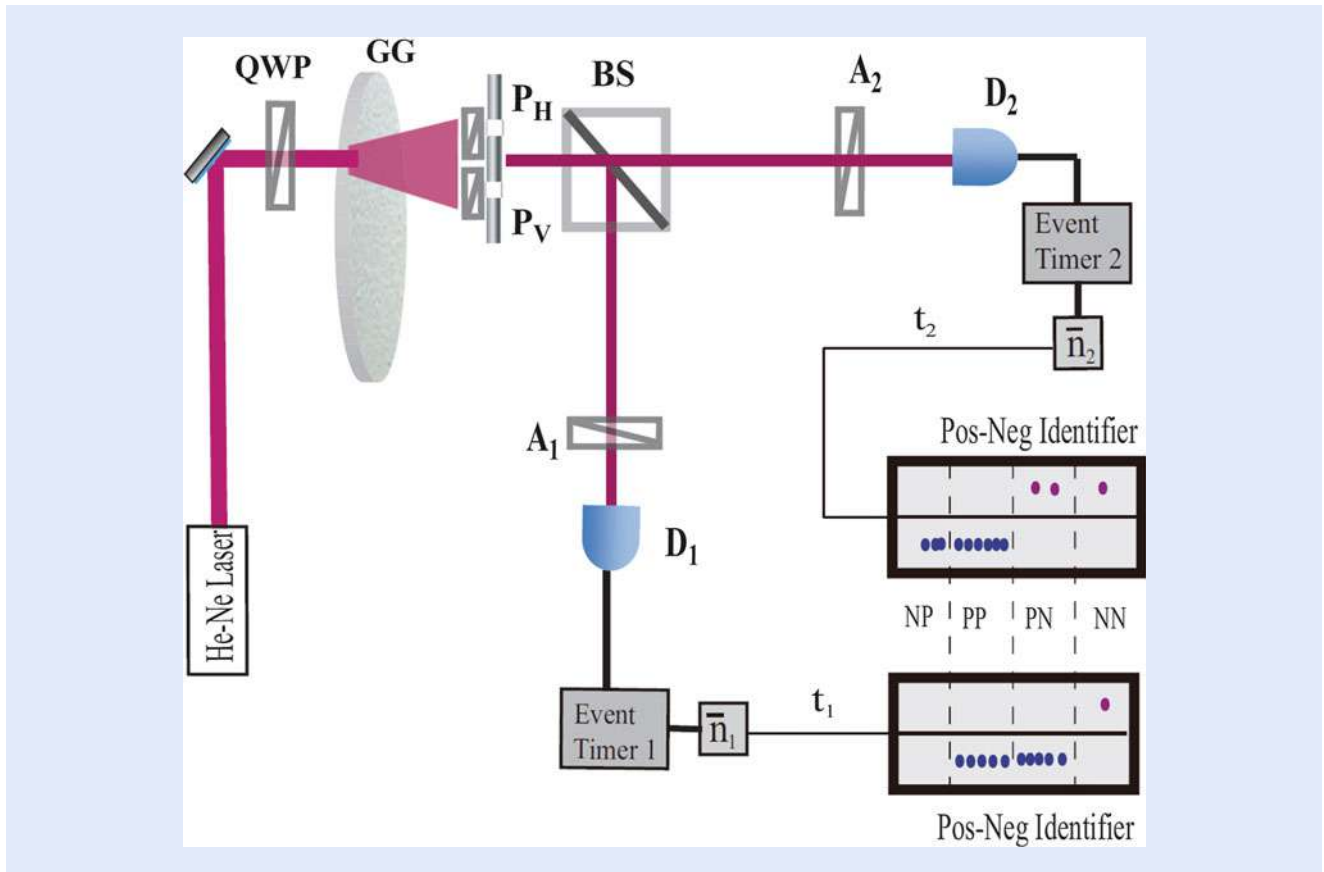
Fig. 16.7 Schematic setup of the experiment: polarization correlation measurement of thermal fields in photon-number fluctuations

in this case than that of the CW pump. One needs to exam the biphoton wavepackets carefully to be sure the superposed probability amplitudes overlap [26, 27].

From the above analysis, we may conclude the EPR-Bohm-Bell correlation is the result of a nonlocal interference: a pair of entangled photons interferes with the pair itself. This peculiar interference involves the superposition of two-photon wavepackets, or two-photon amplitudes, corresponding to different yet indistinguishable alternative ways for a pair of photons to produce a photodetection event at distant space-time coordinates.

16.2.2 Bell State Simulation of Thermal Light

Now we ask what would happen if we replace the entangled photons with a randomly paired photons, or wavepackets, in the thermal state? Can a randomly paired photons in a thermal state simulate the Bell state? The answer is positive. In the following we analyze a recent experiment of Peng et al. in which a Bell-type correlation was observed from the polarization measurement of thermal fields in photon-number fluctuations, indicating the successful simulation of Bell state [28]. Very importantly, the same mechanism can be easily extended to the simulation of a multi-photon GHZ state and N-qubits for $N \gg 2$.

Figure 16.7 schematically illustrates the experimental setup of Peng et al. A large number of circular polarized wavepackets at the single-photon level, such as the mth and the nth, come from a standard pseudo-thermal light source [29] consisting of a circularly polarized 633 nm CW laser beam and a rotating ground

glass (GG). The diameter of the laser beam is ~ 2 mm. The size of the tiny diffusers on the GG is roughly a few micrometers. The randomly distributed wavepackets pass two pinholes P_H and P_V with two linear polarizers oriented at a horizontal polarization $\vec{H}$ ($\theta = 0°$) and vertical polarization $\vec{V}$ ($\theta = 90°$), respectively. The circular polarized wavepackets have 50 % chance to pass the upper pinhole P_H with horizontal polarization and 50 % chance to pass the lower pinhole P_V with vertical polarization. The separation between the two pinholes is much greater than the coherence length of the pseudo-thermal field. Therefore, (1) the $\vec{H}$ polarization and $\vec{V}$ polarization are first-order incoherent and the mixture of the two polarizations results in a unpolarized field; (2) the fluctuations of the $\vec{H}$ polarization and the $\vec{V}$ polarization are completely independent and random without any correlation. A 50–50 non-polarizing beamsplitter (BS) is used to divide the unpolarized thermal field, i.e., the 50–50 mixture of the two polarizations, into arms 1 and 2. Two polarization analyzers A_1, oriented at θ_1, and A_2, oriented at θ_2, followed by two photon counting detectors D_1 and D_2, are placed into arms 1 and 2 for the measurement of the polarization of the wavepackets. The registration time and the number of photodetection events of D_1 and D_2 at each jth time window are recorded, respectively, by two independent but synchronized event timers. The width of the time window, Δt_j, can be adjusted from nanoseconds to milliseconds. For each detector, D_β, $\beta = 1, 2$, at each chosen value of θ_β, the mean photon number, $\tilde{n}_\beta$, is calculated from $\tilde{n}_\beta = \left(\sum_{j=1}^{N} n_{\beta j} \right) / N$, where N is the total number of time windows recorded for each data point in which θ_1 and θ_2 are set at certain chosen values. In our experiments, the total number and the width of the time window were $N \approx 4 \times 10^5$, and $\Delta t_j = 800$ μ s. The mean photon number was chosen $\tilde{n}_1 \sim \tilde{n}_2 \sim 20$. In addition, the counting rate of D_1 and D_2 is monitored to be constants, independent of θ_1 and θ_2. The number fluctuation is then calculated for each time window, $\Delta n_{\beta j} = n_{\beta j} - \tilde{n}_\beta$ [30]:

$$\langle \Delta n_1(\theta_1) \Delta n_2(\theta_2) \rangle = \frac{1}{N} \left[\sum_{j=1}^{N} \Delta n_{1j}(\theta_1) \Delta n_{2j}(\theta_2) \right]. \tag{16.31}$$

Achieving the maximum space-time correlation in photon-number fluctuations, we place D_1 and D_2 at equal longitudinal and transverse coordinates, $z_1 = z_2$ and $\vec{\rho}_1 = \vec{\rho}_2$.

◻ Figure 16.8 reports a typical measurement of the polarization correlation in photon-number fluctuation correlation. In this measurement, we fixed $\theta_1 = 45°$ and rotated θ_2 to a set of different values. The black dots are experimental data, the red sinusoidal curve is the theoretical fitting of $\cos^2(\theta_1 - \theta_2)$ based on Eq. (16.44) with a ~ 92.5 % contrast. For other values of $\theta_1 \neq 45°$ we have observed the same sinusoidal correlation function. ◻ Figure 16.9 reports a measurement of $\langle \Delta n_1(\theta_1) \Delta n_2(\theta_2) \rangle$ by scanning the values of θ_1 and θ_2 (2-D scanning). Based on these measurements, we conclude that our observed polarization correlation is the same as that of the Bell state $|\Phi^{(+)}\rangle$. Apparently, the post-selection measurements of the reported experiment has "entangled" a product state of polarization into the Bell state $|\Phi^{(+)}\rangle$.

To explain the experimental observation, we start from the analysis of chaotic-thermal light. Chaotic-thermal light may come from a natural thermal light source, such as the sun, or from a pseudo-thermal light source, usually consisting of a laser beam, either CW or pulsed, and a fast rotating ground glass containing a large number of tiny scattering diffusers (usually on the order of a few micrometers). For a natural thermal light source, each radiating atom among a large number of
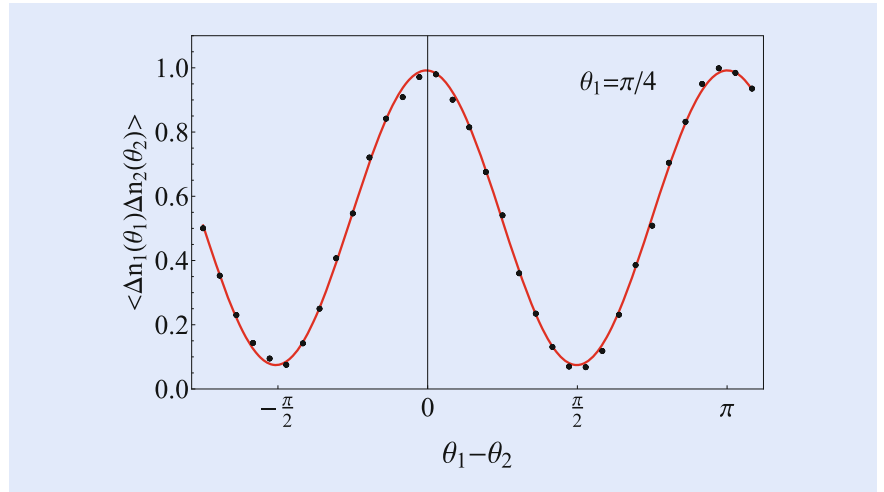
Fig. 16.8 Experimental observation of a Bell correlation with $\sim 92.5\%$ contrast. The *black dots* are experimental data and the *red sinusoidal curve* is a theoretical fitting. The *horizontal axis* labels $\varphi = \theta_1 - \theta_2$ while θ_1 was fixed at 45°, the *vertical axis* reports the normalized photon-number fluctuation correlation $\langle \Delta n_1(\theta_1)\Delta n_2(\theta_2)\rangle$
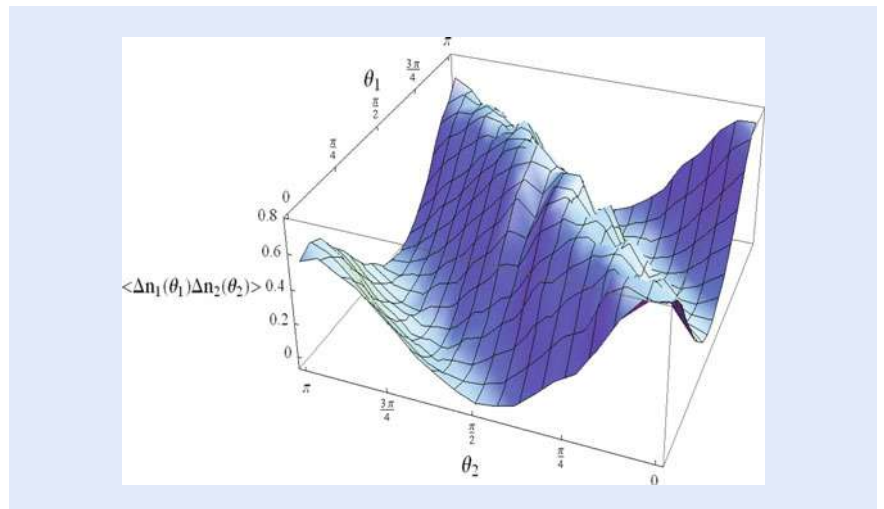
Fig. 16.9 A typical measurement of $\langle \Delta n_1(\theta_1)\Delta n_2(\theta_2)\rangle$ by a 2-*D* scanning of θ_1 and θ_2

randomly distributed and randomly radiated atomic transitions can be considered a sub-source. A photon may be created from an atomic transition, or sub-source, such as the mth atomic transition, or the mth sub-source, at space-time coordinate $(\mathbf{r}_{0m}, t_{0m})$, where $(\mathbf{r}_{0m})$ indicates the spatial coordinate of the mth atomic transition, and t_{0m} is the creation time of the photon. With a pseudo-thermal light source, each tiny scattering diffuser in the ground glass is a sub-source which scatters a wavepacket from the laser beam at space-time coordinate $(\vec{\rho}_{0m}, t_{0m})$ with random phase φ_{0m}, where $(\vec{\rho}_{0m})$ indicates the transverse spatial coordinates of the mth scattering diffuser of the fast rotating ground glass, and t_{0m} is the scattering time of the subfield. It is reasonable to model thermal light, either from a natural thermal light source or from a pseudo-thermal light source, in the coherence state representation [4, 31]:

$$|\Psi\rangle = \prod_m |\{\alpha_m\}\rangle = \prod_{m,\mathbf{k}} |\alpha_m(\mathbf{k})\rangle, \tag{16.32}$$

where m labels the mth photon that is created from the mth atomic transition of a natural thermal source, or the mth wavepacket that is scattered from the mth sub-source of the pseudo-thermal source, and $\mathbf{k}$ is a wavevector. $|\alpha_m(\mathbf{k})\rangle$ is an eigenstate of the annihilation operator with an eigenvalue $\alpha_m(\mathbf{k})$,

$$\hat{a}_m(\mathbf{k})|\alpha_m(\mathbf{k})\rangle = \alpha_m(\mathbf{k})|\alpha_m(\mathbf{k})\rangle. \tag{16.33}$$

Thus, we have

$$\hat{a}_m(\mathbf{k})|\Psi\rangle = \alpha_m(\mathbf{k})|\Psi\rangle. \tag{16.34}$$

The field operator corresponding to the mth subfield at the detector can be written in the following form:

$$\hat{E}_m^{(+)}(\mathbf{r}, t) = \int d\omega \, \hat{a}_m(\omega) g_m(\omega; \mathbf{r}, t) \tag{16.35}$$

with $g_m(\omega; \mathbf{r}, t)$ the Green's function that propagates the ω mode of the mth subfield from the source to $(\mathbf{r}, t)$. A point-like photon counting detector, behind a polarizer oriented at angle $\vec{\theta}$, at space-time coordinate $(\mathbf{r}, t)$ counts the photon number that is polarized along $\vec{\theta}$, $n(\theta; \mathbf{r}, t)$, which is usually written as the sum of mean photon-number $\langle n(\theta; \mathbf{r}, t)\rangle$ and the photon-number fluctuation $\Delta n(\theta; \mathbf{r}, t)$:

$$
\begin{aligned}
n(\theta; \mathbf{r}, t) &= \sum_m \vec{p}_m \langle \alpha_m | \sum_p \hat{E}_p^{(-)}(\mathbf{r}, t) \sum_q \hat{E}_q^{(+)}(\mathbf{r}, t) \sum_n \vec{p}_n |\alpha_n\rangle \\
&= \sum_m (\vec{p}_m \cdot \vec{\theta}) \Psi_m^*(\mathbf{r}, t) \sum_n (\vec{p}_n \cdot \vec{\theta}) \Psi_n(\mathbf{r}, t) \\
&= \sum_m (\vec{p}_m \cdot \vec{\theta}) \Psi_m^*(\mathbf{r}, t) \, (\vec{p}_m \cdot \vec{\theta}) \Psi_m(\mathbf{r}, t) \\
&\quad + \sum_{m \neq n} (\vec{p}_m \cdot \vec{\theta}) \Psi_m^*(\mathbf{r}, t) \, (\vec{p}_n \cdot \vec{\theta}) \Psi_n(\mathbf{r}, t) \\
&\equiv \sum_m \Psi_m^*(\theta; \mathbf{r}, t) \Psi_m(\theta; \mathbf{r}, t) + \sum_{m \neq n} \Psi_m^*(\theta; \mathbf{r}, t) \Psi_n(\theta; \mathbf{r}, t) \\
&= \langle n(\theta; \mathbf{r}, t)\rangle + \Delta n(\theta; \mathbf{r}, t),
\end{aligned}
\tag{16.36}
$$

where $\vec{p}_m$ is the polarization of the mth wavepacket, $|\alpha_m\rangle$ is the state of the mth photon or the mth group of identical photons in the thermal state. In Eq. (16.36) we have introduced the effective wavefunction of a photon or a group of identical photons:

$$\Psi_m(\mathbf{r}, t) = \langle \alpha_m | \hat{E}_m^{(+)}(\mathbf{r}, t)|\alpha_m\rangle = \int d\omega \, a_m(\omega) g_m(\omega; \mathbf{r}, t). \tag{16.37}$$

An effective wavefunction or a wavepacket, corresponding to the classical concept of an electromagnetic subfield $\mathbf{E}_m(\mathbf{r}, t)$ however, represents a very different physical reality. The effective wavefunction represents the "probability amplitude" for a photon or a group of identical photons to produce a photoelectron event at space-time coordinate $(\mathbf{r}, t)$. From Eq. (16.36), we find that the mean photon-number $\langle n(\theta; \mathbf{r}, t)\rangle = \sum_m \Psi_m^*(\theta; \mathbf{r}, t) \Psi_m(\theta; \mathbf{r}, t)$ involves the effective wavefunction of a photon or a wavepacket while the photon-number fluctuation $\Delta n(\theta; \mathbf{r}, t) = \sum_{m \neq n} \Psi_m^*(\theta; \mathbf{r}, t) \Psi_n(\theta; \mathbf{r}, t)$ involves the effective wave functions of two different photons, or a random pair of wavepackets, $m \neq n$. The measurement of mean

photon-number gives the self-coherence of a photon or a group of identical photons while the measurement of photon-number fluctuation gives the mutual-coherence between different photons or different groups of identical photons. In the polarization-based photon counting measurement, the above equation can be used to calculate either the polarization correlation or the space-time correlation.

In general, a Bell type experiment measures the statistical correlation between $n(\theta_1; \mathbf{r}_1, t_1)$ and $n(\theta_2; \mathbf{r}_2, t_2)$. For thermal light, the photon-number correlation can be written as the sum of two contributions [4]:

$$\langle n_1(\theta_1) n_2(\theta_2) \rangle = \langle n_1(\theta_1) \rangle \langle n_2(\theta_2) \rangle + \langle \Delta n_1(\theta_1) \Delta n_2(\theta_2) \rangle$$

$$= \sum_m \Psi_{m1}^* \Psi_{m1} \sum_n \Psi_{n2}^* \Psi_{n2} + \sum_{m \neq n} \Psi_{m1}^* \Psi_{n1} \Psi_{n2}^* \Psi_{m2}. \quad \text{(16.38)}$$

Here, we have shortened the notations of the effective wavefunction: the subindex β, $\beta = 1, 2$, indicates θ_β and $(\mathbf{r}_\beta, t_\beta)$. The first contribution is the result of two independent mean photon-number measurements, the statistics and coherence involves the measurement of single photons only while the second contribution is the result of photon-number fluctuation correlation, the statistics and coherence involves randomly paired photons. A Bell type experiment studies the polarization correlation of a pair of photons, obviously, we need to measure the photon-number fluctuation correlation $\langle \Delta n_1(\theta_1) \Delta n_2(\theta_2) \rangle$. The measurement of $\langle \Delta n_1(\theta_1) \Delta n_2(\theta_2) \rangle$ gives both polarization correlation and space-time correlation between two randomly paired photons. In the Bell type measurements, we usually manage to achieve a maximum correlation in space-time, then test the polarization correlation $\langle \Delta n_1(\theta_1) \Delta n_2(\theta_2) \rangle$ as a function of $\varphi = \theta_1 - \theta_2$ by varying θ_1 and θ_2 to all possible different values.

The effective wavefunctions: Ψ_{m1}^*, Ψ_{n1}, Ψ_{n2}^*, and Ψ_{m2} are calculated in the flowing. In general, each operator of the subfield is identified to be

$$\hat{E}_{m\beta}^{(+)} = \int d\omega \, \hat{a}_m(\omega) g_m(\omega; \mathbf{r}_\beta, t_\beta). \quad \text{(16.39)}$$

Examine the experiment detail, we find

$$g_m(\omega; \mathbf{r}_\beta, t_\beta) = \frac{1}{\sqrt{2}} \left[(\vec{H} \cdot \vec{\theta}_\beta) g_m(\omega; \mathbf{r}_H, t_H) g_H(\omega; \mathbf{r}_\beta, t_\beta) \right.$$
$$\left. + (\vec{V} \cdot \vec{\theta}_\beta) g_m(\omega; \mathbf{r}_V, t_V) g_V(\omega; \mathbf{r}_\beta, t_\beta) \right], \quad \text{(16.40)}$$

where $g_m(\omega; \mathbf{r}_H, t_H)$ and $g_H(\omega; \mathbf{r}_\beta, t_\beta)$ are the Green's functions that propagate the ω mode of the mth subfield from the source to the upper pinhole P_H and from P_H to D_β, respectively. The effective wavefunction $\Psi_{m\beta}$ is thus

$$\Psi_{m\beta} = \int d\omega \, a_m \frac{1}{\sqrt{2}} \left[(\vec{H} \cdot \vec{\theta}_\beta) g_m(\omega; \mathbf{r}_H, t_H) g_{P_H}(\omega; \mathbf{r}_\beta, t_\beta) \right.$$
$$\left. + (\vec{V} \cdot \vec{\theta}_\beta) g_m(\omega; \mathbf{r}_V, t_V) g_{P_V}(\omega; \mathbf{r}_\beta, t_\beta) \right] \quad \text{(16.41)}$$
$$= \frac{1}{\sqrt{2}} \left[\Psi_{mH\beta} + \Psi_{mV\beta} \right],$$

where $\Psi_{mH\beta}$ is the mth wavepacket passes P_H, θ_β, triggers D_β, and $\Psi_{mV\beta}$ the mth wavepacket passes P_V, θ_β, triggers D_β. The normalized photon-number fluctuation correlation is thus

$$\langle \Delta n_1(\theta_1)\Delta n_2(\theta_2)\rangle = \sum_{m,n} \Psi^*_{m1}\Psi_{n1}\Psi_{m2}\Psi^*_{n2}$$

$$= \sum_{m,n} \frac{1}{\sqrt{2}}\left[\Psi^*_{mH1}+\Psi^*_{mV1}\right]\frac{1}{\sqrt{2}}\left[\Psi_{nH1}+\Psi_{nV1}\right] \quad \textbf{(16.42)}$$

$$\times \frac{1}{\sqrt{2}}\left[\Psi_{mH2}+\Psi_{mV2}\right]\frac{1}{\sqrt{2}}\left[\Psi^*_{nH2}+\Psi^*_{nV2}\right]$$
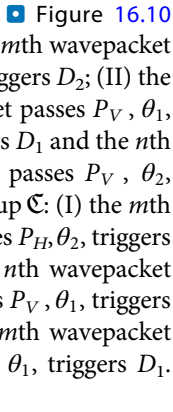
where, for example, Ψ_{mH1} is the mth wavepacket passing through the upper pinhole with $\vec{H}$ polarization contributing to the photodetection event of D_1 at space-time $(\mathbf{r}_1, t_1)$. In this experiment, we have separated the pinholes P_H and P_V beyond the transverse coherence length of the thermal field. Therefore, only four of the above sixteen terms survive from the sum of m and n,
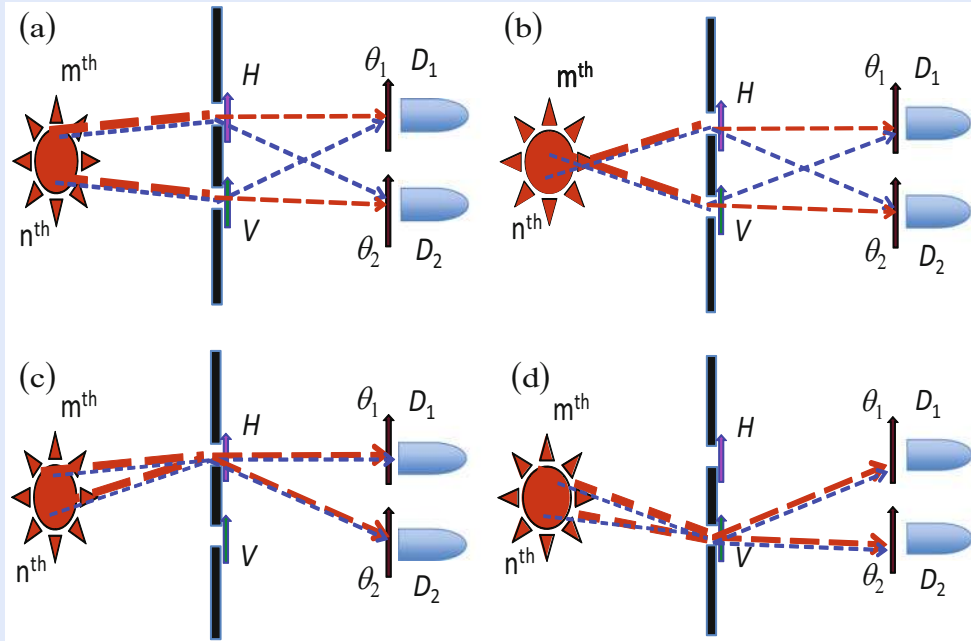
$$\langle \Delta n_1(\theta_1)\Delta n_2(\theta_2)\rangle$$
$$\propto \sum_{m,n}\left[\Psi^*_{mH1}\Psi_{mH2}\Psi_{nH1}\Psi^*_{nH2}+\Psi^*_{mH1}\Psi_{mH2}\Psi_{nV1}\Psi^*_{nV2}\right. \quad \textbf{(16.43)}$$
$$\left.+\Psi^*_{mV1}\Psi_{mV2}\Psi_{nH1}\Psi^*_{nH2}+\Psi^*_{mV1}\Psi_{mV2}\Psi_{nV1}\Psi^*_{nV2}\right]$$

corresponding to an interference effect which involves the joint detection of two wavepackets at two independent photodetectors located a distance apart. Adding the four cross terms that involve the random pair, i.e., the mth and the nth wavepackets, which is observable in the photon-number fluctuations, we obtain

$$\langle \Delta n_1(\theta_1)\Delta n_2(\theta_2)\rangle$$
$$\propto \left[\cos\theta_1\cos\theta_2\cos\theta_1\cos\theta_2+\cos\theta_1\cos\theta_2\sin\theta_1\sin\theta_2\right.$$
$$\left.+\sin\theta_1\sin\theta_2\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\sin\theta_1\sin\theta_2\right] \quad \textbf{(16.44)}$$
$$= |\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2|^2$$
$$= \cos^2(\theta_1-\theta_2),$$

which is the same correlation as that of the Bell state $|\Phi^{(+)}\rangle$.

Under the experimental condition of equal transverse and longitudinal (temporal) coordinates of D_1 and D_2, the observed correlation is the result of four groups of nonlocal superposition between different yet indistinguishable probability amplitudes of a randomly paired photons, or wavepackets. ◘ Figure 16.10 illustrates the four groups of such superposition. Group $\mathcal{A}$: (I) the mth wavepacket passes P_H, θ_1, triggers D_1 and the nth wavepacket passes P_V, θ_2, triggers D_2; (II) the mth wavepacket passes P_H, θ_2, triggers D_2 and the nth wavepacket passes P_V, θ_1, triggers D_1. Group $\mathcal{B}$: (I) the mth wavepacket passes P_V, θ_1, triggers D_1 and the nth wavepacket passes P_H, θ_2, triggers D_2; (II) the mth wavepacket passes P_V, θ_2, triggers D_2 and the nth wavepacket passes P_H, θ_1, triggers D_1. Group $\mathfrak{C}$: (I) the mth wavepacket passes P_H, θ_1, triggers D_1 and the nth wavepacket passes P_H, θ_2, triggers D_2; (II) the mth wavepacket passes P_H, θ_2, triggers D_2 and the nth wavepacket passes P_H, θ_1, triggers D_1. Group $\mathcal{A}$: (I) the mth wavepacket passes P_V, θ_1, triggers D_1 and the nth wavepacket passes P_V, θ_2, triggers D_2; (II) the mth wavepacket passes P_V, θ_2, triggers D_2 and the nth wavepacket passes P_V, θ_1, triggers D_1. Calculating the four superpositions,

Fig. 16.10 There exist four groups of different yet indistinguishable amplitudes of the *m*th and the *n*th wavepacket to produce a joint photodetector event

$$
\begin{aligned}
\left| \mathcal{A}_I(\theta_1, \theta_2) + \mathcal{A}_{II}(\theta_1, \theta_2) \right|^2 &= \left| \Psi_{mH1}\Psi_{nV2} + \Psi_{mH2}\Psi_{nV1} \right|^2 \\
\left| \mathcal{B}_I(\theta_1, \theta_2) + \mathcal{B}_{II}(\theta_1, \theta_2) \right|^2 &= \left| \Psi_{mV1}\Psi_{nH2} + \Psi_{mV2}\Psi_{nH1} \right|^2 \\
\left| \mathcal{C}_I(\theta_1, \theta_2) + \mathcal{C}_{II}(\theta_1, \theta_2) \right|^2 &= \left| \Psi_{mH1}\Psi_{nH2} + \Psi_{mH2}\Psi_{nH1} \right|^2 \\
\left| \mathcal{D}_I(\theta_1, \theta_2) + \mathcal{D}_{II}(\theta_1, \theta_2) \right|^2 &= \left| \Psi_{mV1}\Psi_{nV2} + \Psi_{mV2}\Psi_{nV1} \right|^2.
\end{aligned}
\tag{16.45}
$$

Adding the four cross terms that involve the random pair, i.e., the *m*th and the *n*th wavepackets, which is observable in the photon-number fluctuations, we obtain Eq. (16.44).

16.2.3 Bell's Inequality

In 1964, Bell derived an inequality to distinguish quantum mechanics from local realistic probability theory of hidden variable [32]. In his pioneer work Bell introduced a "more complete specification effected by means of parameter λ" with probability distribution $\rho(\lambda)$ for the classical statistical estimation of the expectation value of the spin correlation measurement $\langle \Psi | (\vec{\sigma}_1 \cdot \hat{\mathbf{a}})(\vec{\sigma}_2 \cdot \hat{\mathbf{b}}) | \Psi \rangle$ of particle-1 and particle-2, such as the spin-1/2 particle pair of Bohm, in the directions $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$, simultaneously and respectively. The quantum mechanical result of this measurement gives

$$E_{ab} = \langle \Psi | (\vec{\sigma}_1 \cdot \hat{\mathbf{a}})\,(\vec{\sigma}_2 \cdot \hat{\mathbf{b}}) | \Psi \rangle = -\hat{\mathbf{a}} \cdot \hat{\mathbf{b}}. \tag{16.46}$$

A special case of this result contains the determinism implicit in this idealized system When the Stern–Gerlach analyzers (SGA) are parallel, we have

$$E_{ab} = \langle \Psi | (\vec{\sigma}_1 \cdot \hat{\mathbf{a}})\,(\vec{\sigma}_2 \cdot \hat{\mathbf{a}}) | \Psi \rangle = -1 \tag{16.47}$$

for all λ and all $\hat{\mathbf{a}}$. Thus, we can predict with certainty the result B by obtaining the result of A. Since $|\Psi\rangle$ does not determine the result of an individual measurement, this fact (via EPR's argument) suggests that there exits a more complete specification of the state by a single symbol λ it may have many dimensions, discrete and/or continuous parts, and different parts of it interacting with either apparatus, etc. Let Λ be the space of λ for an ensemble composed of a very large number of the particle systems. Bell represented the distribution function for the state λ on the space Λ by the symbol $\rho(\lambda)$ and take $\rho(\lambda)$ to be normalized

$$\int_\Lambda \rho(\lambda)\,d\lambda = 1. \tag{16.48}$$

In a deterministic hidden variable theory the observable $[A(\hat{\mathbf{a}})B(\hat{\mathbf{b}})]$ has a defined value $[A(\hat{\mathbf{a}})B(\hat{\mathbf{b}})](\lambda)$ for the state λ.

The locality is defined as follows: a deterministic hidden variable theory is local if for all $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ and all $\lambda \in \Lambda$

$$\big[A(\hat{\mathbf{a}})B(\hat{\mathbf{b}})\big](\lambda) = A(\hat{\mathbf{a}},\lambda)\,B(\hat{\mathbf{b}},\lambda). \tag{16.49}$$

This is, once λ is specified and the particle has separated, measurements of A can depend only upon λ and $\hat{\mathbf{a}}$ but not $\hat{\mathbf{b}}$. Likewise measurements of B depend only upon λ and $\hat{\mathbf{b}}$. Any reasonable physical theory that is realistic and deterministic and that denies action-at-a-distance is local in this sense. For such theories the expectation value of $[A(\hat{\mathbf{a}})B(\hat{\mathbf{b}})]$ is given by

$$\begin{aligned} E(\hat{\mathbf{a}},\hat{\mathbf{b}}) &= \int_\Lambda d\lambda\,\rho(\lambda)\,[A(\hat{\mathbf{a}})B(\hat{\mathbf{b}})](\lambda) \\ &= \int_\Lambda d\lambda\,\rho(\lambda)\,A(\hat{\mathbf{a}},\lambda)\,B(\hat{\mathbf{b}},\lambda), \end{aligned} \tag{16.50}$$

where $E(\hat{\mathbf{a}},\hat{\mathbf{b}}) \equiv E_{ab}$, corresponding to our previous notation. It is clear that Eq. (16.47) can hold if only if

$$A(\hat{\mathbf{a}},\lambda) = -B(\hat{\mathbf{b}},\lambda) \tag{16.51}$$

hold for all $\lambda \in \Lambda$.

Using Eq. (16.51) we calculate the following expectation values, which involves three different orientations of the SGA analyzers:

16

$$E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{c}})$$

$$= \int_\Lambda d\lambda \, \rho(\lambda) \, \left[A(\hat{\mathbf{a}}, \lambda) \, B(\hat{\mathbf{b}}, \lambda) - A(\hat{\mathbf{a}}, \lambda) \, B(\hat{\mathbf{c}}) \right]$$

$$= - \int_\Lambda d\lambda \, \rho(\lambda) \, \left[A(\hat{\mathbf{a}}, \lambda) \, A(\hat{\mathbf{b}}, \lambda) - A(\hat{\mathbf{a}}, \lambda) \, A(\hat{\mathbf{c}}, \lambda) \right] \qquad \text{(16.52)}$$

$$= - \int_\Lambda d\lambda \, \rho(\lambda) \, A(\hat{\mathbf{a}}, \lambda) \, A(\hat{\mathbf{b}}, \lambda) \left[1 - A(\hat{\mathbf{b}}, \lambda) \, A(\hat{\mathbf{c}}, \lambda) \right].$$

Since $A(\hat{\mathbf{a}}, \lambda) = \pm 1$, $A(\hat{\mathbf{b}}, \lambda) = \pm 1$, this expression can be written as

$$\left| E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{c}}) \right| \le \int_\Lambda d\lambda \, \rho(\lambda) \, \left[1 - A(\hat{\mathbf{b}}, \lambda) \, A(\hat{\mathbf{c}}, \lambda) \right], \qquad \text{(16.53)}$$

and consequently,

$$\left| E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{c}}) \right| \le 1 + E(\hat{\mathbf{b}}, \hat{\mathbf{c}}). \qquad \text{(16.54)}$$

This inequality is the first of a family of inequalities which are collectively called "Bell's inequalities."

It is easy to find a disagreement between the quantum mechanics prediction of Eq. (16.46) and the inequality of Eq. (16.54). When we choose $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$, and $\hat{\mathbf{c}}$ to be coplanar with $\hat{\mathbf{c}}$ making an angle of $2\pi/3$ with $\hat{\mathbf{a}}$, and $\hat{\mathbf{b}}$ making an angle of $\pi/3$ with both $\hat{\mathbf{a}}$ and $\hat{\mathbf{c}}$, the quantum prediction gives

$$\left| \left[E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{c}}) \right]_{QM} \right| = 1, \qquad \text{(16.55)}$$

while

$$1 + \left[E(\hat{\mathbf{b}}, \hat{\mathbf{c}}) \right]_{QM} = \frac{1}{2}. \qquad \text{(16.56)}$$

It does not satisfy inequality of Eq. (16.54).

It was soon realized that the Bell's inequality of Eq. (16.54) cannot be tested in a real experiment. Because Eq. (16.47) cannot be realized exactly in a realistic measurement. Any real detector cannot have a perfect quantum efficiency of 100 %, and any real analyzer cannot have a perfect distinguish ratio between orthogonal channels. In 1971, Bell proved a new inequality [33] which includes these concerns by assuming the outcomes of measurement A or B may take one of the following possible results:

$$A(\hat{\mathbf{a}}, \lambda) \text{ or } B(\hat{\mathbf{b}}, \lambda) = \begin{cases} +1 & \text{"spin} - \text{up"} \\ -1 & \text{"spin} - \text{down"} \\ 0 & \text{particle not detected} \end{cases} \qquad \text{(16.57)}$$

For a given state λ, we define the measured values for these quantities by the symbols $\bar{A}(\hat{\mathbf{a}}, \lambda)$ and $\tilde{B}(\hat{\mathbf{b}}, \lambda)$, which satisfy

$$\left| \overline{A}(\hat{\mathbf{a}}, \lambda) \right| \le 1 \qquad \text{and} \qquad \left| \overline{B}(\hat{\mathbf{b}}, \lambda) \right| \le 1. \qquad \text{(16.58)}$$

Following the same definition of locality, the expectation value of $A(\hat{\mathbf{a}})B(\hat{\mathbf{b}})$ is calculated as

$$E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = \int_\Lambda d\lambda\, \rho(\lambda)\, \overline{A}(\hat{\mathbf{a}}, \lambda)\, \overline{B}(\hat{\mathbf{b}}, \lambda). \tag{16.59}$$

Consider a measurement which involves $E(\hat{\mathbf{a}}, \hat{\mathbf{b}})$ and $E(\hat{\mathbf{a}}, \hat{\mathbf{b}}')$

$$
\begin{aligned}
&E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{b}}') \\
&= \int_\Lambda d\lambda\, \rho(\lambda)\, \left[\overline{A}(\hat{\mathbf{a}}, \lambda)\, \overline{B}(\hat{\mathbf{b}}, \lambda) - \overline{A}(\hat{\mathbf{a}}, \lambda)\, \overline{B}(\hat{\mathbf{b}}', \lambda) \right],
\end{aligned}
\tag{16.60}
$$

which can be written in the following form:

$$
\begin{aligned}
&E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{b}}') \\
&= \int_\Lambda d\lambda\, \rho(\lambda)\, \overline{A}(\hat{\mathbf{a}}, \lambda)\, \overline{B}(\hat{\mathbf{b}}, \lambda) \left[1 \pm \overline{A}(\hat{\mathbf{a}}', \lambda)\, \overline{B}(\hat{\mathbf{b}}', \lambda) \right] \\
&\quad - \int_\Lambda d\lambda\, \rho(\lambda)\, \overline{A}(\hat{\mathbf{a}}, \lambda)\, \overline{B}(\hat{\mathbf{b}}', \lambda) \left[1 \pm \overline{A}(\hat{\mathbf{a}}', \lambda)\, \overline{B}(\hat{\mathbf{b}}, \lambda) \right].
\end{aligned}
\tag{16.61}
$$

Applying the triangle theorem, and considering $\rho(\lambda)[1 \pm \overline{A}(\hat{\mathbf{a}}', \lambda)\overline{B}(\hat{\mathbf{b}}', \lambda)]$ and $\rho(\lambda)[1 \pm \overline{A}(\hat{\mathbf{a}}', \lambda)\, \overline{B}(\hat{\mathbf{b}}, \lambda)]$ cannot take negative values, then using inequality in Eq. (16.58), we obtain

$$
\begin{aligned}
\left| E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{b}}') \right| &\leq \int_\Lambda d\lambda\, \rho(\lambda)\, \left[1 \pm \overline{A}(\hat{\mathbf{a}}', \lambda)\, \overline{B}(\hat{\mathbf{b}}', \lambda) \right] \\
&\quad + \int_\Lambda d\lambda\, \rho(\lambda)\, \left[1 \pm \overline{A}(\hat{\mathbf{a}}', \lambda)\, \overline{B}(\hat{\mathbf{b}}, \lambda) \right],
\end{aligned}
\tag{16.62}
$$

or

$$
\left| E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{b}}') \right| \leq \pm \left[E(\hat{\mathbf{a}}', \hat{\mathbf{b}}') + E(\hat{\mathbf{a}}', \hat{\mathbf{b}}) \right] + 2 \int_\Lambda d\lambda\, \rho(\lambda).
\tag{16.63}
$$

We thus derive a measurable inequality

$$-2 \leq E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{b}}') + E(\hat{\mathbf{a}}', \hat{\mathbf{b}}) + E(\hat{\mathbf{a}}', \hat{\mathbf{b}}') \leq 2. \tag{16.64}$$
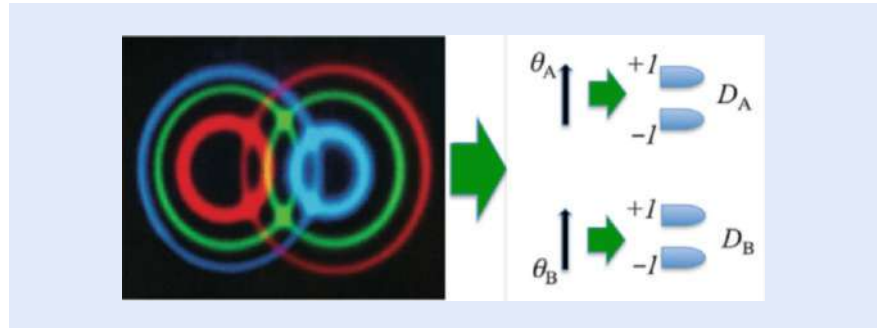
The quantum mechanical prediction of the EPR-Bhom state in a realistic measurement with imperfect detectors, analyzers etc., can be written as

$$\left[E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) \right]_{QM} = C\, \hat{\mathbf{a}} \cdot \hat{\mathbf{b}} \tag{16.65}$$

where $|C| \leq 1$. Suppose we take $\hat{\mathbf{a}}, \hat{\mathbf{a}}', \hat{\mathbf{b}}, \hat{\mathbf{b}}'$ to be coplanar with $\phi = \pi/4$, we can easily find a disagreement between the quantum mechanics prediction and the inequality of Eq. (16.64):

$$\left[E(\hat{\mathbf{a}}, \hat{\mathbf{b}}) - E(\hat{\mathbf{a}}, \hat{\mathbf{b}}') + E(\hat{\mathbf{a}}', \hat{\mathbf{b}}) + E(\hat{\mathbf{a}}', \hat{\mathbf{b}}') \right]_{QM} = 2\sqrt{2}\, C. \tag{16.66}$$

Fig. 16.11 Schematic setup of a Bell's inequality measurement. The expectation value is calculated from Eq. (16.67) which involves the measurement of four joint photodetections of D_A^+ &D_B^+, D_A^-&D_B^+, D_A^+&D_B^-, and D_A^-&D_B^-
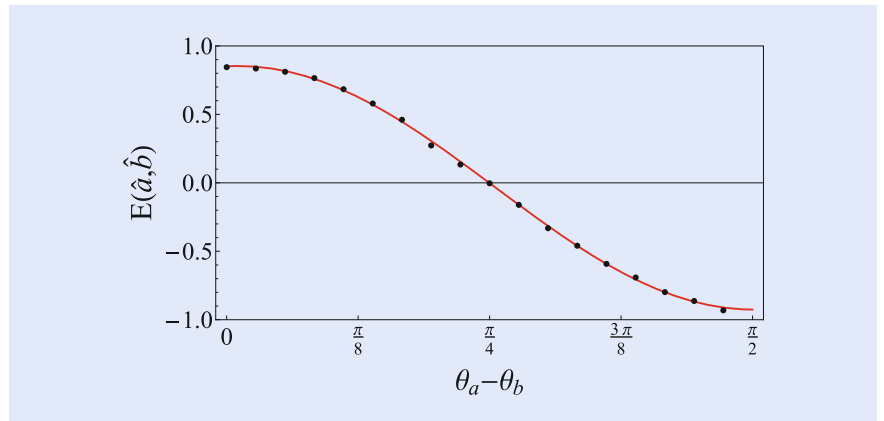
Fig. 16.12 Experimental observation of $E(\theta_A, \theta_B)$ from a typical Bell's inequality measurement. The *black dots* are experimental data and the *red sinusoidal curve* is a theoretical fitting

Although Bell derived his inequalities based on the measurement of spin-1/2 particle pairs, Eqs. (16.54) and (16.64) are not restricted to the measurement of spin-1/2 particle pairs. In fact, most of the historical experimental testing have been the polarization measurements of photon pairs. The photon pairs are prepared in similar states which have been called EPR–Bohm–Bell states, or Bell states in short. Most of the experimental observations violated Bell's inequalities which may have different forms and have their violation occur at different orientations of the polarization analyzers. However, the physics behind the violations is all similar to that of Bell's theorem.

Figure 16.11 is a schematic experimental setup for a Bell's inequality measurement. Since the space of Λ in this measurement is spanned into four regions with classical probabilities P_{ab}, P_{-ab}, P_{a-b}, P_{-a-b} in which A and B have values ± 1, the expectation value evaluation of Eq. (16.50) can be explicitly calculated as

$$E_{ab} = (+1)(+1)P_{ab} + (-1)(+1)P_{-ab} + (+1)(-1)P_{a-b}$$
$$+ (-1)(-1)P_{-a-b} = P_{ab} - P_{-ab} - P_{a-b} + P_{-a-b}, \tag{16.67}$$

P_{ab}, P_{-ab}, P_{a-b} and P_{-a-b}, respectively, are measurable quantities by means of the joint photodetections of D_A^+&D_B^+, D_A^-&D_B^+, D_A^+&D_B^-, and D_A^-&D_B^-. A typical experimental observation of $E(\theta_A, \theta_B)$ from a Bell state is shown in Fig. 16.12. Bell's inequality violation is expected from this measurement.

16.3 Scully's Quantum Eraser

Quantum eraser, proposed by Scully and Drühl in 1982 [9], is another thought experiment challenge the "basic mystery" of quantum mechanics: wave-particle duality. So far, several quantum eraser experiments have been demonstrated with interesting results supporting the ideas of Scully and Drühl [10, 11, 34, 35].

A double-slit type quantum eraser experiment, closing to the original Scully–Drühl thought experiment of 1982, is illustrated in ◘ Fig. 16.13. A pair of entangled photons, photon 1 and photon 2, is excited by a weak laser pulse either from atom A, which is located in slit A, or from atom B, which is located in slit B. Photon 1, propagates to the right, is registered by detector D_0, which can be scanned by a step motor along its x_0-axis for the examination of interference fringes. Photon 2, propagating to the left, is injected into a beamsplitter. If the pair is generated in atom A, photon 2 will follow the A path meeting BSA with 50 % chance of being reflected or transmitted. If the pair is generated in atom B, photon 2 will follow the B path meeting BSB with 50 % chance of being reflected or transmitted. In view of the 50 % chance of being transmitted by either BSA or BSB, photon 2 is detected by either detector D_3 or D_4. The registration of D_3 or D_4 provides which-path information (path A or path B) on photon 2 and in turn provides which-path information for photon 1 because of the entanglement nature of the two-photon state generated by atomic cascade decay. Given a reflection at either BSA or BSB photon 2 will continue to follow its A or B path to meet another 50–50 beamsplitter BS and then be detected by either detectors D_1 or D_2.

The experimental condition was arranged in such a way that no interference is observable in the single counting rate of D_0., i.e., the distance between A and B is large enough to be "distinguishable" for D_0 to learn which-path information of photon 1. However, the "clicks" at D_1 or D_2 will erase the which-path information of photon 1 and help to restore the interference. On the other hand, the "clicks" at D_3 or D_4 record which-path information. Thus, no observable interference is expected with the help of these "clicks." It is interesting to note that both the "erasure" and "recording" of the which-path information can be made as a "delayed choice": the experiment is designed in such a way that L_0, the optical distance between atoms A, B and detector D_0, is much shorter than L_A (L_B), which is the optical distance between atoms A, B and the beamsplitter BSA (BSB) where the "which-path" or "both-paths" "choice" is made randomly by photon 2. Thus, after the annihilation of photon 1 at D_0, photon 2 is still on its way to BSA (BSB), i.e., "which-path" or "both-path" choice is "delayed" compared to the detection of
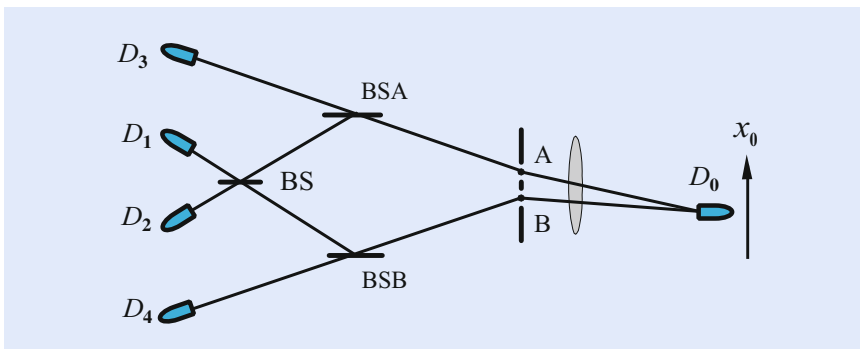
◘ **Fig. 16.13** Quantum erasure: a thought experiment of Scully–Drühl. A pair of entangled photons is emitted from either atom A or atom B by atomic cascade decay. The experimental condition guarantees no interference fringes is observable in the single detector counting rate of D_0. The "clicks" at D_1 or D_2 erase the which-path information, thus helping to restore the interference even after the "click" of D_0. On the other hand, the "clicks" at D_3 or D_4 record which-slit information. Thus, no observable interference is expected with the help of these "clicks"

photon 1. After the annihilation of photon 1, we look at these "delayed" detection events of D_1, D_2, D_3, and D_4 which have constant time delays, $\tau_i \simeq (L_i - L_0)/c$, relative to the triggering time of D_0. L_i is the optical distance between atoms A, B and detectors D_1, D_2, D_3, and D_4, respectively. It was predicted that the "joint-detection" counting rate R_{01} (joint-detection rate between D_0 and D_1) and R_{02} will show an interference pattern as a function of the position of D_0 on its x-axis. This reflects the wave nature (both-path) of photon 1. However, no interference fringes will be observable in the joint detection counting events R_{03} and R_{04} when scanning detector D_0 along its x-axis. This is as would be expected because we have now inferred the particle (which-path) property of photon 1. It is important to emphasize that all four joint detection rates R_{01}, R_{02}, R_{03}, and R_{04} are recorded at the same time during one scanning of D_0. That is, in the present experiment, we "see" both wave (interference) and which-path (particle-like) with the same measurement apparatus.

It should be mentioned that (1) the "choice" in this experiment is not actively switched by the experimentalist during the measurement. The "delayed choice" associated with either the wave or particle behavior of photon 1 is "randomly" made by photon 2. The experimentalist simply looks at which detector D_1, D_2, D_3 or D_4 is triggered by photon 2 to determine either wave or particle properties of photon 1 after the annihilation of photon 1; (2) the photo-detection event of photon 1 at D_0 and the delayed choice event of photon 2 at BSA (BSB) are space-like separated events. The "coincidence" time window is chosen to be much shorter than the distance between D_0 and BSA (BSB). Within the joint-detection time window, it is impossible to have the two events "communicating."

16.3.1 Random Delayed Choice Quantum Eraser One

Kim et al. realized the above random delayed choice quantum eraser in 2000 [10]. The schematic diagram of the experimental setup of Kim et al. is shown in ◘ Fig. 16.14. Instead of atomic cascade decay, SPDC is used to prepare the entangled two-photon state.

In the experiment, a 351. 1 nm Argon ion pump laser beam is divided by a double-slit and directed onto a type-II phase matching nonlinear crystal BBO at regions A and B. A pair of 702. 2 nm orthogonally polarized signal-idler photon is
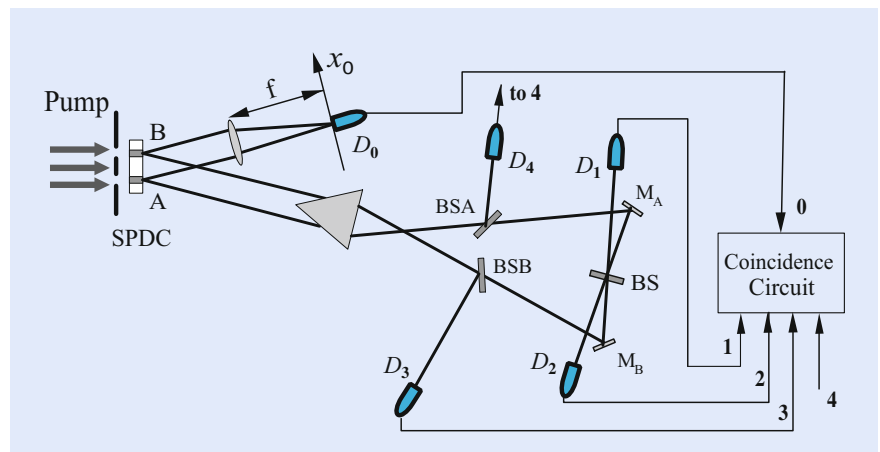

◘ **Fig. 16.14** Delayed choice quantum eraser: Schematic of an actual experimental setup of Kim et al. Pump laser beam is divided by a double-slit and makes two regions A and B inside the SPDC crystal. A pair of signal-idler photons is generated either from the A or B region. The "delayed choice" to observe either wave or particle behavior of the signal photon is made randomly by the idler photon about 7. 7 ns after the detection of the signal photon

generated either from region A or region B. The width of the region is about $a = 0.3$ mm and the distance between the center of A and B is about $d = 0.7$ mm. A Glen-Thompson prism is used to split the orthogonally polarized signal and idler. The signal photon (photon 1, coming either from A or B) propagates through lens LS to detector D_0, which is placed on the Fourier transform plane of the lens. The use of lens LS is to achieve the "far field" condition, but still keep a short distance between the slit and the detector D_0. Detector D_0 can be scanned along its x-axis by a step motor for the observation of interference fringes. The idler photon (photon 2) is sent to an interferometer with equal-path optical arms. The interferometer includes a prism PS, two 50–50 beamsplitters BSA, BSB, two reflecting mirrors M_A, M_B, and a 50–50 beamsplitter BS. Detectors D_1 and D_2 are placed at the two output ports of the BS, respectively, for erasing the which-path information. The triggering of detectors D_3 and D_4 provides which-path information for the idler (photon 2) and, in turn, which-path information for the signal (photon 1). The detectors are fast avalanche photodiodes with less than 1 ns rise time and about 100 ps jitter. A constant fractional discriminator is used with each of the detectors to register a single photon whenever the leading edge of the detector output pulse is above the threshold. Coincidences between D_0 and D_j ($j = 1, 2, 3, 4$) are recorded, yielding the joint detection counting rates R_{01}, R_{02}, R_{03}, and R_{04}.

In the experiment, the optical delay $(L_{A, B} - L_0)$ is chosen to be $\simeq 2.3$ m, where L_0 is the optical distance between the output surface of BBO and detector D_0, and L_A (L_B) is the optical distance between the output surface of the BBO and the beamsplitter BSA (BSB). This means that any information (which-path or both-path) one can infer from photon 2 must be at least 7.7 ns later than the registration of photon 1. Compared to the 1 ns response time of the detectors, 2.3 m delay is thus enough for "delayed erasure." Although there is an arbitrariness about when a photon is detected, it is safe to say that the "choice" of photon 2 is delayed with respect to the detection of photon 1 at D_0 since the entangled photon pair is created simultaneously.

◘ Figure 16.15 reports the joint detection rates R_{01} and R_{02}, indicating the regaining of standard Young's double-slit interference pattern. An expected π
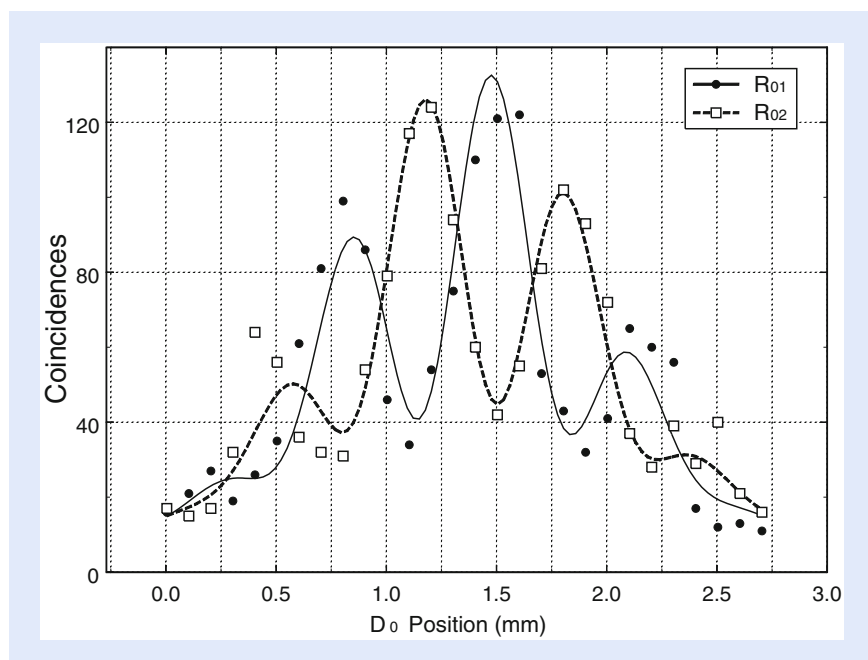
◘ **Fig. 16.15** Joint detection rates R_{01} and R_{02} against the x coordinates of detector D_0. Standard Young's double-slit interference patterns are observed. Note the π phase shift between R_{01} and R_{02}. The *solid line* and the *dashed line* are theoretical fits to the data

16
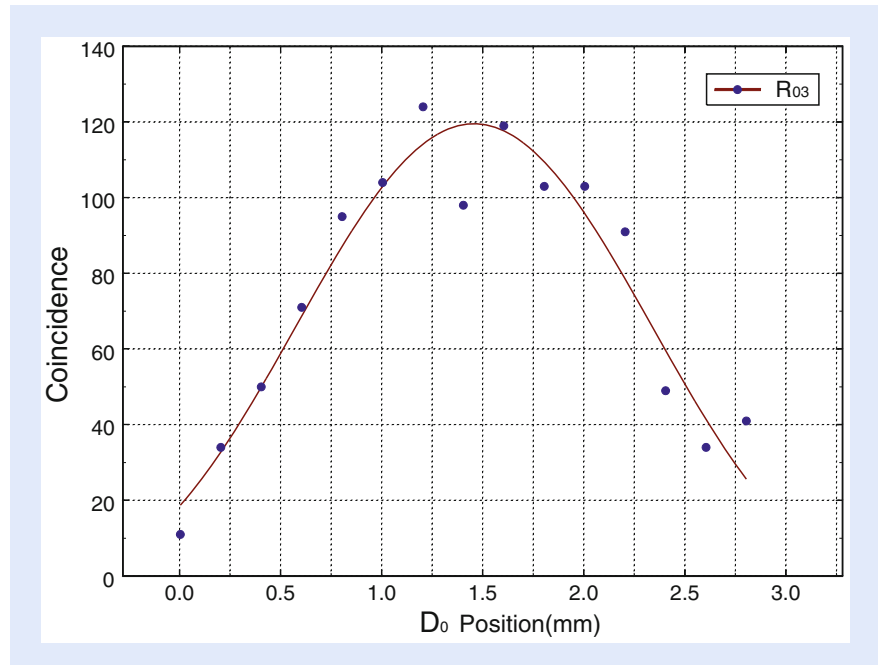
● **Fig. 16.16** Joint detection counting rate of R_{03}. Absence of interference is clearly demonstrated. The *solid line* is a sinc-function fit

phase shift between the two interference patterns is clearly shown in the measurement. The single detector counting rates of D_0 and D_1 are recorded simultaneously. Although interference is observed in the joint detection counting rate, there is no significant modulation in any of the single detector counting rate during the scanning of D_0. R_0 is a constant during the scanning of D_0. The absence of interference in the single detector counting rate of D_0 is simply because the separation between slits A and slit B is much greater than the coherence length of the single field.

● Figure 16.16 reports a typical R_{03} (R_{04}), joint detection counting rate between D_0 and "which-path detector" D_3 (D_4). An absence of interference is clearly demonstrated. The fitting curve of the experimental data indicates a sinc-function like envelope of the standard Young's double slit interference-diffraction pattern. Two features should bring to our attention that (1) there is no observable interference modulation as expected, and (2) the curve is different from the constant single detector counting rate of D_0.

The experimental result is surprising from a classical point of view. The result, however, is easily explained in the contents of quantum theory. In this experiment, there are two kinds of very different interference phenomena: single-photon interference and two-photon interference. As we have discussed earlier, single-photon interference is the result of the superposition between single-photon amplitudes, and two-photon interference is the results of the superposition between two-photon amplitudes. Quantum mechanically, single-photon amplitude and two-photon amplitude represent very different measurements and, thus, very different physics.

In this regard, we analyze the experiment by answering the following questions:

(1) Why is there no observable interference in the single-detector counting rate of D_0?

This question belongs to single-photon interferometry. The absence of interference in single-detector counting rate of D_0 is very simple: the separation

between slit A and slit B is much greater than the coherence length of the signal filed.

(2) Why is there observable interference in the joint detection counting rate of D_{01} and D_{02}?

This question belongs to two-photon interferometry. Two-photon interference is very different from single-photon interference. Two-photon interference involves the addition of different yet indistinguishable two-photon amplitudes. The coincidence counting rate R_{01}, again, is proportional to the probability P_{01} of joint detecting the signal-idler pair by detectors D_0 and D_1,

$$R_{01} \propto P_{01} = \langle \Psi | E_0^{(-)} E_1^{(-)} E_1^{(+)} E_0^{(+)} | \Psi \rangle = \left| \langle 0 | \ E_1^{(+)} E_0^{(+)} | \Psi \rangle \right|^2 . \tag{16.68}$$

To simplify the mathematics, we use the following "two-mode" expression for the state, bearing in mind that the transverse momentum δ-function will be taken into account.

$$|\Psi\rangle = \varepsilon \left[a_s^\dagger a_i^\dagger e^{i\varphi_A} + b_s^\dagger b_i^\dagger \ e^{i\varphi_B} \right] |0\rangle$$

where ε is a normalization constant that is proportional to the pump field and the nonlinearity of the SPDC crystal, φ_A and φ_B are the phases of the pump field at A and B, and $a_j^\dagger (b_j^\dagger), j = s, i$, are the photon creation operators for the lower (upper) mode in ◘ Fig. 16.14.

In Eq. (16.68), the fields at the detectors D_0 and D_1 are given by

$$E_0^{(+)} = a_s \ e^{ikr_{A0}} + b_s \ e^{ikr_{B0}}$$
$$E_1^{(+)} = a_i \ e^{ikr_{A1}} + b_i \ e^{ikr_{B1}} \tag{16.69}$$

where $r_{Aj} (r_{Bj}), j = 0, 1$ are the optical path lengths from region A (B) to the jth detector. Substituting the biphoton state and the field operators into Eq. (16.68),

$$R_{01} \propto \left| \ e^{i(kr_A + \varphi_A)} + e^{i(kr_B + \varphi_B)} \right|^2 = |\Psi_A + \Psi_B|^2$$
$$= 1 + \cos \left[k(r_A - r_B) \right] \simeq \cos^2 \left(x_0 \pi d / \lambda z_0 \right) \tag{16.70}$$

where $r_A = r_{A0} + r_{A1}, r_B = r_{B0} + r_{B1}$; Ψ_A and Ψ_B are the two-photon effective wave functions of path A and path B, representing the two different yet indistinguishable probability amplitudes to produce a joint photodetection event of D_0 and D_1, indicating a two-photon interference. In Dirac's language: a signal-idler photon pair interferes with the pair itself.

To calculate the diffraction effect of a single-slit, again, we need an integral of the effective two-photon wavefunction over the slit width (the superposition of infinite number of probability amplitudes results in a click-click joint detection event):

$$R_{01} \propto \left| \ \int_{-a/2}^{a/2} dx_{AB} \ e^{-ik \ r(x_0, \ x_{AB})} \ \right|^2 \cong \text{sinc}^2 (x_0 \pi a / \lambda z_0) \tag{16.71}$$

where $r(x_0, x_{AB})$ is the distance between points x_0 and x_{AB}, x_{AB} belongs to the slit's plane, and the far-field condition is applied.

Repeating the above calculations, the combined interference-diffraction joint detection counting rate for the double-slit case is given by

$$R_{01} \propto \text{sinc}^2(x_0 \pi a / \lambda z_0) \cos^2(x_0 \pi d / \lambda z_0). \tag{16.72}$$

If the finite size of the detectors is taken into account, the interference visibility will be reduced.

(3) Why is there no observable interference in the joint detection counting rate of R_{03} and R_{04}?

This question belongs to two-photon interferometry. From the view of two-photon physics, the absence of interference in the joint detection counting rate of R_{03} and R_{04} is obvious: only one two-photon amplitude contributes to the joint detection events.

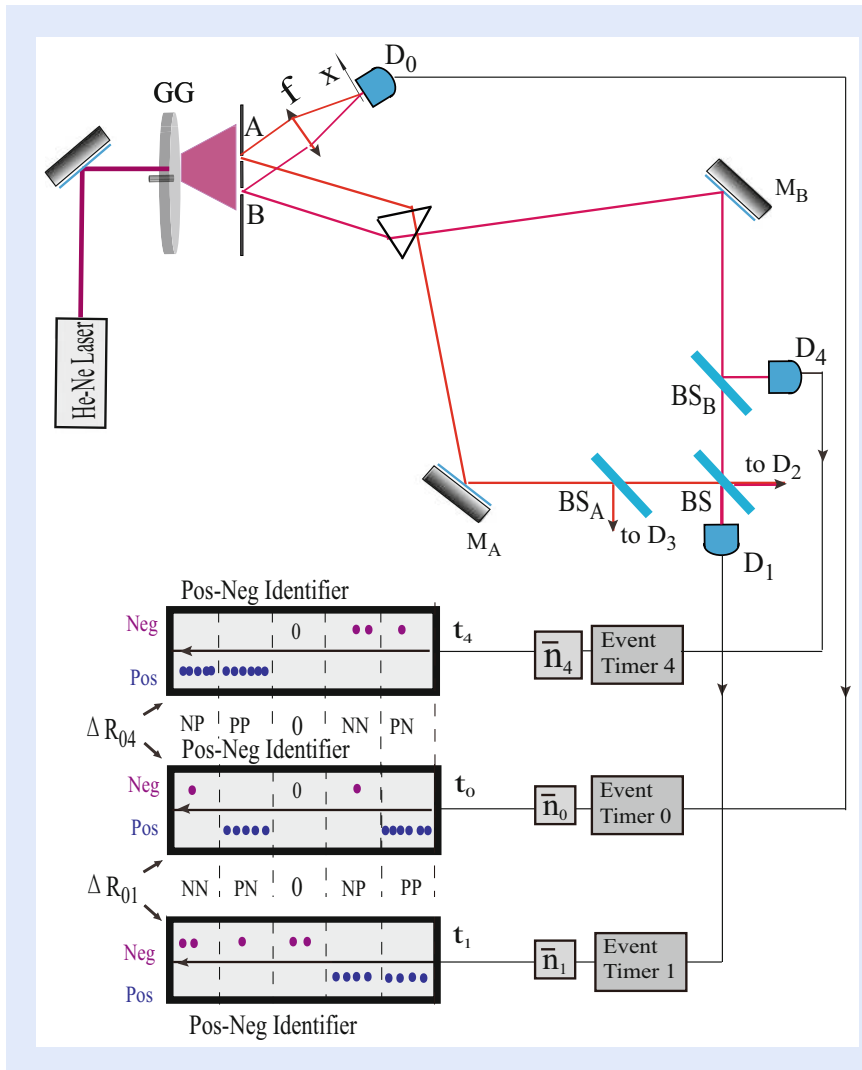### 16.3.2 Random Delayed Choice Quantum Eraser Two

Now we ask, again, what would happen if we replace the entangled photons with a randomly paired photons, or wavepackets, in thermal state? Can a randomly paired photons in thermal state erase the which-path information? The answer is, again, positive. A random delayed choice quantum eraser using randomly paired photons in thermal state has been demonstrated by Peng et al. recently [11].

The experiment setup of Peng et al. is schematically illustrated in ◘ Fig. 16.17. The experimental setup is almost the same as that of the experiment of Kim et al. of 2000, except the photon source and the coincidence measurement: the randomly paired photons, or wavepackets are created from a standard pseudo-thermal source and the purse this quantum eraser measures the photon-number fluctuation-correlation of thermal light. Thermal light has a peculiar "spatial coherence" property: the fluctuation of the measured photon-numbers, or intensities, correlated within its spatial coherence area only. When the measurements are beyond its coherence area, the photon-numbers fluctuation correlation vanish:

$$\langle \Delta n_A \Delta n_B \rangle \propto \left| G^{(1)}(\vec{\rho}_A, \vec{\rho}_B) \right|^2, \tag{16.73}$$

where $\Delta n_j, j = A, B$, is the photon-number fluctuation at $(\vec{\rho}_j, t_j)$ of the double-slit plane, $G^{(1)}(\vec{\rho}_A, \vec{\rho}_B)$ is the first-order spatial coherence function of the thermal field. The spatial coherence of thermal light guarantees the photon-numbers fluctuate correlatively only when $|\vec{\rho}_A - \vec{\rho}_B| < l_c$, where l_c is the spatial coherence length. In this experiment, we choose $|\vec{\rho}_A - \vec{\rho}_B| \gg l_c$. Under this condition, we have achieved $\langle \Delta n_A \Delta n_{A'} \rangle \neq 0$, $\langle \Delta n_B \Delta n_{B'} \rangle \neq 0$ but $\langle \Delta n_A \Delta n_B \rangle = 0$. Note, again, here $(\vec{\rho}_j)$ is on the double-slit plane, see ◘ Fig. 16.17. This peculiar property of thermal light together with the photon-number fluctuation-correlation measurement between D_0 and D_3 (or D_4) provides the which-path information. It is interesting, the which-slit information is erasable in the fluctuation-correlation measurement between D_0 and D_1 (or D_2).

The experimental setup in ◘ Fig. 16.17 can be divided into four parts: a thermal light source, a Young's double-slit interferometer, a Mach–Zehnder-like interferometer, and a photon-number fluctuation-correlation measurement circuit. (1) The light source is a standard pseudo-thermal source [29] which consists of a He-Ne laser beam (~ 2 mm diameter) and a rotating ground glass (GG). Within the ~ 2 mm diameter spot, the ground glass contains millions of tiny diffusers. A large number of randomly distributed sub-fields, or wavepacket, are scattered from millions of randomly distributed tiny diffusers with random phases. The pseudothermal field then passes a double-slit which is about 25 cm away from

Fig. 16.17 Schematic of a random delayed choice quantum eraser. The He-Ne laser beam spot on the rotating ground glass has a diameter of ~ 2 mm. A double-slit, with slit-width 150 μm and slit-separation 0. 7 mm, is placed ~ 25 cm away from the GG. The spatial coherence length of the pseudo-thermal field on the double-slit plane is calculated from Eq. (16.73), $l_c = \lambda/\Delta\theta \sim 160$ μ, which guarantees the two fields E_A and E_B are spatially incoherent. Under this experimental condition, the photon number fluctuates correlatively only within slit A or slit B. All beamsplitters are non-polarizing and 50/50. The two fields from the two slits may propagate to detector D_0 which is transversely scanned on the focal plan of lens f for observing the interference pattern of the double-slit interferometer; and may also pass a long a Mach–Zehnder-like interferometer and finally reach at D_1 or D_4 (D_2 or D_3). A positive-negative fluctuation-correlation protocol is followed to evaluate the photon-number fluctuation-correlations from the coincidences between D_0-D_1 and D_0-D_4 (or D_0-D_2 and D_0-D_3)

the GG. The spatial coherence length of the pseudo-thermal field on the double slit plane is calculated from Eq. (16.73), $l_c = \lambda/\Delta\theta \sim 160$ μ, which guarantees the spatial incoherence of the two fields E_A and E_B that passing through slit A and slit B, respectively. Under this experimental condition, the photon-numbers fluctuate correlatively only within slit-A or slit-B. We therefore learn the which-slit information in a photon number-fluctuation correlation measurement. (2) The double-slit has a slit-width 150 μm, and a slit-separation 0. 7 mm (distance between the center of two slits). A lens, f, is placed following the double-slit. On the focal-plane of the lens a scannable point-like photodetector D_0 is used to learn

16

the which-slit information or to observe the Young's double-slit interference pattern. (3) The Mach–Zehnder-like interferometer and the photodetectors D_1, D_2 are used to "erase" the which-slit information. Simultaneously, the joint-detection between D_0 and D_3 or D_4 is used to "learn" the which-slit information. All five photodetectors are photon-counting detectors working at single-photon level. The Mach–Zehnder-like interferometer has three beamsplitters, BS, BS_A, and BS_B, all of them are 50/50 non-polarizing beamsplitters. Moreover, the detectors are fast avalanche photodiodes with rise time less than 1 ns, and the path delay between BS_A or BS_B, and D_0 is ≈ 1.5 m which ensure that, at each joint-detection measurement, when a photon chooses to be reflected (read which-way) or transmitted (erase which-way) at BS_A or BS_B, it is already 5 ns later than the annihilation of its partner at D_0. Comparing the 1 ns rise time, we are sure this is a "delayed choice" made by that photon. (4) The photon-number fluctuation-correlation circuit consists of five synchronized "event-timers" which record the registration times of D_0, D_1, D_2, D_3, and D_4. A positive-negative fluctuation identifier follows each event-timer to distinguish "positive-fluctuation" Δn^+, from "negative-fluctuation" Δn^-, for each photodetector within each coincidence time window. The photon-number fluctuation-correlations of D_0-D_1: $\Delta R_{01} = \langle \Delta n_0 \Delta n_1 \rangle$ and D_0-D_4: $\Delta R_{04} = \langle \Delta n_0 \Delta n_4 \rangle$ are calculated, accordingly and respectively, based on their measured positive-negative fluctuations. The detailed description of the photon-number fluctuation-correlation circuit can be found in [30].

The experimental observation of ΔR_{04} is reported in ▣ Fig. 16.18. The data excludes any possible existing interferences. This measurement means the coincidences that contributed to ΔR_{04} must have passed through slit B. ▣ Figure 16.19 reports a typical experimental result of ΔR_{01}: a typical double-slit interference-diffraction pattern. The 100 % visibility of the sinusoidal modulation indicates complete erasure of the which-slit information.

Assuming a random pair of sub-fields at single-photon level, such as the mth and nth wavepackets, is scattered from the mth and the nth sub-sources located at transverse coordinates $\vec{\rho}_{0m}$ and $\vec{\rho}_{0n}$ of the ground glass and fall into the coincidence time windows of D_0-D_1 and D_0-D_4, the mth wavepacket may propagate to the double-slit interferometer and the nth wavepacket may pass through the Mach–Zehnder, or vice versa. Under the experimental condition of spatial incoherence between E_A and E_B, the which-slit information is learned from the photon-number
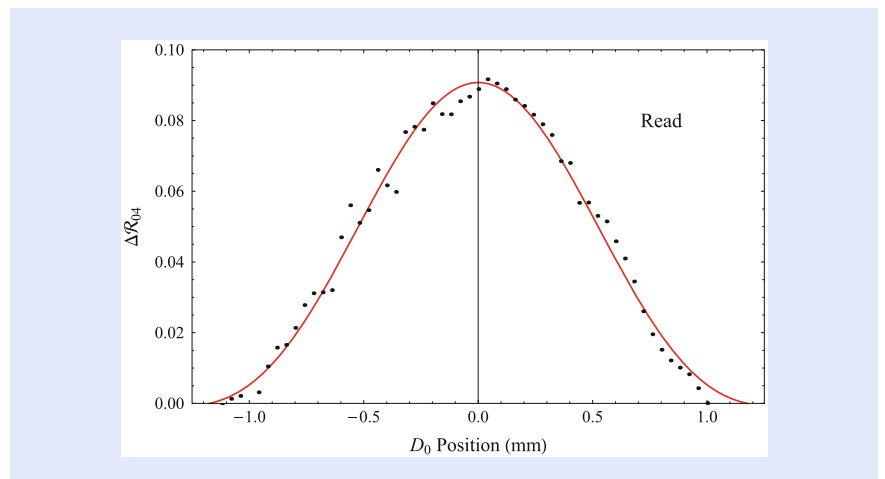
▣ **Fig. 16.18** The measured ΔR_{04} by scanning D_0 on the observation plane of the Young's double-slit interferometer. The *black dots* are experimental data, the *red line* is the theoretical fitting with Eq. (16.81)
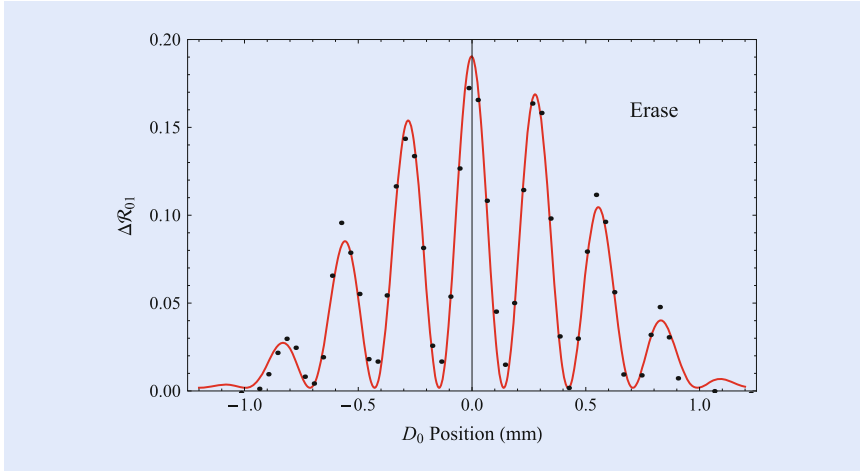
■ **Fig. 16.19** The measured ΔR_{01} as a function of the transverse coordinate of D_0. The *black dots* are experimental data, the *red line* is the theoretical fitting with Eq. (16.82)

fluctuation-correlation measurements $\Delta R_{04} = \langle \Delta n_0 \Delta n_4 \rangle = \langle \Delta n_{B0} \Delta n_{B4} \rangle$ of D_0-D_4, and no interference is observable by scanning D_0. It is interesting that the which-slit information are erasable in the photon-number fluctuation-correlation measurements of $\Delta R_{01} = \langle \Delta n_0 \Delta n_1 \rangle$ of D_0-D_1, resulting in a reappeared interference pattern as a function of the scanning coordinate of D_0.

The field operator at detector D_0 can be written in the following form in terms of the subfields:

$$
\begin{aligned}
\hat{E}^{(+)}(\mathbf{r}_0, t_0) &= \hat{E}_A^{(+)}(\mathbf{r}_0, t_0) + \hat{E}_B^{(+)}(\mathbf{r}_0, t_0) \\
&= \sum_m \left[\hat{E}_{mA}^{(+)}(\mathbf{r}_0, t_0) + \hat{E}_{mB}^{(+)}(\mathbf{r}_0, t_0) \right] \\
&= \sum_m \int d\mathbf{k}\, \hat{a}_m(\mathbf{k}) [g_m(\mathbf{k}; \mathbf{r}_A, t_A) g_A(\mathbf{k}; \mathbf{r}_0, t_0) \\
&\qquad\qquad + g_m(\mathbf{k}; \mathbf{r}_B, t_B) g_B(\mathbf{k}; \mathbf{r}_0, t_0)].
\end{aligned}
\tag{16.74}
$$

where $g_m(\mathbf{k}; \mathbf{r}_s, t_s)$ is a Green's function which propagates the mth subfield from the mth sub-source to the sth slit ($s = A, B$). $g_s(\mathbf{k}; \mathbf{r}_0, t_0)$ is another Green's function that propagates the field from the sth slit to detector D_0. It is easy to notice that, although there are two ways a photon can be detected at D_0, due to the first order incoherence of E_A and E_B, there should be no interference at the detection plane.

D_4 (D_3) in the experiment can only receive photons from slit B (slit A), so the field operator is then:

$$
\begin{aligned}
\hat{E}^{(+)}(\mathbf{r}_4, t_4) &= \sum_m \hat{E}_{mB}^{(+)}(\mathbf{r}_4, t_4) \\
&= \sum_m \int d\mathbf{k}\, \hat{a}_m(\mathbf{k}) g_m(\mathbf{k}; \mathbf{r}_B, t_B) g_B(\mathbf{k}; \mathbf{r}_4, t_4).
\end{aligned}
\tag{16.75}
$$

The detector D_1 (D_3), however, can receive photons from both slit A and slit B through the Mach–Zehnder-like interferometer, so the field operator has two terms:

$$
\hat{E}^{(+)}(\mathbf{r}_1, t_1) = \sum_m \left[\hat{E}_{mA}^{(+)}(\mathbf{r}_1, t_1) + \hat{E}_{mB}^{(+)}(\mathbf{r}_1, t_1) \right]
$$

$$
= \sum_m \int d\mathbf{k} \, \hat{a}_m(\mathbf{k})[g_m(\mathbf{k}; \mathbf{r}_A, t_A)g_A(\mathbf{k}; \mathbf{r}_1, t_1) \tag{16.76}
$$

$$
+ g_m(\mathbf{k}; \mathbf{r}_B, t_B)g_B(\mathbf{k}; \mathbf{r}_1, t_1)].
$$

Based on the state of Eq. (16.32) and the field operators of Eqs. (16.74–16.76), we apply the Glauber-Scully theory [25, 36] to calculate the photon-number fluctuation-correlation or the second-order coherence function $G^{(2)}(\mathbf{r}_0, t_0; \mathbf{r}_\alpha, t_\alpha)$ from the coincidence measurement of D_0 and D_α,($\alpha = 1, 2, 3, 4$):

$$
G^{(2)}(\mathbf{r}_0, t_0; \mathbf{r}_\alpha, t_\alpha)
$$

$$
= \left\langle \left\langle \Psi | E^{(-)}(\mathbf{r}_0, t_0) E^{(-)}(\mathbf{r}_\alpha, t_\alpha) E^{(+)}(\mathbf{r}_\alpha, t_\alpha) E^{(+)}(\mathbf{r}_0, t_0) | \Psi \right\rangle \right\rangle_{Es}
$$

$$
= \left\langle \left\langle \Psi | \sum_m E_m^{(-)}(\mathbf{r}_0, t_0) \sum_n E_n^{(-)}(\mathbf{r}_\alpha, t_\alpha) \sum_q E_q^{(+)}(\mathbf{r}_\alpha, t_\alpha) \sum_p E_p^{(+)}(\mathbf{r}_0, t_0) | \Psi \right\rangle \right\rangle_{Es}
$$

$$
= \sum_m \psi_m^*(\mathbf{r}_0, t_0)\psi_m(\mathbf{r}_0, t_0) \sum_n \psi_n^*(\mathbf{r}_\alpha, t_\alpha)\psi_n(\mathbf{r}_\alpha, t_\alpha) \tag{16.77}
$$

$$
+ \sum_{m,n} \psi_m^*(\mathbf{r}_0, t_0)\psi_n(\mathbf{r}_0, t_0)\psi_n^*(\mathbf{r}_\alpha, t_\alpha)\psi_m(\mathbf{r}_\alpha, t_\alpha)
$$

$$
= \langle n_0 \rangle \langle n_\alpha \rangle + \langle \Delta n_0 \Delta n_\alpha \rangle.
$$

Here $\psi_m(\mathbf{r}_\alpha, t_\alpha)$ is the effective wavefunction of the mth subfield at $(\mathbf{r}_\alpha, t_\alpha)$. In the case of $\alpha = 1, 2$

$$
\psi_m(\mathbf{r}_\alpha, t_\alpha) = \psi_{mA\alpha} + \psi_{mB\alpha}
$$

$$
= \int d\mathbf{k}\alpha_m(\mathbf{k})[g_m(\mathbf{k}; \mathbf{r}_A, t_A)g_A(\mathbf{k}; \mathbf{r}_\alpha, t_\alpha) + g_m(\mathbf{k}; \mathbf{r}_B, t_B)g_B(\mathbf{k}; \mathbf{r}_\alpha, t_\alpha)]. \tag{16.78}
$$

This shows that the measured effective wavefunction $\psi_m(\mathbf{r}_\alpha, t_\alpha)$ is the result of a superposition between two alternative amplitudes in terms of path-A and path-B, $\psi_{m\,\alpha} = \psi_{mA\,\alpha} + \psi_{mB\,\alpha}$. When $\alpha = 4$ (or $\alpha = 3$), the effective wavefunction has only one amplitude

$$
\psi_m(\mathbf{r}_4, t_4) = \psi_{mB4} = \int d\mathbf{k}\alpha_m(\mathbf{k})g_m(\mathbf{k}; \mathbf{r}_B, t_B)g_B(\mathbf{k}; \mathbf{r}_4, t_4). \tag{16.79}
$$

From Eq. (16.77) and the measurement circuit in ◻ Fig. 16.17, it is easy to find that what we measure in this experiment is the photon-number fluctuation-correlation:

$$
\langle \Delta n_0 \Delta n_\alpha \rangle = \sum_{m,n} \psi_m^*(\mathbf{r}_0, t_0)\psi_n(\mathbf{r}_0, t_0)\psi_n^*(\mathbf{r}_\alpha, t_\alpha)\psi_m(\mathbf{r}_\alpha, t_\alpha). \tag{16.80}
$$

We thus obtain

$$
\Delta R_{04} \propto \langle \Delta n_0 \Delta n_4 \rangle = \sum_{n \neq m} \psi_{mB0}^* \psi_{nB0} \psi_{nB4}^* \psi_{mB4} \propto \mathrm{sinc}^2(x\pi a/\lambda f), \tag{16.81}
$$

indicating a diffraction pattern which agrees with the experimental observation of ◻ Fig. 16.18.

In the case of $\alpha = 1, 2$, we obtain

$$
\begin{aligned}
\Delta R_{01} \quad &\propto \quad \langle \Delta n_0 \Delta n_1 \rangle \\
&\propto \quad \sum_{n \neq m} \Big[\psi_{mA0}^* \psi_{nA0} \psi_{nA1}^* \psi_{mA1} + \psi_{mB0}^* \psi_{nB0} \psi_{nB1}^* \psi_{mB1} \\
&\qquad\qquad + \psi_{mA0}^* \psi_{nB0} \psi_{nB1}^* \psi_{mA1} + \psi_{mB0}^* \psi_{nA0} \psi_{nA1}^* \psi_{mB1} \Big] \\
&\propto \quad \mathrm{sinc}^2(x\pi a/\lambda f) \, \cos^2(x\pi d/\lambda f),
\end{aligned}
\tag{16.82}
$$

which agrees with the experimental observation in ◘ Fig. 16.19.

16.4 Popper's Experiment

Popper's original thought experiment is schematically shown in ◘ Fig. 16.20 [12]. A point source S, positronium as Popper suggested, is placed at the center of the experimental arrangement from which entangled pair of particle 1 and particle 2 are emitted in opposite directions along the respective positive and negative x-axes towards two screens A and B. There are slits on both screens parallel to the y-axis and the slits may be adjusted by varying their widths Δy. Beyond the slits on each side stand an array of Geiger counters for the joint measurement of the particle pairs as shown in the figure. The entangled pair could be emitted to any direction in 4π solid angles from the point source. However, if particle 1 is detected in a certain direction, particle 2 is then known to be in the opposite direction due to the momentum conservation of the quanta pair.

First, let us imagine the case in which slits A and B are both adjusted very narrowly. In this circumstance, particle 1 and particle 2 experience diffraction at slit A and slit B, respectively, and exhibit greater Δp_y for smaller Δy of the slits.
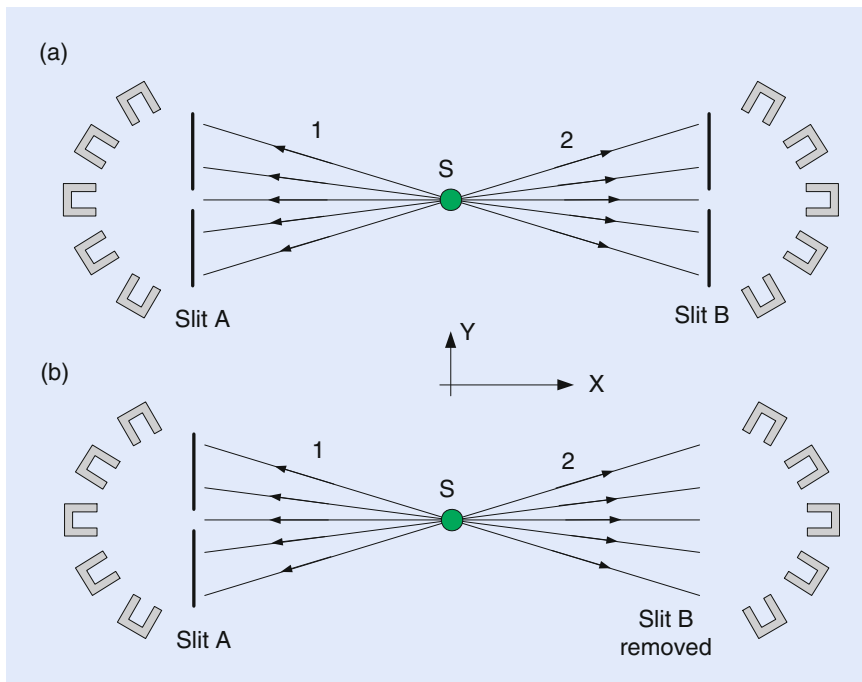
◘ **Fig. 16.20** Popper's thought experiment. An entangled pair of particles are emitted from a point source with momentum conservation. A narrow slit on screen A is placed in the path of particle 1 to provide the precise knowledge of its position on the y-axis and this also determines the precise y-position of its twin, particle 2, on screen B. (**a**) Slits A and B are both adjusted very narrowly. (**b**) Slit A is kept very narrow and slit B is left wide open

There seems to be no disagreement in this situation between Copenhagen and Popper.

Next, suppose we keep slit A very narrow and leave slit B wide open. The main purpose of the narrow slit A is to provide the precise knowledge of the position y of particle 1 and this subsequently determines the precise position of its twin (particle 2) on side B through quantum entanglement. Now, Popper asks, in the absence of the physical interaction with an actual slit, does particle 2 experience a greater uncertainty in Δp_y due to the precise knowledge of its position? Based on his beliefs, Popper provides a straightforward prediction: *particle 2 must not experience a greater Δp_y unless a real physical narrow slit B is applied*. However, if Popper's conjecture is correct, this would imply the product of Δy and Δp_y of particle 2 could be smaller than h ($\Delta y \, \Delta p_y < h$). This may pose a serious difficulty for Copenhagen and perhaps for many of us. On the other hand, if particle 2 going to the right does scatter like its twin, which has passed though slit A, while slit B is wide open, we are then confronted with an apparent *action-at-a-distance*!

The use of a *point source* in Popper's proposal has been criticized historically as the fundamental error Popper made. It is true that a point source can never produce a pair of entangled particles which preserves EPR correlation in momentum as Popper expected. However, notice that a *point source* is *not* a necessary requirement for Popper's experiment. What is required is a precise position-position EPR correlation: if the position of particle 1 is precisely known, the position of particle 2 is 100 % determined. Ghost imaging is a perfect tool to achieve this.

16.4.1 Popper's Experiment One

In 1999, Popper's experiment was realized by Y.H. Kim et al. [13] with the help of biphoton ghost imaging [37]. ◘ Figure 16.21 is a schematic diagram that is useful for comparison with the original Popper's thought experiment. It is easy to see that this is a typical ghost imaging experimental setup. An entangled photon pair is used to image slit A onto a distant image plane of "screen" B. In the setup, s_o is chosen to be twice the focal length of the imaging lens LS, $s_o = 2f$. According to the Gaussian thin lens equation, an equal size "ghost" image of slit A appears on the two-photon image plane at $s_i = 2f$. The use of slit A provides a precise knowledge of the position of photon 1 on the y-axis and also determines the precise y-position of its twin, photon 2, on screen B by means of the biphoton ghost imaging. The experimental condition specified in Popper's experiment is then achieved: when slit A is adjusted to a certain narrow width and slit B is wide open, slit A provides precise knowledge about the position of photon 1 on the y-axis up to an accuracy Δy which equals the width of slit A, and the corresponding ghost image of pinhole A at screen B determines the precise position y of photon 2 to within the same accuracy Δy. Δp_y of photon 2 can be independently studied by measuring the width of its "diffraction pattern" at a certain distance from "screen" B. This is obtained by recording coincidences between detectors D_1 and D_2 while scanning detector D_2 along its y-axis, which is behind screen B at a certain distance.

◘ Figure 16.22 is a conceptual diagram to connect the modified Popper's experiment with biphoton ghost imaging. In this unfolded ghost imaging setup, we assume the entangled signal-idler photon pair holds a perfect EPR correlation in momentum with $\delta(\mathbf{k}_s + \mathbf{k}_i) \sim 0$, which can be easily realized in a large transverse sized SPDC. In this experiment, we have chosen $s_o = s_i = 2f$. Thus, an equal size ghost image of slit A is expected to appear on the image plane of screen B.
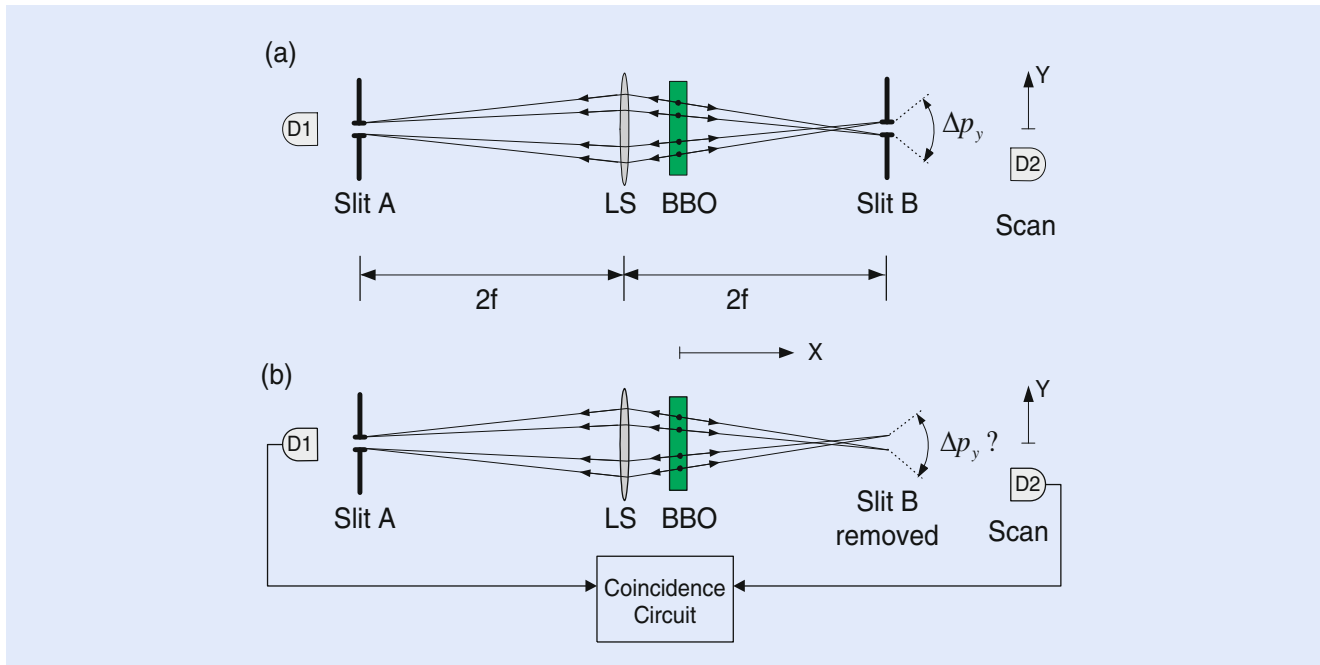
Fig. 16.21 Modified version of Popper's experiment. An entangled photon pair is generated by SPDC. A lens and a narrow slit A are placed in the path of photon 1 to provide the precise knowledge of its position on the y-axis and also to determine the precise y-position of its twin, photon 2, on screen B by means of biphoton ghost imaging. Photon counting detectors D_1 and D_2 are used to scan in y-directions for joint detections. (**a**) Slits A and B are both adjusted very narrowly. (**b**) Slit A is kept very narrow and slit B is left wide open
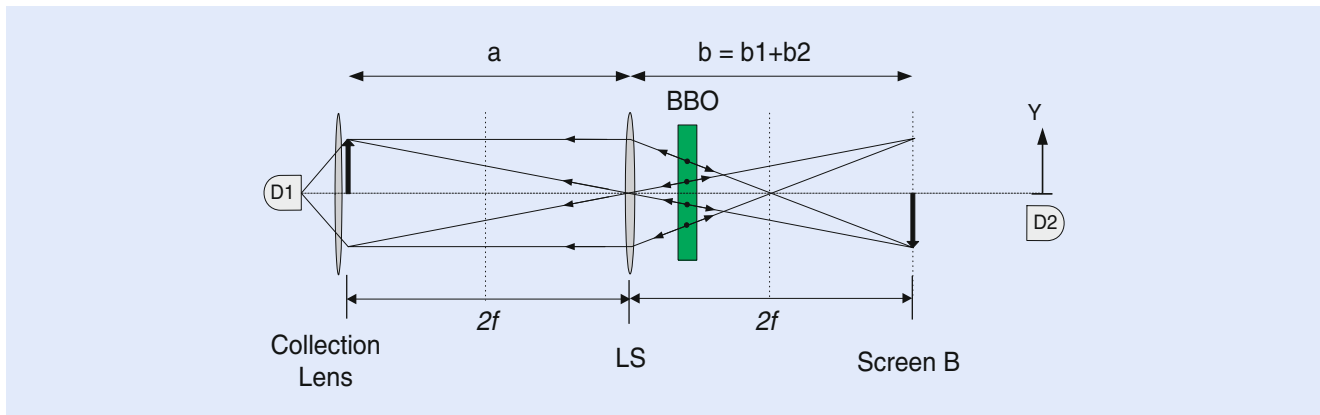
Fig. 16.22 An unfolded schematic of ghost imaging. We assume the entangled signal-idler photon pair holds a perfect momentum correlation $\delta(\mathbf{k}_s + \mathbf{k}_i) \sim 0$. The locations of the slit A, the imaging lens LS, and the ghost image must be governed by the Gaussian thin lens equation. In this experiment, we have chosen $s_o = s_i = 2f$. Thus, the ghost image of slit A is expected to be the same size as that of slit A

The detailed experimental setup is shown in ◘ Fig. 16.23 with indications of the various distances. A CW Argon ion laser line of $\lambda_p = 351.1$ nm is used to pump a 3 mm long beta barium borate (BBO) crystal for type-II SPDC to generate an orthogonally polarized signal-idler photon pair. The laser beam is about 3 mm in diameter with a diffraction limited divergence. It is important not to focus the pump beam so that the phase-matching condition, $\mathbf{k}_s + \mathbf{k}_i = \mathbf{k}_p$, is well reinforced in the SPDC process, where $\mathbf{k}_j$ ($j = s, i, p$) is the wavevectors of the signal (s), idler (i), and pump (p) respectively. The collinear signal-idler beams, with $\lambda_s = \lambda_i = 702$ .2 $nm = 2\lambda_p$ are separated from the pump beam by a fused quartz dispersion prism, and then split by a polarization beam splitter PBS. The signal beam
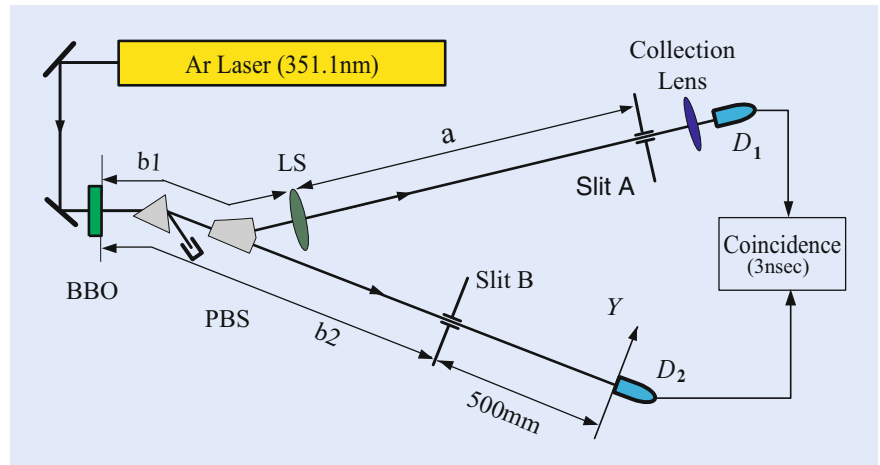
◘ Fig. 16.23 Schematic of the experimental setup. The laser beam is about 3 mm in diameter. The "phase-matching condition" is well reinforced. Slit A (0.16 mm) is placed 1000 mm $= 2f$ behind the converging lens, LS ($f = 500$ mm). The one-to-one ghost image (0.16 mm) of slit A is located at B. The optical distance from LS in the signal beam taken as back through PBS to the SPDC crystal ($b_1 = 255$ mm) and then along the idler beam to "screen B" ($b_2 = 745$ mm) is 1000 mm $= 2f$ ($b = b_1 + b_2$)

(photon 1) passes through the converging lens LS with a 500 mm focal length and a 25 mm diameter. A 0.16 mm slit is placed at location A which is 1000 mm ($s_o = 2f$) behind the lens LS. A short focal length lens is used with D_1 for collecting all the signal beam that passes through slit A. The point-like photon counting detector D_2 is located 500 mm behind "screen B." "Screen B" is the image plane defined by the Gaussian thin equation. Slit B, either adjusted as the same size as that of slit A or opened completely, is placed to coincide with the ghost image. The output pulses from the detectors are sent to a coincidence circuit. During the measurements, the bucket detector D_1 is fixed behind slit A while the point detector D_2 is scanned on the y-axis by a step motor.

Measurement 1 Measurement 1 studied the case in which both slits A and B were adjusted to be 0.16 mm. The y-coordinate of D_1 was chosen to be 0 (center) while D_2 was allowed to scan along its y-axis. The circled dot data points in ◘ Fig. 16.24 show the *coincidence* counting rates against the y-coordinates of D_2. It is a typical single-slit diffraction pattern with $\Delta y \, \Delta p_y = h$. Nothing is special in this measurement except that we have learned the width of the diffraction pattern for the 0.16 mm slit and this represents the minimum uncertainty of Δp_y. We should emphasize at this point that the *single* detector counting rate of D_2 as a function of its position y is basically the same as that of the coincidence counts except for a higher counting rate.

Measurement 2 The same experimental conditions were maintained except that slit B was left wide open. This measurement is a test of Popper's prediction. The y-coordinate of D_1 was chosen to be 0 (center) while D_2 was allowed to scan along its y-axis. Due to the entangled nature of the signal-idler photon pair and the use of coincidence measurement circuit, only those twins which have passed through slit A and the "ghost image" of slit A at screen B with an uncertainty of $\Delta y = 0.16 \, mm$ (which is the same width as the real slit B we have used in measurement 1) would contribute to the coincidence counts through the joint detection of D_1 and D_2. The diamond dot data points in ◘ Fig. 16.24 report the measured coincidence counting rates against the y coordinates of D_2. The measured width of the pattern is narrower than that of the diffraction pattern shown in measurement 1. It is also interesting to notice that the single detector counting rate of D_2 keeps constant in
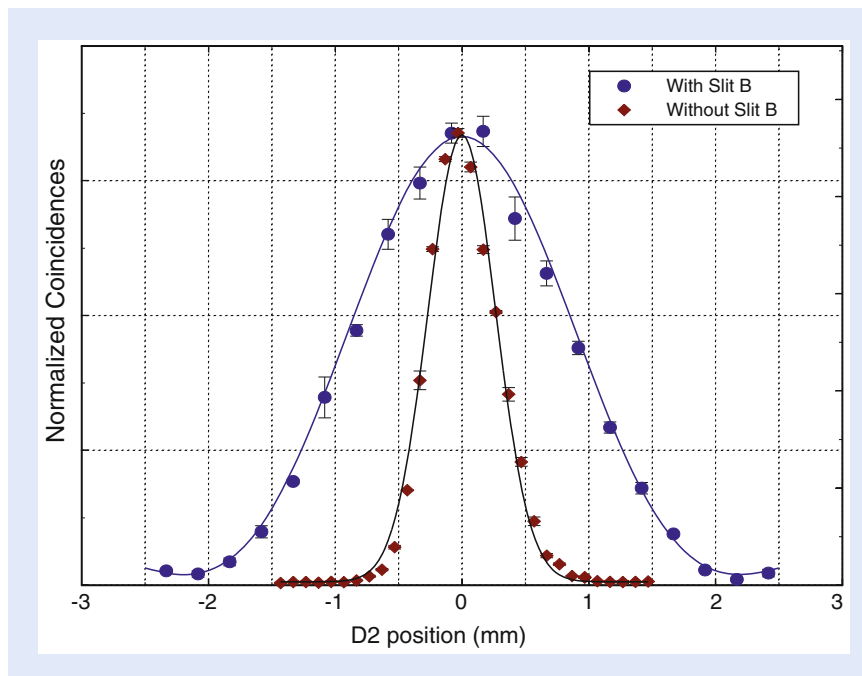
Fig. 16.24 The observed coincidence patterns. The *y*-coordinate of D_1 was chosen to be 0 (*center*) while D_2 was allowed to scan along its *y*-axis. *Circled dot points*: *Slit A = Slit B = 0.16* mm. *Diamond dot points*: *Slit A = 0.16* mm *, Slit B wide open*. The width of the sinc-function curve fitted by the *circled dot points* is a measure of the minimum Δp_y determined by a 0.16 mm slit. The fitting curve for the *diamond dots* is numerical result of Eq. (16.83), indicating a *blurred* ghost image of silt A

the entire scanning range, which is very different from that in measurement 1. The experimental data has provided a clear indication of $\Delta y\,\Delta p_y < h$ in the joint measurements of the entangled photon pairs.

Given that $\Delta y\,\Delta p_y < h$, is this a violation of the uncertainty principle? Does quantum mechanics agree with this peculiar experimental result? If quantum mechanics does provide a solution with $\Delta y\,\Delta p_y < h$ for photon 2, we would indeed be forced to face a paradox as EPR had pointed out in 1935.

Quantum mechanics does provide a solution that agrees with the experimental result. However, it is not the solution for photon 2. Instead, it is for a joint measurement of the entangled photon pair.

We now examine the experimental results with the quantum mechanical calculation by adopting the formalisms from the ghost image experiment with two modifications:

Case (I): - slits A = 0. 16 mm, slit B = 0. 16 mm.

This is the experimental condition for measurement one: slit B is adjusted to be the same as slit A. There is nothing surprise for this measurement. The measurement simply provides us the knowledge of Δp of photon 2 after the diffraction coursed by slit B of $\Delta y = 0.16$ mm. The experimental data shown in Fig. 16.24 agrees with the calculation. Notice that slit B is about 745 mm far away from the 3 mm two-photon source, the angular size of the light source is roughly the same as $\lambda/\Delta y$, $\Delta\theta \sim \lambda/\Delta y$, where $\lambda = 702$ nm is the wavelength and $\Delta y = 0.16$ mm is the width of the slit. The calculated diffraction pattern is very close to that of the "far-field" Fraunhofer diffraction of a 0.16 mm single-slit.

16

Case (II): - slit A = 0. 16 mm, slits $B \sim \infty$ (wide open).

Now we remove slit B from the ghost image plane. The calculation of the transverse effective two-photon wavefunction and the second-order correlation is the same as that of the ghost image except the observation plane of D_2 is moved from the image plane a distance of 500 mm behind. The two-photon image of slit A is located at a distance $s_i = 2f = 1000$ mm ($b_1 + b_2$) from the imaging lens, in this measurement D_2 is placed at $d = 1500$ mm from the imaging lens. The measured pattern is simply a "blurred" two-photon image of slit A. The "blurred" two-photon image can be calculated from Eq. (16.83)

$$\Psi(\vec{\rho}_o, \vec{\rho}_2) \propto \int_{lens} d\vec{\rho}_l \, G\left(\mid \vec{\rho}_2 - \vec{\rho}_l \mid, \frac{\omega}{cd}\right) G\left(\mid \vec{\rho}_l \mid, \frac{\omega}{cf}\right) G\left(\mid \vec{\rho}_l - \vec{\rho}_o \mid, \frac{\omega}{cs_o}\right)$$

$$\propto \int_{lens} d\vec{\rho}_l \, G\left(\mid \vec{\rho}_l \mid, \frac{\omega}{c}[\frac{1}{s_o} + \frac{1}{d} - \frac{1}{f}]\right) e^{-i\frac{\omega}{c}(\frac{\vec{\rho}_o}{s_o} + \frac{\vec{\rho}_i}{d}) \cdot \vec{\rho}_l} \tag{16.83}$$

where d is the distance between the imaging lens and D_2. In this measurement, D_2 was placed 500 mm behind the image plane, i.e., $d = s_i + 500$ mm. The numerical calculated "blurred" image, which is narrower than that of the diffraction pattern of the 0.16 mm slit B, agrees with the measured result of ◘ Fig. 16.24 within experimental error.

The measurement does show a result of $\Delta y \, \Delta p_y < h$. The measurement, however, has nothing to do with the uncertainty relation that governs the behavior of photon 2 (the idler). Popper and EPR were correct in the prediction of the outcomes of their experiments. Popper and EPR, on the other hand, made the same error by applying the results of two-particle physics to the explanation of the behavior of an individual particle.

In both the Popper and EPR experiments, the measurements are *joint detection* between two detectors applied to entangled states. Quantum mechanically, an entangled two-particle state only provides *the precise knowledge of the correlations of the pair*. The behavior of *photon 2* observed in the joint measurement is conditioned upon the measurement of its twin. A quantum must obey the uncertainty principle but the *conditional behavior* of a quantum in an entangled biparticle system is different in principle. We believe paradoxes are unavoidable if one insists the *conditional behavior* of a particle is the *behavior* of the particle. This is the central problem in the rationale behind both Popper and EPR. $\Delta y \, \Delta p_y \geq h$ is not applicable to the conditional behavior of either *photon 1* or *photon 2* in the experiments of Popper and EPR.

The behavior of photon 2 being conditioned upon the measurement of photon 1 is well represented by the two-photon amplitudes. Each of the *straight lines* in ◘ Fig. 16.22 corresponds to a two-photon amplitude. Quantum mechanically, the superposition of these two-photon amplitudes is responsible for a "click-click" measurement of the entangled pair. A "click-click" joint measurement of the two-particle entangled state projects out certain two-particle amplitudes, and only these two-particle amplitudes are featured in the quantum formalism. In the above analysis we never consider photon 1 or photon 2 *individually*. Popper's question about the momentum uncertainty of photon 2 is then inappropriate. The correct question to ask in these measurements should be: what is the uncertainty of Δp_y for the signal-idler *pair* which are localized within $\Delta y = 0.16$ mm at "screen" A with and without slit B? This is indeed the central point for Popper's experiment.

Once again, the demonstration of Popper's experiment calls our attention to the important message: the physics of the entangled two-particle system must inherently be very different from that of individual particles.

16.4.2 Popper's Experiment Two

In fact, the nonfactorizable, point-to-point image-forming correlation is not only the property of entangled photon pairs; it can also be realized in the joint-detection of a randomly paired photons in thermal state. In 2005, 10 years after the first ghost imaging experiment, a near-field lensless ghost imaging experiment that uses chaotic-thermal radiation source was demonstrated by Valencia et al. [38]. This experiment opened a door for the realization of Popper's thought experiment through the joint measurement of randomly paired photons in thermal state.

With the help of a novel joint detection scheme, namely the photon-number fluctuation correlation (PNFC) circuit [30], which distinguishes the positive and negative photon-number fluctuations measured by two single-photon counting detectors, and calculates the correlation between them, we were able to produce the ghost image of an object at a distance with 100 % visibility. By modifying the 1999 Kim-Shih experiment with a different light source and a lensless configuration, Peng and Shih realized Popper's thought experiment again in 2015 [14]. ◘ Figure 16.25 is an unfolded schematic, in which a large enough angular sized thermal source produces an equal-sized ghost image of slit-A at the plane $d_B = d_A$. The ghost image of slit-A can be verified by scanning the point-like photodetector D_B in the plane of slit-B. This ghost image provides the value of Δy through the
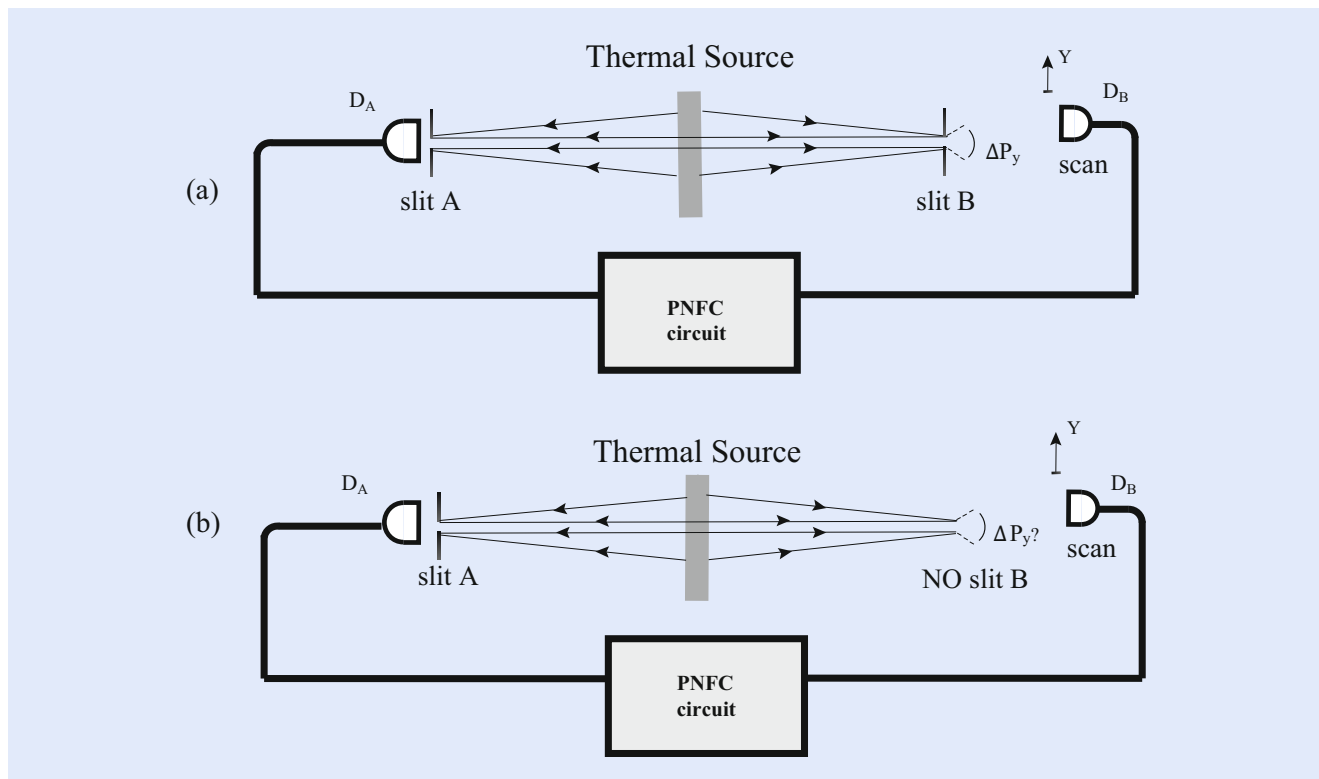

◘ **Fig. 16.25** Unfolded schematic of Popper's experiment with thermal light. The lensless ghost imaging setup with PNFC protocol produces an equal sized 100 % visibility ghost image of slit-A at the position of slit-B. Detector D_B is scanning transversely in the y direction to measure the photon-number fluctuation correlation with D_A when (**a**) Slit-A and slit-B are adjusted both very narrowly, and (**b**) Slit-A is kept very narrow and slit-B is left wide open

16
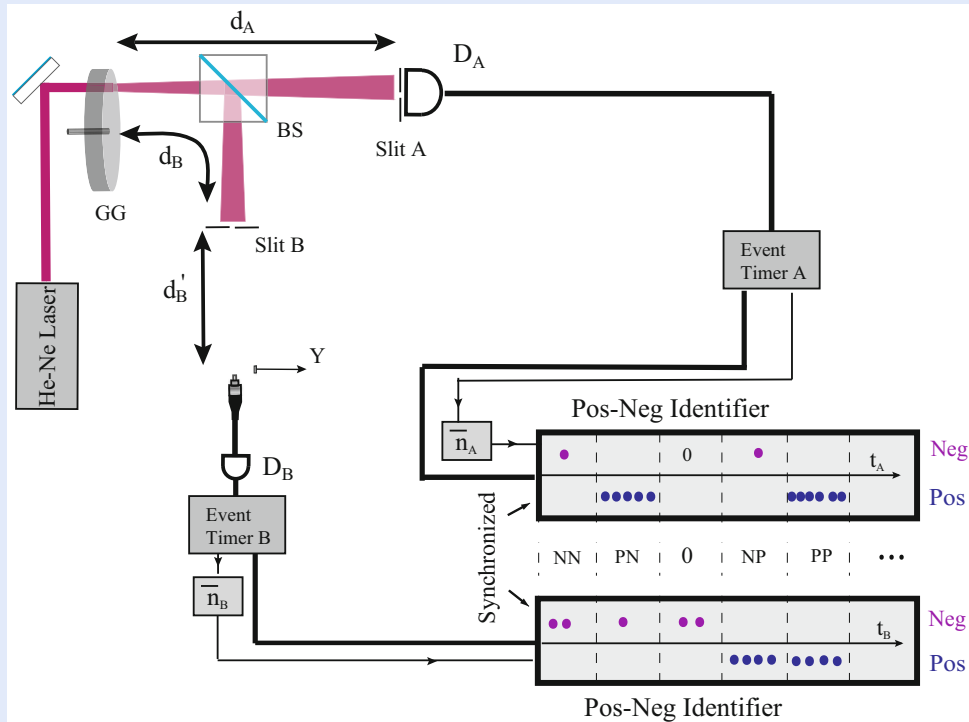
Fig. 16.26 Schematic of the experimental setup. A rotating ground glass (GG) is employed to produce pseudo-thermal light. BS is a 50/50 non-polarizing beam splitter. After BS, the transmitted beam passes through slit-A (0.15 mm) and collected by a "bucket" detector D_A which is put right after the slit. The reflected beam passes slit-B, which can be adjusted to be the same width as that of slit-A or wide open, and then reaches the scanning detector D_B. The distances from slit-A and slit-B to the source are the same ($d_A = d_B = 400$ mm). The distance from the scanning fiber tip of D_B to the plane of slit-B is $d_B' = 900$ mm. A PNFC protocol is followed to evaluate the photon-number fluctuation correlations from the coincidences between D_A and D_B

correlation measurement. Again, the question of Popper is: Do we expect to observe a diffraction pattern that satisfies $\Delta p_y \Delta y > h$? To answer this question, we again make two measurements following Popper's suggestion. Measurement-I is illustrated in the upper part of ◘ Fig. 16.25. In this measurement, we place slit-B, which has the same width as that of slit-A, coincident with the 1:1 ghost image of slit-A and measure the diffraction pattern by scanning D_B along the y-axis in far-field. In this measurement, we learn the value of Δp_y due to the diffraction of a real slit of Δy. Measurement-II is illustrated in the lower part of ◘ Fig. 16.25. Here, we open slit-B completely, scanning D_B again along the same y-axis to measure the "diffraction" pattern of the 1:1 ghost image with the same width as slit-A. By comparing the observed pattern width in measurement-II with that of measurement-I, we can examine Popper's prediction.

The experimental details are shown in ◘ Fig. 16.26. The light source is a standard pseudo-thermal source, consisting of a He-Ne laser beam and a rotating ground glass (GG). A 50/50 beamsplitter (BS) is used to split the pseudo-thermal light into two beams. One of the beams illuminates a single slit, slit-A, of width $D = 0.15$ mm located $d_A \sim 400$ mm from the source. A "bucket" photodetector D_A is placed right behind slit-A. An equal-sized ghost image of slit-A is then observable from the positive-negative photon-number fluctuation correlation measurement between the "bucket" detector D_A and the transversely scanning point-like photodetector D_B, if D_B is scanned on the ghost image plane located at $d_B = d_A = 400$ mm. In this experiment, however, D_B is scanned on a plane that is located $d_B' \sim 900$ mm behind the ghost image plane, to measure the "diffraction"
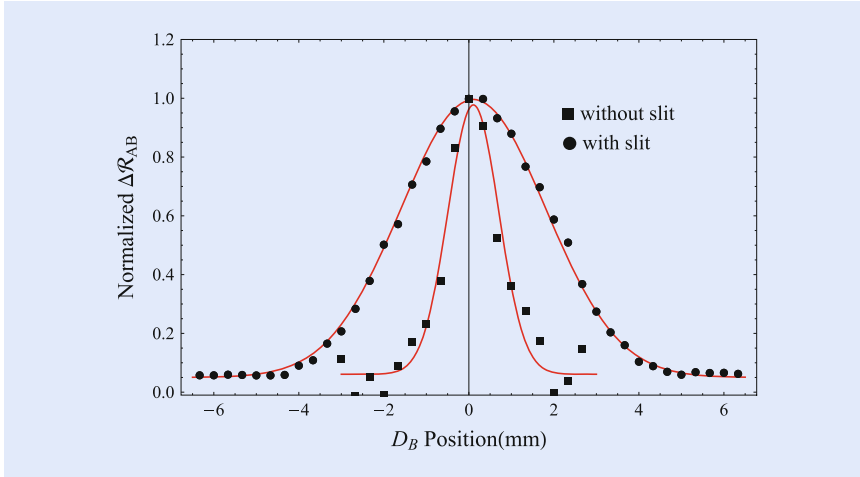
⬛ **Fig. 16.27** The observed diffraction patterns. *Circles*: slit-A and slit-B are both adjusted for 0. 15 mm. *Squares*: slit-A is 0.15 mm, slit-B is wide open. The width of the curve without the slit is almost three times narrower than that of the curve with slit, agreeing well with the theoretical predictions from Eqs. (16.89) and (16.91)

pattern of the ghost image. The output pulses from the two single-photon counting detectors are then sent to a PNFC circuit, which starts from two Pos-Neg identifiers follow two event-timers distinguish the "positive-fluctuation" Δn^+, from the "negative-fluctuation" Δn^-, measured by D_A and D_B, respectively, within each coincidence time window. The photon-number fluctuation-correlations of D_A-D_B: $\Delta R_{AB} = \langle \Delta n_A \Delta n_B \rangle$ is calculated, accordingly and respectively, based on their measured positive-negative fluctuations. The detailed description of the PNFC circuit can be found in [30].

The experiment was performed in two steps after confirming the 1:1 ghost image of slit-A. In measurement-I, we place slit-B ($D = 0.15$ mm) coincident with the ghost image and move D_B to a plane at $d_B' \sim 900$ mm to measure the diffraction pattern of slit-B. In measurement-II, we keep the same experimental condition as that of measurement-I, except slit-B is set wide open.

⬛ Figure 16.27 reports the experimental results. The circles show the normalized photon-number fluctuation correlation from the PNFC protocol against the position of D_B along the y-axis for Popper's measurement-I. As expected, we observed a typical single-slit diffraction pattern giving us the uncertainty in momentum, Δp_y^{real}. The squares show the experimental observation from the PNFC for Popper's measurement-II, when slit-B is wide open. The measured curves agree well with our theoretical fittings. We found the width of the curve representing no physical slit is much narrower than that of the real diffraction pattern, which agrees with Popper's prediction.

Similar to our early analysis in Bell state and in quantum eraser, we chose the coherent state representation for the calculation of the joint photodetection counting rate of D_A and D_B which is proportional to the second-order coherence function $G_{AB}^{(2)}$:

$$
\begin{aligned}
G_{AB}^{(2)} = \Big\langle \langle \hat{E}^{(-)}(\vec{\rho}_A, z_A, t_A) \hat{E}^{(-)}(\vec{\rho}_B, z_B, t_B) \\
\times \hat{E}^{(+)}(\vec{\rho}_A, z_B, t_B) \hat{E}^{(+)}(\vec{\rho}_A, z_A, t_A) \rangle_{\text{QM}} \Big\rangle_{\text{Es}},
\end{aligned}
\tag{16.84}
$$

where $E^{(+)}(\vec{\rho}_j, z_j, t_j)$ $(E^{(-)}(\vec{\rho}_j, z_j, t_j))$ is the positive (negative) field operator at space-time coordinate $(\vec{\rho}_j, z_j, t_j)$, $j = A, B$, with $(\vec{\rho}_j, z_j, t_j)$ the transverse,

longitudinal, and time coordinates of the photodetection event of D_A or D_B. Note, in the Glauber-Scully theory [25, 36], the quantum expectation and classical ensemble average are evaluated separately, which allows us to examining the two-photon interference picture before ensemble averaging.

The field at each space-time point is the result of a superposition among a large number of subfields propagated from a large number of independent, randomly distributed and randomly radiating sub-sources of the entire chaotic-thermal source,

$$
\begin{aligned}
\hat{E}^{(\pm)}(\vec{\rho}_j, z_j, t_j) &= \sum_m \hat{E}^{(\pm)}(\vec{\rho}_{0m}, z_{0m}, t_{0m}) g_m(\vec{\rho}_j, z_j, t_j) \\
&\equiv \sum_m \hat{E}_m^{(\pm)}(\vec{\rho}_j, z_j, t_j),
\end{aligned}
\tag{16.85}
$$

where $\hat{E}^{(\pm)}(\vec{\rho}_{0m}, z_{0m}, t_{0m})$ is the mth subfield at the source coordinate $(\vec{\rho}_{0m}, z_{0m}, t_{0m})$, and $g_m(\vec{\rho}_j, z_j, t_j)$ is the optical transfer function that propagates the mth subfield from coordinate $(\vec{\rho}_{0m}, z_{0m}, t_{0m})$ to $(\vec{\rho}_j, z_j, t_j)$. We can write the field operators in terms of the annihilation and creation operators:

$$
\hat{E}_m^{(+)}(\vec{\rho}_j, z_j, t_j) = C \int d\mathbf{k} \, \hat{a}_m(\mathbf{k}) \, g_m(\mathbf{k}; \vec{\rho}_j, z_j, t_j),
\tag{16.86}
$$

C is a normalization constant, $g_m(\mathbf{k}; \vec{\rho}_j, z_j, t_j)$, $j = A, B$, is the optical transfer function for mode $\mathbf{k}$ of the mth subfield propagated from the mth sub-source to the jth detector, and $\hat{a}_m(\mathbf{k})$ is the annihilation operator for the mode $\mathbf{k}$ of the mth subfield.

Substituting the field operators and the state, in the multi-mode coherent representation, into Eq. (16.84), we then write $G_{AB}^{(2)}$ in terms of the superposition of a large number of effective wavefunctions, or wavepackets:

$$
\begin{aligned}
&G^{(2)}(\vec{\rho}_B, z_B, t_B; \vec{\rho}_B, z_B, t_B) \\
&= \left\langle \sum_{m,n,p,q} \psi_m^*(\vec{\rho}_A, z_A, t_A) \psi_n^*(\vec{\rho}_B, z_B, t_B) \psi_p(\vec{\rho}_B, z_B, t_B) \psi_q(\vec{\rho}_A, z_A, t_A) \right\rangle_{Es} \\
&= \left\langle \sum_{m,n} \left| \psi_m(\vec{\rho}_A, z_A, t_A) \psi_n(\vec{\rho}_B, z_B, t_B) + \psi_n(\vec{\rho}_A, z_A, t_A) \psi_m(\vec{\rho}_B, z_B, t_B) \right|^2 \right\rangle_{Es} \\
&= \left\langle \sum_m \left| \psi_m(\vec{\rho}_A, z_A, t_A) \right|^2 \sum_n \left| \psi_n(\vec{\rho}_B, z_B, t_B) \right|^2 \right. \\
&\quad \left. + \sum_{m \neq n} \left[\psi_m^*(\vec{\rho}_A, z_A, t_A) \psi_m(\vec{\rho}_B, z_B, t_B) \psi_n(\vec{\rho}_A, z_A, t_A) \psi_n^*(\vec{\rho}_B, z_B, t_B) \right] \right\rangle_{Es} \\
&\equiv \langle n_A \rangle \langle n_B \rangle + \langle \Delta n_A \Delta n_B \rangle.
\end{aligned}
\tag{16.87}
$$

with

$$
\psi_s(\vec{\rho}_j, z_j, t_j) = \int d\mathbf{k} \, \alpha_s(\mathbf{k}) e^{i\varphi_{0s}} g_s(\mathbf{k}; \vec{\rho}_j, z_j, t_j),
$$

where $s = m, n, p, q$, $j = A, B$, and the phase factor $e^{i\varphi_{0s}}$ represents the random initial phase of the mth subfield. In Eq. (16.87), we have completed the ensemble average in terms of the random phases of the subfields, i.e. φ_{0s}, and kept the nonzero terms only. Equation (16.87) indicates the second-order coherence function is the result of a sum of a large number of subinterference patterns, each

subpattern indicates an interference in which a random pair of wavepackets interfering with the pair itself. For example, the mth and the nth wave packets have two different yet indistinguishable alternative ways to produce a joint photodetection event, or a coincidence count, at different space-time coordinates: (1) the mth wavepacket is annihilated at D_A and the nth wavepacket is annihilated at D_B; (2) the mth wavepacket is annihilated at D_B and the nth wavepacket is annihilated at D_A. In quantum mechanics, the joint detection probability of D_A and D_B is proportional to the normal square of the superposition of the above two probability amplitudes. We name this kind of superposition "nonlocal interference." The superposition of the two amplitudes for each random pair results in an interference pattern, and the addition of these large number of interference patterns yields the nontrivial correlation of the chaotic-thermal light.

The cross interference term in Eq. (16.87) indicates the photon-number fluctuation correlation $\langle \Delta n_A \Delta n_B \rangle$:

$$
\begin{aligned}
&\left\langle \Delta n_A(\vec{\rho}_A, z_A, t_A) \Delta n_B(\vec{\rho}_B, z_B, t_B) \right\rangle_{Es} \\
&= \left\langle \sum_{m \neq n} \left[\psi_m^*(\vec{\rho}_A, z_A, t_A) \psi_n(\vec{\rho}_A, z_A, t_A) \right] \left[\psi_m(\vec{\rho}_B, z_B, t_B) \psi_n^*(\vec{\rho}_B, z_B, t_B) \right] \right\rangle_{Es} \\
&\simeq \left\langle \sum_m \psi_m^*(\vec{\rho}_A, z_A, t_A) \psi_m(\vec{\rho}_B, z_B, t_B) \sum_n \psi_n(\vec{\rho}_A, z_A, t_A) \psi_n^*(\vec{\rho}_B, z_B, t_B) \right\rangle_{Es} .
\end{aligned}
$$
(16.88)

In measurement-I, the optical transfer functions that propagate the fields from the source to D_A and D_B are

$$
g_m(\vec{\kappa}, \omega; \vec{\rho}_A, z_A = d_A) = \frac{-i\omega e^{i(\omega/c)z_A}}{2\pi c d_A} \int d\vec{\rho}_s f(\vec{\rho}_s) \, e^{i\vec{\kappa} \cdot \vec{\rho}_s} G(|\vec{\rho}_s - \vec{\rho}_o|)_{[\omega/(cd_A)]},
$$

and

$$
\begin{aligned}
&g_n(\vec{\kappa}, \omega; \vec{\rho}_B, z_B = d_B + d'_B) \\
&= \frac{-\omega^2 e^{i(\omega/c)z_B}}{(2\pi c)^2 d_B d'_B} \int d\vec{\rho}_s \int d\vec{\rho}_i f(\vec{\rho}_s) e^{i\vec{\kappa} \cdot \vec{\rho}_s} G(|\vec{\rho}_s - \vec{\rho}_i|)_{[\omega/(cd_B)]} \\
&\qquad\qquad t(\vec{\rho}_i) G(|\vec{\rho}_i - \vec{\rho}_B|)_{[\omega/(cd'_B)]},
\end{aligned}
$$

where $\vec{\rho}_s$ is defined on the output plane of the source and $f(\vec{\rho}_s)$ denotes the aperture function of the source. We also assumed a perfect "bucket" detector D_A, which is placed at the object plane of slit-A ($\vec{\rho}_A = \vec{\rho}_o$), in the following calculation. $\vec{\rho}_i$ is defined on the ghost image plane, which coincides with the plane of slit-B, and $\vec{\rho}_B$ is defined on the detection plane of D_B, $t(\vec{\rho}_i)$ is the aperture function of slit-B. The function $G(|\alpha|)_{[\beta]}$ is the Gaussian function $G(|\alpha|)_{[\beta]} = e^{\frac{\beta}{2}|\alpha|^2}$. The measured fluctuation correlation can be calculated from Eq. (16.88)

$$
\Delta R_{AB} = \int d\vec{\rho}_o |t(\vec{\rho}_o)|^2 \mathrm{sinc}^2 [\omega_0 D \frac{\vec{\rho}_B}{2cd'_B}] \equiv C' \times \mathrm{sinc}^2 [\omega_0 D \frac{\vec{\rho}_B}{2cd'_B}],
$$
(16.89)

where $t(\vec{\rho}_o)$ is the aperture function of slit-A. The above calculation indicates a product between a constant C', which is from the integral on the "bucket" detector D_A, and a first order diffraction pattern of slit-B. With our experimental setup, the width of the diffraction pattern is estimated to be ~ 4 mm, which agrees well with the experimental observation, as shown in ▫ Fig. 16.27.

In measurement-II, with slit-B wide open, the field at D_B becomes

$$g_n(\vec{\kappa}, \omega; \vec{\rho}_B, z_B) = \frac{-i\omega e^{i(\omega/c)z_B}}{2\pi c z_B} \int d\vec{\rho}_s f(\vec{\rho}_s) e^{i\vec{\kappa}\cdot\vec{\rho}_s} G(|\vec{\rho}_s - \vec{\rho}_B|)_{[\omega/(cz_B)]}.$$

We first check if a ghost image of slit-A is present when scanning D_B in the ghost image plane of $d_B = d_A$. The photon-number fluctuation correlation is calculated to be

$$\Delta R_{AB} = \int d\vec{\rho}_o |t(\vec{\rho}_o)|^2 \sin c^2 \left[\frac{\omega_0 a}{c d_A} |\vec{\rho}_o - \vec{\rho}_B| \right]$$
$$= |t(\vec{\rho}_o)|^2 \otimes \sin c^2 \left[\frac{\omega_0 a}{c d_A} |\vec{\rho}_o - \vec{\rho}_B| \right] \approx |t(\vec{\rho}_B)|^2. \tag{16.90}$$

Note, we have placed D_A right behind slit-A and thus $\vec{\rho}_A = \vec{\rho}_o$. This suggests an equal-sized 100 % visibility ghost image on the plane of $d_B = d_A$.

When we move D_B away from the ghost image plane to the far-field plane of $d_B + d_{B'}$, the photon-number fluctuation correlation becomes:

$$\Delta R_{AB} = \int d\vec{\rho}_o |t(\vec{\rho}_o)|^2 \tilde{\mathcal{F}}_s^2 (m\vec{\rho}_o - \vec{\rho}_{B.} = |t(\vec{\rho}_o)|^2 \otimes \tilde{\mathcal{F}}_s^2 (m\vec{\rho}_o - \vec{\rho}_B), \tag{16.91}$$

where $\tilde{\mathcal{F}}_s$ is the Fourier transform of the defocused pupil function $\mathcal{F}_s = f(\vec{\rho}_s) e^{-i(\omega_0/2c\mu)\vec{\rho}_s 2}$ and μ, m are defined as $1/\mu = 1/d_A - 1/(d_B + d_{B'})$, $m = (d_B + d_{B'})/d_A$, respectively. The measured result of measurement-II is thus a convolution between the aperture function of slit-A, $t(\vec{\rho}_o)$, and the correlation function $\tilde{\mathcal{F}}_s (m\vec{\rho}_o - \vec{\rho}_{B.}$, resulting in a "blurred" image of slit-A. With our experimental setup, the width of the "diffraction" pattern is estimated to be ~ 1.4 mm, which is almost three times narrower than the diffraction pattern of measurement-I and agrees well with the experimental observation, as shown in ◘ Fig. 16.27. Compared with the Kim-Shih experimental result [13], we can see that although the number varies due to different experimental parameters, we have obtained a very similar result: the measured width of the "diffraction pattern" in measurement-II is much narrower than that of the diffraction pattern in measurement-I.

The above analysis indicates that the experimental observations are reasonable from the viewpoint of the coherence theory of light. The important physics we need to understand is to distinguish the first-order coherent effect and the second-order coherent effect, even if the measurement is for thermal light. In Popper's measurement-I, the fluctuation correlation is the result of first-order coherence. The joint measurement can be "factorized" into a product of two first-order diffraction patterns. After the integral of the "bucket" detector, which turns the diffraction pattern of slit-A into a constant, the joint measurement between D_A and D_B is a product between a constant and the standard first-order diffraction pattern of slit-B. There is no question the measured width of the diffraction pattern satisfies $\Delta p_y \Delta y \geq h$. In Popper's measurement-II when slit-B is wide open or removed, the measurement can no longer be written as a product of single-photon detections but as a non-separable function, i.e., a convolution between the object aperture function and the photon-number fluctuation correlation function of randomly paired photons, or the second-order coherence function of the thermal field. We thus consider the observation of $\Delta p_y \Delta y < h$ the result of the second-order coherence of thermal field which is caused from nonlocal interference: a randomly paired photon interferes with the pair itself at a distance by means of a joint photodetection event between D_A and D_B. The result of nonlocal two-photon interference does not contradict the uncertainty principle that governs the

behavior of single photons. Again, the observation of this experiment is not a violation of the uncertainty principle. The observation of $\Delta p_y \Delta y < h$ from thermal light, however, may reveal a concern about nonlocal interference.

16.5 Conclusion

This chapter reviewed three types of optical tests of the foundations of quantum theory: (1) EPR-Bohm-Bell correlation and Bell's inequality; (2) Scully's quantum eraser; (3) Popper's experiment. The results of these experiments are very interesting. On one hand, the experimental observations confirm the predictions of EPR-Bell, Scully, and Popper. On the other hand, the calculations from quantum theory perfectly agree with the experimental data. Moreover, apparently, the experimental observations do not lead to any "violations" of the principles of quantum mechanics. One important conclusion we may draw from these optical tests is that all the observations are the results of multi-photon interference: a group of photons interferes with the group itself at distance. The nonlocal multi-photon interference phenomena may never be understood in classical theory, however, it is legitimate in quantum mechanics. The superposition principle of quantum theory supports the superposition of multi-photon amplitudes, whether the photons are entangled or randomly grouped and despite the distances between these individual photodetection events. Perhaps we must accept the probabilistic nature of the "wavefunction" associated with a quantum or a group of quanta. Although a photon does not have a "wavefunction," we have developed the concept of an effective wavefunction for a photon and for a group of photons which have similar physical meanings as that of the wavefunction of a particle or the wavefunction of a group of particles. In terms of the superposition, although the effective wavefunction plays the same role as that of the electromagnetic wave, apparently, the effective wavefunction is different from the electromagnetic field in nature. Any efforts attempting to physically equal the two concepts would trap us in the question posed by Einstein: how long does it take for the energy on the other side of the 2-lightyear diameter sphere to arrive at the detector? Is it possible god of the quantum world does play dice?

References

1. Wheeler JA (1982) A delayed choice experiment. Maryland lectures collection
2. Einstein A, Podolsky B, Rosen N (1935) Can quantum-mechanical description of physical reality be considered complete? Phys Rev 47:777
3. Bohm D (1951) Quantum theory. Prentice-Hall, Yew York
4. Shih YH (2011) An introduction to quantum optics: photon and biphoton physics. CRC press, Taylor & Francis, London

5. Feynman RF, Leighton RB, Sands ML (1965) Lectures on physics. Addison-Wesley, Reading
6. Bohr N (1928) Das Quantenpostulat und die neuere Entwicklung der Atomistik. Naturwissenschaften 16:245
7. Dirac P (1930) The principle of quantum mecanics. Oxford University Press, Oxford
8. Wheeler JA, Zurek WH (1983) Quantum theory and measurement. Princeton University Press, Princeton
9. Scully MO, Druhl H (1982) Quantum eraser: a proposed photon correlation Experiment concerning observation and "delayed choice" in quantum mechanics. Phys Rev A 25:2208
10. Kim YH, Yu SP, Kulik SP, Shih YH, Scully MO (2000) A delayed "choice" quantum eraser. Phys Rev Lett 84:1
11. Peng T, Chen H, Shih YH, Scully MO (2014) Delayed-choice quantum eraser with thermal light. Phys Rev Lett 112:180401
12. Popper K (1934) Zur Kritik der Ungenauigkeitsrelationen. Naturwissenschaften 22:807
13. Kim YH, Shih YH (1999) Experimental realization of Popper's experiment: violation of the uncertainty principle? Found Phys 29:1849
14. Peng T, Simon J, Chen H, French R, Shih YH (2015) Popper's experiment with randomly paired photons in thermal state. Euro Phys Lett 109:14003
15. Clauser JF, Shimony A (1978) Bell's theorem. Experimental tests and implications. Rep Prog Phys 41:1881. An excellent review on Bell measurement before the introduction of SPDC
16. Freedman SJ, Clauser JF (1972) Experimental test of local hidden-variable theories. Phys Rev Lett 28:938
17. Clauser JF (1976) Experimental investigation of a polarization correlation anomaly. Phys Rev Lett 36:1223
18. Fry ES, Thompson PC (1976) Experimental test of local hidden-variable theories. Phys Rev Lett 37:465
19. Aspect A, Grangier P, Roger G (1981) Experimental tests of realistic local theories via Bell's theorem. Phys Rev Lett 47:460
20. Aspect A, Grangier P, Roger G (1982) Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: a new violation of Bell's inequalities. Phys Rev Lett 49:91
21. Alley CO, Shih YH (1986) In: Namiki M et al (ed) Foundations of quantum mechanics in the light of new technology. Physical Society of Japan, Tokyo; Shih YH, Alley CO (1988) New type of Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by optical parametric down conversion. Phys Rev Lett 61:2921
22. Kwiat PG, Mattle K, Weinfurter H, Zeilinger A, Sergienko AV, Shih YH (1995) New high-intensity source of polarization-entangled photon pairs. Phys Rev Lett 75:4337
23. Klyshko DN (1988) Photons and nonlinear optics. Gorden and Breach, New York
24. Rubin MH, Klyshko DN, Shih YH, Sergienko AV (1994) The theory of two-photon entanglement in type-II optical parametric down conversion. Phys Rev A 50:5122
25. Glauber RJ (1963) The quantum theory of optical coherence. Phys Rev 130:2529
26. Kim YH, Berardi V, Chekhova MV, Shih YH (2001) Anti-correlation effect in femtosecond-pulse pumped type-II spontaneous parametric down-conversion. Phys Rev A 64:R011801
27. Kim YH, Kulk SP, Shih YH (2001) Bell state preparation using pulsed non-degenerate two-photon entanglement. Phys Rev A 63:R060301
28. Peng T, Shih YH (2015) Bell correlation of thermal fields in photon-number fluctuations. Europhys Lett 112: 60006
29. Martienssen W, Spiller E (1964) Coherence and fluctuations in light beams. Am J Phys 32:919
30. Chen H, Peng T, Shih YH (2013) 100% correlation of chaotic thermal light. Phys Rev A 88:023808
31. Glauber RJ (1963) Coherent and incoherent states of the radiation field. Phys Rev 131:2766
32. Bell JS (1964) On the Einstein Podolsky Rosen paradox. Physics 1:195
33. Bell JS (1987) Speakable and unspeakable in quantum mechanics. Cambridge University Press, Cambridge
34. Herzog TJ, Kwiat PG, Weinfurter H, Zeilinger A (1995) Complementarity and the quantum eraser. Phys Rev Lett 75:3034
35. Walborn SP, Terra Cunha MO, Padua S, Monken CH (2002) Double-slit quantum erasure. Phys Rev A 65:033818
36. Scully MO, Zubairy MS (1997) Quantum optics. Cambridge University Press, Cambridge
37. Pittman TB, Shih YH, Strekalov DV, Sergienko AV (1995) Optical imaging by means of two-photon entanglement. Phys Rev A 52:R3429
38. Valencia A, Scarcelli G, D'Angelo M, Shih YH (2005) Two-photon imaging with thermal light. Phys Rev Lett 94:063601

Quantum Mechanical Properties of Light Fields Carrying Orbital Angular Momentum

Robert W. Boyd and Miles J. Padgett

R.W. Boyd (✉)
Department of Physics and School of Electrical Engineering and Computer Science, University of Ottawa,
Ottawa, ON, Canada

Max Planck Centre for Extreme and Quantum Photonics, University of Ottawa, Ottawa, ON, Canada

Department of Physics and Astronomy, The Institute of Optics, University of Rochester, Rochester, NY, USA

School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
e-mail: boydrw@mac.com

M.J. Padgett
School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
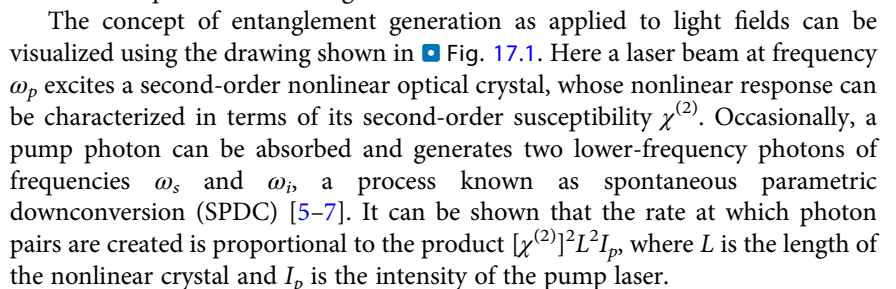e-mail: Miles.Padgett@glasgow.ac.uk

17.1 Introduction

There is a growing appreciation of the importance of quantum mechanics and in particular of quantum information science both for understanding the nature of the world in which we live and in the development of new technologies for communication and imaging. It is in this spirit that this chapter is written. The chapter deals with structured light fields, especially fields that carry orbital angular momentum (OAM), and their application to communication systems.

The topic of structured light fields in the quantum domain is intimately related to the topic of quantum imaging [1]. Quantum imaging is a discipline that studies quantum aspects of image formation and that uses quantum properties of light to produce images. Imaging techniques inherently require the ability to encode massive amounts of information in a light field. The quantum aspects of optical images can therefore be a key resource for quantum information and communication systems.

In this chapter we will review several examples of the quantum properties of structured light field. In broad concept, there are two sorts of quantum states that appear in our examples. One sort is a "single-photon" state, a state in which it is known that there is one and only one photon present in the field of interest. Such a state shows strong quantum properties because, for example, if you send such a beam onto a beamsplitter, the photon will emerge in one of the two output ports but not half and half in both. The other sort of quantum state of interest in this chapter is an entangled state of two photons. In fact, the concept of entanglement is one of the great mysteries of quantum mechanics. The term entanglement and the first explicit description of this phenomenon were introduced by Schrödinger in 1935 [2]. Here is quote from his paper:

"When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, viz. by endowing each of them with a representative of its own. I would not call that one but rather the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought."

This concept of entanglement leads to what today is often called "quantum weirdness," a term that arises from the seemingly paradoxical effects that can occur in an entangled system. Entanglement has, for instance, played a key role in compelling laboratory demonstrations [3, 4] of the nonlocality of quantum phenomena. However, this weirdness has in fact turned into an asset. This entanglement weirdness leads to effects such as quantum teleportation and certain forms of quantum communication with guaranteed security. One of the easiest ways to produce quantum entanglement uses nonlinear optical methods, and it is these methods that therefore give the field of photonics a special and elevated position in the arena of quantum technologies.

The concept of entanglement generation as applied to light fields can be visualized using the drawing shown in ◘ Fig. 17.1. Here a laser beam at frequency ω_p excites a second-order nonlinear optical crystal, whose nonlinear response can be characterized in terms of its second-order susceptibility $\chi^{(2)}$. Occasionally, a pump photon can be absorbed and generates two lower-frequency photons of frequencies ω_s and ω_i, a process known as spontaneous parametric downconversion (SPDC) [5–7]. It can be shown that the rate at which photon pairs are created is proportional to the product $[\chi^{(2)}]^2 L^2 I_p$, where L is the length of the nonlinear crystal and I_p is the intensity of the pump laser.

By conservation of energy, the condition $\omega_p = \omega_s + \omega_i$ must be satisfied, as illustrated in part (b) of the figure, furthermore these two new photons are
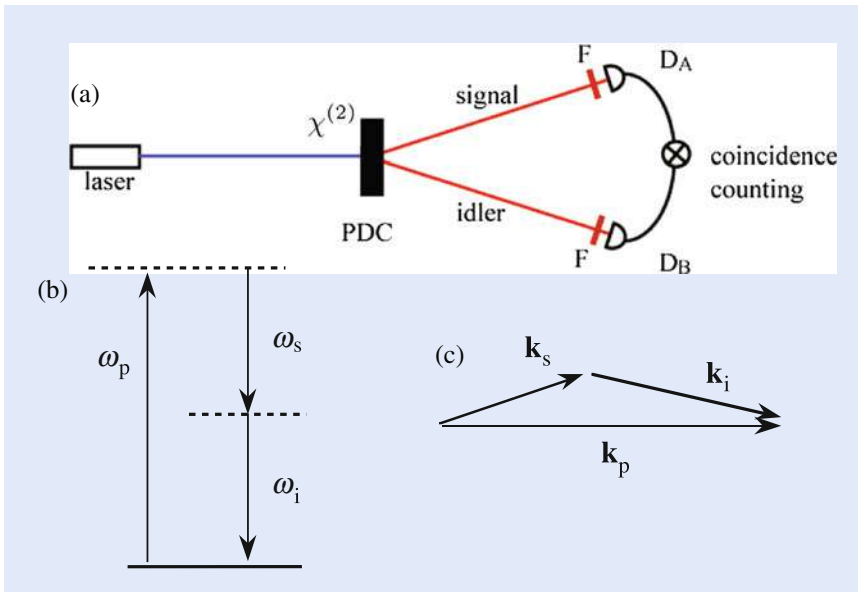
🔵 **Fig. 17.1** (**a**) Schematic illustration of the process of spontaneous parametric downconversion (SPDC). A laser beam excites a second-order ($\chi^{(2)}$) nonlinear optical crystal, leading to the generation of pairs of photons conventionally known as signal and idler photons. This process must obey the conservation of both energy (**b**) and momentum (**c**). These conditions lead to quantum correlations known as entanglement between the signal and idler photons, as discussed further in the text

generated at the same position (i.e., position correlated). However, photon momentum must also be conserved in this generation process, as illustrated in part (c) of the figure, and hence the two new photons are generated with opposite transverse momentum components (i.e., momentum anti correlated). It is these simultaneous conditions on position and momentum that lead to the paradox of Einstein et al. [8] and the concept of quantum entanglement [2].

The photons created by SPDC form entangled pairs, and in fact these photons can be entangled simultaneously in more than one pair of degrees of freedom. The possible types of entanglement that are often studied are

- position and transverse momentum
- angular position and orbital angular momentum
- time and energy
- polarization in different measurement bases.

Examples of the first two types of entanglement will be presented later in this chapter. Here we present a brief discussion of the other two types of entanglement.

By time-energy entanglement, one means that if one measures, for example, the energy of the signal photon, one is able to predict with certainty that the energy of the idler photon will be given by $\hbar\omega_i = \hbar\omega_p - \hbar\omega_s$. However, if one instead chooses to measure the moment of time at which the signal photon is emitted, one will always find that the idler photon is emitted at exactly the same moment. It seems that the product of uncertainty in tightness in the correlation of energies multiplied the uncertainty in the correlation of times can be arbitrarily small and certainly smaller than the value $\frac{1}{2}\hbar$ that one might have envisaged from the naive application of uncertainty relations [9, 10]. The situation is the essence of entanglement: the resolution of this seeming paradox is that a measurement that one performs on the signal photon results in a restriction of our ability to predict the properties of the idler photon, even if that idler photon is arbitrarily distant from

the signal photon measurement location. This nonlocality which applies to entangled systems leads to the phrase "spooky action at a distance."

Polarization entanglement can be similarly described. Under certain circumstances [6], each of the photons emitted by SPDC will be unpolarized, that is a complete statistical mixture of two orthogonal polarization states. However, for any one particular measurement the polarization of the signal photon will be found to have a defined value; one says that the measurement process projects the polarization state unto one of the polarization eigenstates. Furthermore, one finds that the idler photon will always be projected onto a polarization state that is orthogonal to that of the first photon (due to the conservation of angular momentum in the down conversion process). This type of entanglement is conveniently described in terms of the Bell states, which have played a key role in the understanding of many of the conceptual foundations of quantum mechanics. Details can be found in the excellent textbook of Gerry and Knight [11].

17.2 How Much Information Can One Photon Carry?

In classical optical telecommunication systems, many photons are required to transmit one bit of information. But it can be interesting to turn this question around and ask it differently: How much information can be carried by a single photon? Perhaps surprising to some is that research conducted over the last decade shows that there is no fundamental limit to the amount of information that can be carried by a single photon.

We start this section by giving a specific example of the ability to transfer many bits of information for each photon; this example will be developed in greater depth in the following sections. Laboratory procedures now exist for switching between single-photon states in any one of the Laguerre–Gaussian modes (Eq. (17.3)) of light using, for example, liquid crystal-based spatial light modulators [12] or a digital micromirror devices (DMD) [13]. It is crucial to recall that the Laguerre–Gaussian modes constitute an infinite set of basis functions. Thus, to the extent that one can perform OAM encoding and decoding with high efficiency, there is no limit to the amount of information that can be carried by a single photon.

The ability to encode more than one bit per photon is, of course, not restricted to the Laguerre–Gaussian light beams. More generally, the transverse degree of freedom of the light field offers a means to carry and manipulate quantum information. An example of multi-bit information transfer relating to imaging is provided in an experiment performed by Broadbent et al. [14]. A schematic of this experiment is shown in �“ Fig. 17.2. Part (a) of the figure shows a multiplexed hologram of objects A and B. By saying that the hologram is multiplexed, we mean that different write-beam directions are used to form the interference fringes for each object. Part (b) of the figure shows the read-out stage. It makes use of entangled photons created by parametric downconversion in a BiBO crystal. One of these photons falls onto the trigger detector, which heralds the presence of the photon in the other arm. This photon falls onto an object in its arm, which could be either object A or B. This photon is diffracted from the hologram into the path of either detector A or B, depending on which object is placed in this arm. In this manner, one can determine with high reliability which object is located in this arm, even though only one photon is used to make this determination. Quantitative results are presented in the paper. It is shown that the likelihood of a misidentification (that is, for example, that the photon is detected by detector B when in fact object A is present) is less than 1 %.

The experiment just described shows that one can discriminate between two objects using single-photon illumination. A subsequent experiment [15]
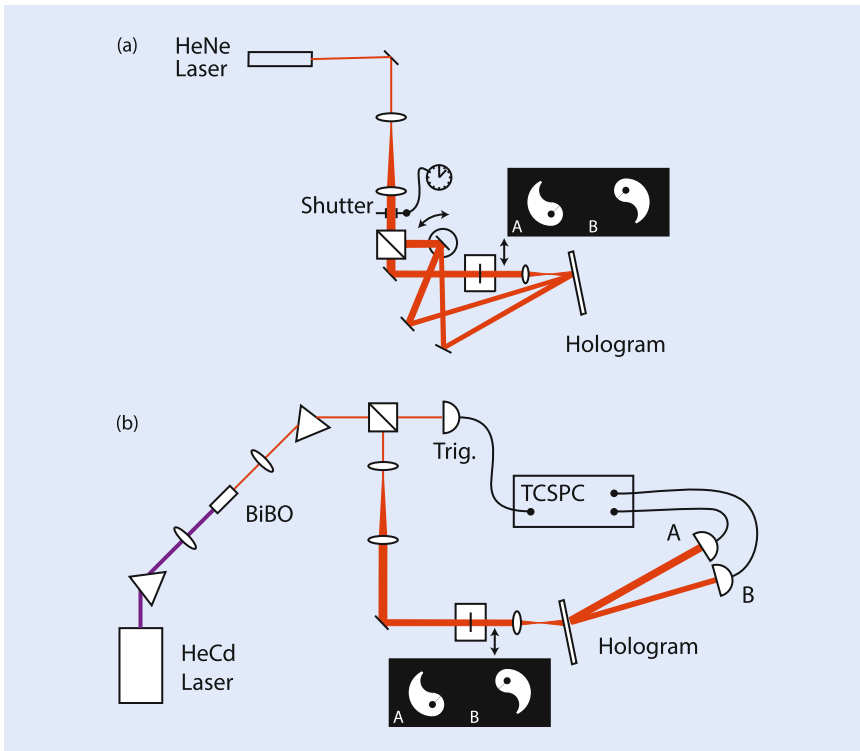
■ **Fig. 17.2** Configuration of the "single-photon imaging" experiment of Broadbent et al. [14] described in the text. Part (**a**) shows the procedure for writing a multiplexed hologram, and part (**b**) shows the read-out stage, which operates at the single-photon level. The TCPSC is a time-correlated single-photon counter

demonstrated the ability to discriminate among four objects, again using only single-photon illumination. For this experiment a "quantum ghost-imaging" protocol [16] is used. The setup is shown schematically in ■ Fig. 17.3. Spatially entangled photons are again created by the process of parametric downconversion. One of these photons illuminates one of the four test objects (only two are displayed in the diagram to avoid clutter) and the other falls onto a multiplexed hologram, where it is diffracted into one of four output ports. Coincidence events between the reference detector R and one of the detectors A, B, etc. are recorded. In this figure, DM denotes a dichroic mirror for blocking the pump laser and IF is an interference filter with a 10 nm bandwidth, centered at 727.6 nm.

The results of this experiment are shown in ■ Fig. 17.4. Note that input object a produces counts predominantly in detector A, and similarly for the other three object-detector combinations. The data are displayed using two different normalization conventions. In part (a), data for each object-detector combination are normalized by the maximum coincidence count for the corresponding object. In part (b), the T/A ratio is calculated by dividing the total coincidences by the accidental coincidences for each object-detector combination. Part (c) of the figure shows the four test objects. These results show that one can reliably discriminate among four objects even when they are illuminated with weak light at a single photon level. However, in this experiment the detection efficiency was low, and thus more than one photon needed to illuminate the object in order to make an unambiguous determination. In fact, for the sort of simple, multiplexed hologram used in this experiment the maximum detection efficiency (that is, assuming lossless optical elements and unit quantum efficiency detectors) is equal to $1/N$ (where N is the number of objects). However, there seems to be no reason in principle [17] why a hologram could not be designed to give a maximum detection efficiency of unity.
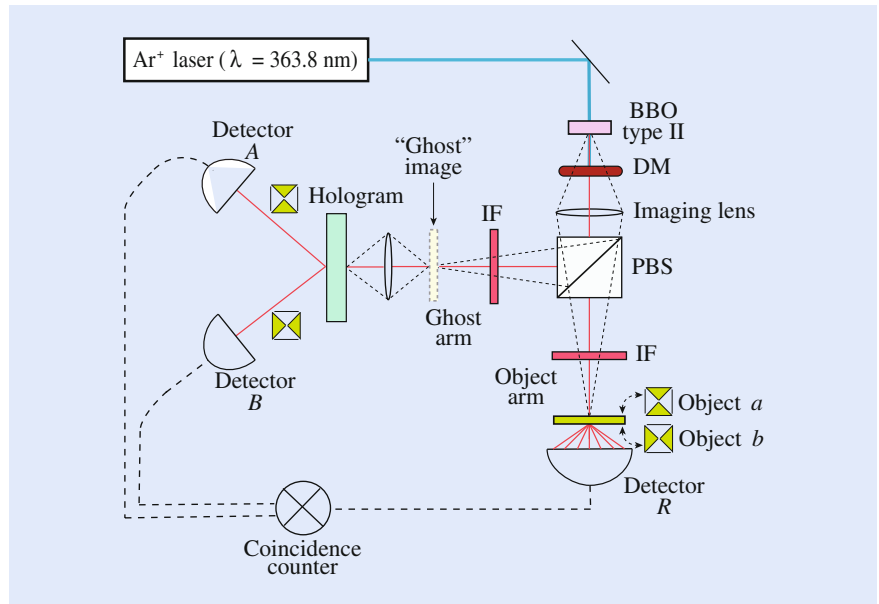
Fig. 17.3 Configuration of an experiment to demonstrate the discrimination among four different objects at the single-photon level, as described by Malik et al. [15]
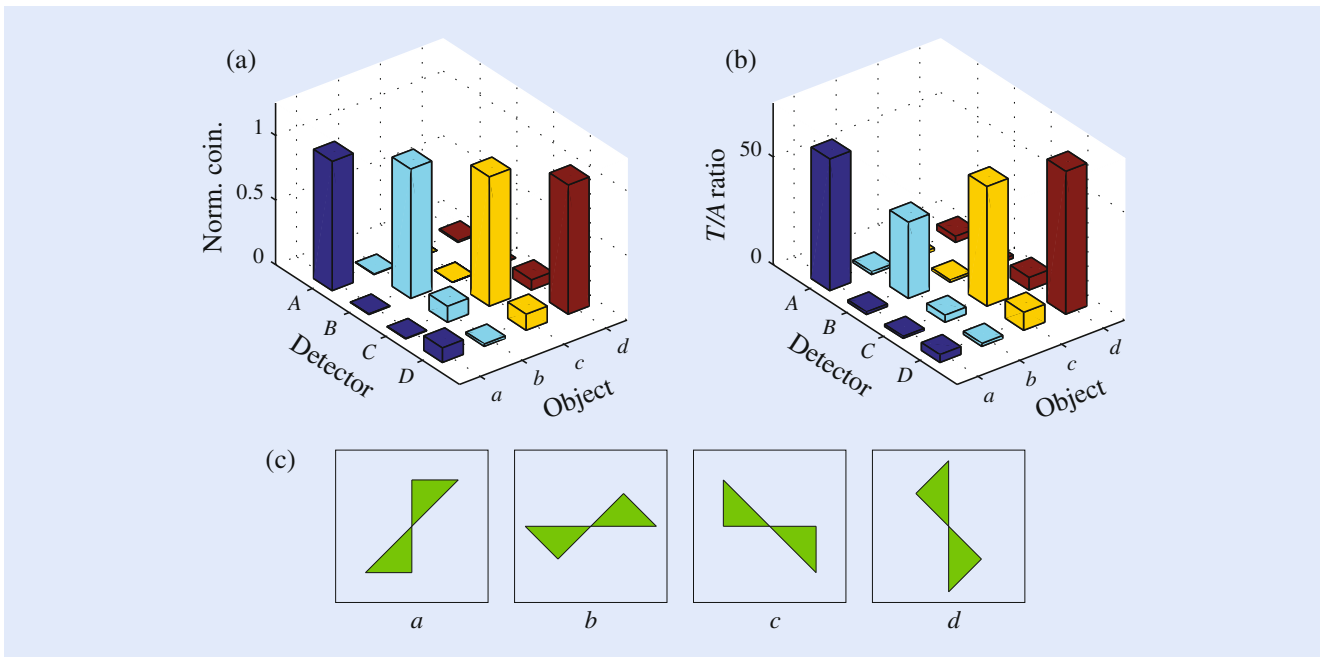

Fig. 17.4 Results of the single-photon ghost imaging experiment described in Fig. 17.3, which can distinguish between four different (non-overlapping) objects [15]

We note that this sort of ghost imaging relies upon correlations between photon pairs. If detector R registers a photon, we know with certainly that this photon possessed the transverse mode structure given by the transmission function of the object in its path. Since this detected photon is entangled with the photon in the other arm, this detected photon must therefore acquire the same conjugate mode structure, and thus be diffracted by the hologram into a specific output port.

17.3 Light Beams that Carry Orbital Angular Momentum

We turn now to another example of a structured light field, namely one carrying orbital angular momentum, which displays interesting quantum properties that can lead to important applications. First, we consider a light field of the form

$$E(r, t) = u(x, y)e^{i\ell\phi}e^{i(kz-\omega t)} \tag{17.1}$$

Here $u(x, y)$ is some function of the transverse coordinates x and y, ℓ is a positive or negative integer, $k = \omega/c$ is the propagation constant, z is the longitudinal coordinate, and ω is the angular frequency. We assume propagation through vacuum. It is well known that such a field carries angular momentum of amount $\ell\hbar$ per photon [18]. For this reason, ℓ is often referred to as the OAM quantum number or OAM mode index. This contribution to the angular momentum is referred to as orbital angular momentum (OAM), distinguishing it from spin angular momentum, which is associated with circular polarization of a light field. These two contributions are additive, and in the paraxial limit considered here independent of each other.

We can understand why the field given by Eq. (17.1) carries angular momentum with the help of the sketch in part (a) of ▫ Fig. 17.5. We see that such a field possesses a wavefront structure in the form of a helix, and that the phase at each point advances in the azimuthal direction at a rate proportional to the value of l. One might well imagine that a small particle placed in such a beam would experience a radiation pressure in the direction of phasefront normal and hence a force with an azimuthal component that induces the object to begin to rotate around the beam axis, and in fact this is just what has been observed experimentally [19, 20].

Equation (17.1) shows that the a light field will carry OAM for any transverse mode function $u(x, y)$. However, some specific mode functions are especially important in the utilization of structured light fields. One such example is that of the Laguerre–Gaussian modes, and we will now briefly explore their properties. The paraxial approximation to the wave equation $(\nabla^2 - \partial^2/\partial t^2)E(x, y, z) = 0$ gives us the paraxial wave equation, which is written in the cartesian coordinate system as

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + 2ik\frac{\partial}{\partial z} \right)E(x, y, z) = 0. \tag{17.2}$$

The paraxial wave equation is satisfied by the Laguerre–Gaussian modes, a family of orthogonal modes that have a well-defined orbital angular momentum. The field amplitude, in cylindrical coordinates, $LG_p^l(\rho, \phi, z)$ of a normalized Laguerre–Gaussian mode is given by

$$\begin{aligned} G_p^\ell(\rho, \phi, z) = \ & \sqrt{\frac{2p!}{\pi(|\ell|+p)!}}\frac{1}{w(z)}\left[\frac{\sqrt{2}\rho}{w(z)}\right]^{|\ell|}L_p^\ell\left[\frac{2\rho^2}{w^2(z)}\right]\exp\left[-\frac{\rho^2}{w^2(z)}\right] \\ & \times\exp\left[-\frac{ik^2\rho^2 z}{2(z^2+z_R^2)}\right]\exp\left[i(2p+|\ell|+1)tan^{-1}\left(\frac{z}{z_R}\right)\right]e^{-i\ell\phi}, \end{aligned} \tag{17.3}$$

where k is the wave-vector magnitude of the field, z_R the Rayleigh range, $w(z)$ the radius of the beam at z, ℓ is the azimuthal quantum number, and p is the radial quantum number. L_p^l is the associated Laguerre polynomial.
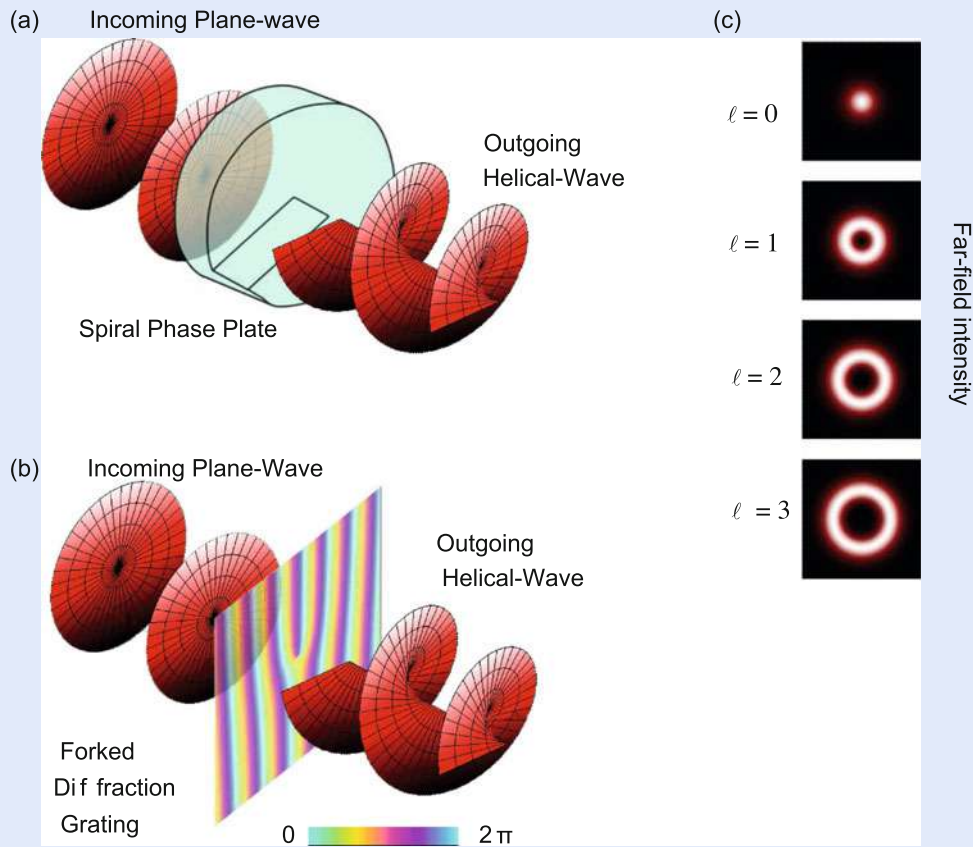
Fig. 17.5 Two methods of producing a light beam that carries orbital angular momentum. (**a**) The conceptually simplest way to form a beam carrying OAM is to pass a plane wave beam through a spiral phase plate, an optical element whose thickness increases linearly with the azimuthal angle. After transmission through such an element, a incident plane wave is transformed into a light beam with helical phasefronts. The height of the phase step controls the azimuthal index ℓ of the transmitted beam. (**b**) Alternatively, one can replace the phase plate with its holographic equivalent with a phase or amplitude structure in the form of a pitchfork as shown. The first-order diffracted beam will have helical phasefronts with an azimuthal index given by the number of dislocations in the pitchfork. These holograms are conveniently created by using a spatial light modulator. (**c**) Examples of some OAM beams produced by these approaches

Some methods for the production of beams that carry OAM are sketched in Fig. 17.5. The use of a spiral phase plate is shown in part (a) [19]. Another important method for generating OAM light beams (part b) is to impress a specially designed computer-generated hologram (CGH) taking the form of a forked diffraction grating [12, 20–23] or a digital micromirror device (DMD) [13]. If a beam with nearly plane wave fronts, such as a Gaussian laser beam, is made to fall onto such a CGH, the diffracted light will acquire the desired form of a beam carrying OAM. Another means to form beams carrying OAM is through the use of a device known as a q-plate [24, 25]. This device is a birefringent phase plate in which the orientation of the birefringent axes varies uniformly as a function of azimuthal position around the axis of the plate. Such a device acts as a spin angular momentum to OAM converter, that is, the OAM carried by the output beam depends on the polarization state of the input beam. A q-plate can thus serve as a quantum interface between polarization-encoded quantum light states and OAM-encoded quantum light states.

17.4 Fundamental Quantum Studies of Structured Light Beams

In ◼ Sect. 17.1 of this chapter we noted that the process of spontaneous parametric downconversion can lead to entanglement in several different degrees of freedom, including position-momentum [26], time-energy [9, 10], polarization [3, 4], and superpositions of OAM modes [27] or explicitly angle-OAM [28]. In this section we provide a brief account of work aimed at studying these various types of entanglement.

In Einstein, Podolski, and Rosen's (EPR's) classic paper [8], they argued against the completeness of quantum mechanics. Their argument was based on the situation of two particles that were strongly correlated both in position and momentum. Later, David Bohm [29] restated this argument in terms of two particles entangled in their spin (or polarization), and it was this spin-version of the EPR paradox that was treated by John Bell in devising his celebrated Bell inequalities. In the ensuing decades, most subsequent work [3, 4] has concentrated on the polarization of the EPR paradox. In 1990 Rarity and Tapster [30] extended the Bell violation to one based on measurement of phase and momentum. Howell et al. [26] later performed an experimental investigation in which they studied the original (i.e., Einstein et al. [8]) position-momentum version of the EPR paradox. Some of their results are summarized in ◼ Fig. 17.6. Their experimental procedure is as follows. Photons entangled in position and momentum were created by type-II parametric downconversion in a BBO nonlinear crystal, and the two photons were separated by a polarizing beamsplitter (PBS) and traveled over separate paths.
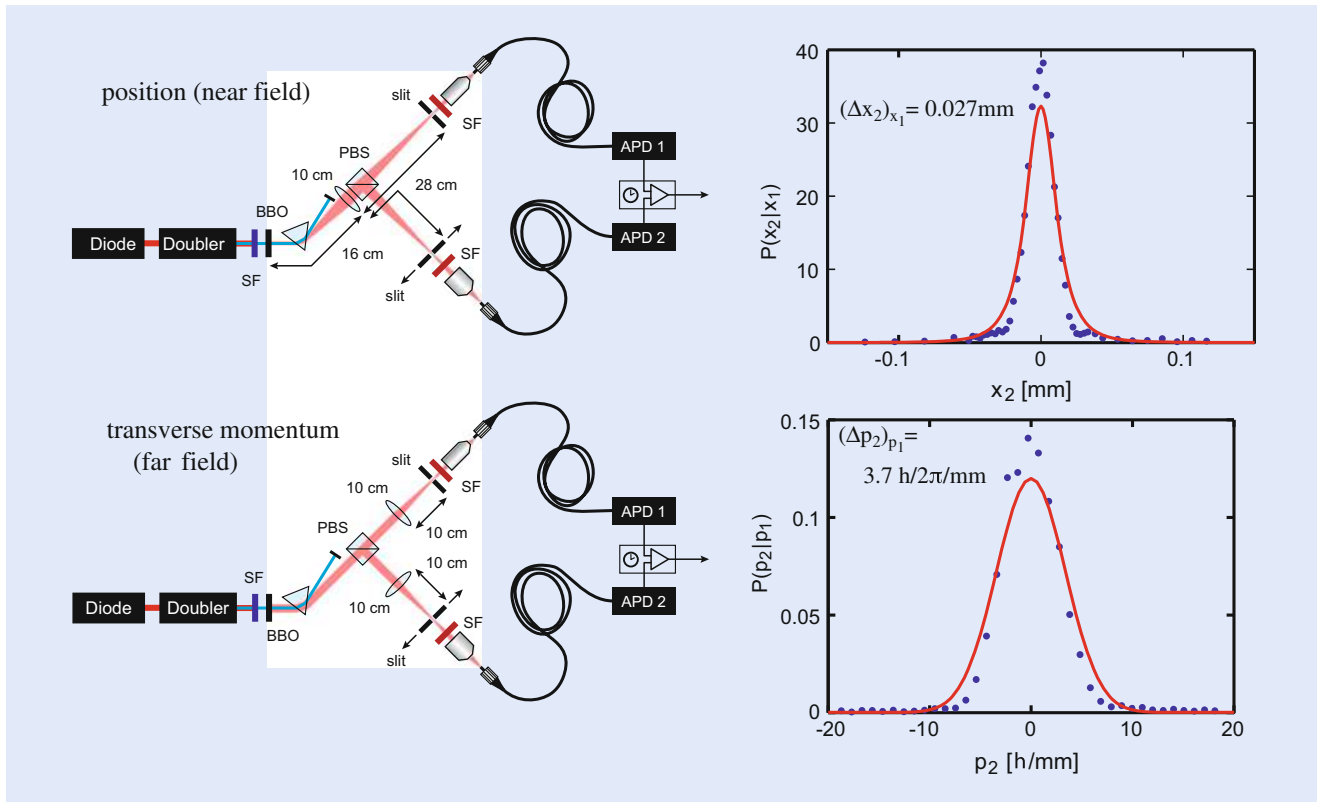
◼ **Fig. 17.6** Laboratory setups (*left*) and measured conditional count rates (*right*) for a laboratory demonstration of the EPR effect for position and transverse momentum variables. The notation $P(x_2 \mid x_1)$ means that probability of measuring one photon at position x_1 conditioned on the other photon being detected at position x_1, and analogously for $P(p_2 \mid p_1)$. The measured conditional uncertainty product is $0.1\hbar$, which violates the Heisenberg uncertainty relation for independent particles [26]

Either the birthplace (i.e., position) of each photon or its transverse momentum could be determined by placing a slit followed by an area detector either in a focal plane of the crystal or in its far field, respectively. Coincidence counts between the two detectors were measured, and the conditional count rates associated with position and momentum are shown in the graphs on the right-hand side of the figure. The measured conditional uncertainty product is found to be $(\Delta x_2)_{x_1}(\Delta p_2)_{p_1} = 0.1\hbar$, which is $\approx$ five times smaller than which might be expected for the uncertainty principle as applied to independent particles.

There has also been considerable interest in studies of time-energy entanglement. For example, Ali-Khan et al. [10] have developed a protocol for quantum key distribution (QKD) that can encode as much as ten bits of information onto a single photon. In a separate study, Jha et al. [10] have studied time-energy entanglement controlled by a geometrical (Berry) phase on the Poincare sphere instead of by using a dynamical phase. The ability to manipulate entanglement by means of a geometrical phase could have important consequences for quantum information technology, because polarization controllers can be much more stable than translation stages needed to actively control optical path lengths.

We next turn to a description of angle-OAM entanglement. We first note that angle and OAM form a Fourier transform pair [22, 31]:

$$A_\ell = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\psi(\phi)\exp(-i\ell\phi)d\phi \tag{17.4}$$

$$\psi(\phi) = \sum_{\ell=-\infty}^{\ell=\infty} A_\ell \exp(i\ell\phi) \tag{17.5}$$

where A_ℓ is the amplitude a OAM state ℓ and where $\psi(\phi)$ represents the azimuthal dependence of the complex beam amplitude. One might well expect this result based on classical reasoning. However, Jha et al. [32] showed theoretically that a similar Fourier relation holds between the photons of an entangled photon pair produced by a down-conversion source. They also experimentally demonstrated the characteristic OAM sideband structure that this Fourier relationship implies (◘ Fig. 17.7).

In a related experiment, Jha et al. [33] studied the correlations between the OAM values of two entangled photons after each had passed through separated double-slit apertures. Strong, non-classical, correlations were observed in the resulting interference pattern. These authors also showed that under their experimental conditions the visibility of this interference pattern was numerically equal to the concurrence of the two-photon state, a measure of the degree of entanglement between the two photons. A measured visibility of between 85 and 92 % quantifies the nonclassical entanglement of the photons produced by their two-photon source.

In still another related experiment, Leach et al. [28] performed an experiment that is the direct analog of the Einstein–Podolsky–Rosen (EPR) gedankenexperiment, but in the angle-OAM degrees of freedom. Part of the motivation for this study is that, unlike the continuous and unbounded variables in position-momentum entanglement considered by EPR, angular position is a periodic variable leading to a discrete OAM spectrum. As such, the analysis of EPR for angles involves subtleties similar to the issue of the existence of photon-number photon-phase uncertainty relation [34]. The details of this experiment are presented in ◘ Fig. 17.8.

There has also been great recent interest in harnessing the radial modes of Laguerre–Gaussian (LG) beams in addition to the azimuthal modes that we have
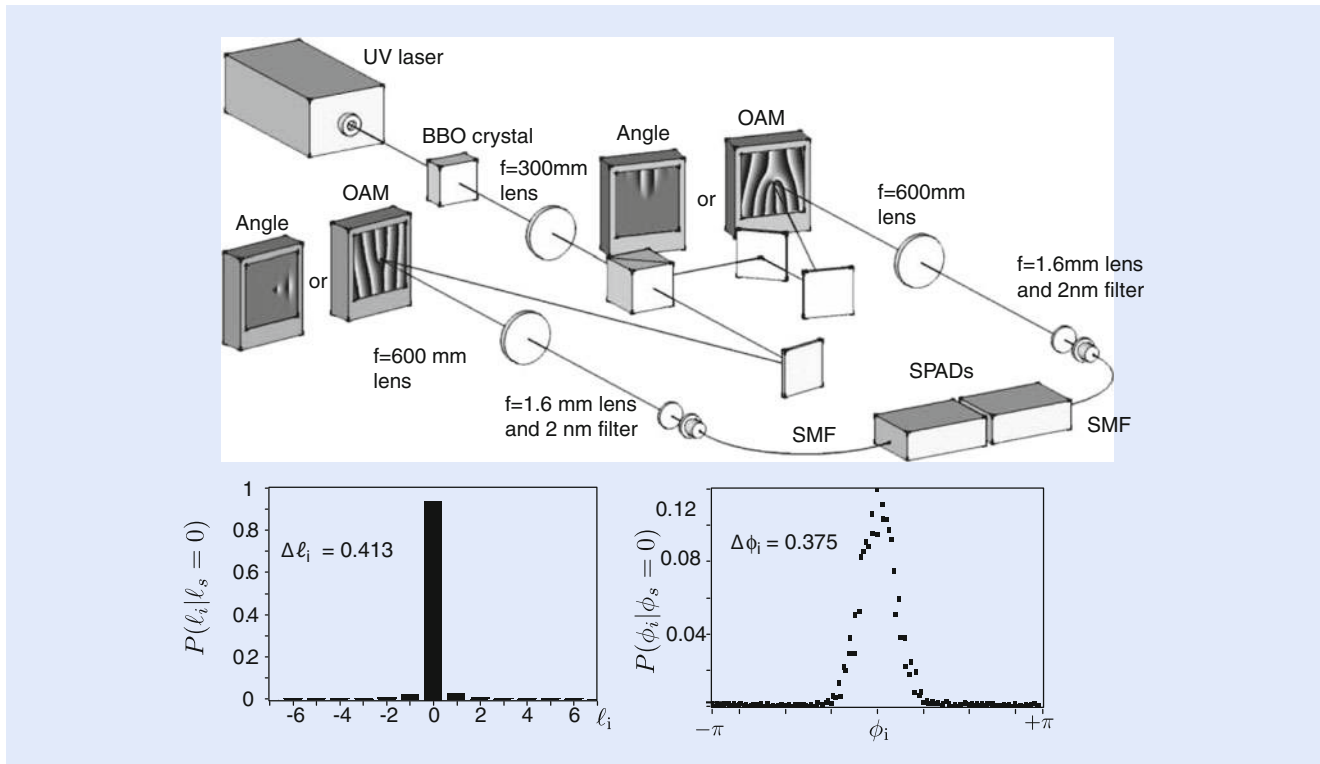
Fig. 17.7 Laboratory setup (*top*) and conditional probability of detection (*bottom*) for a laboratory demonstration of the EPR effect for angular position and orbital angular momentum variables. The conditional uncertainty product is 0.024, which violates the Heisenberg uncertainty relation for independent particles [28]
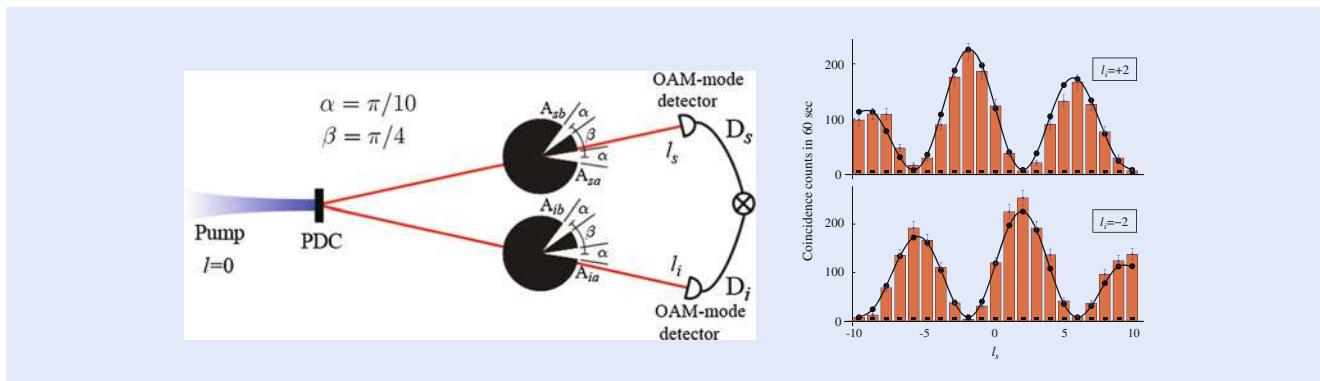
Fig. 17.8 (*left*) Laboratory setup to study angular two-photon interference and angular two-qubit states. In this experiment, an entangled two-photon state is created by parametric downconversion, and each photon falls onto a different aperture having the form of a double angular slit. The OAM content of each photon is then measured, and correlations between the two outputs are calculated. (*right*) Some of the results of this experiment. Here the coincident count rate is plotted as a function of the OAM of the signal photon for two different values of the OAM of the idler photon. The high visibility of the interference fringes is an indication of the high level of entanglement between the two photons [33]

primarily discussed up to now. One reason for this interest is to increase the information capacity of a light beam of a given restricted diameter. We note that the LG modes of Eq. (17.3) depend on two indices, the azimuthal index ℓ and the radial index p. But there are also further subtleties involved in exploiting the radial distribution, related to the fact that the radial coordinate ρ ranges from 0 to ∞, unlike the azimuthal coordinate ϕ, which ranges from 0 to 2π. Recently, Karimi et al. [35] presented a theoretical analysis of the operator nature of the radial

degree of freedom. Moreover, Karimi et al. [36] have studied the dependence of Hong–Ou–Mandel interference on the transverse structure of the interfering photons.

17.5 Secure Quantum Communication with More than One Bit Per Photon

We now turn to an application for the OAM of light in the field of cryptography and secure communication. To put this application topic in context, we first review the use of a one-time pad in cryptography (Shannon [37]). We consider the situation in which one party, A (or Alice), wants to communicate securely to another party, B (or Bob). We assume that by pre-arrangement the two parties share the same string of random binary digits known as the key, that no one else has access to this key, and that this string of digits is at least as long as the message that Alice wants to send. Alice encodes her message by first placing it into a binary format and then performing a binary add of the i-th digit in her message with the i-th digit in the one-time pad. This encoded message is then sent over a public channel. It is provably true that this message contains no useful information except to someone who has possession of the secret key. Bob then decodes the message by performing a binary add with his copy of the key, thus obtaining the original message that Alice sent.

A difficulty with implementing the one-time pad method is the distribution of the secret key between Alice and Bob. This is especially difficult if the two parties are not and cannot be in the same place, where the key can simply be handed from one to the other. When not in the same place, a procedure proposed by Bennet and Brassard in 1984 (know as the BB-84 protocol) can be used to distribute the key in an entirely secure manner. In brief (some of the details are provided below), Alice sends the key one element at a time, and each digit is encoded in the quantum state of a single photon. If an eavesdropper (Eve) intercepts and measures this photon and then tried to send an exact replacement photon of it, she will certainly fail, because the laws of quantum mechanics prohibit her from determining full knowledge of the quantum state of a photon in a single measurement. The impossibility of doing so results from the celebrated "no-cloning" theorem of quantum mechanics [38]. Secure communication through use of the BB-84 protocol of quantum key distribution (QKD) is now a commercial reality (see, for example the website ▶ http://www.idquantique.com).

Work in which the present authors have participated involves extending the BB-84 protocol so that more than one bit of information can be carried by each photon. Such a procedure may be used to increase the secure bit rate of a quantum communication system. To achieve the goal of transmitting more than one bit per photon, we encode information in the transverse degree of freedom of the light field. For the transverse degree of freedom one can choose any complete set of orthonormal modes. In keeping with the context of this chapter we consider encoding in OAM modes such as Laguerre–Gaussian (LG) modes. In the original QKD proposal of Bennett and Brassard, information is encoded in the polarization degree of freedom of an individual photon. As a result, only one bit of information could be impressed onto each photon. In contrast, when using OAM, there is no limit to how many bits of information can be impressed onto a single photon, as the LG modes span an infinite-dimensional state space. As mentioned above, one motivation for doing this is that rate of data transmission is thereby increased. Another more subtle motivation is that the security of the protocol can be increased by encoding information within a higher-dimensional state space.

The system that we envisage is illustrated in broad scope in ◻ Fig. 17.9. It consists of a sender, Alice, and a receiver, Bob. Alice impresses information onto
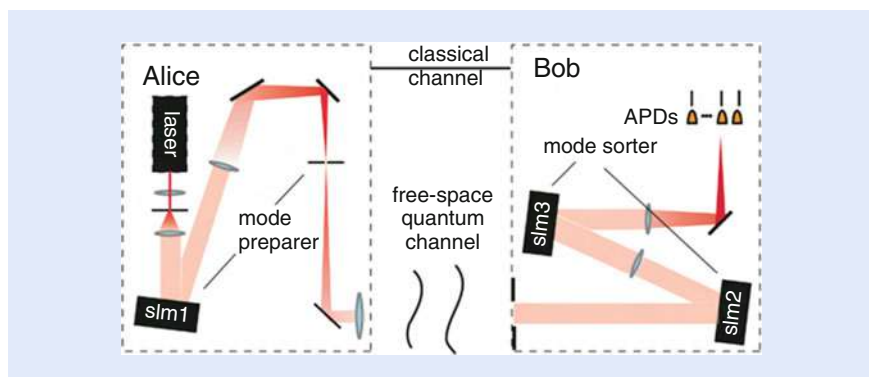
■ **Fig. 17.9** System schematic of the baseline QKD protocol of Mirhosseini et al. [41]. A sender (A or Alice) impresses information onto an individual photon through use of a spatial light modulator (SLM). This photon is then sent to the receiver (B or Bob) through a free-space link, where it may experience degradation by means of atmospheric turbulence. The receiver then determines the quantum state of this photon
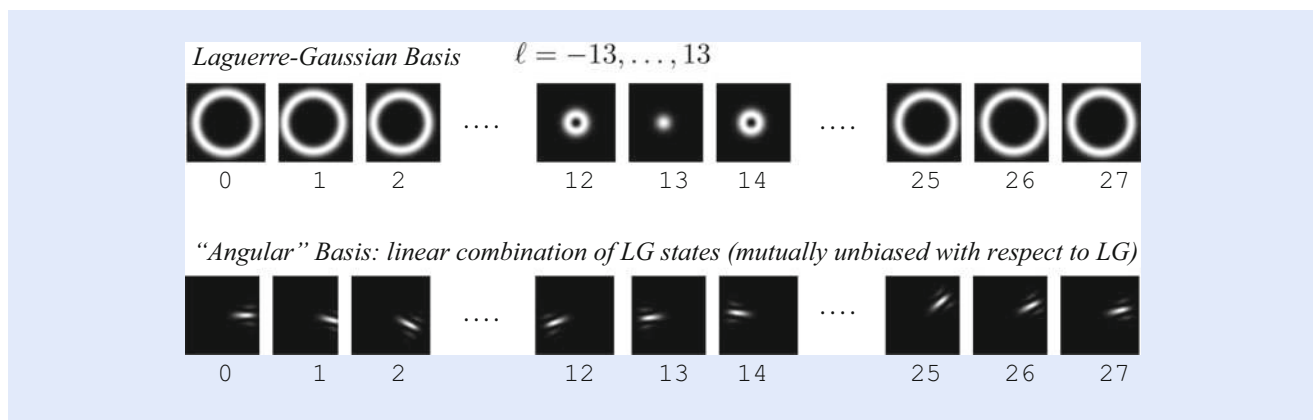
■ **Fig. 17.10** The LG basis (*top*) and a linear combination of the LG states (*bottom*) that constitutes the angular basis (AB). The information is encoded by launching individual photons that have been prepared in one of these modes [40]

the transverse degree of freedom of individual photons through the use of a spatial light modulator (SLM). Bob then randomly guesses which basis (OAM or angle) Alice might be using and makes a measurement of the quantum state of the received photon in this basis. The procedure for ensuring the security of the transmission is a generalization of that of the BB84 protocol and is described in the review of Gisin and Thew [39]. In the remainder of the present section we describe in more detail our laboratory procedure and present some laboratory data.

The BB84 QKD protocol entails Alice sending each photon in a randomly chosen basis. At least two mutually unbiased bases (MUBs) must be used. Certain advantages accrue from using more than two MUBs. It is known that the maximum number B_{max} of MUBs is related to the dimension D of the state space by $B_{max} = D + 1$. In our laboratory investigations we use the minimum number of MUBs, $B = 2$. We choose this value for convenience and to maximize our data transmission rate. Our two basis sets are illustrated in ■ Fig. 17.10. One basis is comprised of the LG states themselves. The other basis is composed of a linear combination of the LG states of the form
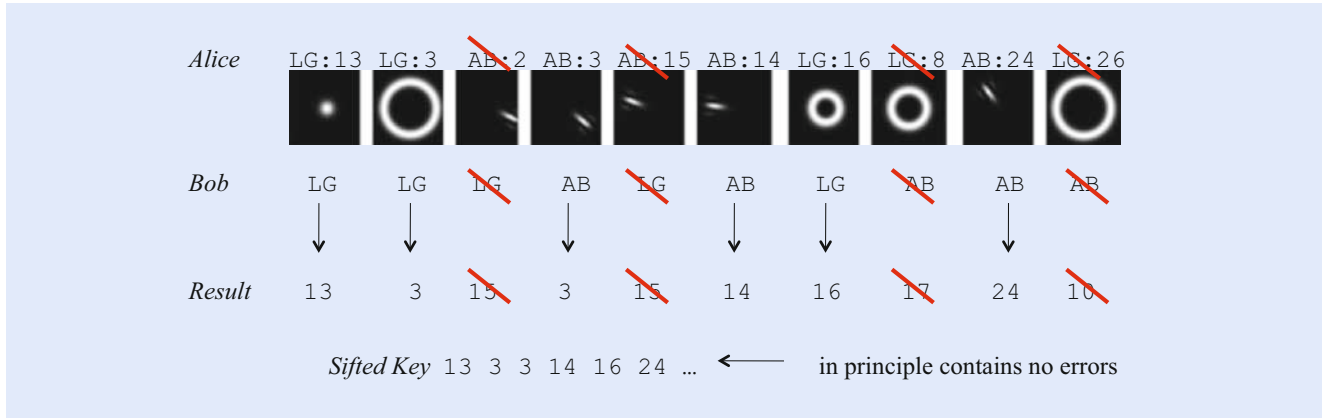
○ **Fig. 17.11** Example of a proposed implementation of a generalized BB84 protocol in a high-dimensional (27-dimensions as illustrated) state space [40]

$$\Psi_{AB}^{N} = \frac{1}{\sqrt{27}} \sum_{\ell=-13}^{13} LG_{\ell,0}\exp(2\pi i\ell/27). \tag{17.6}$$

From this expression we obtain the states shown in the lower row of the figure, which is referred to as the angular basis (AB).

An example of the implementation of this protocol is shown in ○ Fig. 17.11. In this example, Alice is attempting to send the string of numbers 13, 3, 2, 3, 15, 14, 16, 8, 24, 26 to Bob. For each transmitted photon, Alice chooses randomly between the LG basis and the AB basis. Also, for each transmitted photon Bob chooses randomly between the OAM and AB bases. After the transmission of the entire data train is complete, Alice and Bob publicly disclose the basis they used for each measurement. If they used different bases (which occurs on average half of the time), they discard the results of that measurement. The remaining data string is known as the sifted data, and this data should contain no errors. Any error in this data string could be the result of measurement errors or to the presence of an eavesdropper. For reasons of extreme caution, one must ascribe all errors to the presence of an eavesdropper. To test for errors, Alice and Bob sacrifice some fraction of their data for public comparison. If errors are detected, they conclude that an eavesdropper is present and take appropriate corrective measures.

We have implemented this BB84-type protocol in our laboratory. ○ Figure 17.11 shows how Alice forms each of the basis states. Basically, she programs a spatial light modulator (SLM) to convert an individual photon in a plane-wave state into one of the desired LG or AB modes [41]. The upper row shows the LG basis and the lower row shows the angular AB basis. The panel on the left shows representative examples of the pattern displayed on the SLM. The panels on the right show examples of the field distribution written onto the light field. These frames show actual laboratory results, although read out with intense classical light, not with single photons (○ Fig. 17.12).

Special considerations apply to the configuration of the receiver, or Bob. He is presented with a single photon and needs to determine its quantum state. Thus, he is allowed to perform only one measurement to determine in which of a large number of quantum states the received photon resides. This sorting task has eluded the scientific community until very recently, when Berkhout et al. [40] demonstrated a means for performing this task. Their approach is illustrated in ○ Fig. 17.13. The key element of this approach is the ability to map the azimuthal phase distribution of an incident mode onto a linear phase distribution at the
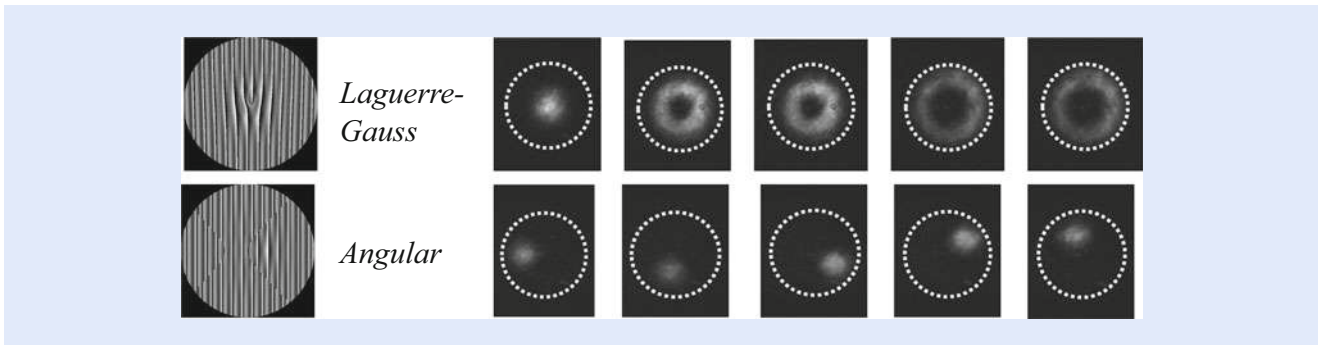
■ **Fig. 17.12** Illustration of the procedure for producing light fields in one of the Laguerre–Gaussian or angular basis states, shown for the case of a five-state bases (D = 5). The *dotted circles* in the panels on the right denote the aperture of the transmitting optics [Unpublished laboratory results of M.N. O'Sullivan]
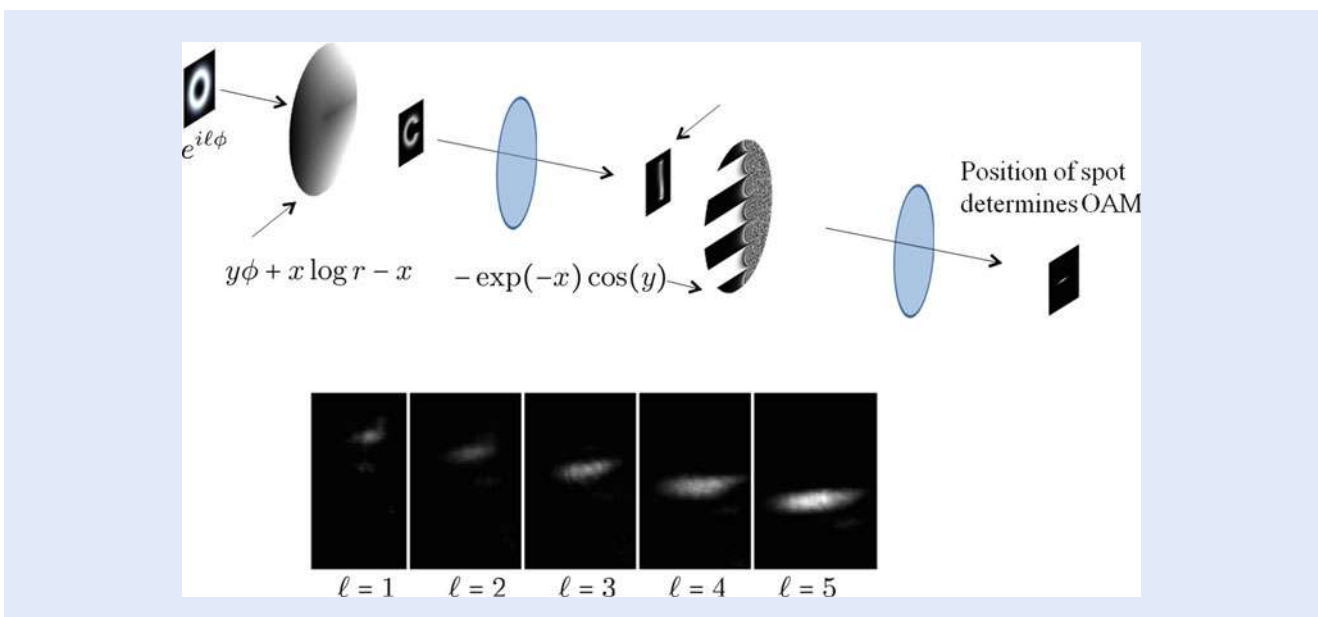
■ **Fig. 17.13** The angular-to-linear reformatter (Glasgow mode-sorter). (*top*) Physical layout of the reformatter. (*bottom*) Some results showing the performance of the reformatter used as a sorter. Note that the vertical position of the light beam at the output of the sorter depends on the OAM value ℓ of the beam (Unpublished data from the Boyd laboratory)

output of the device. Of course, a linear phase ramp in one cartesian dimension is simply a wavefront tilt, and leads to a shift in the position of the beam in the far field. It turns out that one can determine analytically the form of the phase function that needs to be applied to a light field to perform this mapping. In their original implementation of this sorting procedure, Berkhout et al. [42] applied this phase mapping through the use of an SLM. In a more recent work they have fabricated refractive elements that perform this same function but with much higher conversion efficiency than those based on diffraction from an SLM.

Some laboratory results validating the performance of this sorter are shown in ■ Fig. 17.14. These results demonstrate our ability to discriminate among various quantum states in either the LG or angular basis. In each basis we include only four basis states. This limitation is due to the number of photodetectors (APDs) available to us. We see no fundamental limit to our ability to distinguish among all of the states in our protocol, 27 in this particular situation. We see that there is a small amount of crosstalk among the various channels.
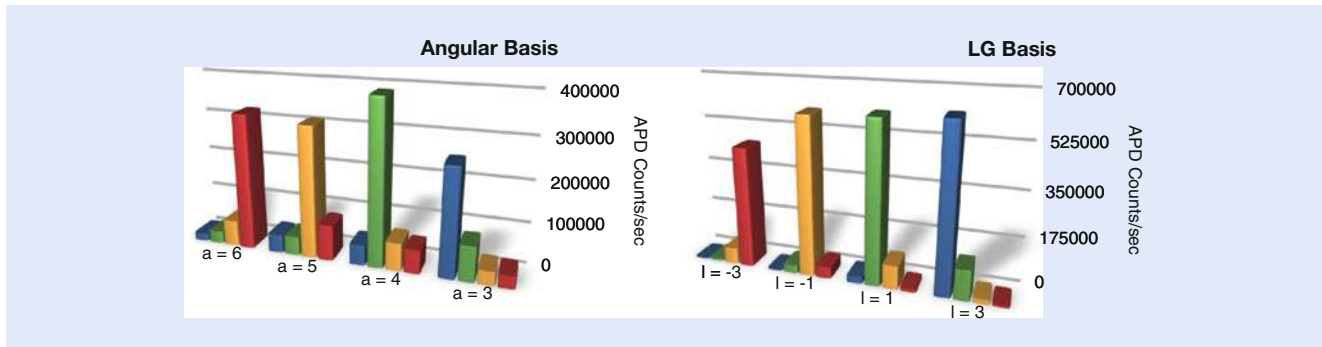
Fig. 17.14 Laboratory data demonstrating Bob's ability to discriminate among various quantum states in either the LG or angular basis through the use of the Glasgow mode sorter. Note that discrimination is good but not perfect; there is cross-talk among the channels (Unpublished data from the Boyd laboratory)
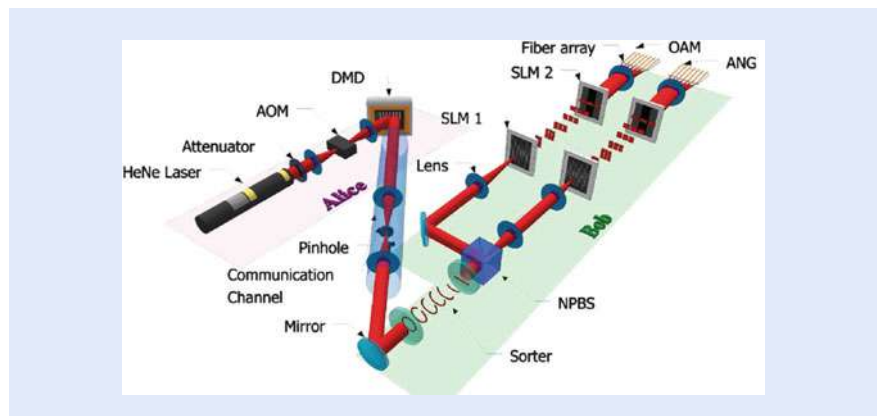
17

Fig. 17.15 Experimental setup of Mirhosseini et al. [43]. This system uses the OAM and angle bases to implement a QKD system. Alice encodes information in either the OAM or angle basis (chosen randomly), and Bob performs a measurement after making a further random choice of basis. Data obtained when they use different bases is later discarded, in a process known as sifting. If there is no eavesdropper, there should in concept be no errors in this data. Alice and Bob can test for the presence of an eavesdropper by the following procedure. They openly disclose a subset of this data, and check to see if any errors are present. The presence of errors suggests the presence of an eavesdropper

Using the procedures described here, we recently performed a realistic demonstration of quantum key distribution based on OAM encoding [43]. Our experimental setup is shown in ◘ Fig. 17.15 and is composed of the various components described above. Alice prepares state to be sent by first carving out pulses from a highly attenuated He-Ne laser through the use of an AOM. Then spatial mode information is impressed on these pulses with a digital micromirror device (DMD). Bob's mode sorter and fan-out elements map the OAM modes and the ANG modes onto separated spots that are collected by an array of fibers and sent to individual APD detectors.

Some of the results of this demonstration are shown in ◘ Fig. 17.16. The top row (left) shows the string of numbers sent by Alice and the top row right shows the string of numbers received by Bob. Note that the strings are not identical; due to various transmission and detection errors, some of the symbols are not detected as transmitted. In the figure errors are marked in red and are underlined. As a test of their system, Alice and Bob could publicly disclose these results to determine the fractional error rate. However, in an operating system, Alice and Bob would want
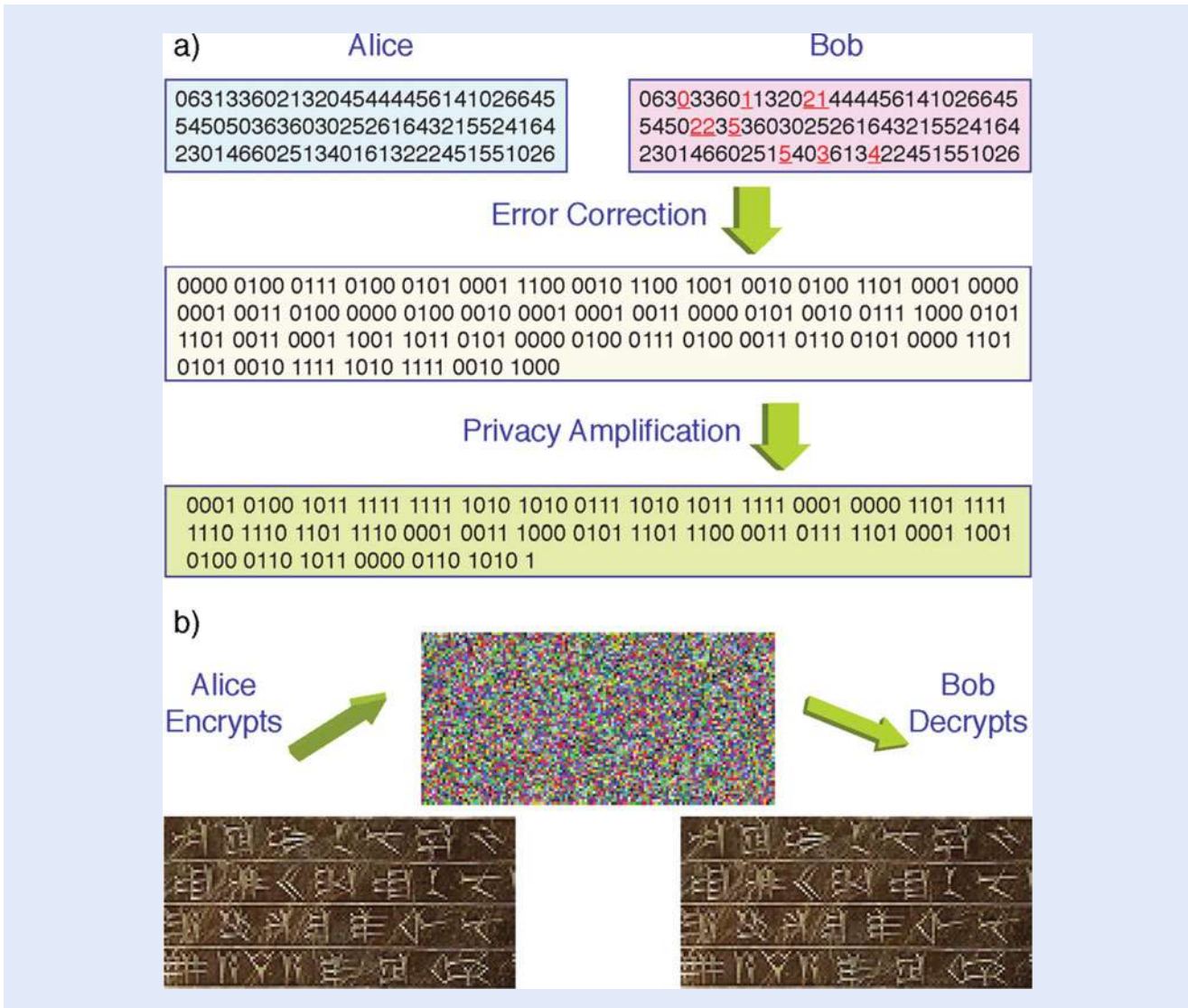
● **Fig. 17.16** Experimental results from the study of Mirhosseini et al. [43] (**a**) Example of a random sifted key from the experiment. The spatial modes are mapped to numbers between 0 and 6, and errors are marked in *red* and are *underlined*. Each symbol is converted into a three digit binary number first and the binary key is randomized before the error-correction. Privacy amplification minimizes Eve's information by shortening the key length. (**b**) Alice encrypts the secret message (in this case an image of an ancient Persian tablet) using the shared secure key and Bob subsequently decrypts it

to sacrifice only a small fraction of their data to test system security. Alice and Bob therefore employ an error correction algorithm and a method known as privacy amplification (which shortens the length of their shared string) to decrease the number of errors in the shared strings. They end up with a shared key that contains essentially no errors. (For this reason we show the string constituting the shared key only once.) As a graphic demonstration of the use of this procedure, we also show how it could be used for the secure transmission of an image. The image is separated into pixels which are then digitized and transmitted using the secret key shared by Alice and Bob. An eavesdropper who intercepted the signal would see only the noisy pattern that is also displayed.

17

17.6 Summary and Conclusions

In this chapter we have presented a review of the quantum mechanical properties of spatially structured light fields, paying special attention to light fields that carry orbital angular momentum (OAM). We have considered both the conceptual understanding of the quantum features of these light fields and the use of these quantum features for applications. We describe how to produce spatially entangled light fields by means of the nonlinear optical process of spontaneous parametric downconversion. We address the question of how much information can be encoded onto a single photon. As an example, we review a recent experiment that demonstrated the ability to discriminate among four target objects using only one photon for illumination. We also present a description of the concept of the OAM of light, and we describe means to generate and detect OAM. We then present a brief survey of some recent studies of the fundamental quantum properties of structured light beams. Much of this work is aimed at studying the nature of entanglement for the complementary variables of angular position and OAM. Finally, as a real-world application, we describe a secure communication system based on quantum key distribution (QKD). This key distribution system makes use of encoding information in the OAM modes of light and hence is able to transmit more than one bit of information per photon.

References

1. Kolobov MI (ed) (2006) Quantum imaging. Springer, New York
2. Schrödinger E (1935) Discussion of probability between separated systems. Proc Camb Phys Soc 31:555
3. Freedman SJ, Clauser JF (1972) Experimental test of local hidden-variable theories. Phys Rev Lett 28(938):938–941
4. Aspect A, Grangier P, Roger G (1981) Experimental tests of realistic local theories via Bell's theorem. Phys Rev Lett 47(7):460–463

5. Burnham DC, Weinberg DL (1970) Observation of simultaneity in parametric production of optical photon pairs. Phys Rev Lett 25:84

6. Kwiat PG, Mattle K, Weinfurter H, Zeilinger A, Sergienko AV, Shih Y (1995) New high-intensity source of polarization-entangled photon Pairs. Phys Rev Lett 75:4337

7. Ling A, Lamas-Linares A, Kurtsiefer C (2008) Absolute emission rates of spontaneous parametric down-conversion into single transverse Gaussian modes. Phys Rev A 77:043834

8. Einstein A, Podolsky B, Rosen N (1935) Can quantum-mechanical description of physical reality be considered complete? Phys Rev 47:777

9. Ali-Khan I, Broadbent CJ, Howell JC (2007) Large-alphabet quantum key distribution using energy-time entangled bipartite States. Phys Rev Lett 98:060503

10. Jha AK, Malik M, Boyd RW (2008) Exploring energy-time entanglement using geometric phases. Phys Rev Lett 101:180405

11. Gerry C, Knight PL (2005) Introductory quantum optics. Cambridge University Press, Cambridge

12. Curtis JE, Koss BA, Grier DG (2002) Dynamic holographic optical tweezers. Opt Commun 207:169–175

13. Mirhosseini M, Magaña-Loaiza OS, Chen C, Rodenburg B, Malik M, Boyd RW (2013) Rapid generation of light beams carrying orbital angular momentum. Opt Exp 21:30204

14. Broadbent CJ, Zerom P, Shin H, Howell JC, Boyd RW (2009) Discriminating orthogonal single-photon images. Phys Rev A 79:033802

15. Malik M, Shin H, O'Sullivan M, Zerom P, Boyd RW (2010) Quantum ghost image discrimination with correlated photon Pairs. Phys Rev Lett 104:163602

16. Strekalov DV, Sergienko AV, Klyshko DN, Shih YH (1995) Observation of two-photon "Ghost" interference and diffraction. Phys Rev Lett 74:3600

17. Miller DAB (2013) Self-configuring universal linear optical component. Photon Res 1(1):1. ► http://doi.org/10.1364/PRJ.1.000001

18. Allen L, Beijersbergen MW, Spreeuw RJC, Woerdman JP (1992) Orbital angular-momentum of light and the transformation of Laguerre-Gaussian laser modes. Phys Rev A, 45(11):8185–8189

19. He H, Friese M, Heckenberg N, Rubinsztein-Dunlop H (1995) Direct observation of transfer of angular momentum to absorptive particles from a laser beam with a phase singularity. Phys Rev Lett 75:826–829

20. O'Neil AT, MacVicar I, Allen L, Padgett MJ (2002) Intrinsic and extrinsic nature of the orbital angular momentum of a light beam. Phys Rev Lett 88(5):053601

21. Leach J, Dennis MR, Courtial J, Padgett MJ (2005) Vortex knots in light. New J Phys 7:55. ► http://doi.org/10.1088/1367-2630/7/1/055

22. Yao E, Franke-Arnold ES, Courtial J, Barnett SM, Padgett MJ (2006) Fourier relationship between angular position and optical orbital angular momentum. Opt Exp14:9071

23. Bolduc E, Bent N, Santamato E, Karimi E, Boyd RW (2013) Exact solution to simultaneous intensity and phase masking with a single phase-only hologram. Opt Lett 38:3546

24. Marrucci L, Manzo C, Paparo D (2006) Optical spin-to-orbital angular momentum conversion in inhomogeneous anisotropic media. Phys Rev Lett 96:1605

25. Karimi E, Piccirillo B, Nagali E, Marrucci L, Santamato E (2009) Efficient generation and sorting of orbital angular momentum eigenmodes of light by thermally tuned q-plates. Appl Phys Lett 94:1124

26. Howell JC, Bennink RS, Bentley SJ, Boyd RW (2004) Realization of the Einstein-Podolsky-Rosen paradox using momentum and position-entangled photons from spontaneous parametric down conversion. Phys Rev Lett 92:210403

27. Mair A, Vaziri A, Weihs G, Zeilinger A (2001) Entanglement of the orbital angular momentum states of photons. Nature 412(6844):313–316

28. Leach J, Jack B, Romero J, Jha AK, Yao AM, Franke-Arnold S, Ireland DG, Boyd RW, Barnett SM, Padgett MJ (2010) Quantum correlations in optical angle-orbital angular momentum variables. Science 329:662

29. Bohm D (1951). Quantum theory. Prentice-Hall, Englewood Cliffs, p 29, and Chapter 5 Section 3, and Chapter 22 Section 19

30. Rarity JG, Tapster PR (1990) Experimental violation of Bell?s inequality based on phase and momentum. Phys Rev Lett 64(21):2495–2498

31. Pors JB, Aiello A, Oemrawsingh SSR, van Exter MP, Eliel ER, Woerdman JP (2008) Angular phase-plate analyzers for measuring the dimensionality of multimode fields. Phys Rev A 77:033845

32. Jha AK, Jack B, Yao E, Leach J, Boyd RW, Buller GS, Barnett SM, Franke-Arnold S, Padgett MJ (2008) Fourier relationship between the angle and angular momentum of entangled photons. Phys Rev A 78:043810

33. Jha AK, Leach J, Jack B, Franke-Arnold S, Barnett SM, Boyd RW, Padgett MJ (2010) Angular two-photon interference and angular two-qubit states. Phys Rev Lett 104:010501

34. Pegg DT, Barnett SM (1997) Tutorial review - quantum optical phase. J Mod Opt 44:225–264

35. Karimi E, Boyd RW, de la Hoz P, de Guise H, Rehacek J, Hradil Z, Aiello A, Leuchs G, Sánchez-Soto LL (2014) Radial quantum number of Laguerre-Gaussian modes. Phys Rev A89:063813

36. Karimi E, Giovannini D, Bolduc E, Bent N, Miatto FM, Padgett MJ, Boyd RW (2014) Exploring the quantum nature of the radial degree of freedom of a photon via Hong-Ou Mandel interference. Phys Rev A 89:013829

37. Shannon C (1949) Communication theory of secrecy systems. Bell Syst Tech J 28 (4):656–715. As is often the case, the idea of the one-time pad appears to have been invented independently several times in the past. Shannon presents a good review of communication security based on the one-time pad

38. Wootters WK, Zurek WH (1982) A single quantum cannot be cloned. Nature 299:802–803

39. Gisin N, Thew R (2007) Quantum communications. Nat Photon 1:165

40. Boyd RW, Jha A, Malik M, O'Sullivan C, Rodenburg B, Gauthier DJ (2011) Quantum key distribution in a high-dimensional state space: exploiting the transverse degree of freedom of the photon. In: Proceedings of the SPIE 7948, 79480L

41. Gruneisen MT, Miller WA, Dymale RC, Sweiti AM (2008) Holographic generation of complex fields with spatial light modulators: application to quantum key distribution. Appl Opt 47:A32

42. Berkhout GCG, Lavery MPJ, Courtial J, Beijersbergen MW, Padgett MJ (2010) Efficient sorting of orbital angular momentum states of light. Phys Rev Lett 105:153601

43. Mirhosseini M, Magaña-Loaiza OS, O'Sullivan MN, Rodenburg B, Malik M, Lavery MPJ, Padgett MJ, Gauthier DJ, Boyd RW (2015) High-dimensional quantum cryptography with twisted light. New J Phys 17:033033

17

Quantum Communication with Photons

Mario Krenn, Mehul Malik, Thomas Scheidl, Rupert Ursin, and Anton Zeilinger

M. Krenn (✉) • M. Malik • T. Scheidl • A. Zeilinger (✉)
Vienna Center for Quantum Science and Technology (VCQ), Faculty of Physics, University of Vienna,
Boltzmanngasse 5, A-1090 Vienna, Austria

Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences (ÖAW),
Boltzmanngasse 3, A-1090 Vienna, Austria
e-mail: mario.krenn@univie.ac.at; anton.zeilinger@univie.ac.at

R. Ursin
Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences (ÖAW),
Boltzmanngasse 3, A-1090 Vienna, Austria

18.1 Introduction

Ever since its inception, quantum physics has changed our understanding of the fundamental principles of nature. Apart from their impact on all fields of academic research, these insights have merged together with the field of information science to create the novel field of quantum information. Quantum information science provides qualitatively new concepts for communication, computation, and information processing, which are much more powerful than their classical counterparts. Quantum information is an intriguing example where purely fundamental and even philosophical research can lead to new technologies. The developments in this young field recently experienced a worldwide boom—as is evidenced by the increasing number of quantum information centers being founded in countries all over the world. Although its long-term industrial applications cannot be clearly anticipated, it is clear that quantum information science entails a huge potential economic impact. For reasons of space we limit ourselves to polarization and orbital angular momentum (OAM) as information carrying degrees of freedom.

18.1.1 The Quantum Bit

In classical information and computation science, information is encoded in the most fundamental entity, the bit. Its two possible values **0** and **1** are physically realized in many ways, be it simply by mechanical means (as a switch), in solids by magnetic or ferroelectric domains (hard drives), or by light pulses (optical digital media). All of these methods have one thing in common—one state of the device mutually excludes the simultaneous presence of the other—the switch is either **on** or **off** (◘ Fig. 18.1).

The superposition principle entails one of the most fundamental aspects of quantum physics, namely to allow the description of a physical system as being in a probabilistic combination of its alternative states. This so-called *superposition of states* not only provides all predictions for the outcome of a physical measurement, but it also has drastic consequences for the nature of the physical state that we ascribe to a system. Its most important direct implication is the so-called

classical bit   qubit
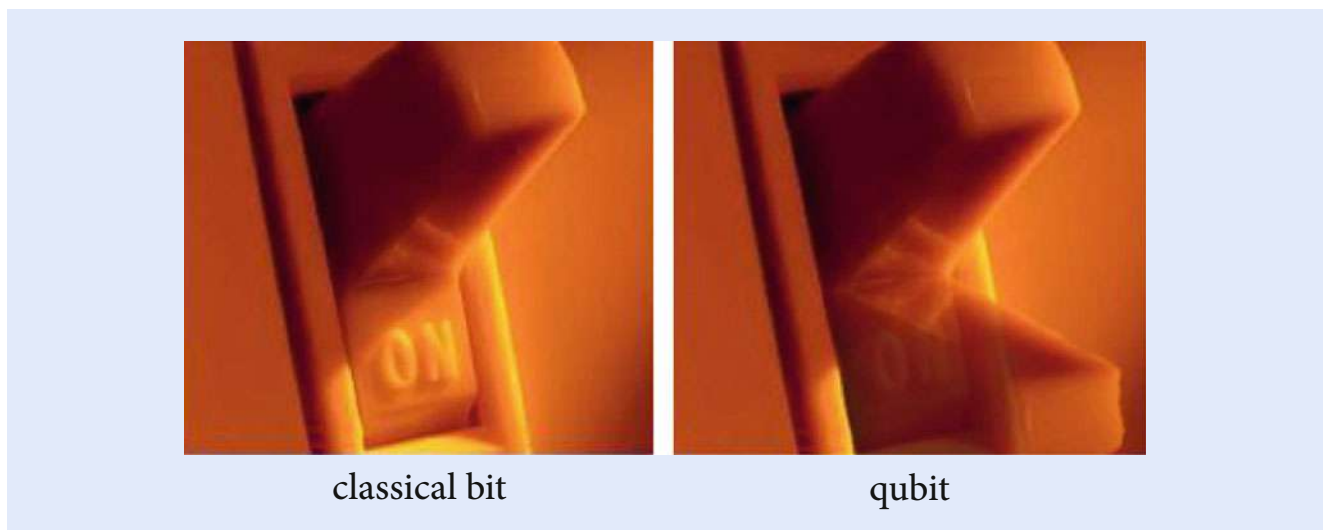
◘ **Fig. 18.1** An illustration of the difference between a classical bit and a qubit. The classical bit is always in a well-defined state while the qubit can also exist in a superposition of orthogonal states (copyright University of Vienna)

no-cloning theorem, which states that it is impossible to obtain a perfect copy of a qubit in an unknown state without destroying the information content of the original. The no-cloning theorem is the basis for the security of all quantum communication schemes described in the following sections, and will be explained later in more detail.

A qubit can be realized in many different physical systems such as atoms, ions, and super-conducting circuits. The most prominent physical realization of a qubit in view of a potential global-scale quantum communication network is with photons. Using photons, the two values of a bit, **0** and **1**, can be encoded in many different ways. One possibility is to use two orthogonal polarization states of a single photon, referred to as a polarization qubit. In the latter case, one can ascribe the horizontal polarization state of the photon with the logical value **0** and the vertical polarization state with the value of **1**. Any arbitrary polarization state can be obtained via a superposition of the horizontal and vertical state. The advantage of using photonic polarization qubits is that they can be easily generated, controlled, and manipulated with rather simple linear optical devices like wave plates. Furthermore, since photons rarely exhibit interaction with the environment they are the best candidates for long-distance free-space transmission as would be required in a future network involving ground-to-space links.

To fully understand a qubit, it is important to distinguish between a coherent superposition and a mixture of possible states. For its use in quantum communication, it is important that a photon exists in a coherent superposition of its possible states. For example, a polarization qubit being in a coherent superposition of horizontal and vertical polarizations (with a certain phase relation) can be understood as a photon polarized diagonally at $+45°$. A polarizer set at this angle will always transmit such a photon with 100 % probability (and zero probability when set to $-45°$). However, a photon in a mixture (incoherent superposition) of horizontal and vertical polarization states will be transmitted with 50 % probability.

Quantum superpositions, however, are not limited to just two possible states. The information carried by a photon is potentially enormous. While polarization is necessarily a two-level (qubit) property, other degrees of freedom of a photon such as its spatial or temporal structure can have many orthogonal levels. For example, a photon can exist in a coherent superposition of different paths coming out of a multi-port beam splitter. These types of superpositions are referred to as "high-dimensional" by virtue of their ability to encode large amounts of information. Consider a photon that is carrying a complicated image, such as that shown in ◨ Fig. 18.2. This image can be decomposed in terms of any orthonormal basis of spatial modes. The number of modes required for a complete description of this image dictates the number of levels, or dimensionality of this photon. One such basis is the set of Laguerre–Gaussian modes, which are described by a photon carrying a twisted wavefront. The phase structure of such a photon winds from
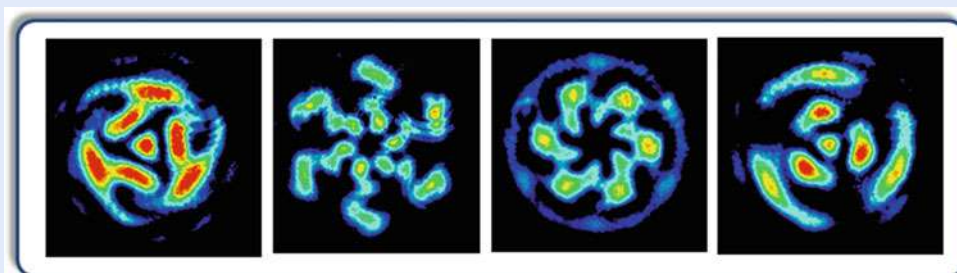
◨ **Fig. 18.2** Some types of higher-order spatial modes, which can carry more information than one bit per photon (Image by Mario Krenn, copyright University of Vienna)

0 to 2π azimuthally around the optical axis, with the number of twists dictating the photon state dimensionality. Using such high-dimensional degrees of freedom of a photon for encoding surely increases the amount of information one can send per photon. However, a more subtle advantage of doing this is found in quantum communication—not only can one vastly increase the information capacity of quantum communication systems, one can also increase their security. This point is discussed in detail later in this chapter.

18.1.2 Entanglement

The principle of superposition also holds for states containing several qubits. This allows for multi-qubit systems, which can only be described by joint properties. Such states are called *entangled*, describing the fact that none of the particles involved can be described by an individual quantum state [6, 19, 70]. This is equivalent to the astonishing property of entangled quantum systems, that all of their information content is completely entailed in the correlations between the individual subsystems and none of the subsystems carry any information on their own. For example, when performing measurements on only one of two entangled qubits, the outcome will be perfectly random, i.e., it is impossible to obtain information about the entangled system. However, since the entangled state consists of two qubits, the correlations shared between them must consist of two bits of classical information. As a consequence, these two bits of information can only be obtained when the outcomes of the individual measurements on the separate subsystems are compared (see ◘ Fig. 18.3).

Another intriguing feature of entangled states is that a measurement on one of the entangled qubits instantaneously projects the other one onto the corresponding perfectly correlated state, thereby destroying the entanglement. Since these perfect correlations between entangled qubits are in theory independent of the distance between them, the entanglement is in conflict with the fundamental concepts of classical physics—locality (i.e., distant events cannot interact faster than the speed of light) and realism (i.e., each physical quantity

◘ **Fig. 18.3** If one could entangle a pair of dices with respect to their numbers, one can encode the message **7** by using their entanglement. None of the dices would carry this information on its own and a local measurement of the dice will result in a completely random result (without revealing the information). However, the results are perfectly correlated to add up to **7** for every joint measurement on the two dices. Note that a rolling dice corresponds to a six-dimensional quDit (where D stands for Dimension), which was prepared in a way unknown to us, and which is about to be measured in one out of six orthogonal bases (copyright University of Vienna)

that can be predicted with certainty corresponds to an ontological entity, a so-called element of reality) [6]. This has led to various philosophical debates about whether quantum mechanics can serve as a complete description of reality. However, there have been many experiments performed addressing this issue, and to date each of them has confirmed the predictions of quantum mechanics [4, 25, 66, 68, 84]. One should note that while here we focus on polarization and orbital-angular momentum entanglement, light can be entangled in its other degrees of freedom as well, such as time-frequency [24, 33] and position-momentum [11, 37, 45].

18.1.3 Mutually Unbiased Bases

One fascinating concept in quantum mechanics is the possibility to encode quantum information in different ways. In the simple example of the polarization of light, there are three bases in which one can encode one bit of information (see ◘ Fig. 18.4). These are the horizontal and vertical (H/V) basis, the diagonal and anti-diagonal (D/A) basis, and the left- and right-circular (L/R) basis. One can encode a bit in the H/V basis by considering 0 to be horizontal polarization and 1 to be vertical polarization. If a photon encoded in either H or V polarization is measured in any of the other two bases, its information cannot be extracted. For example, in the case of measurements made in the D/A basis, in 50 % of the cases, a diagonally polarized photon will be observed; in the other cases, the photon will be measured as anti-diagonally polarized. This property is the main ingredient for quantum cryptography, as we will see later. Furthermore, in higher-dimensional systems, fundamental properties of mutually unbiased bases are still open questions that are significant for quantum communication.

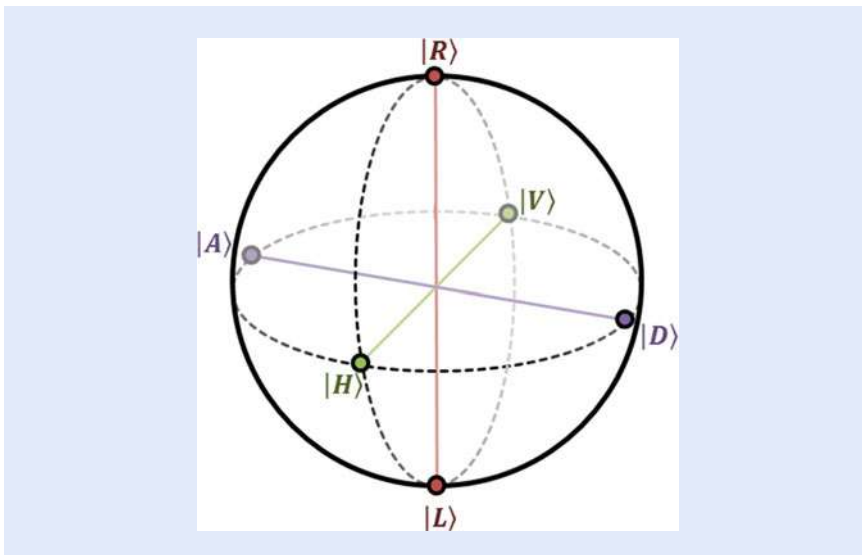◘ **Fig. 18.4** The Bloch-sphere: Graphical representation of a two-dimensional qubit. There are three mutually unbiased bases—three ways of encoding information in different ways. In the case of polarization, they correspond to horizontal and vertical (*violet*), diagonal and anti-diagonal (*green*) and right- and left-circular (*red*) polarization (Image by Mario Krenn, copyright University of Vienna)

18.1.4 Faster-than-Light Communication and the No-Cloning Theorem

As discussed above, two entangled photons are connected even though they can be spatially separated by hundreds of kilometers. The measurement of the first photon immediately defines the state of the second photon. Can one use that to transmit information faster than the speed of light? If Alice and Bob share an entangled state and measure their respective photon in the same mutually unbiased basis (for instance, in the horizontal/vertical basis), they will always find the same result. However, whether they detect a horizontal or vertical photon is intrinsically random—there is no way that Alice could influence the outcome of Bob. Regardless, there could exist a workaround, as shown in ◘ Fig. 18.5. Alice could use her choice of measurement basis to convey information: either horizontal/vertical (H/V) if she wants to transmit 0 or diagonal/antidiagonal (D/A) if she wants to send 1. When she does this, Bob's photon is immediately defined in that specific basis. If Bob could now clone his photon, he could make several measurements in both bases and find out in which of the two bases his photon is well defined: If Alice measured in the H/V basis and finds a H outcome, all of Bob's measurements in the H/V basis will be H. However, his measurements in the D/A basis will show 50 % diagonal and 50 % antidiagonal. Thus, he knows that Alice has chosen the H/V basis, and thereby transmitted the bit value 0.

Unfortunately, there is one problem with that protocol: It cannot exist. In 1982, Wootters and Zurek found that quantum mechanics forbids one to perfectly clone a quantum state [86]. This profound result originates from a simple property of quantum mechanics, namely the linear superposition principle. We can inspect what a potential cloning-operation $\hat{C}$ would do. We use an input quantum state, and an undefined second photon $|X\rangle$. After the cloning operation, the second photon should have the polarization property of the first photon. This is how our cloning machine would act on states in the H/V-basis:

$$\hat{C}\left(|H\rangle|X\rangle\right) = |H\rangle|H\rangle \tag{18.1}$$

$$\hat{C}\left(|V\rangle|X\rangle\right) = |V\rangle|V\rangle \tag{18.2}$$

The cloning-machine should work in every basis, thus we inspect what happens when we try to clone a diagonally polarized photon $|D\rangle$. Note that a diagonally
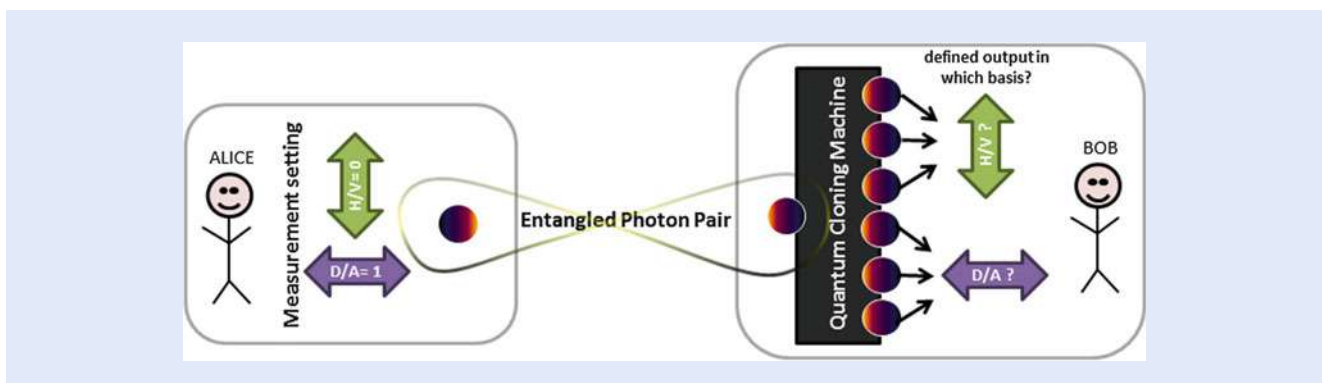

◘ **Fig. 18.5** Visualization of a faster-than-light quantum communication protocol, if (!) quantum states could be cloned: Alice and Bob share an entangled photon pair. By choosing the measurement basis between horizontal/vertical or diagonal/antidiagonal polarization, Alice projects the whole state into an eigenstate of that basis. This means that Bob's state is also defined in that basis. To find the basis chosen by Alice, Bob would need to measure more than one photon. If he could perfectly clone his photon, he could find the basis, and receive the information faster than light. Unfortunately, this is prohibited by the no-cloning theorem, a fundamental rule in quantum mechanics (Image by Mario Krenn, copyright University of Vienna)

polarized photon can be expressed in the H/V basis as a coherent superposition of a horizontal and a vertical part $|D\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$. The quantum cloning machine acts as

$$
\begin{aligned}
\hat{C}\left(|D\rangle|X\rangle\right) &= \hat{C}\left(\frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)|X\rangle\right) \\
&= \frac{1}{\sqrt{2}}(\hat{C}|H\rangle|X\rangle + \hat{C}|V\rangle|X\rangle) \\
&= \frac{1}{\sqrt{2}}(|H\rangle|H\rangle + |V\rangle|V\rangle)
\end{aligned}
\tag{18.3}
$$

The last line in Eq. (18.3) was obtained by using Eqs. (18.1) and (18.2) for the cloning operator $\hat{C}$. The result is an entangled state that cannot be factorized into $|D\rangle\,|D\rangle$. If one were to measure either of the entangled photons individually, the result would be random, and certainly not $|D\rangle$. From this simple example it is clear that quantum cloning is not possible. This property prohibits faster-than-light communication, but it opens the door to many different quantum secret sharing protocols, such as quantum cryptography.

18.1.5 Quantum Communication Schemes

The counterintuitive quantum principles of superposition and entanglement are not only the basis of acquiring a deeper understanding of nature, but also enable new technologies that allow one to perform tasks which are not possible by classical means. When speaking about such "quantum technologies," we refer to technologies that make explicit use of these kinds of quantum properties that do not have a classical analog. Quantum information science and quantum communication are important ingredients in future quantum information processing technologies. They enable the transfer of a quantum state from one location to another. All quantum communication schemes have in common that two or more parties are connected via both a classical communication channel and a quantum channel (i.e., a channel over which quantum systems are transmitted). Typically, measurements are performed on the individual quantum (sub-) systems and the measurement bases used for every measurement are communicated via the classical channel. Here, we focus on quantum communication with discrete variables. However, we should mention that there exists a parallel branch of quantum communication that is based on continuous variables, where extensive theoretical and experimental work has been performed. More information on this field can be found in [83] and references therein.

18.1.6 Quantum Key Distribution

If two parties want to share a secret message, they have two options: the first possibility is to share a random key that is the size of the message that needs to be encrypted with it (shown in ◘ Fig. 18.6). The sender, let's call her Alice, performs a simple logical operator (an *exclusive or*, XOR) of the message with the key, and gets the cipher. The cipher can only be read if the key is known. The receiver of the encrypted text, whom we will call Bob, can use the key to undo Alice's operation, which gives him the original message. The challenge lies in Alice and Bob having to share the entire secret key.
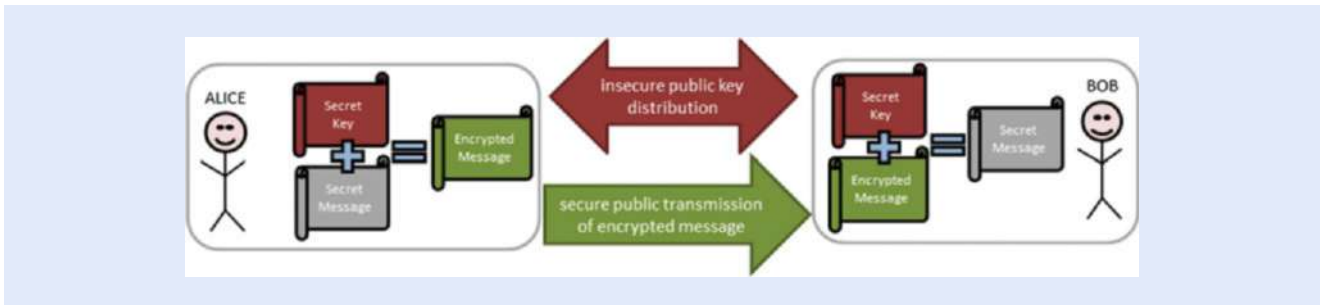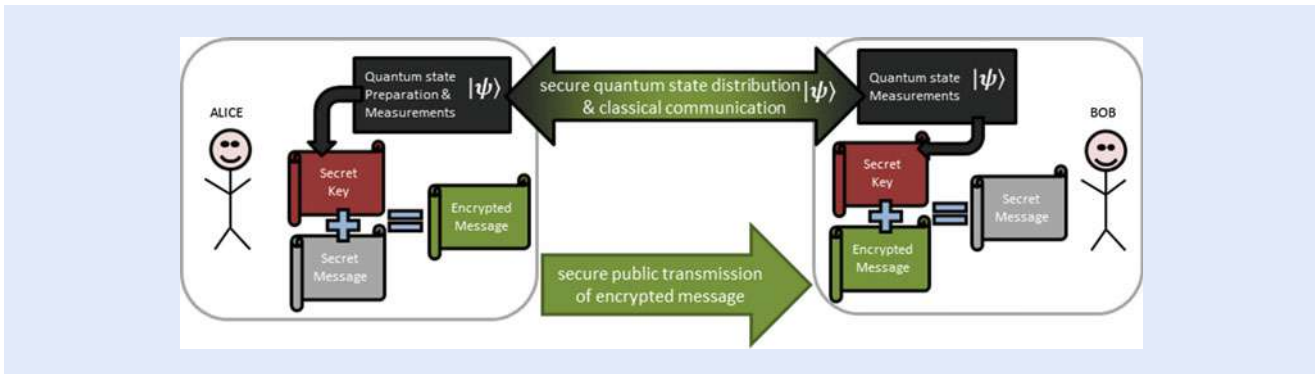
Fig. 18.6 Scheme of a classical symmetric cryptographical system. Alice wants to send a secret message to Bob. In order to do so, Alice and Bob have to share a secret key. With this key, they can distribute messages securely. The bottleneck is the distribution of the key. This problem is solved by quantum cryptography (Image by Mario Krenn, copyright University of Vienna)
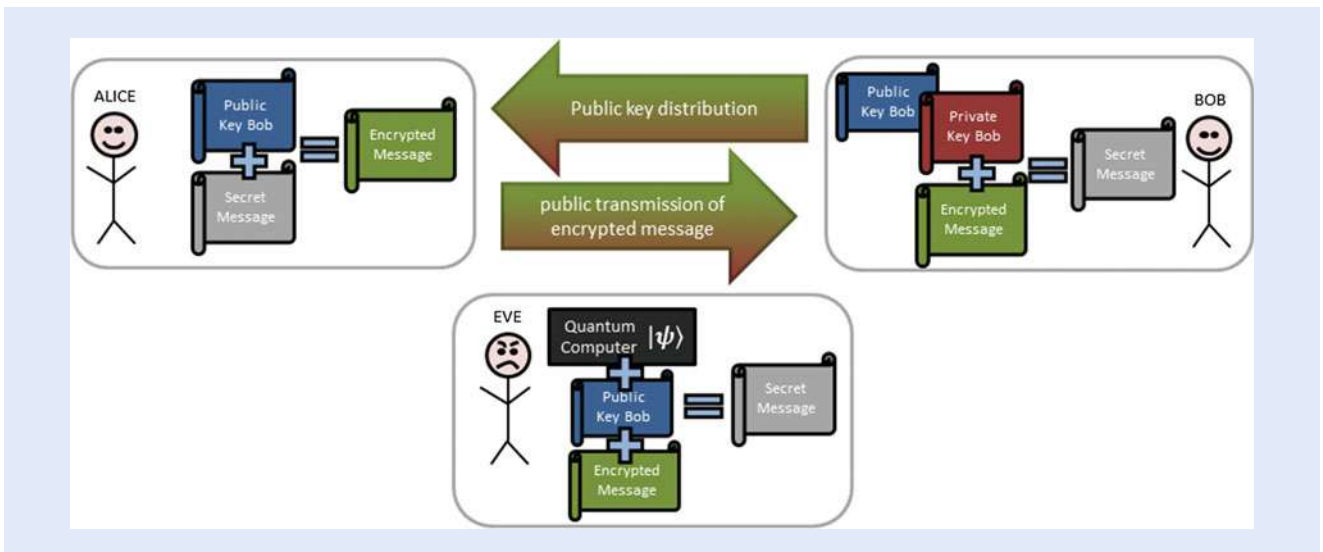
Fig. 18.7 Scheme of a classical asymmetric cryptographical system. Alice wants to send a secret message to Bob. In order to do so, Bob prepares a public and private key. Alice can then prepare an encrypted message for Bob with his public key. Usually, the message can only be decrypted by Bob with his private key. However, a powerful enough eavesdropper (for example, one with a quantum computer!) can infer Bob's private key from the public key, and can thus break the encryption protocol (Image by Mario Krenn, copyright University of Vienna)

The alternative is a public–private key cryptography. This method, invented in the 1970s, is based on the computational complexity of finding the prime factors of large numbers. Again, Alice wants to send a secret message to Bob. Now Bob creates a pair of keys, a private and a public one. Everybody who has Bob's public key can encrypt messages for him. However, only Bob can decrypt those messages with his private key. However, it has been discovered by Peter Shor in 1994 that a quantum computer could factor prime numbers significantly faster than classical computers [71]. It would allow an eavesdropper to read the secret message with only the information that is distributed publicly (see Fig. 18.7). One possible way to circumvent this problem is quantum key distribution.

Quantum key distribution (QKD) allows two authorized parties to establish a secret key at a distance. The generation of this secret key is based on the same quantum physical principles that a quantum computer relies on. In contrast to classical cryptography, QKD does not simply rely on the difficulty of solving a mathematical problem (such as finding the prime powers of a large number). Therefore, even a quantum computer could not break the key. QKD consists of two phases (see Fig. 18.8). In the first phase the two communicating parties, usually called Alice and Bob, exchange quantum signals over the quantum channel

🔲 **Fig. 18.8** Scheme of a quantum cryptographical system. Alice wants to send a secret message to Bob. In order to do so, a secret key is established over public (quantum) channels. Alice prepares a quantum state and transmits it to Bob. By making appropriate measurements, Alice and Bob can obtain a shared secret key. Alice then encrypts the message with this key and sends it to Bob; Bob can decrypt it with his copy of the key. Eavesdropping attempts during the key transmission appear as errors in the measurement results, allowing the presence of an eavesdropper to be detected (Image by Mario Krenn, copyright University of Vienna)

and perform measurements, obtaining a raw key (i.e., two strongly correlated but nonidentical and only partly secret strings). In the second phase, Alice and Bob use the classical channel to perform an interactive post-processing protocol, which allows them to distill two identical and completely secret (known only to themselves) strings, which are two identical copies of the generated secret key. The classical channel in this protocol needs to be authenticated: this means that Alice and Bob identify themselves; a third person can listen to the conversation but cannot participate in it. The quantum channel, however, is open to any possible manipulation from a third person. Specifically, the task of Alice and Bob is to guarantee security against an adversarial eavesdropper, usually called Eve, tapping on the quantum channel and listening to the exchanges on the classical channel.

In this context security explicitly means that a nonsecret key is never used: either the authorized parties can indeed create a secret key, or they abort the protocol. Therefore, after the transmission of the quantum signals, Alice and Bob must estimate how much information about raw keys has leaked out to Eve. Such an estimate is obviously impossible in classical communication: if someone is tapping on a telephone line, or when Eve listens to the exchanges on the classical channel, the communication goes on unmodified. This is where quantum physics plays a crucial role: in a quantum channel, leakage of information is quantitatively related to a degradation of the communication. The origin of security of QKD can be traced back to the fundamental quantum physical principles of superposition and no-cloning. If Eve wants to extract some information from the quantum states, this is a generalized form of measurement, which will usually modify the state of the system. Alternatively, if Eve's goal is to have a perfect copy of the state that Alice sends to Bob, she will fail due to the no-cloning theorem, which states that one cannot duplicate an unknown quantum state while keeping the original intact. In summary, the fact that security can be based on general principles of physics allows for unconditional security, i.e. the possibility of guaranteeing security without imposing any restriction on the power of the eavesdropper.

The first quantum cryptography scheme was published by Bennett and Brassard in 1984 [8] and is known today as the BB84 protocol. It requires four different qubit states that form two complementary bases (i.e., if the result of a measurement can be predicted with certainty in one of the two bases, it is completely undetermined in the other). These states are usually realized with four linear polarization states of a photon forming two complementary bases, for, e.g., horizontal (H), vertical (V), diagonal (D), and anti-diagonal (A). As illustrated in 🔲 Fig. 18.9, Alice sends single photons to Bob, which were prepared randomly

18
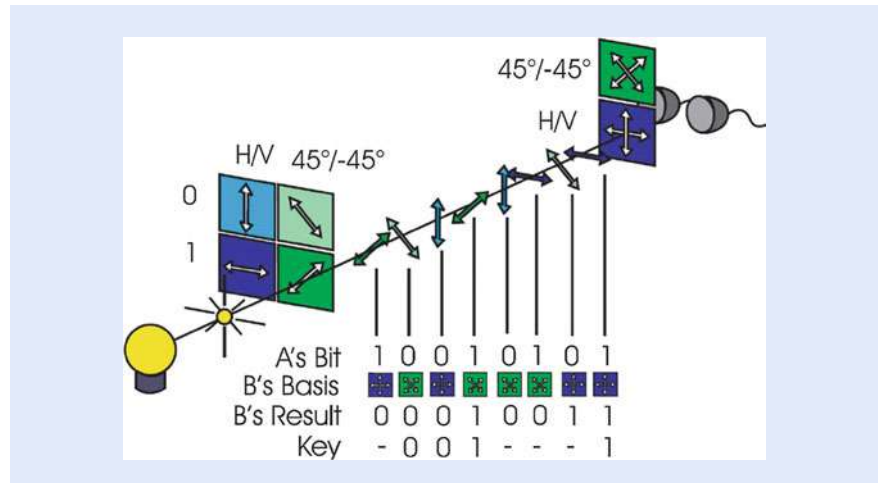
Fig. 18.9 An illustration of the coherent state BB84 protocol. Alice sends polarized single photons, prepared randomly in either of two complementary bases. Bob measures them, again randomly in one of the two bases. After publicly announcing their choice of bases, they obtain the sifted key from their data (Copyright Univ. of Vienna)

in any of the four polarization states and records the state of any sent photon. Bob receives and analyzes them with a two-channel analyzer, again randomly in one of the two complementary bases H/V or D/A. He records his measurement results together with the corresponding measurement basis. After enough photons have been transmitted, Bob communicates publicly with Alice and tells her which photons actually arrived and in which basis it was measured, but does not reveal the measurement result. In return, Alice tells Bob when she has used the same bases to prepare them, because only in these cases Bob obtains the correct result. Assigning the binary value **0** to H and D and the value **1** to V and A leaves Alice and Bob with an identical set of **0** s and **1** s. This set is called the sifted key.

The security of the key distribution is based on the fact that a measurement of an unknown quantum system will (in most cases) disturb the system: If Alice's and Bob's sifted keys are perfectly correlated (which can be proven by comparing a small subset of the whole sifted key via classical communication), no eavesdropper tried to listen to the transmission and the key can be used for encoding a confidential message using the one-time pad (i.e., a specific key is exactly as long as the message to be encrypted and this key is only used once). In practical systems, however, there will always be some inherent noise due to dark counts in the detectors and transmission errors. As it cannot be distinguished whether the errors in the sifted key come from noise in the quantum channel or from eavesdropping activity, they all must be attributed to an eavesdropping attack. If the error is below a certain threshold, Alice and Bob can still distill a final secret key using classical protocols for error correction and privacy amplification. If the error is above the threshold, the key is discarded and a new distribution has to be started.

In contrast to the *single-photon* protocols described above, entanglement based QKD uses entangled photon pairs to establish the secure key [20]. Let's assume that Alice and Bob share a polarization entangled two-photon state. Due to the perfect polarization correlations between entangled photons, Alice and Bob will always obtain the same result, when they measure the polarization state of their photon in the same measurement basis. Since both measure randomly in one of the two complementary bases (just as in the BB84 protocol), they have to publicly communicate after they have finished their measurements, which photons they actually detected and in which basis it was measured. Again, they discard those results in which they disagreed in the measurement basis and finally end up with

an identical set of **0** s and **1** s—the sifted key. Just as in the BB84 protocol, Alice and Bob authenticate their keys by openly comparing (via classical communication) a small subset of their keys and evaluating the bit error rate.

There are two big advantages in using entangled photons for implementing the QKD protocol. First, the randomness of the individual measurement results is intrinsic to the entangled state and therefore the randomness of the final key is ascertained. Second, an eavesdropper cannot mimic an entangled state by sending single photons in correlated polarization states simultaneously to Alice and Bob. Hence, when using a subset of the transmitted photon pairs to examine the entanglement between them, secure communication is possible even though the operator of the entangled photon source might not be trustworthy.

18.1.7 Quantum Teleportation

Quantum teleportation is a process by which the state of a quantum system is transferred onto another distant quantum system without ever existing at any location in between [10]. In contrast to what is often wrongly stated, this does not even in principle allow for faster-than-light communication or transport of matter. This becomes clearer when considering the entire three-step protocol of quantum teleportation (an illustration is shown in ◘ Fig. 18.10).

First, it is necessary that Alice (the sender) and Bob (the receiver) share a pair of entangled qubits (qubits 2 and 3 in the figure). Next, Alice is provided with a third qubit (qubit 1), the state of which she wants to teleport and which is unknown to her. In the last step, Alice destroys any information about the state of qubit 1 by performing a so-called Bell-state measurement (BSM) between qubits 1 and 2. As a consequence of this measurement and due to the initial entanglement between qubit 2 and 3, qubit 3 is instantaneously projected onto the same state as qubit 1. However, the teleportation protocol only works in cases, where the BSM resulted in exactly one out of four possible random outcomes. As a consequence, Bob needs to be notified by Alice about the outcome of the BSM in order to being able to identify the successful teleportation events. This requires classical communication between Alice and Bob and essentially limits the speed of information transfer within the teleportation protocol to the speed of the classical communication channel.

Quantum teleportation is an essential prerequisite for a so-called quantum repeater. A quantum repeater will be an important building block in a future
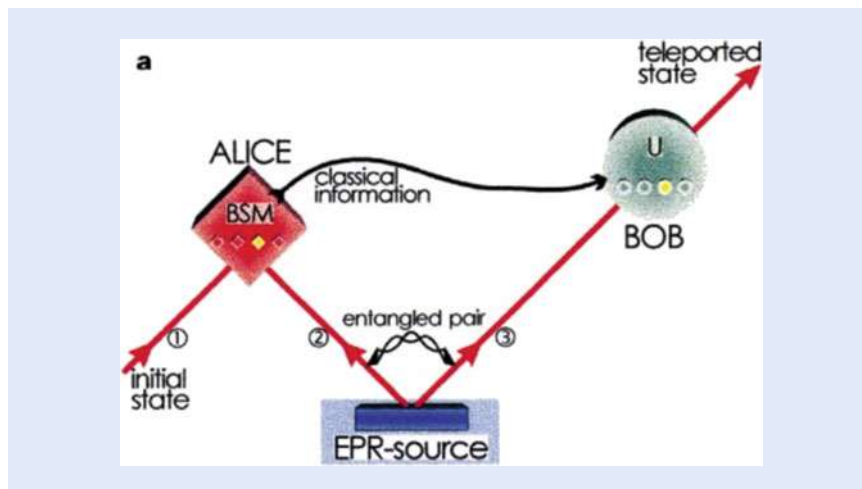

◘ **Fig. 18.10** Quantum state teleportation scheme. Picture taken from [14]

network, since it allows to interconnect different network nodes. In a quantum repeater, two particles of independent entangled pairs are combined within a BSM, such that the entanglement is relayed onto the remaining two particles. This process is called entanglement swapping and will eventually allow to overcome any distance limitations in a global-scale network. However, in order to efficiently execute entanglement swapping, it has to be supplemented with an entanglement purification step requiring quantum memories.

## 18.2 Long Distance Quantum Communication

18.2.1 Ground-Based Long-Distance Experiments

Quantum physics was invented to describe nature at the microscopic level of atoms and light. It remains an open question to what extent these laws are applicable in the macroscopic domain. In this respect, numerous ongoing research efforts pursue the goal of extending the distance between entangled quantum systems. They aim at investigating whether there are any possible fundamental limitations to quantum entanglement and if it is feasible to establish a global-scale quantum communication network in the future. In the past years, several free-space quantum communication experiments have been performed by several groups over various distances [5, 58, 63, 69, 76, 90], studying the feasibility of different quantum communication protocols over large distances. Starting with fairly short free-space links in the order of a few kilometers, the range was quickly extended up to today's world-record distance of 144 km, held by some of the authors of this article.

One of the first experiments using a 144 km free-space link between the Canary Islands of La Palma and Tenerife was performed by Ursin et al. in 2007 [76]. In this experiment (see ◪ Fig. 18.11), a source of entangled photon pairs was installed in La Palma at the top of the volcano mountain Roque de los Muchachos at an altitude of 2400 m.

One of the photons of an entangled pair was detected locally, while the other photon was sent to Tenerife. There, the optical ground station (OGS) of the European Space Agency (ESA), located at the Observatory del Teide at an altitude of 2400 m, was used as the receiving telescope for the photons coming from La Palma. After analyzing the polarization correlations between the associated photons on both islands, the scientists could verify that the photons are still entangled even though they have been separated by 144 km. Additionally, the same group implemented quantum key distribution protocols based on both entangled and single photons [69, 76]. On the one hand, the results of these experiments addressed a question of fundamental physical interest that entanglement can survive global-scale separations between the entangled particles. On the other hand, it verified that the OGS in Tenerife, which was originally built for laser communication with satellites, is also suitable to faithfully receive entangled photons. In combination, these results demonstrate the general feasibility for potential future space-based quantum communication experiments, thus setting the cornerstone for fundamental physical research as well as for potential applications of quantum mechanical principles in future network scenarios.

The achievements of these experiments were based on a combination of advanced techniques, laying the cornerstone for the Austrian researchers for a whole range of continuative activities employing the same free-space link between La Palma and Tenerife. In 2008, Fedrizzi et al. [21] generated entangled photon pairs in La Palma and sent both photons to Tenerife. The authors could verify entanglement between the photons detected in Tenerife and also implemented an entanglement based QKD protocol. This experiment was an important step
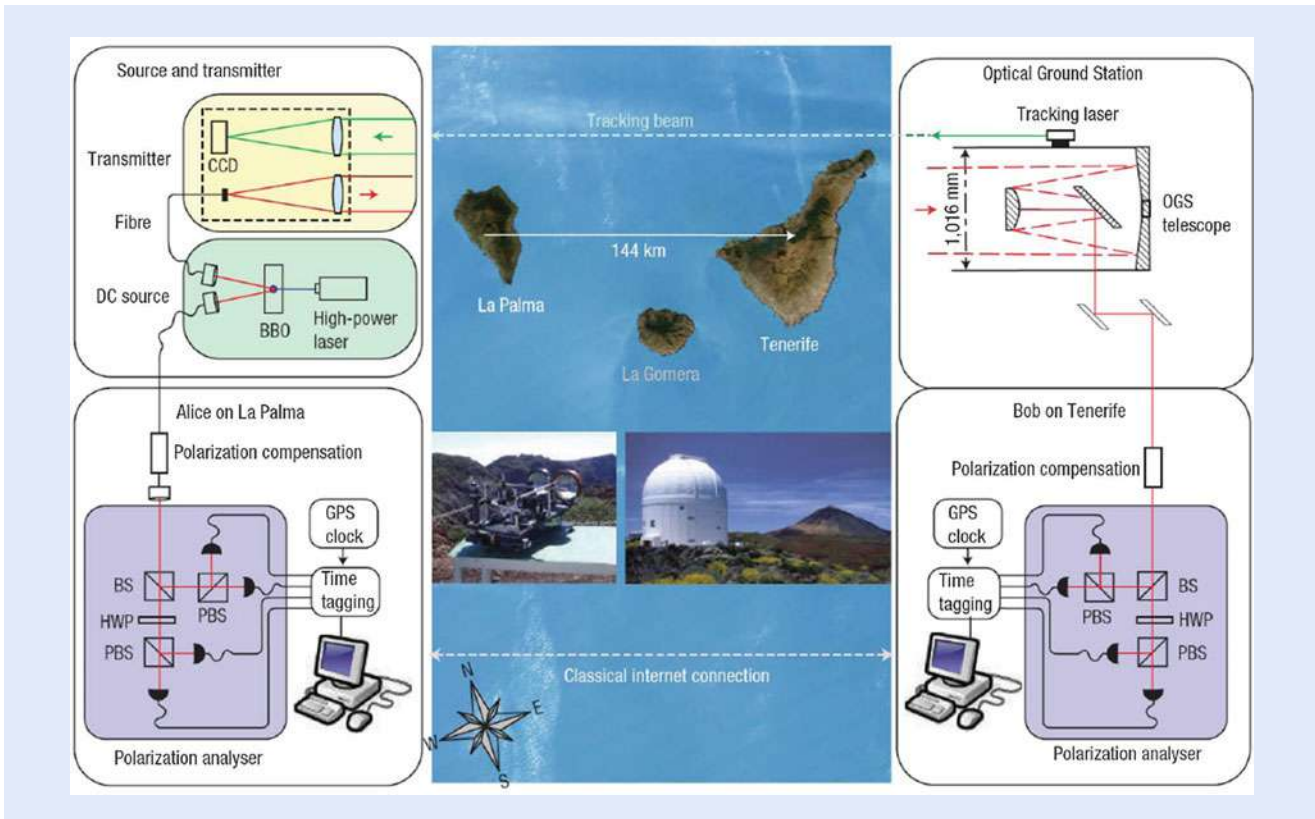
🔲 **Fig. 18.11** An illustration of the experimental setup in the inter-island experiment from Ursin et al., distributing entangled photons over 144 km between La Palma and Tenerife. Figure taken from [76]

towards a potential future quantum communication network, because with respect to the transmission loss, their experimental configuration was equivalent to a basic future network scenario, where entangled pairs are transmitted from a satellite to two separate receiving stations on ground.

The long-distance experiments of our group so far involved only two photons. However, quantum communication protocols like teleportation or entanglement swapping, as described earlier, require more than two photons and will be of utmost importance in a future network. Its experimental implementation, however, is substantially more complex than the two-photon protocols, necessitating a step back regarding the communication distance (🔲 Fig. 18.12).

In 2010, a group of Chinese researchers were the first to report on a long-distance free-space quantum teleportation experiment [34], demonstrating this protocol outside the shielded laboratory environment. They implemented a variant of the teleportation scheme described earlier and teleported the quantum states of photons over a distance of 16 km. This achievement triggered a race between the Austrian and Chinese groups to push the distance record for teleportation even further. It lasted until 2012 that the Chinese group reported on a successful demonstration of quantum teleportation over a 97 km free-space link across the Qinghai lake [90]. But it was only 8 days later that also the Austrian group with the results of their work on long-distance quantum teleportation between La Palma and Tenerife, reporting a new distance record of 143 km [47].

The communication distances spanned in these experiments were in fact more challenging than expected for a satellite-to-ground link and thus the results of both groups proved the feasibility of quantum repeaters in a future space- and ground-based worldwide quantum internet. Together with a reliable quantum memory, these results set the benchmark for an efficient quantum repeater at the heart of a global quantum-communication network (🔲 Fig. 18.13).
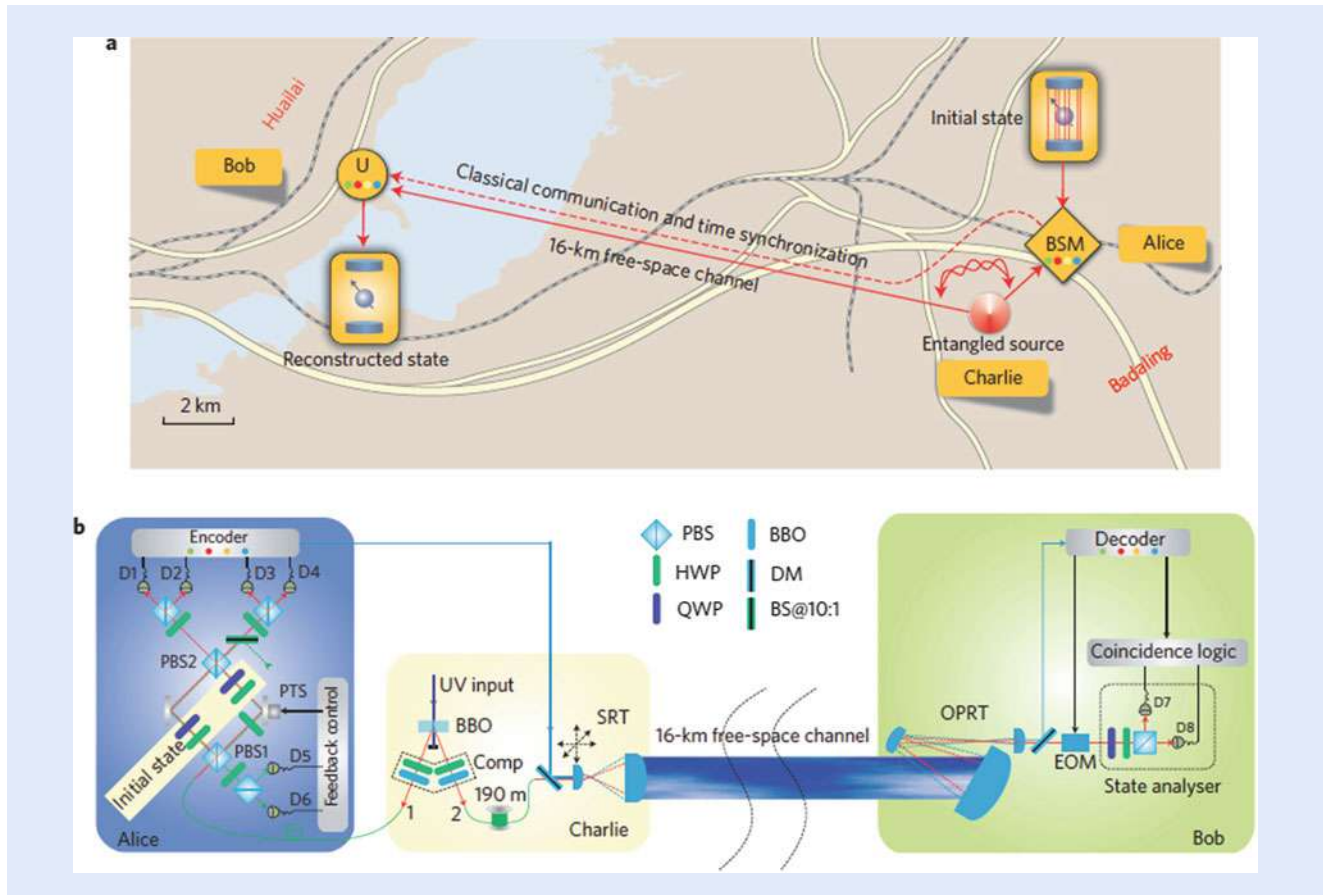
Fig. 18.12 Free-space quantum teleportation experiment over 16 km in 2010 in China. (**a**) A bird's-eye view of the experiment. (**b**) The details of the quantum optical experiment at Alice's and Bob's station. Figure taken from [34], where [34] is Experimental free-space quantum teleportation

18.2.2 Space-Based Quantum Communication

The experiments described above represent the state of the art of long-distance quantum communication. Significantly longer distances are no longer possible on ground, since the curvature of the earth will then prevent direct line of sight links. The logical next step is to bring quantum technology into space and several international research initiatives in Europe, Singapore, China, USA, and the Canada are currently pursuing related projects.

It is a clear vision of the science community to establish a worldwide quantum communication network with all the advantages over its classical counterpart described above. That requires significantly expanding the distances for distributing quantum systems beyond the capabilities of terrestrial experiments and can only be realized by tackling the additional challenge of bringing the concepts and technologies of quantum physics to a space environment. Long-distance quantum communication experiments have been underway for some time sending single photons through long optical fibers. The first scientific demonstration, still in the shielded laboratory, was conducted in the late 1990s. The question to be answered at that time was if the peculiar and fragile laboratory experiments can also be executed facing harsh real-world environmental conditions as are present in optical telecommunication networks.

There are limitations for high-speed quantum communication in optical fibers. For example, the maximum speed of generating, preparing, and detecting single photons is on the order of a few Mbit per second using state-of-the-art high speed
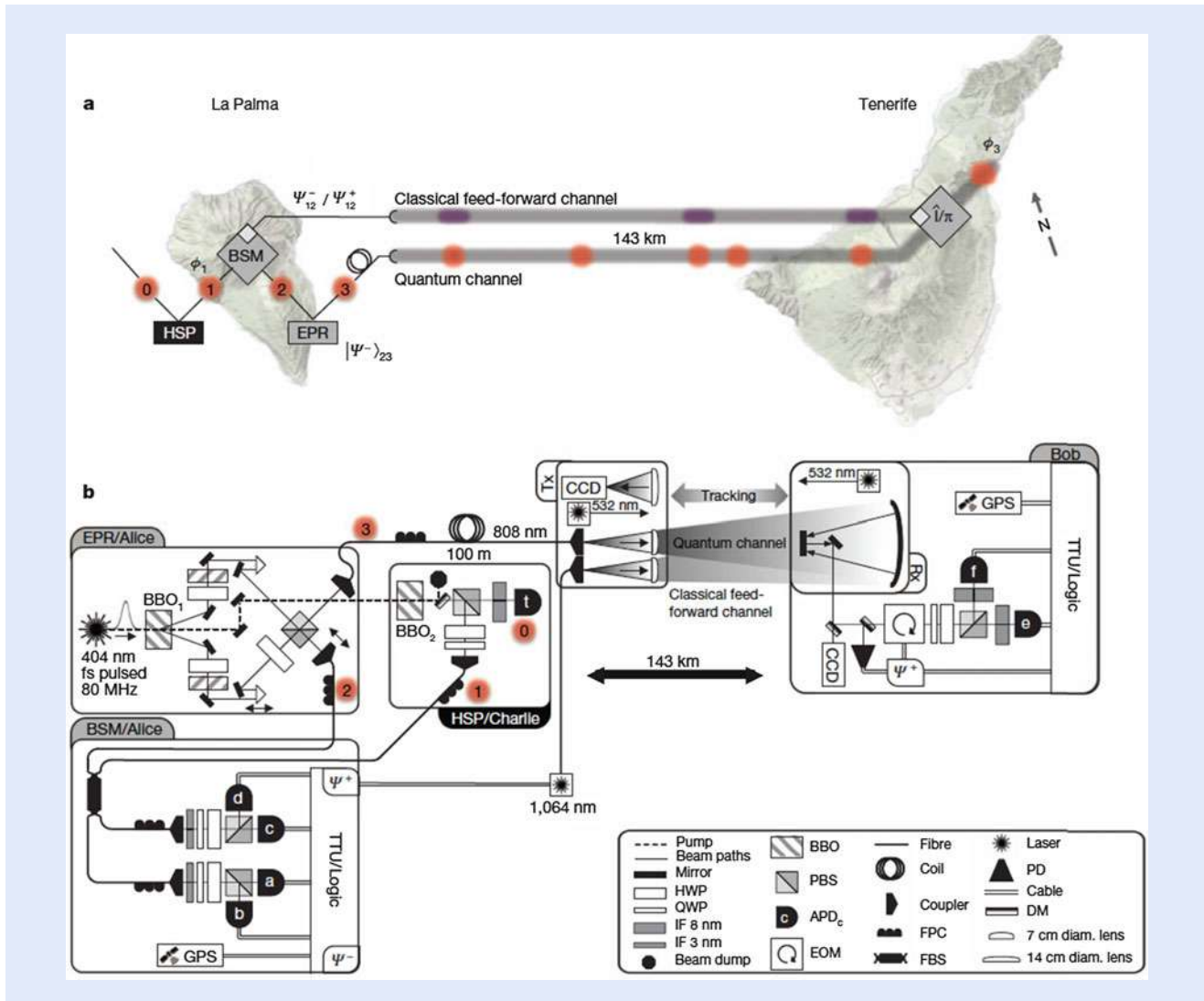
🔲 **Fig. 18.13** Free-space quantum teleportation experiment over 143 km in 2012 by the Austrian group, conducted at the Canary Islands. (**a**) A bird's-eye view of the experiment. (**b**) The details of the quantum optical experiment at the stations in La Palma and Tenerife. Figure taken from [47], where [47] is Quantum teleportation over 143 kilometres using active freed-forward

electronics. Due to the combination of noise in real detector-devices and transmission loss in the optical fiber, the distance over which quantum information can be communicated is restricted to a few 100 km [79]. Hence, for bridging distances on a global scale using optical fiber networks, the implementation of the so-called quantum repeaters is paramount. Quantum repeaters are the quantum analog to classical optical amplifiers making global fiber communication as of today yet feasible. Quantum repeaters are a theoretical concept proposed in 1998 [16] and require as basic building blocks the concepts of quantum teleportation and quantum memories. Specifically, the combination of both is highly complex from a technological point of view, such that the development of a quantum repeater is yet in the early stages. The second solution to bridge distances on a global scale is to use satellite-to-earth and inter-satellite optical free-space connections [76].

🔲 Figure 18.14 depicts a typical space-mission scenario for the distribution of entanglement from a transmitter terminal to two receiver stations (Alice and Bob). The quantum source installed on the transmitter emits pairs of photons in a desired entangled state. The photon pairs exhibit strong correlations in time,
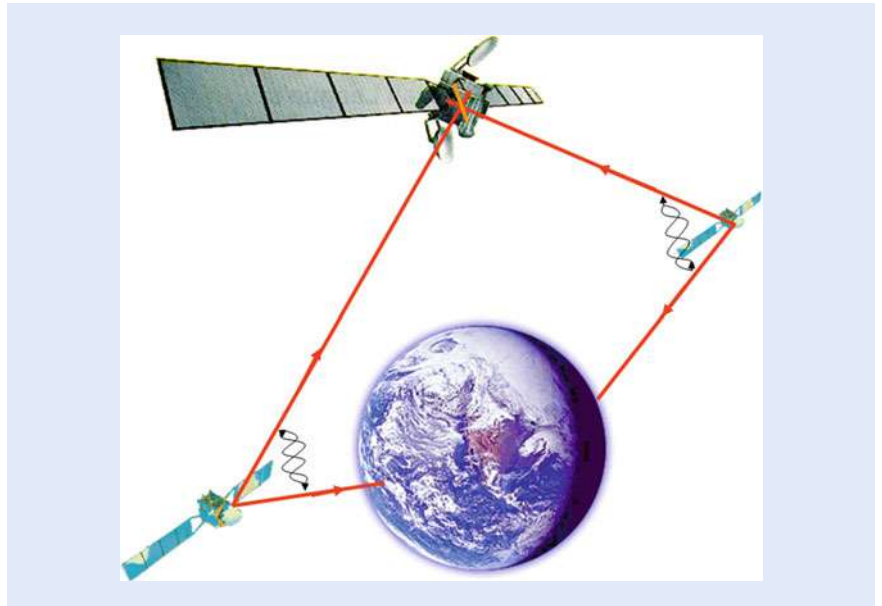
Fig. 18.14 A vision: Global Quantum Communication via satellites connecting any point on ground requiring optical ground station (taken from [77])

and entanglement in the degree of freedom in which the quantum information is encoded. The single photons comprising each of these entangled pairs are sent to Alice and Bob via free-space communication links (quantum links) established between the satellites and an optical ground station. The photons are collected via telescopes at the receiver terminals, where Alice and Bob each perform quantum measurements on their respective photons. Before initializing the transfer of information, the transmitter must establish a separate standard communication channel with Alice and Bob. This classical communications channel is subsequently used to send information about which basis state the measurements were performed on a given pair. The detection time of every arriving photon is recorded using fast single-photon detectors, and detection events that comprise an entangled pair are identified by means of their temporal correlations. The identification of photon pairs by their detection times requires the transmitter and receiver modules to establish and maintain a synchronized time basis, which can be achieved using an external reference, or autonomously via the classical communication link. Once the pair-detection events have been identified, Alice and Bob can reveal their stronger-than-classical correlations by communicating the bases of the quantum measurements performed on each photon pair via the classical communication channel.

Distributing entangled photon pairs over long-distance links and revealing their quantum correlations is an immensely challenging task from a technological point of view, in particular due to the fact that, as a result of unavoidable losses in the quantum link, only a fraction of the photons emitted by the transmitter actually arrive at the receiver modules. The main sources contributing to losses along the optical transmission channel are atmospheric absorption and scattering, on the one hand, and diffraction, telescope pointing errors, and atmospheric turbulence, which all lead to beam broadening and thus limit the fraction of photons collected by the receiver aperture, on the other. Typical losses in such scenarios are in the order of -30 to -40 dB.

Nevertheless, in order to achieve feasible pair-detection rates at such huge link losses requires a very bright source of entangled photon pairs as well as minimizing losses in the transmission channel and the receivers. Note that since correlated

photon pairs are identified by their arrival times, there is an upper limit to how effective the photon production rate can mitigate against link loss. Once the time between two successive pair emissions at the source decreases below the timing jitter of the detectors, these two successive photons can no longer be distinguished from each other, such that as a result the quantum bit error ratio (QBER) will be increased.

The pairs detected by the two terminals will ultimately comprise of photons steaming not only from the entangled photons source (the signal) but also from unavoidable sources of uncorrelated background photons (the noise). The background is from stray light the detector might see and the intrinsic dark counts of the photon avalanche detectors in use. The background can be mitigated to a certain extent by using very narrow-band filters, allowing only those photons to be guided to the detector, who are at the wavelength of the quantum source in use. Also the common timing of the entangled photons is useful to mitigate noise pair counts.

Entangled photon sources maintaining both their high brightness and the quality of the emitted quantum state will have to be manufactured in a very reliable and stable manner to survive the launch of the satellite as well as the harsh space environment (radiation). The first research and development projects funded by the European Space Agency were dedicated to the nonlinear periodically poled potassium titanyl phosphate (ppKTP) crystal, which is used in state-of-the-art entangled photon sources. Additionally, the implementation of the rather complex structure of lenses and beam-splitters is addressed in these studies and radiation effects on single photon detectors have already been investigated in detail [35]. These first attempts do show that a quantum mission based on state-of-the-art technology is feasible and requires the integration into commercially available space-laser terminals as a next step.

As outlined above, quantum communication provides a novel way of information transfer. Even though it is still under development, it has the potential to become our future technology for communication and computation. The first proposed experiments in space will serve as a very good platform to test these concepts and could pave the way for follow-up industrial systems. From a very long-term perspective it is highly interesting to test quantum mechanics at distances on the order of millions of km, and even beyond. Furthermore, an ultimate experiment regarding the role of randomness and humans free-will could be performed by two individuals, separated by at least one light second, who each measures entangled particles and separately chooses the setting of their analyzer. To extend the scale of quantum mechanical states over astronomical distances might provide us with a suitable insight on the link between gravitation, quantum mechanics, and even more. Clearly, these experiments require advances in technology not even foreseeable today. Nevertheless, the proposed experiments are a major step in investigating these fundamental questions as well as enhancing the technology for the society's benefit.

18.3 Higher Dimensions

So far, we have focused only on qubits, which are quantum mechanical two-level systems. This is a natural choice, as all of our classical data storage, transmission, and processing are based on classical two-level systems that encode zeros and ones. There are only a very few exotic exceptions, such as the *Setun* computer built in Soviet union in the late 1950s, which used trinary logic.

However, if one were to look at nature's way of encoding and processing information, one would be surprised to find that it uses a higher-level system: DNA (deoxyribonucleic acid) uses four types of nucleobase (adenine, guanine,

cytosine, and thymine) to encode information. Three nucleobases together encode one amino acid, the basis of biological life. If nature—optimized over hundreds of million of years through evolution—uses a higher-level system for encoding information, we see no reason why one shouldn't investigate its use in quantum information as well!

There are two types of high-dimensional systems that depend on whether one considers discrete or continuous parameters. An example of a continuous degree-of-freedom (DoF) is the position (or likewise, the momentum) of a photon. Quantum correlations in this DoF have been used for interesting new types of imaging schemes such as quantum ghost imaging, where the image of the object can only be seen in the correlations of the photons [50, 60, 73]. A different, even more counterintuitive quantum imaging procedure was recently demonstrated where an object was imaged without ever detecting the photons which were in contact with the imaged object [46].

In some scenarios, a discrete basis is more advantageous. In classical communications or data storage, for example, information is encoded either as a 0 or a 1; fractional numbers in between are not used. The same is true for quantum communication or quantum computation, even with larger alphabets. A natural basis that uses a discrete DoF of a photon is its orbital angular momentum, which is presented in the next section. Other possible bases can be constructed by the discretization of continuous parameters such as position or wavelength.

18.3.1 Twisted Photons

If one investigates the spatial profile of a laser beam with a camera, one usually finds that it has a Gaussian shape. However, that is only a special case of a much more complex family of fundamental spatial structures or modes. One very convenient set of modes are the so-called Laguerre–Gaussian modes [3, 57, 88]. In ▣ Fig. 18.15, the intensity and phase structure of a Gaussian mode ($\ell = 0$) compared to Laguerre–Gaussian modes ($|\ell| > 0$) are shown.

In contrast to its polarization, which is a property related to its spin angular momentum, a photon with a Laguerre–Gaussian mode structure can also carry orbital angular momentum (OAM). The spin and orbital-angular momenta have
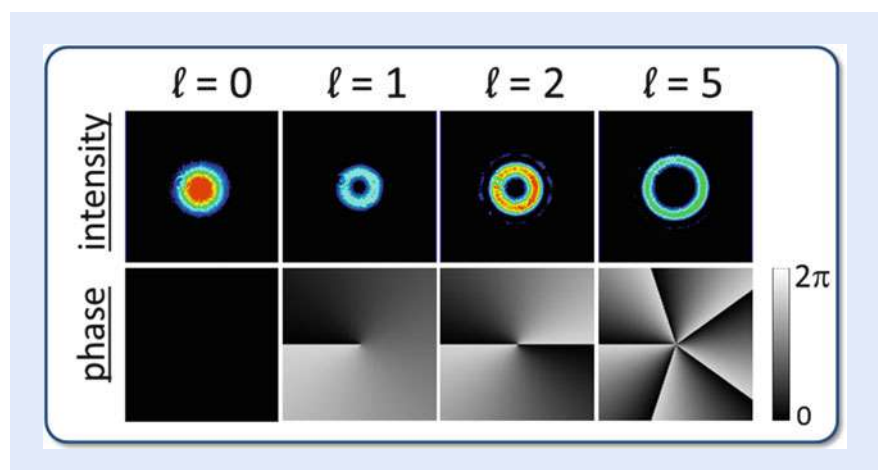

▣ **Fig. 18.15** Intensities and phase information of orbital angular momentum beams. The intensity is collected with a camera. The OAM = 0 mode is the well-known Gaussian distribution. OAM larger than 0 show a ring, or doughnut structure. The *lower line* shows that these structures have a twisted phase-front, with $2\pi \ell$ phase-change in a ring. In the center, they have a phase singularity—also known as Vortex. The vortex is the reason why there is no intensity in the center (Image by Mario Krenn, copyright University of Vienna)

distinct physical properties: if a laser beam with circular polarization illuminates a small particle, the particle will start to rotate around its own axis. However, if a beam with orbital angular momentum shines on a particle, it starts to rotate around the external orbit defined by the laser beam [29]. Surprisingly, the OAM of photons and its connection to Laguerre–Gauss modes was identified only recently in 1992 [2].

Interestingly, the OAM quantum number of a photon can theoretically take on any integer number between $-\infty$ and ∞. This allows one to encode a huge amount of data onto a single photon [17, 26]. In classical communications, this can improve the data rates enormously. Recent experiments have demonstrated data transmission of 100 Tbit/s by using the OAM of light together with other DoFs [30, 82]. In quantum communication, secret sharing protocols have been developed that use OAM modes as an alphabet for encoding [28, 48, 54, 56, 80]. Not only do such protocols offer an increased data rate, they also provide an improved level of security against eavesdropping attacks [31, 81].

18.3.2 High-Dimensional Entanglement

Earlier in this chapter, entanglement was explained in the context of photon polarization, which is a two-level system. In such systems, the separated photon pair can share one bit of information in a nonlocal manner, referred to as an entangled bit or "ebit" (◘ Fig. 18.16).

However, if we consider larger dimensional systems such as the OAM of photons, one can easily imagine that a pair of photons entangled in their OAM could share much more information than photons entangled in their polarization. Such modes get bigger in size as the OAM quantum number ℓ is increased. Thus, the amount of information carried by them is only limited by the size of the optical devices used, or more generally, by the size of the universe itself! A natural question that arises is whether there exists a limit to the amount of information that can be non-locally shared between two entangled photon pairs. This question

◘ **Fig. 18.16** Two classical 100-sided dice. If one were to roll them, it is very unlikely that they would both show the same number. However, were they high-dimensionally entangled, they would both always show the same number. Note: such a metaphor for quantum entanglement is limited in that one cannot visualize the results of correlated measurement outcomes in super-position bases. This is key for distinguishing entanglement from classical correlations (Image by Mario Krenn, copyright University of Vienna)

is being investigated in several laboratories around the world [1, 13, 18, 27, 39, 52, 55, 61, 65, 67, 75, 78]. These efforts have confirmed that two distant photons can be entangled in hundred and more dimensions of their spatial mode structure. This means that by measuring the first photon of the entangled pair, one will observe one definite result out of the hundred possible outcomes. This immediately tells us the outcome of a similar measurement on the second, distant photon. However, the strangeness lies in the fact that the two photons did not have a definite value before they were measured. Only when the first photon is observed does the common state become a reality, and the second photon gets a defined value.

Photons entangled in their orbital angular momentum also enable the possibility to explore more complex types of entanglement that is not possible with two-dimensional entangled states. Recent state-of-the-art experiments have shown the entanglement of eight photons [89], nine superconducting circuits [36], and fourteen ions [42]. However, these experiments have singularly focused on increasing the number of particles entangled, while remaining in a two-dimensional space for each particle. The OAM of light was recently used to create the first entangled state where both the number of particles and the number of dimensions were greater than two [51]. This state involved three photons asymmetrically entangled in their OAM: two photons resided in a three-dimensional space, while one photon lived in two dimensions. These experiments have been designed by a computer algorithm [41]. Interestingly, this asymmetric structure only appears when one considers multi-particle entanglement in dimensions greater than two [32]. Such states also enable a novel "layered" quantum communication protocol. For example, if three parties were to share the state described above, all three would have access to one bit of secure information, allowing them to generate a secure random key for sharing information. However, part of the time, two of the parties would have access to another bit of secure information. This would allow them to share an additional layer of information unknown to the third party in the communication scheme. This protocol can be generalized to include multiple layers of information shared asymmetrically amongst many different parties.

18.3.3 Mutually Unbiased Bases in High Dimensions

Earlier in this chapter we have learned that for 2-dimensional systems, three unbiased bases exist. For larger dimensions, one finds more of these unbiased bases: in 3 dimensions there are 4 bases, in 4 dimensions there are 5 bases. In fact, it is known that for every prime-power dimension (with $d = p^n$), the number of MUBs is (d+1). That means, in dimension d, there are (d+1) different ways to encode information. Now there is one very surprising fact: If the dimension of the space is not a prime-power, it is not known how many MUBs there are. The first of those cases is dimension $2 \cdot 3 = 6$ [7, 85]. Numerical search has only found 3 MUBs, and it is a conjecture that there are only 3 MUBs. It is fascinating because it means that in 5 dimensions, there are more ways to encode information in different ways than in 6 dimensions, even though intuitively one might think that a larger space allows for more ways to embed information in different ways. This is crucial for quantum communication, because the number of MUBs is directly connected to the robustness (against noise and eavesdropping-attacks) of the protocol. The more different ways of encoding the information, the more secure the system is.

18.3.4 High-Dimensional Quantum Key Distribution

Quantum cryptography based on photons carrying OAM is similar to the schemes developed for polarization that are explained earlier in this chapter. High-dimensional analogs to the BB84 and Ekert QKD protocols have been developed that use OAM for encoding [49]. Similar to polarization-based QKD, OAM-based QKD requires measurements to be performed in mutually unbiased bases to guarantee security against eavesdropping. The earliest such protocol was demonstrated with photons entangled in three dimensions of their OAM ($\ell = 0, +1$, and -1) [28]. The high-dimensionally entangled photon pairs were produced in a BBO crystal and sent to two separate stations, where basis transformations were randomly performed by two holograms mounted on moving motorized stages at each station. The photons were then probabilistically split into three paths where their OAM content was measured by three additional holograms. In this manner, a three-dimensional key was generated with an error rate of 10 %. Security was verified by testing for the presence of entanglement via a high-dimensional Bell inequality.

One of the challenges in using OAM modes for quantum communication is the ability to sort single photons carrying OAM. The QKD scheme described above used beam splitters and holograms to projectively measure the OAM content of the single photons. This resulted in a scheme that was photon-inefficient, i.e. only one out of every nine photons was actually used for communication. While techniques for efficiently sorting the OAM of single photons existed, they relied on N cascaded Mach–Zehnder interferometers for sorting $N + 1$ OAM modes [44]. Thus, the use of such a device in a quantum communication scheme was impractical due to issues of complexity and stability. However, in 2010, the group of Miles Padgett developed a refractive device that could sort the OAM of a single photon [12]. This device "unwrapped" the helical wavefront of an OAM mode, transforming it into a plane wave with a tilted wavefront. The amount of tilt was proportional to the OAM quantum number ℓ, allowing these modes to be separated by a simple lens. This device provided a diffraction-limited sorting efficiency of 75 %, which was improved to 93 % by the addition of two additional holographic transformations [53].

The development of this device allowed photon-efficient OAM-based quantum communication schemes to be realized in the laboratory. Recently, a BB84 protocol using a seven-dimensional OAM alphabet was performed which made heavy use of the OAM sorter discussed above [54]. Additionally, a digital micro-mirror device (DMD) was used to generate OAM modes at a rate of 4 kHz, which is much faster than the rates attainable with spatial light modulators. The key was encoded in the OAM basis as well as the mutually unbiased of the so-called angular modes (ANG), as shown in ◼ Fig. 18.17b. Using this scheme, Alice and Bob were able to communicate securely at a rate of 2.05 bits per sifted photon. Their generated key had an error rate of approximately 10 %, which was below the bounds for security against coherent attacks in a seven-dimensional QKD link. This experiment served as a proof-of-principle demonstration of OAM-based QKD. Several technological improvements (discussed in Ref. [54]) will be required to take such a scheme into the real world.

18.3.5 Large Quantum Number Entanglement

Twisted photons not only allow access to a very large state space, but also give access to very high quantum numbers. Photons can carry $\ell\hbar$ of angular momentum, and ℓ can be arbitrarily large. Usually, quantum phenomena are only

18
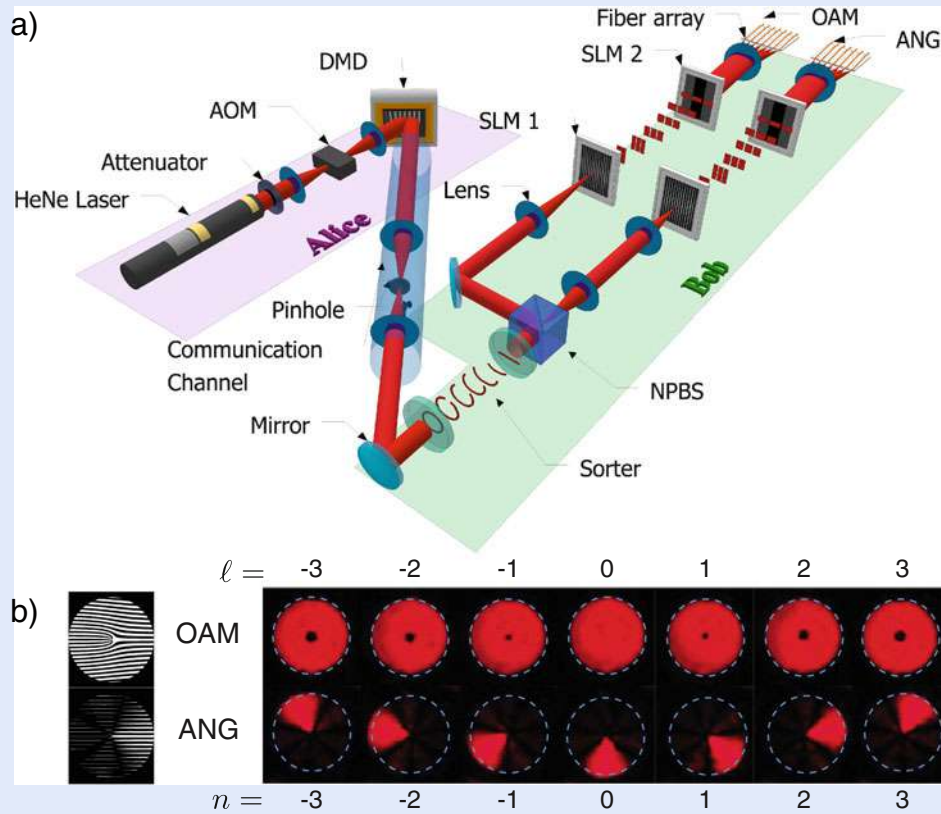
Fig. 18.17 (**a**) An OAM-based BB84 scheme for quantum key distribution. Alice encodes a random key in a seven-dimensional alphabet consisting of OAM modes using a high-speed digital micro-mirror device (DMD). Bob sorts these modes using an OAM sorter and four additional holograms implemented on spatial light modulators (SLMs). Using this scheme, Alice and Bob are able to communicate with a channel capacity of 2.05 bits per sifted photon. (**b**) CCD images showing the intensity profiles of the seven-dimensional alphabet in the OAM basis, as well as the mutually unbiased basis of angular (ANG) modes. Examples of binary holograms for generating these modes are shown on the *left* (figure adapted from Ref. [54])

observed in the microscopic world. Here, however, with twisted photons it is possible to create entanglement between photons that differ by a very large amount of angular momentum. Theoretically, there is no upper limit of the number of angular momentum, which would give rise to the possibility of entanglement of macroscopic values of angular momentum.

With this method, it was possible to show that two photons with a difference of $600\hbar$ can be entangled [22]. If the first photon carries $300\hbar$ of angular momentum, the second carried $-300\hbar$, and vice versa. While being entangled in a two-dimensional subspace, it was the largest quantum number difference achieved. In that experiment, a spatial light modulator has been used, which can be seen in ▣ Fig. 18.18. Recently, using novel methods to encode very large angular momentum at single photons, it was able to show entanglement of photons with a quantum number difference of 10,000$\hbar$ [23].

An important question that needs to be answered is the definition of *macroscopic angular momentum*, and which phenomena might arise from that. For example, there are predictions that photons close to a black hole change their angular momentum [74]. As black holes are purely general relativistic objects, and entanglement is a purely quantum mechanical phenomenon, a deeper investigation into these effects will be exciting.
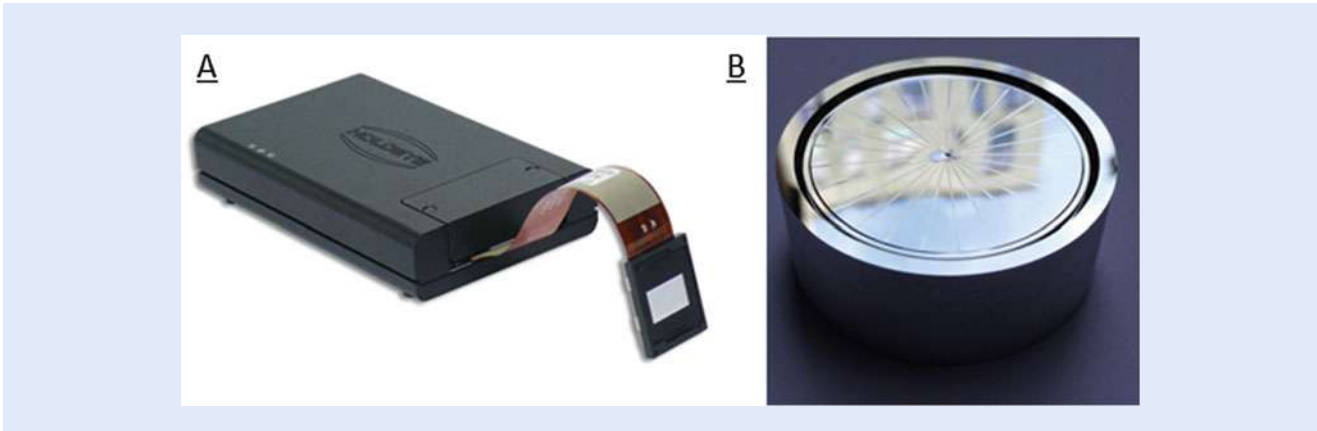
⬛ **Fig. 18.18** Different ways to create photons with large angular momentum. (**a**): A spatial light modulator consists of a liquid crystal display. The display consists of roughly 1000 × 1000 pixels, which performs phase shifts from zero to 2π. The flexibility allows to create arbitrary phase structure, thus arbitrary structures of the modulated light. However, due to their finite resolution, there is an upper limit of roughly $300\hbar$. (**b**): A different method that can create angular momentum of up to $10.000\hbar$ are fixed phase holograms built out of aluminium. In compensation for the lower flexibility, the holograms can be produced very precise, which is responsible for the much larger possible angular momentum. (Image by Robert Fickler, copyright University of Vienna)

⬛ **Fig. 18.19** Receiver at the Hedy Lamarr quantum communication telescope for the first free-space long-distance entanglement distribution experiment with a high-dimensional degree of freedom (Image by Mario Krenn, copyright University of Vienna)

18.3.6 **Long-Distance Transmission of Twisted Photons**

In a quantum communication scenario, the encoded information needs to be distributed between two parties. Usually one would think that optical fibers are the ideal solutions. Unfortunately, the information in twisted photons is not conserved in propagation through conventional fibers: Different modes mix in fibers, therefore the output is different than the input. Although recent advances show that special fibers can be used to transmit the first higher-order OAM modes for more than 1 km [15], and reach a classical communication rate in the order of Terabit, this technology is still in its infancy. Specifically, it hasn't been used in the realm of quantum physics yet. An alternative method is the transmission through free-space. In the case of earth-to-satellite quantum communication, this is the only possibility in any case (⬛ Fig. 18.19).

If long-distance transmission is considered, immediately the influence of atmospheric turbulence has to be taken into account. Varying pressure and temperature influence the structure of twisted photons. The question is: How much? While many mathematical and lab-scale studies have been performed, experimental investigation of that question is rare. Only recently, the first classical [38, 43] and quantum communication [40] experiments have been performed over free-space intra-city link of 3 km distance. Those results show that quantum entanglement with twisted photons can be distributed over larger distances, and the quality can be improved with technology that is already implemented in lab-scale experiments [62, 64, 87]. As such, it could be a reliable way to distribute high-dimensional entanglement in a future quantum network.

18.4 Conclusion

The possibility to share secret messages is of utmost importance for our society. From simple things like sending emails which can't be read by an eavesdropper to the transmission of highly sensitive information between governments that needs to be secure for decades—cryptography plays a key role in ensuring privacy, economic stability, and stable relations between countries worldwide.

As we have seen, classical cryptographic systems are vulnerable to various types of eavesdropping attacks. The problem is that either the secret key needs to be transmitted over insecure channels or (in a public–private cryptography system) the security relies on mathematical conjectures that specific properties are difficult to calculate. Furthermore, quantum computing algorithms can significantly reduce the required time to find solutions for such problems (finding prime factors of large numbers, or calculating a discrete logarithm). On top of all this, back-doors can be implemented into these algorithms such that they perform as expected, but the creator of the algorithm obtains additional information. Such attacks have been widely discussed in connection with a weak generator for pseudo random numbers certified by NIST [59, 72].

The need for overcoming these problems posed by classical asymmetric cryptographic systems has led to the development of a field called post-quantum-cryptography. There, problems which are believed to be more difficult than factoring large numbers are used to prepare a public and private key. Such methods are not practically used yet because of performance issues and unclear results on their security. While there are no classical or quantum algorithms to solve such problems yet, it is only conjectured that they are difficult to solve—a breakthrough in (quantum) complexity theory or novel kind of computations might only shift the problem into the future.

The only unconditionally secure encryption requires a random key with the same size as the message, a so-called one-time pad. The question is, how can such a key be distributed securely? Quantum key distribution provides a solution to that question, by exploiting quantum mechanical properties of individual particles. Several newly founded companies already provide small-scale quantum key distribution systems, such as ID Quantique in Switzerland, MagiQ Technologies in the USA, QuintessenceLabs in Australia, or SeQureNet in France.

As shown in this chapter, fundamental investigations test the feasibility of global quantum networks, on the order of 100 km on the Earth's surface, as well as between ground and space. A second path of research focuses on more complex quantum states, to improve data-rates and robustness against noise and eavesdropping attacks. The experiments discussed in this chapter form only a small subset of experimental efforts currently in progress around the world. It is clear that we are perched on the edge of a quantum communication revolution that will change information security and how we understand privacy for years to come.

References

1. Agnew M, Leach J, McLaren M, Roux FS, Boyd RW (2011) Tomography of the quantum state of photons entangled in high dimensions. Phys Rev A 84(6):062,101
2. Allen L, Beijersbergen M, Spreeuw R, Woerdman JP (1992) Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. Phys Rev A 45 (11):8185–8189
3. Allen L, Padgett MJ, Babiker M (1999) The orbital angular momentum of light. Prog Opt (39):291–372
4. Aspect A, Dalibard J, Roger G (1982) Experimental test of bell's inequalities using time-varying analyzers. Phys Rev Lett 49(25):1804
5. Aspelmeyer M, Böhm HR, Gyatso T, Jennewein T, Kaltenbaek R, Lindenthal M, Molina-Terriza G, Poppe A, Resch K, Taraba M et al (2003) Long-distance free-space distribution of quantum entanglement. Science 301(5633):621–623
6. Bell J (1964) On the einstein-podolsky-rosen paradox. Physics 1(3):195–200
7. Bengtsson I (2007) Three ways to look at mutually unbiased bases. In: AIP Conf Proc 889, 40. ▶ http://dx.doi.org/10.1063/1.2713445
8. Bennett C, Brassard G (1984) Quantum cryptography: public key distribution and coin tossing. In: Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing. Bangalore, p 175
9. Bennett C, Wiesner S (1992) Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. Phys Rev Lett 69(20):2881–2884
10. Bennett CH, Brassard G, Crépeau C, Jozsa R, Peres A, Wootters WK (1993) Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. Phys Rev Lett 70(13):1895–1899
11. Bennink R, Bentley S, Boyd RW, Howell J (2004) Quantum and classical coincidence imaging. Phys Rev Lett 92(3):033,601
12. Berkhout GCG, Lavery MPJ, Courtial J, Beijersbergen MW, Padgett MJ (2010) Efficient sorting of orbital angular momentum states of light. Phys Rev Lett 105(15):153,601
13. Bolduc E, Gariepy G, Leach J (2016) Direct measurement of large-scale quantum states via expectation values of non-Hermitian matrices. Nat Commun 7
14. Bouwmeester D, Pan JW, Mattle K, Eibl M, Weinfurter H, Zeilinger A (1997) Experimental quantum teleportation. Nature 390(6660):575–579
15. Bozinovic N, Yue Y, Ren Y, Tur M, Kristensen P, Huang H, Willner AE, Ramachandran S (2013) Terabit-scale orbital angular momentum mode division multiplexing in fibers. Science 340(6140):1545–1548

16. Briegel HJ, Dür W, Cirac JI, Zoller P (1998) Quantum repeaters: the role of imperfect local operations in quantum communication. Phys Rev Lett 81(26):5932
17. Čelechovskỳ R, Bouchal Z (2007) Optical implementation of the vortex information channel. New Journal of Physics 9(9):328
18. Dada AC, Leach J, Buller GS, Padgett MJ, Andersson E (2011) Experimental high-dimensional two-photon entanglement and violations of generalized Bell inequalities SI. Nat Phys 7(9):677–680
19. Einstein A, Podolsky B, Rosen N (1935) Can quantum-mechanical description of physical reality be considered complete? Phys Rev 47(10):777–780
20. Ekert AK (1991) Quantum cryptography based on Bell's theorem. Phys Rev Lett 67 (6):661–663
21. Fedrizzi A, Ursin R, Herbst T, Nespoli M, Prevedel R, Scheidl T, Tiefenbacher F, Jennewein T, Zeilinger A (2009) High-fidelity transmission of entanglement over a high-loss free-space channel. Nat Phys 5(6):389–392
22. Fickler R, Lapkiewicz R, Plick WN, Krenn M, Schaeff C, Ramelow S, Zeilinger A (2012) Quantum entanglement of high angular momenta. Science 338(6107):640–643
23. Fickler R, Campbell GT, Buchler BC, Lam PK, Zeilinger A (2016) Quantum entanglement of angular momentum states with quantum numbers up to 10010. arXiv preprint arXiv:1607.00922.
24. Franson J (1989) Bell inequality for position and time. Phys Rev Lett 62(19):2205–2208
25. Freedman SJ, Clauser JF (1972) Experimental test of local hidden-variable theories. Phys Rev Lett 28(14):938
26. Gibson G, Courtial J, Padgett MJ, Vasnetsov M, Pas'ko V, Barnett SM, Franke-Arnold S (2004) Free-space information transfer using light beams carrying orbital angular momentum. Opt Express 12(22):5448–5456
27. Giovannini D, Romero J, Leach J, Dudley A, Forbes A, Padgett MJ (2013) Characterization of high-dimensional entangled systems via mutually unbiased measurements. Physical Rev Lett 110(14):143,601
28. Groblacher S, Jennewein T, Vaziri A, Weihs G, Zeilinger A (2006) Experimental quantum cryptography with qutrits. New J Phys 8:75
29. He H, Friese M, Heckenberg N, Rubinsztein-Dunlop H (1995) Direct observation of transfer of angular momentum to absorptive particles from a laser beam with a phase singularity. Phys Rev Lett 75(5):826
30. Huang H, Xie G, Yan Y, Ahmed N, Ren Y, Yue Y, Rogawski D, Tur M, Erkmen B, Birnbaum K et al (2013) 100 tbit/s free-space data link using orbital angular momentum mode division multiplexing combined with wavelength division multiplexing. In: Optical Fiber Communication Conference, Optical Society of America, paper OTh4G.5
31. Huber M, Pawłowski M (2013) Weak randomness in device-independent quantum key distribution and the advantage of using high-dimensional entanglement. Phys Rev A 88 (3):032,309
32. Huber M, de Vicente J (2013) Structure of multidimensional entanglement in multipartite systems. Phys Rev Lett 110(3):030,501
33. Jha AK, Malik M, Boyd RW (2008) Exploring energy-time entanglement using geometric phase. Phys Rev Lett 101(18):180,405
34. Jin XM, Ren JG, Yang B, Yi ZH, Zhou F, Xu XF, Wang SK, Yang D, Hu YF, Jiang S, et al (2010) Experimental free-space quantum teleportation. Nat Photonics 4(6):376–381
35. Kaiser KH, Aulenbacher K, Chubarov O, Dehn M, Euteneuer H, Hagenbuck F, Herr R, Jankowiak A, Jennewein P, Kreidel HJ et al (2008) The 1.5 gev harmonic double-sided microtron at mainz university. Nucl Instrum Methods Phys Res Sect A Accelerators Spectrom Detect Assoc Equip 593(3):159–170
36. Kelly J, Barends R, Fowler AG, Megrant A, Jeffrey E, White TC, Sank D, Mutus JY, Campbell B, Chen Y, Chen Z, Chiaro B, Dunsworth A, Hoi IC, Neill C, O'Malley PJJ, Quintana C, Roushan P, Vainsencher A, Wenner J, Cleland AN, Martinis JM (2015) State preservation by repetitive error detection in a superconducting quantum circuit. Nature 519(7541):66–69
37. Klyshko DN (1988) A simple method of preparing pure states of an optical field, of implementing the Einstein–Podolsky–Rosen experiment, and of demonstrating the complementarity principle. Sov Phys Usp 31(1):74–85
38. Krenn M, Fickler R, Fink M, Handsteiner J, Malik M, Scheidl T, Ursin R, Zeilinger A (2014) Communication with spatially modulated light through turbulent air across Vienna. New J Phys 16(11):113,028
39. Krenn M, Huber M, Fickler R, Lapkiewicz R, Ramelow S, Zeilinger A (2014) Generation and confirmation of a (100 x 100)-dimensional entangled quantum system. Proc Natl Acad Sci 111:6243
40. Krenn M, Handsteiner J, Fink M, Fickler R, Zeilinger A (2015) Twisted photon entanglement through turbulent air across vienna. Proc Natl Acad Sci 112(46):14197–14201

41. Krenn M, Malik M, Fickler R, Lapkiewicz R, Zeilinger A (2016) Automated search for new quantum experiments. Phys Rev Lett 116(9):090405

42. Lanyon BP, Zwerger M, Jurcevic P, Hempel C, Dür W, Briegel HJ, Blatt R, Roos CF (2014) Experimental violation of multipartite bell inequalities with trapped ions. Phys Rev Lett 112(10):100,403

43. Lavery MP, Heim B, Peuntinger C, Karimi E, Magaña-Loaiza OS, Bauer T, Marquardt C, Boyd RW, Padgett M, Leuchs G et al (2015) Study of turbulence induced orbital angular momentum channel crosstalk in a 1.6 km free-space optical link. In: CLEO: Science and Innovations, Optical Society of America, paper STu1L.4

44. Leach J, Padgett MJ, Barnett SM, Franke-Arnold S (2002) Measuring the orbital angular momentum of a single photon. Phys Rev Lett 88:257,901

45. Leach J, Warburton RE, Ireland DG, Izdebski F, Barnett SM, Yao AM, Buller GS, Padgett MJ (2012) Quantum correlations in position, momentum, and intermediate bases for a full optical field of view. Phys Rev A 85(1):013,827

46. Lemos GB, Borish V, Cole GD, Ramelow S, Lapkiewicz R, Zeilinger A (2014) Quantum imaging with undetected photons. Nature 512(7515):409–412

47. Ma XS, Herbst T, Scheidl T, Wang D, Kropatschek S, Naylor W, Wittmann B, Mech A, Kofler J, Anisimova E et al (2012) Quantum teleportation over 143 kilometres using active feed-forward. Nature 489(7415):269–273

48. Mafu M, Dudley A, Goyal S, Giovannini D, McLaren M, Padgett MJ, Konrad T, Petruccione F, Lütkenhaus N, Forbes A (2013) Higher-dimensional orbital-angular-momentum-based quantum key distribution with mutually unbiased bases. Phys Rev A 88(3):032,305

49. Malik M, Boyd RW (2014) Quantum imaging technologies. Riv Nuovo Cimento 37:273

50. Malik M, Shin H, O'Sullivan MN, Zerom P, Boyd RW (2010) Quantum ghost image identification with correlated photon pairs. Phys Rev Lett 104(16):163,602

51. Malik M, Erhard M, Huber M, Krenn M, Fickler R, Zeilinger A (2016) Multi-photon entanglement in high dimensions. Nat Photonics 10(4):248–252

52. McLaren M, Agnew M, Leach J, Roux FS, Padgett MJ, Boyd RW, Forbes A (2012) Entangled bessel-gaussian beams. Opt Express 20(21):23,589–23,597

53. Mirhosseini M, Malik M, Shi Z, Boyd RW (2013) Efficient separation of the orbital angular momentum eigenstates of light. Nat Commun 4:2781

54. Mirhosseini M, Magaña-Loaiza OS, O'Sullivan MN, Rodenburg B, Malik M, Lavery MPJ, Padgett MJ, Gauthier DJ, Boyd RW (2015) High-dimensional quantum cryptography with twisted light. New J Phys 17(3):033,033

55. Molina-Terriza G, Vaziri A, Řeháček J, Hradil Z, Zeilinger A (2004) Triggered qutrits for quantum communication protocols. Phys Rev Lett 92(16):167,903

56. Molina-Terriza G, Vaziri A, Ursin R, Zeilinger A (2005) Experimental quantum coin tossing. Phys Rev Lett 94(4):40,501

57. Molina-Terriza G, Torres JP, Torner L (2007) Twisted photons. Nat Phys 3(5):305–310

58. Peng CZ, Yang T, Bao XH, Zhang J, Jin XM, Feng FY, Yang B, Yang J, Yin J, Zhang Q et al (2005) Experimental free-space distribution of entangled photon pairs over 13 km: towards satellite-based global quantum communication. Phys Rev Lett 94(15):150,501

59. Perlroth N (2013) Government announces steps to restore confidence on encryption standards. New York Times. Available at: ▶ bits.blogs.nytimes.com/2013/09/10/govern ment-announces-steps-to-restore-confidence-on-encryption-standards/

60. Pittman T, Shih Y, Strekalov D, Sergienko A (1995) Optical imaging by means of two-photon quantum entanglement. Phys Rev A 52(5):R3429–R3432

61. Pors J, Oemrawsingh S, Aiello A, Van Exter M, Eliel E, Woerdman J et al (2008) Shannon dimensionality of quantum channels and its application to photon entanglement. Physical review letters 101(12):120,502

62. Ren Y, Xie G, Huang H, Ahmed N, Yan Y, Li L, Bao C, Lavery MP, Tur M, Neifeld MA et al (2014) Adaptive-optics-based simultaneous pre-and post-turbulence compensation of multiple orbital-angular-momentum beams in a bidirectional free-space optical link. Optica 1(6):376–382

63. Resch K, Lindenthal M, Blauensteiner B, Böhm H, Fedrizzi A, Kurtsiefer C, Poppe A, Schmitt-Manderbach T, Taraba M, Ursin R et al (2005) Distributing entanglement and single photons through an intra-city, free-space quantum channel. Opt Express 13(1):202–209

64. Rodenburg B, Mirhosseini M, Malik M, Magaña-Loaiza OS, Yanakas M, Maher L, Steinhoff NK, Tyler GA, Boyd RW (2014) Simulating thick atmospheric turbulence in the lab with application to orbital angular momentum communication. J Phys 16(3):033,020

65. Romero J, Giovannini D, Franke-Arnold S, Barnett S, Padgett M (2012) Increasing the dimension in high-dimensional two-photon orbital angular momentum entanglement. Phys Rev A 86(1):012,334

66. Rowe MA, Kielpinski D, Meyer V, Sackett CA, Itano WM, Monroe C, Wineland DJ (2001) Experimental violation of a bell's inequality with efficient detection. Nature 409 (6822):791–794

67. Salakhutdinov V, Eliel E, Löffler W (2012) Full-field quantum correlations of spatially entangled photons. Phys Rev Lett 108(17):173,604

68. Scheidl T, Ursin R, Kofler J, Ramelow S, Ma XS, Herbst T, Ratschbacher L, Fedrizzi A, Langford NK, Jennewein T et al (2010) Violation of local realism with freedom of choice. Proc Natl Acad Sci 107(46):19,708–19,713

69. Schmitt-Manderbach T, Weier H, Fürst M, Ursin R, Tiefenbacher F, Scheidl T, Perdigues J, Sodnik Z, Kurtsiefer C, Rarity JG et al (2007) Experimental demonstration of free-space decoy-state quantum key distribution over 144 km. Phys Rev Lett 98(1):010,504

70. Schrödinger E (1935) Die gegenwärtige situation in der quantenmechanik. Naturwissenschaften 23(49):823–828

71. Shor PW (1994, November) Algorithms for quantum computation: Discrete logarithms and factoring. In Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on (pp. 124-134). IEEE

72. Shumow D, Ferguson N (2007) On the possibility of a back door in the nist sp800–90 dual ec prng. In: Proc. Crypto, vol 7

73. Strekalov D, Sergienko A, Klyshko D (1995) Observation of two-photon "Ghost" interference and diffraction. Phys Rev Lett 74:3600

74. Tamburini F, Thidé B, Molina-Terriza G, Anzolin G (2011) Twisting of light around rotating black holes. Nat Phys 7(3):195–197

75. Torres JP, Deyanova Y, Torner L, Molina-Terriza G (2003) Preparation of engineered two-photon entangled states for multidimensional quantum information. Phys Rev A 67(5):052,313

76. Ursin R, Tiefenbacher F, Schmitt-Manderbach T, Weier H, Scheidl T, Lindenthal M, Blauensteiner B, Jennewein T, Perdigues J, Trojek P et al (2007) Entanglement-based quantum communication over 144 km. Nat Phys 3(7):481–486

77. Ursin R, Jennewein T, Kofler J, Perdigues JM, Cacciapuoti L, de Matos CJ, Aspelmeyer M, Valencia A, Scheidl T, Acin A et al (2009) Space-quest, experiments with quantum entanglement in space. Europhys. News 40(3):26–29

78. Vaziri A, Weihs G, Zeilinger A (2002) Experimental two-photon, three-dimensional entanglement for quantum communication. Phys Rev Lett 89(24):240,401

79. Waks E, Zeevi A, Yamamoto Y (2002) Security of quantum key distribution with entangled photons against individual attacks. Phys Rev A 65(5):052,310

80. Walborn S, Lemelle D, Almeida M, Ribeiro P (2006) Quantum Key Distribution with Higher-Order Alphabets Using Spatially Encoded Qudits. Phys Rev Lett 96(9):090,501

81. Wang C, Deng F, Li Y, Liu X, Long G (2005) Quantum secure direct communication with high-dimension quantum superdense coding. Phys Rev A 71(4):–

82. Wang J, Yang JY, Fazal IM, Ahmed N, Yan Y, Huang H, Ren Y, Yue Y, Dolinar S, Tur M, Willner AE (2012) Terabit free-space data transmission employing orbital angular momentum multiplexing. Nat Phot 6(7):488–496

83. Weedbrook C, Pirandola S, García-Patrón R, Cerf NJ, Ralph TC, Shapiro JH, Lloyd S (2012) Gaussian quantum information. Rev Mod Phys 84(2):621–669

84. Weihs G, Jennewein T, Simon C, Weinfurter H, Zeilinger A (1998) Violation of bell's inequality under strict einstein locality conditions. Phys Rev Lett 81(23):5039

85. Wieśniak M, Paterek T, Zeilinger A (2011) Entanglement in mutually unbiased bases. J Phys 13(5):053,047

86. Wootters WK, Zurek WH (1982) A single quantum cannot be cloned. Nature 299 (5886):802–803

87. Xie G, Ren Y, Huang H, Lavery MP, Ahmed N, Yan Y, Bao C, Li L, Zhao Z, Cao Y, et al (2015) Phase correction for a distorted orbital angular momentum beam using a zernike polynomials-based stochastic-parallel-gradient-descent algorithm. Opt Lett 40 (7):1197–1200

88. Yao A, Padgett MJ (2011) Orbital angular momentum: origins, behavior and applications. Adv Opt Photon 3(2):161–204

89. Yao XC, Wang TX, Xu P, Lu H, Pan GS, Bao XH, Peng CZ, Lu CY, Chen YA, Pan JW (2012) Observation of eight-photon entanglement. Nat Phot 6(4):225–228

90. Yin J, Ren JG, Lu H, Cao Y, Yong HL, Wu YP, Liu C, Liao SK, Zhou F, Jiang Y, et al (2012) Quantum teleportation and entanglement distribution over 100-kilometre free-space channels. Nature 488(7410):185–188

18

Wave-Particle Dualism in Action

Wolfgang P. Schleich

W.P. Schleich (✉)
Institut für Quantenphysik and Center for Integrated Quantum Science and Technology (IQST), Universität Ulm, D-89069 Ulm, Germany

Institute for Quantum Science and Engineering (IQSE), Texas A&M University Institute for Advanced Study (TIAS), College Station 77843-4242, TX, USA

Department of Physics and Astronomy, Texas A&M University, College Station 77843-4242, TX, USA

19.1 Introduction

The wave-particle dualism, that is the wave nature of particles and the particle nature of light together with the uncertainty relation of Werner Heisenberg and the principle of complementarity formulated by Niels Bohr represent pillars of quantum theory. We provide an introduction into these fascinating yet strange aspects of the microscopic world and summarize key experiments confirming these concepts so alien to our daily life.

» "It looks strange and it looks strange and it looks very strange; and then suddenly it doesn't look strange at all and you can't understand what made it look strange in the first place."

Gertrude Stein

The opening quote refers to modern art but might as well refer to the light quantum, that is the photon. Indeed, in his lecture entitled "Delayed choice experiment and the Bohr-Einstein-Dialogue" on June 5, 1980 in a joint session of the American Philosophical Society and the Royal Society John Archibald Wheeler notes:

» "The quantum, the most revolutionary principle in all of science and the strangest continues today to unfold its wonders and raise every deeper questions about the relation between man and the universe."

The year of the light constitutes an excellent opportunity to review the progress in our understanding of the light quantum and its idiosyncrasies made possible only recently thanks to novel experimental techniques of addressing and manipulating single particles.

19.1.1 The Strange Photon

Although we have learned a lot we still lack the full picture. In particular, there is still no unique answer to the long-standing question: "What is a photon?"

In the present essay we of course do not answer this deep question either but illuminate one important aspect of the photon that on first sight looks very strange that is the wave-particle dualism. Indeed, according to the quantum theory of radiation the photon is a wave *and* a particle at the same time and their respective distinct features manifest themselves in countless phenomena. The double-slit experiment with individual photons is one of them.

The ultimate goal of our article is to discuss a rather special double-slit experiment based on two entangled photons which seems to show simultaneously the wave and the particle nature of light. Such a behavior which is strictly forbidden by quantum theory and, in particular, by the principle of complementarity makes the photon even stranger. However, a closer look at the details of the light generation reveals that there is no violation of quantum mechanics, and in the words of G. Stein: "...suddenly it doesn't look strange at all."

19.1.2 Overview

In order to lay the foundations for our study we first recall important concepts of quantum mechanics such as the wave nature of matter, the uncertainty principle, complementarity, and the quantum eraser. We then focus on a brief description and an elementary analysis of this experiment.

Our article is organized as follows: In ▪ Sect. 19.2 we focus on *the* trademark of quantum mechanics, that is, discrete events and yes/no answers arising from

measurements of single particles. Closely associated with this notion is the wave nature of particles discussed in more detail in ◘ Sect. 19.3. Here we consider not only matter waves but also light waves.

We then dedicate ◘ Sect. 19.4 to a historical overview starting with the Heisenberg uncertainty principle and arriving via the formulation of the principle of complementarity at the delayed-choice experiment and the game of twenty questions in its surprise version. In ◘ Sect. 19.5 we turn to the Bohr–Einstein dialogue on the recoiling double-slit and the quantum eraser.

Finally, ◘ Sect. 19.6 is devoted to the discussion of the double-slit experiment using two entangled photons suggesting "which-path" information while observing at the same time interference. We explain these rather counter-intuitive results by considering an elementary model. In particular, we demonstrate that the mutually exclusive scattering arrangements involve different atoms. Therefore, there is no contradiction to the principle of complementarity. We conclude in ◘ Sect. 19.7 by summarizing our results and by providing ideas for further research.

19.2 From the Macro- to the Microcosmos

The transition from the macroscopic to the microscopic world, that is from our daily life to that of an electron orbiting a nucleus, is not as smooth as the limit of classical mechanics of a particle moving with a large velocity to that with a small velocity, or vice versa. In the present section we provide an elementary introduction into some peculiarities of the quantum world, in particular the importance of single events. This fact which emerges from an elementary gedanken experiment suggests that trajectories of particles do not exist in the microscopic world.

19.2.1 Atommechanik

Newtonian mechanics is extremely successful and describes correctly the motion of macroscopic bodies, such as cars, trains, planes, and even planets. Indeed, the description of the motion of the earth around the sun on a Kepler orbit has been a great triumph of classical mechanics.

Of course there are deviations from Newtonian mechanics, for example, due to special relativity when the velocity of the moving object approaches the speed c of light, or due to general relativity, when the curvature of spacetime is no more negligible. An example for the latter is the perihelion shift of mercury.

Why not apply the Newtonian concept of planetary motion which has worked so beautifully for the macro-cosmos to problems of the microscopic world, such as a hydrogen atom. In complete analogy to the earth–sun system we now consider the motion of a single electron around the proton. The resulting Rutherford model of hydrogen supplemented by the appropriate quantization conditions of the actions as proposed by Niels Bohr, Arnold Sommerfeld, and William Wilson gives us a first glimpse of the inner workings of the atom.

However, the early success of "Atommechanik" as this field was called quickly faded. There were too many features of the atom this theory could not explain. Only quantum mechanics developed by W. Heisenberg, Erwin Schrödinger, and Paul Adrien Maurice Dirac could provide a complete and consistent picture.

What is the crucial element not included in Atommechanik? What is the unique feature distinguishing the macro- from the micro-world? Where is the borderline between them, as asked in ◘ Fig. 19.1?
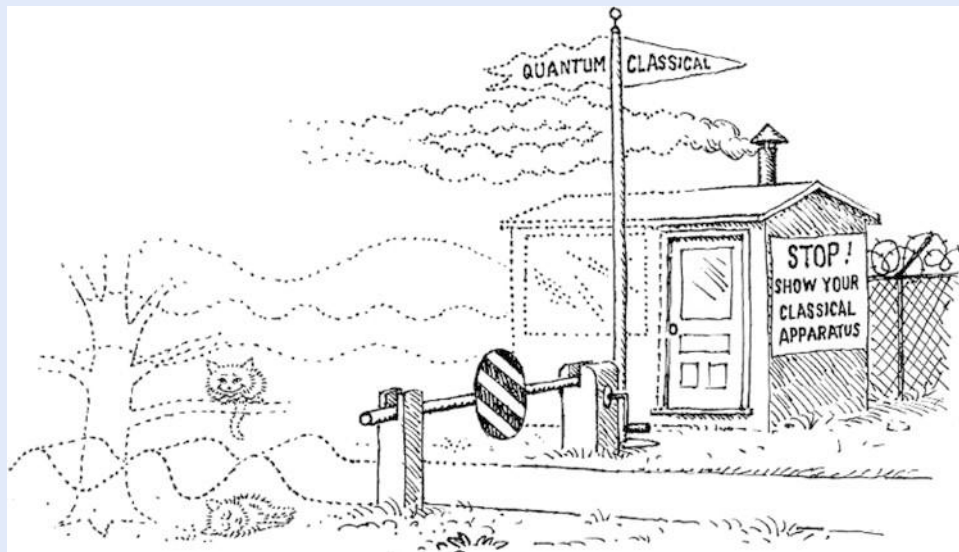
Fig. 19.1 The interface of the classical and the quantum world depicted as the border between two countries with either well-defined structures such as the right part of the guard house, or fuzzy ones represented, for example, by the Schrödinger cat sitting alive in the tree *and* laying dead in the grass. [Taken from Zurek WH (1991) Decoherence and the transition from quantum to classical. Physics Today 44:36–44]

19

19.2.2 Single Events and Probabilities

In order to provide at least some partial answers to these questions we now consider gedanken experiments which are extremely popular and very helpful in quantum theory. Gedanken experiments whose outcome is predicted by quantum mechanics are constructed to emphasize certain alien aspects of the underlying theory and can be performed in our brain without ever really going to a laboratory.

We illustrate the concept of a gedanken experiment using a specific example. How to determine the motion of an electron in an atom around the nucleus?

Since the electron has an average separation from the nucleus which is of the order of a few Bohr radii we cannot just simply take a camera and take pictures of the electron, or look at it with a microscope. The only way to gain more information is to send a probe into the atom.

In the discussion of what defines the borderline between the microscopic and the macroscopic world one quantity stands out most clearly and allows us to make such a decisive cut: it is Planck's constant $\hbar$. For a given object we can compare its angular momentum $\mathbf{J}$ to $\hbar$. When $\mathbf{J}$ is of the order of $\hbar$ we certainly deal with a problem from the microscopic world.

In classical mechanics, that is in the mechanics of macroscopic bodies we observe trajectories. At every instance of time we can determine uniquely the position of the body. The positions at different times form a world line in spacetime.

However, there are no continuous trajectories for a quantum particle. This feature originates from the discreteness of the particles and reflects the fact that we are trying to learn something about the properties of the microscopic world. Since we do not have a microscope with a resolving power large enough to observe the electron in the atom we have to send a probe from the outside into the atom. By measuring the change of that probe induced by the interaction with the electron we learn something about the electron.

When we use a single particle as a probe we get one bit of information from the detection of the scattered probe. In order to obtain more information we have to repeat this experiment many times. In this way these scattering events, each

obtained from single quantum probes, provide us with information about the inside of the atom.

This analysis brings out most clearly that we do not see the electron in the atom move around the nucleus as suggested by the Bohr–Sommerfeld–Wilson atom model, but rather find *probabilities* that the electron had been at a certain position. Obviously, the scattering events do not tell us with certainty the locations where the electron was at a given time.

19.2.3 Single Clicks Reconstruct the Microcosmos

In summary, the microscopic world is only accessible through probes which have to be of the same size as the elements of the microscopic world that we want to investigate. Hence, we probe quantum objects by single microscopic particles.

From every scattering event we gain one bit of information and complete our picture of the microscopic world by recording a multitude of single events, that is, single clicks. Once more we are reminded of a quote by J. A. Wheeler who summarized this situation in his poetic style:

» "Do we not do better to recognize that what we call existence consists of countably many iron posts of observations between which we fill in by an elaborate papier-mâché construction of imagination and theory?"

19.3 Double-Slit Experiments with Light and Matter

Wave-like aspects of light have already been observed around 1660. For example, Francesco Grimaldi noticed that when light passes a narrow slit in a wall the edges of the narrow band of brightness are slightly blurred suggesting that light diffracts.

However, it was only in the beginning of the twentieth century that a similar revolution took place for matter. Up to that moment electrons, atoms, or molecules were considered particles. However, the experiments of Clinton Joseph Davisson and Lester Halbert Germer in 1926 who scattered electrons from a nickel crystal brought out most clearly that also matter displays wave features as proposed earlier by Louis-Victor Pierre Raymond de Broglie.

In the present section we first recall the transition from the corpuscular theory of light due to Isaac Newton to the wave interpretation of Thomas Young. We then briefly review various double-slit experiments with matter waves and conclude by emphasizing subtleties associated with this arrangement.

19.3.1 Light: Corpuscle Versus Wave

More than 200 years ago, Th. Young demonstrated the wave nature of light. However, he did not use slits in an opaque screen as widely believed but rather pinholes. Despite this fact this famous experiment carries the name *double-slit experiment*.

Before his impressive demonstration the dominance of the corpuscular theory of light proposed by I. Newton had suppressed any wave theory. The following quote from Th. Young's article may illustrate this strong influence of I. Newton even almost 100 years later:

» "In making some experiments on the fringes of colors accompanying shadows, I have found so simple and so demonstrative a proof of the general law of the interference of two portions of light, which I have already

endeavored to establish, that I think it right to lay before the Royal Society, a short statement of the facts which appear to me so decisive. The proposition on which I mean to insist at the present, is simply this, that fringes of colors are produced by the interference of two portions of light; and I think it will not be denied by the most prejudiced, that the assertion is proved by the experiments I am about to relate, which may be repeated with great ease, whenever the sun shines, and without any other apparatus than is at hand to every one."

He continues his critique of the corpuscular theory by stating:

» "Those who are attached to the Newton theory of light, or to hypotheses of modern opticians, founded on views still less enlarged, would do well to endeavor to imagine anything like an explanation of these experiments, derived from their own doctrines; and, if they fail in the attempt, to refrain at least from idle declamation against a system which is founded on the accuracy of its application to all these facts, and to a thousand others of a similar nature."

It is amusing that the quantum theory of radiation brings back the particle aspect of light in the form of the photon, that is, the quantized excitation of a mode of the radiation field. In this way I. Newton and his corpuscular theory were vindicated after all.

19.3.2 Matter: Particle Versus Wave

Next we turn to the wave nature of matter which under appropriate conditions can also manifest itself in interference fringes in the far field of a double-slit. Here we discuss "slits in space" as well as "slits in time."

Slits in Space

In ◻ Fig. 19.2 we show the essential ingredients of a double-slit experiment for matter waves consisting of a particle source, an opaque screen with two slits, and a detector in the far field. We assume that the source emits one particle at a time and there is a long delay between two successive emissions. In this case there is only a single particle between the source and the detector at a time.

Each particle can only go either through the upper or the lower slit. After many particles have passed the slits, we should observe a double-hump distribution where the two maxima correspond to the two slits. However, numerous experiments clearly show that under appropriate conditions this by classical notions motivated picture is incorrect.
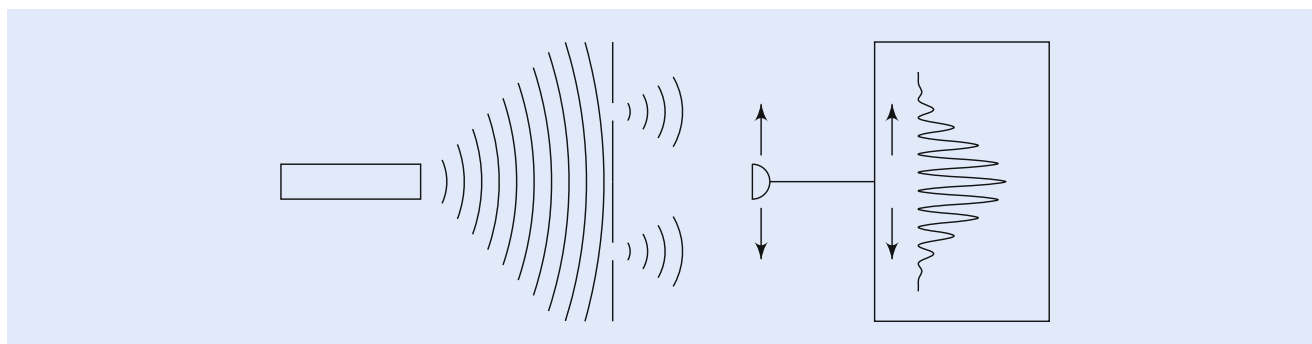
◻ **Fig. 19.2** Elementary building blocks of a double-slit experiment for particles involving a source of particles (*left*), a screen with two slits (*middle*), and a detector in the far field (*right*). The particles to be scattered could be electrons, neutrons, atoms, or rather large molecules. We observe an oscillatory count rate (*far right*) as the detector moves along an axis parallel to the screen demonstrating the existence of matter waves
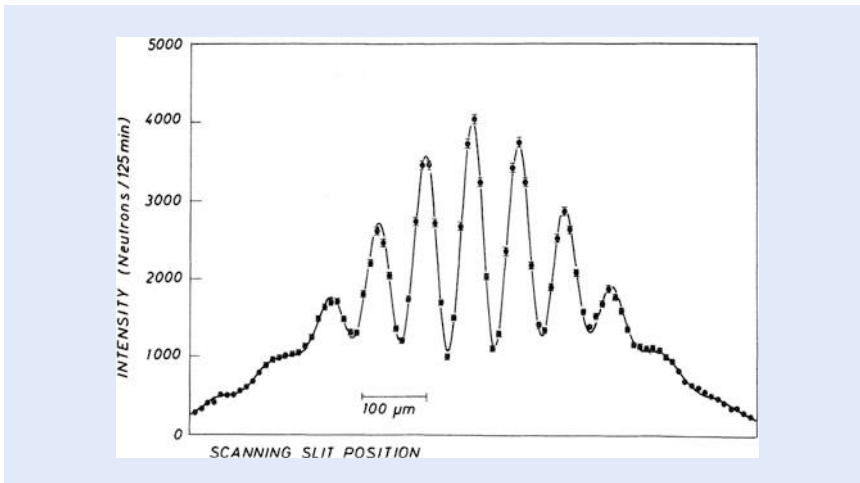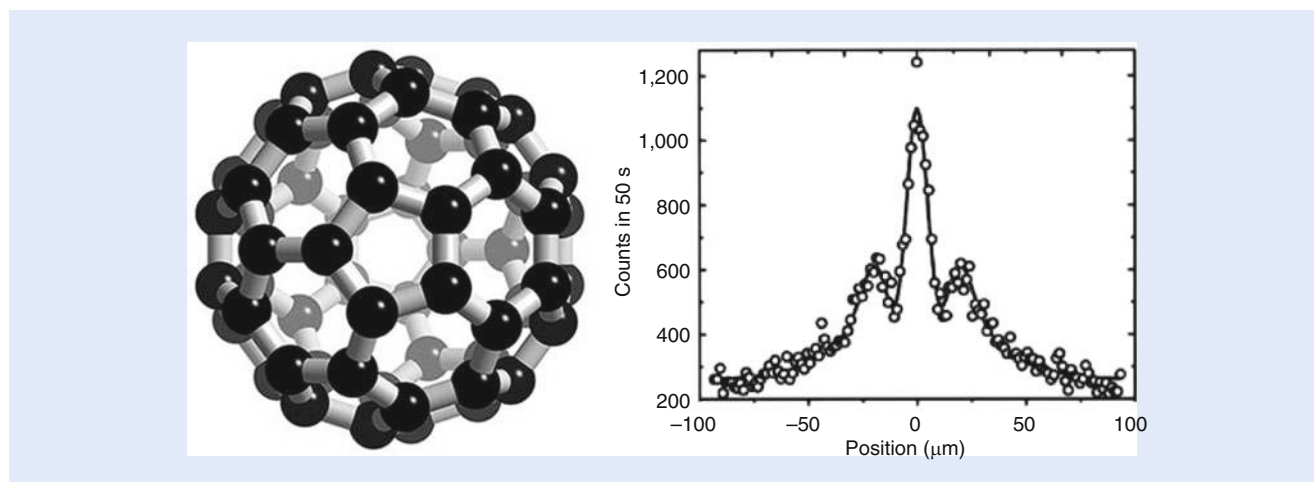
■ **Fig. 19.3** Interference pattern of a double-slit experiment with cold neutrons. [Taken from Zeilinger A et al. (1988) Single and double-slit diffraction of neutrons. Rev Mod Phys 60:1067–1073]
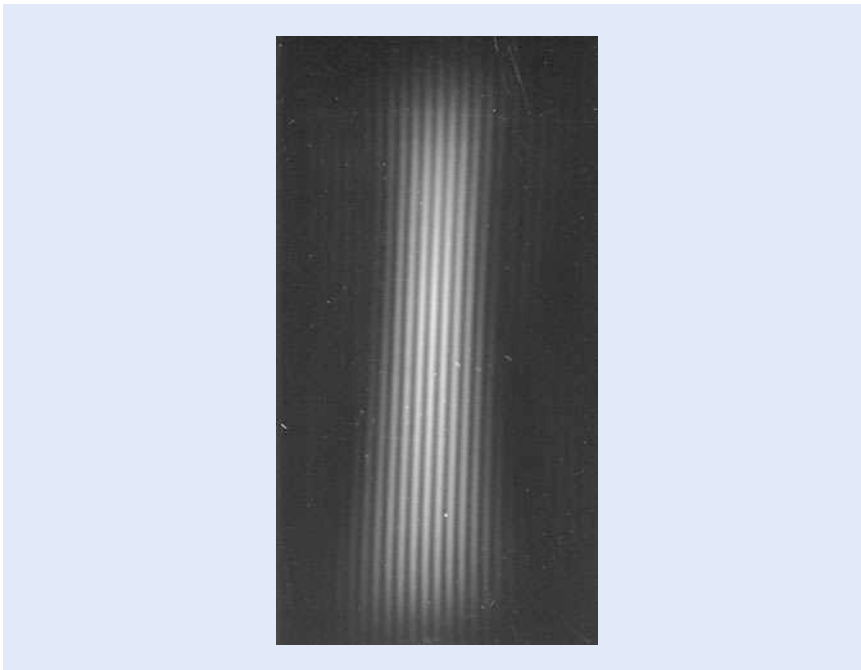
■ **Fig. 19.4** Distribution of electrons scattered from two slits in a screen. [Taken from Jönsson C (1961) Elektroneninterferenzen an mehreren künstlich hergestellten Feinspalten. Z f Phys 161:454–474, for an English translation see: Jönsson C (1974) Electron Diffraction at Multiple Slits. Am J Phys 42(1), 4–11]

■ Figure 19.3 depicts the intensity pattern of neutrons in the far field of a mechanical double-slit which displays interference fringes. This effect is quite remarkable when we recognize the count rates on the vertical axis. At a maximum of the fringe we find approximately 4000 counts per 125 min. This rate corresponds to two neutrons going through the apparatus per second. Since the velocity of the neutrons was 200 m/sec there was never more than one neutron in the apparatus.

Similar experiments have been performed earlier by Claus Jönsson in the group of Gottfried Möllenstedt at the Universität Tübingen with electrons. ■ Figure 19.4

Fig. 19.5 Double-slit interference pattern (*right*) of fullerene molecules (*left*) which are regular structures of 60 carbon atoms in the shape of a soccer ball. For this reason the fullerene molecule is sometimes jokingly referred to as soccerballium. When individual molecules are sent one at a time through a double-slit the pattern found on a screen in the far field shows clear interference fringes. [Taken from Arndt M et al. (1999) Wave particle duality of C_{60} molecules. Nature 401:680–682]

19

shows the count rate of a double-slit experiment with electrons and the interference fringes are clearly visible.

It is interesting to note that in September 2002 the journal *Physics World* reported a poll concerning the top ten most beautiful experiments in physics. The Jönsson experiment was the number one.

This phenomenon of matter–wave interference is not limited to neutrons or electrons. Even bigger objects such as the fullerene molecule C_{60} exhibit an interference pattern, as shown in ◘ Fig. 19.5.

Slits in Time

A rather intriguing version of a double-slit experiment with electrons was carried out by the group of the late Herbert Walther using ultra-short laser pulses. Here the interference appears in the time rather than the space domain.

Light pulses in the femto-second regime consist of a few optical cycles and can ionize single electrons in atoms, as shown in ◘ Fig. 19.6. However, this process only occurs when the associated electric field is above a threshold. Since the pulses are short the intensity necessary for ionization exists only during one or two time periods with an extension of an atto-second. By shifting the envelope of the pulse relative to the oscillation we can control the time window of ionization and create in this way a single- or a double-slit type of excitation of the atom.

Indeed, the double-slit situation appears when the sub-cycle pulse contains *two* narrow time windows in which the atom can be ionized. Therefore, the electron which reaches the detector with a well-defined momentum results from field ionization either in the first, or in the second time window. As long as we cannot decide in principle in which one the electron was born the two ionization paths in time must interfere. Since only a single electron is ionized in each event the interference takes place on the level of individual particles. Moreover, it manifests itself in the energy spectrum of the electron shown in the bottom of ◘ Fig. 19.6. If there is only one time window for ionization representing a single-slit situation there is no interference.
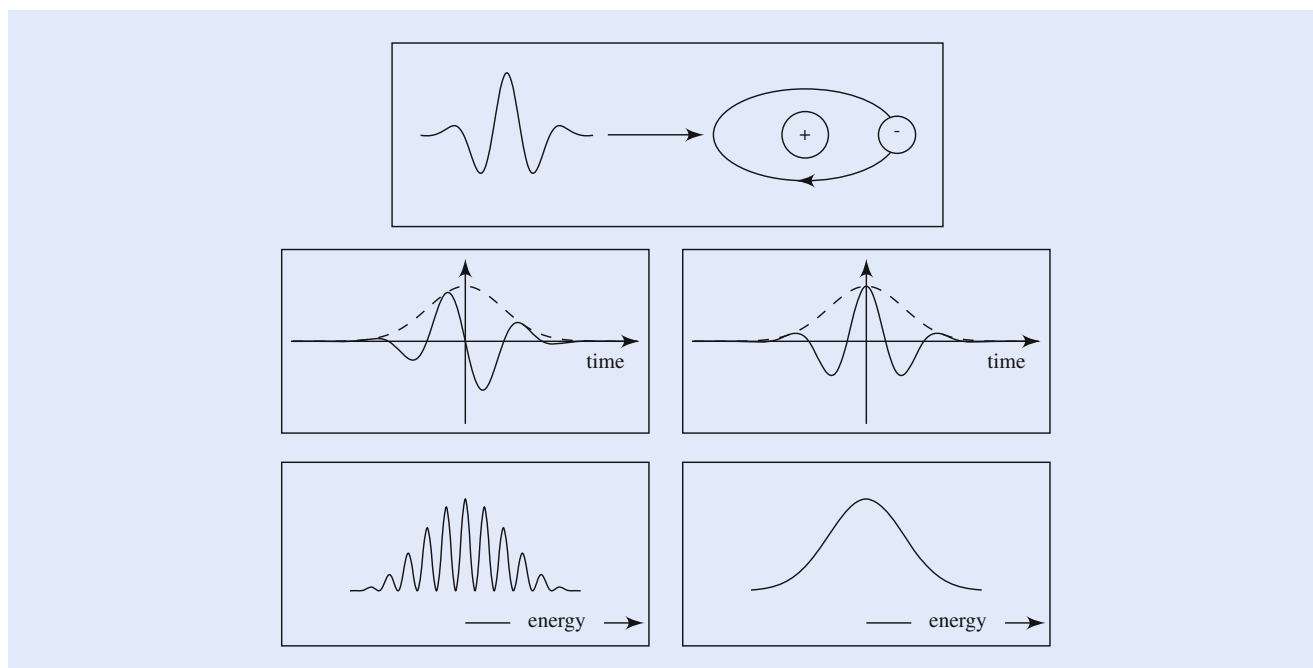
🔹 **Fig. 19.6**   Double-slit experiment in time. A sub-cycle laser pulse ionizes an electron in an atom (*top*). Depending on the phase of the oscillation relative to the envelope (*middle*) we either have two maxima (*left*) or a single dominant maximum (*right*) that can ionize the electron. In case of two maxima we cannot identify which one ionized the electron. As a result, the ionization current (*bottom*) displays oscillations (*left*) as a function of energy due to the indistinguishable excitation paths. In contrast, no interference arises for a single intensity maximum (*right*). [After Lindner F et al. (2005) Attosecond double-slit experiment. Phys Rev Lett 95:040401]

19.3.3 The Mystery of the Double-Slit Experiment

The physics of the double-slit experiment has occupied physicists since the early days of quantum mechanics. The rather paradoxical phenomenon of an interference pattern for particles can be expressed most vividly by the following situation borrowed from a wild-west movie.

Imagine a person shooting bullets towards a screen with two slits. Due to the wave nature of the bullets the interference pattern in the far field of the slits enjoys positions where no bullet will ever hit. Hence, a person standing behind the screen at one of these zeros of the fringes is safe. However, if one of the holes gets closed the interference and, hence, the zeros cease to exist. As a consequence, the next bullet might kill the person.

This discussion of the double-slit experiment also brings out most clearly the importance of the single event emphasized already in 🔹 Sect. 19.2.3 as a building block of our conception of the microscopic world. Indeed, we send individual particles through the apparatus, one at a time. Each particle is detected after it has passed the slits and will either hit the screen at this position or at another position. After we have sent many such particles through the apparatus we have a histogram, that is, a number of counts at every position on the screen which will not be uniform but will show oscillations. From individual counts we have built up a continuous distribution.

19.4 Complementary Views of an Uncertain World

In the present section we lay the ground work for the discussion of the two-photon double-slit experiment presented in ◘ Sect. 19.6 by briefly summarizing pioneering articles related to the principle of complementarity. We admit that this selection is rather biased and mainly motivated by our own considerations. We start from the seminal paper by W. Heisenberg introducing the uncertainty principle and then turn to N. Bohr's introduction of complementarity culminating in the concept of a delayed-choice experiment and the game of twenty questions in its surprise version.

19.4.1 The Uncertainty Principle

It was during a stay at Copenhagen in the institute of N. Bohr in 1927 that W. Heisenberg concluded that the discreteness of quantum mechanics and, in particular, the non-existence of a continuous trajectory of a particle make it impossible to determine simultaneously with arbitrary accuracy its position and its momentum. We quote from his seminal paper in which he considers the use of the Compton effect, that is, the scattering of a photon from an electron to determine its position and momentum:

» "At the instant at which the position of the electron is known, its momentum therefore can be known up to magnitudes which correspond to that discontinuous change. Thus, the more precisely the position is determined the less precisely the momentum is known, and conversely. In this circumstance we see a direct physical interpretation of the equation $\mathbf{pq} - \mathbf{qp} = -i\hbar$. Let q_1 be the precision with which the value q is known (q_1 is, say, the mean error of q), therefore here the wavelength of the light. Let p_1 be the precision with which the value p is determinable: that is here the discontinuous change of p in the Compton effect. Then, according to the elementary laws of the Compton effect p_1 and q_1 stand in the relation

$$p_1 q_1 \sim h. \tag{19.1}$$

That this relation (19.1) is a straight-forward mathematical consequence of the rule $pq - qp = -i\hbar$ will be shown below.[1]"

It is interesting to note that in this article W. Heisenberg does not use the notation Δq and Δp for the uncertainties but rather q_1 and p_1. Moreover, it is also amusing that he applies Schrödinger wave functions rather than matrices to

---

1 "In dem Moment, in dem der Ort des Elektrons bekannt ist, kann daher sein Impuls nur bis auf Größen, die jener unstetigen Änderung entsprechen, bekannt sein; also je genauer der Ort bestimmt ist, desto ungenauer ist der Impuls bekannt und umgekehrt; hierin erblicken wir eine direkte anschauliche Erläuterung der Relation $\mathbf{pq} - \mathbf{qp} = \frac{h}{2\pi i}$. Sei q_1 die Genauigkeit, mit der der Wert q bekannt ist (q_1 ist etwa der mittlere Fehler von q), also hier die Wellenlänge des Lichtes, p_1 die Genauigkeit, mit der der Wert p bestimmbar ist, also hier die unstetige Änderung von p beim Comptoneffekt, so stehen nach elementaren Formeln des Comptoneffekts p_1 und q_1 in der Beziehung

$$p_1 q_1 \sim h. \tag{19.1}$$

Daß diese Beziehung (19.1) in direkter mathematischer Verbindung mit der Vertauschungsrelation $pq - qp = \frac{h}{2\pi i}$ steht, wird später gezeigt werden."

illustrate the consequences of the uncertainty principle. The only remnants of matrix mechanics are the non-commuting operators **q** and **p**.

Heisenberg submitted his manuscript during a skiing vacation of N. Bohr who upon his return pointed out various mistakes and brought to light a deeper concept. As a result, W. Heisenberg felt obliged to include the following note added in proof:

» "After the conclusion of the foregoing paper, more recent investigations of Bohr have led to a point of view which permits an essential deepening and sharpening of the analysis of quantum-mechanical correlations attempted in this work. In this connection Bohr has brought to my attention that I have overlooked essential points in the course of several discussions in this paper. Above all, the uncertainty in our observation does not arise exclusively from the occurrence of discontinuities, but is tied directly to the demand that we ascribe equal validity to the quite different experiments which show up in the corpuscular theory on one hand, and in the wave theory on the other hand.[2]"

According to N. Bohr the uncertainty principle does not arise from the discontinuities but from the choice of the wave versus particle description demanded by the specific experimental setup. Hence, he supports the idea of an uncertainty relation but identifies a different origin of it.

19.4.2 The Birth of Complementarity

The article by N. Bohr summarizing his point of view appeared a year later, that is in 1928, with the title "The quantum postulate and the recent development of atomic theory." It was based on a lecture he gave on September 16, 1927 in Como at the International Congress of Physics in commemoration of the centenary of the death of Alessandro Volta. The reason for this delay originated from an unusual twist of events associated with his original manuscript on complementarity, his passport, and his train to Como.[3]

In his typical style N. Bohr draws attention to the fundamental difference between the classical and the quantum world when he states:

» "The very nature of the quantum theory thus forces us to regard the space-time coordination and the claim of causality, the union of which characterises the classical theories, as complementary but exclusive features of the description, symbolising the idealisation of observation and definition respectively."

Here the words "complementary" and "exclusive" enter the stage of physics for the first time.

---

2   "Nach Abschluß der vorliegenden Arbeit haben neuere Untersuchungen von Bohr zu Gesichtspunkten geführt, die eine wesentliche Vertiefung und Verfeinerung der in dieser Arbeit versuchten Analyse der quantenmechanischen Zusammenhänge zulassen. In diesem Zusammenhang hat mich Bohr darauf aufmerksam gemacht, daß ich in einigen Diskussionen dieser Arbeit wesentliche Punkte übersehen hatte. Vor allem beruht die Unsicherheit in der Beobachtung nicht ausschließlich auf dem Vorkommen von Diskontinuitäten, sondern hängt direkt zusammen mit der Forderung, den verschiedenen Erfahrungen gleichzeitig gerecht zu werden, die in der Korpuskulartheorie einerseits, der Wellentheorie andererseits zum Ausdruck kommen."

3   For the details of this amusing story see the commentary by Leon Rosenfeld on page 85 of the book by J. A. Wheeler and W. H. Zurek listed in Further Reading.

Moreover, N. Bohr has a clear picture how the observer changes the microscopic world by the intrusion necessary for his measurement. Indeed, he writes:

» "...the measurement of the positional coordinates of a particle is accompanied not only by a finite change in the dynamical variables, but also the fixation of its position means a complete rupture in the causal description of its dynamical behaviour, while the determination of its momentum always implies a gap in the knowledge of its spatial propagation. Just this situation brings out most strikingly the complementary character of the description of atomic phenomena which appears as an inevitable consequence of the contrast between the quantum postulate and the distinction between object and agency of measurement, inherent in our very idea of observation."

Obviously N. Bohr associates with the act of the measurement physical effects on the system to be measured. We shall return to this aspect in ▣ Sect. 19.5.
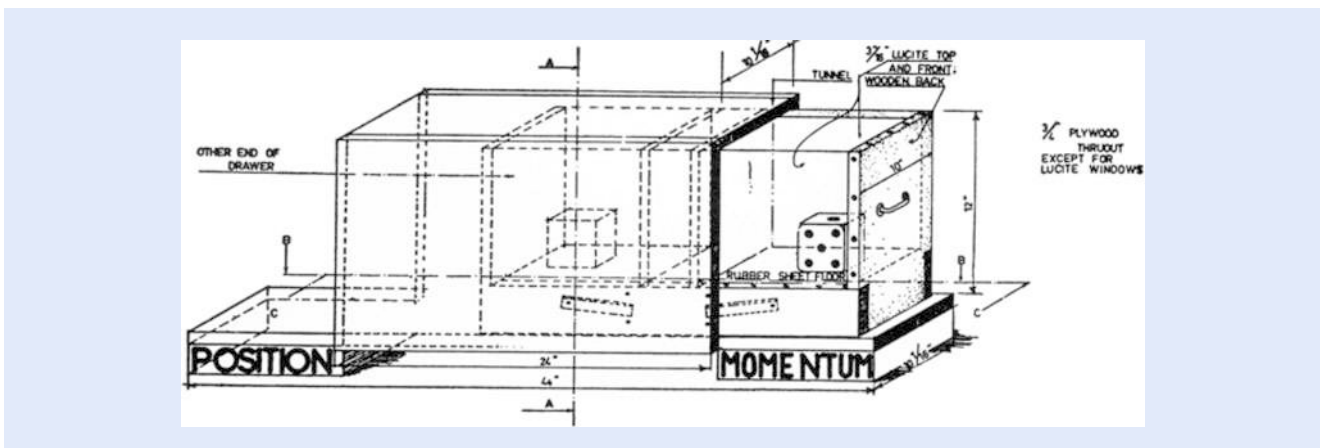
19.4.3 A Mechanical Model of Complementarity?

In 1939 at the world exhibition in New York the University of Copenhagen presented a mechanical model illustrating the principle of complementarity. ▣ Figure 19.7 shows a sketch of this device originating from J. A. Wheeler.

An unusual filing cabinet contains a drawer which can be pulled out in the front as well as in the back and which is divided into two compartments each of which contains a die. The task is to read the number shown on the top of *both* dice.

However, there is a slight problem. We cannot observe both dice simultaneously. When we pull the drawer to the front we can see only the die in the front compartment. When we push the drawer through to the back we can observe only the one in the back.

Of course, we could first copy down the number on the top of the first die and then pull the drawer out on the other side to have a look at the other die. However, a devilish device hidden underneath the floor of the drawer, such as a little hammer, is set into action as the drawer slides through the cabinet. Due to the kick imparted onto the floor the die whose number has just been recorded gets knocked over.

▣ **Fig. 19.7** Mechanical model of the principle of complementarity designed by the University of Copenhagen. Reading the numbers on the top faces of the dice stored in the two different compartments of a drawer in a filing cabinet corresponds to measuring two conjugate variables such as position and momentum, or path and interference. Unfortunately, this mechanical model misses the central lesson of quantum mechanics: There is no number on the dice until we make an observation. [Taken from Wheeler JA (1994) At Home in the Universe. AIP Press]

Although we can now record the number shown on the die confined to the back compartment it does not even make sense to do so. In our attempt to obtain information about the back die we have lost the information about the front one.

Unfortunately, we face the same dilemma if we start from the back die and move the drawer forward. Again the hammer is set in action and makes the knowledge we have just obtained redundant.

Obviously, the top faces of the dice play the role of two complementary quantities, such as position and momentum, and on first sight this model seems to illustrate in an impressive way the principle of complementarity. However, it lacks the fundamental ingredient of quantum mechanics summarized by J. A. Wheeler in the pregnant phrase:

» "No elementary quantum phenomenon is a phenomenon until it is a recorded phenomenon, brought to a close by an irreversible act of amplification."

According to quantum mechanics it is not a meaningful question to ask: What would the numbers on the dice have been, if we measured them. They do not exist until they are observed.

19.4.4 No Existence Without Measurement

The unusual property of quantum mechanical observables such as position or momentum, or components of angular momentum to take on a definite value only after observation comes out most clearly in the delayed-choice experiment. When we inject a single particle, one at a time, into the upper entrance of the Mach–Zehnder interferometer shown in ◘ Fig. 19.8 we expect the particle to either go on
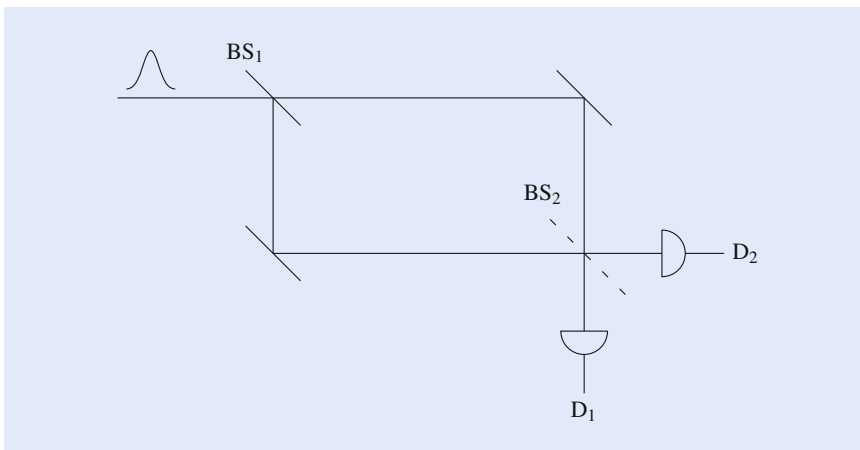
◘ **Fig. 19.8** Mach–Zehnder interferometer in a delayed-choice mode. A particle entering the interferometer in one of the two ports at the upper left beam splitter BS_1 either takes the high or the low road. When the beam splitter BS_2 at the lower right corner is inserted the two paths interfere at the two detectors D_1 and D_2. However, when we remove BS_2 we obtain "which-path" information. We can delay our decision concerning BS_2 till the particle hits it. Loosely speaking, at the moment of our decision we force the particle to retrace its tracks and change its history. Indeed, when BS_2 is part of the interferometer as the particle passes BS_1 and we remove BS_2 later we force the particle to alter its nature. Before our decision it was supposed to display interference properties, but afterwards it needs to provide us with path information. In this arrangement the particle was supposed to move along a *single* path rather than on *both* paths. This paradoxical situation can only be resolved by the assumption that the particle does not have any path whatsoever until we observe it, nor can it display any interference fringes until we measure them. In the language of J. A. Wheeler: "The particle is a great smoky dragon that is only sharp where it enters the interferometer and where it leaves the interferometer biting the detector"

the upper or the lower road. When we leave out the beam splitter at the lower right corner we can detect the path of the particle, because only one of the two detectors will respond.

However, when we do insert the beam splitter and have many particles, one at a time, pass through the interferometer we will find all particles in one of the exit ports provided we adjust the arm length appropriately. This behavior is a consequence of the wave nature of matter. Indeed, due to destructive interference no particles are in the other port.

So far we have only discussed another manifestation of the complementarity principle. However, we can now go one step further, and use our knowledge of the time at which the particle entered the interferometer. Moreover, if we know the initial velocity of the particle we can predict the time at which it will impinge on the second beam splitter. Within the period defined by the entrance of the particle into and its exit from the Mach–Zehnder interferometer we can now decide if we want to insert the second beam splitter or not. In this way we make a *delayed* choice between our ultimate observation of interference or "which-path."

We start from a situation where the second beam splitter is present when the particle enters the interferometer. In the language of the macroscopic world the particle has to display its interference nature and has to move on both paths.

This interpretation runs into problems when in the last moment before the particle hits the second beam splitter, we take it out. Now, we are asking for path information and the particle had to go on one path only. However, by that time the particle has almost reached the beam splitter. It can therefore not go back and retrace its tracks. Our procrastination in making a decision, that is, the delayed choice of interference versus "which-path" highlights the idea that in the microscopic world the properties of particles are not well-defined until they are observed.

We emphasize that many delayed-choice experiments with light and matter waves have been performed. They clearly show that the delay has no influence on the observation. We observe the features that we choose to observe. As J. A. Wheeler puts it:

» "The past has no existence, except as it is recorded in the presence."

19.4.5 Our Questions Create the Microscopic World

The delayed-choice experiment demonstrates in a striking way that observables of a quantum system have no existence until we observe them. Hence, our measurement has an enormous influence on how we view the microscopic world. In order to emphasize this aspect J. A. Wheeler has coined the phrase "participatory universe," whose meaning stands out most clearly in the game of twenty questions in its surprise version.

A group of friends sends one victim out of the room while his/her remaining colleagues agree on a word to be guessed. After the person has returned he/she is allowed to ask twenty questions. The answers must be given truthfully with "yes" or "no" and after this question-answer-period the person is confronted with the challenge to produce the word.

However, when once the turn came to J. A. Wheeler to be the victim he found upon reentering the room his friends with a grin on their faces. He knew something was up.

He started by asking: "Is it a cloud?" A quick response came: "No!" The second question: "Is it a car?" Now the answer took a little bit longer. His friends had to think about it and finally they answered: "No!" The more questions he asked the longer it took them to answer.

This hard work on their part was difficult to understand because a word had been agreed on and all that had to be done was to see if his guess was correct. Nevertheless, it took even longer as they approached the final trial answers.

Finally J. A. Wheeler had to make a decision. Challenging one of his opponents he put forward one final question: "Is it a bear?"

Again the challenged had to think for a long time before he eventually admitted: "Yes, you are right!" – Laughter broke out in the room.

How come his friends had to think at least as hard as he? The answer to this question originates from the fact that when he had left the room they had decided not to agree on a word at all. However, their individual answers would have to be consistent with each other and he would only win if his guess was consistent with the chain of their answers.

As a result of these new rules the game was as difficult for them as it was for him. No word existed in the room until it was challenged by the observer who became the "creator" of the word.

This game of twenty questions in its surprise version encapsulates the crucial point of quantum mechanics: The microscopic world does not exist until we observe it.

19.5 Physical Disturbance Versus Correlations

In our discussion of the principle of complementarity in ▣ Sect. 19.4.2 the second quote of N. Bohr shows that he associated with a measurement of the microscopic world a physical disturbance. His point of view stands out most clearly in the Bohr–Einstein discussion of the recoiling double-slit designed by Albert Einstein to obtain "which-path" information together with interference fringes. This dialogue started at the Solvay meeting of 1927 and continued for almost 30 years.

We dedicate this section to a brief introduction of this gedanken experiment which later has been analyzed by William Wootters and Wojciech Hubert Zurek using the formalism of modern quantum mechanics and, in particular, of joint measurements. We conclude by highlighting the key ingredients of the quantum eraser developed in various forms by Marlan Orvil Scully and coworkers.

19.5.1 Recoiling Double-Slit

With his friend the philosopher Harald Høffding, N. Bohr frequently discussed the double-slit experiment and Høffding asked: "Where can the particle be said to be?" Bohr answered in the familiar Hamlet way: "To be? To be? What does it mean 'to be'?"

What does it mean to talk about a particle going through the upper or lower slit, or through both slits if we do not make a measurement to prove our claim? But how can we make such a measurement?

The proposal of A. Einstein for such a measurement of "which-path" information *and* interference involves a *movable* rather than a *fixed* screen. By measuring the momentum transfer of the scattering particle on the slit, and the interference fringes in the far field A. Einstein argued that in principle we can observe simultaneously position and momentum with arbitrary accuracy. However, N. Bohr showed that this claim is not correct since the momentum transfer of the scattering particle wipes out the fringes. Measurements of this type are still limited by the uncertainty principle.

In 1979 W. Wootters and W. H. Zurek revisited this arrangement of a recoiling double-slit and demonstrated that the interference pattern is surprisingly sharp

even when the trajectories have been determined with a fairly high accuracy. In their analysis the entanglement between the center-of-mass motions of the scattering particle and the slit plays a crucial role.

19.5.2 **Quantum Eraser**

We emphasize that the mechanism for the destruction of the interference fringes discussed in the preceding subsection relies heavily on random phase disturbances. Indeed, the key argument is always the physical transfer of momentum which leads to an uncontrollable phase disturbance and wipes out the fringes. We now analyze two situations which show that this notion is not correct.

Double-Slit with Two Atoms

A new era in the analysis of the double-slit experiment started in 1982 with the proposal of the quantum eraser illustrated in ▫ Fig. 19.9. Here the two mechanical slits are replaced by two identical atoms excited by the incident radiation. Each atom decays with the emission of a photon which is detected in the far field. When
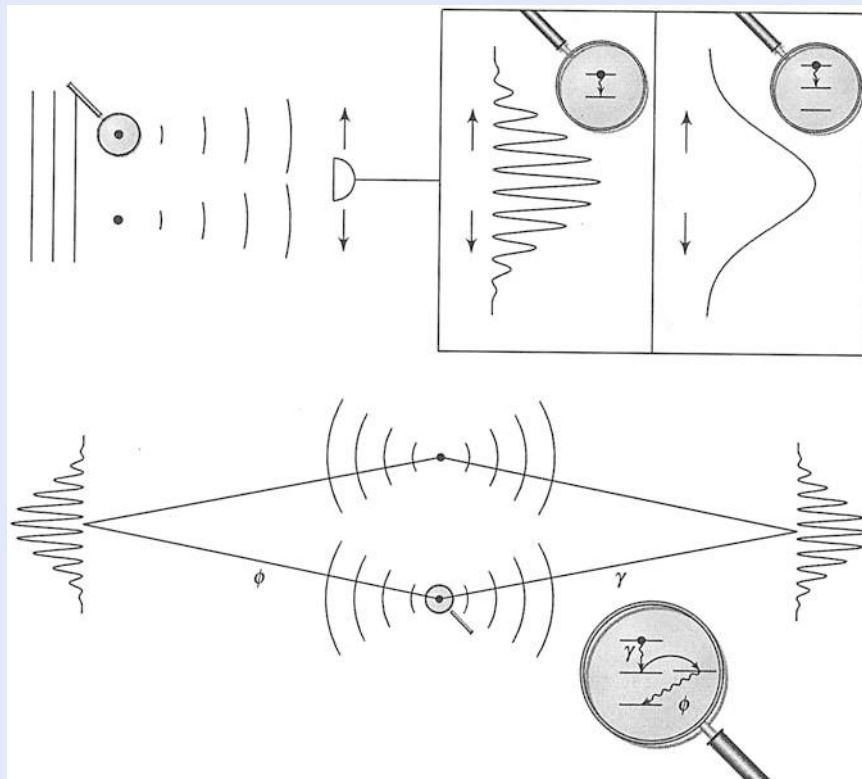
▫ **Fig. 19.9** Realization of a double-slit experiment based on the scattering of light from two atoms (*top*) and the concept of the quantum eraser (*bottom*). In this arrangement we replace the two slits in the screen by two atoms in the absence of any screen and scatter one light quantum from both atoms which are initially in the ground state. After the scattering event both atoms are again in their respective ground states provided we deal with *two*-level atoms. Since in this case it is impossible to tell which atom scattered the quantum we observe interference in the far field (*top-left*). For two *three*-level atoms one will always remain in the long-living intermediate state which provides us with "which-path" information and no interference occurs (*top right*). However, the fringes reemerge when instead we use two *four*-level atoms (*bottom*) together with a joint measurement between the two emitted quanta γ and ϕ. In this case we have erased the "which-path" information of the scattering since both atoms have again returned to their ground states. It is this process of erasing the "which-path" information which recreates the interference fringes

the radiation is so weak that only one of the two atoms gets excited at a time the paths of excitation can either interfere giving rise to fringes in the far field, or leave "which-path" information in the atom creating a smooth intensity pattern.

This decisive difference is dictated by the internal structure of the atoms. In order to bring this fact out most clearly we first consider two *two*-level atoms which are initially in their ground states. After the excitation by and subsequent emission of the photon both atoms are again in their ground states. As a result, we cannot tell from the final arrangement which atom has scattered the light and interference fringes occur.

However, when we use two *three*-level atoms with a long-living intermediate state the scattering path can be reconstructed from the final internal state. Indeed, the atom that has been excited and has reemitted the photon will be left in the middle state, whereas the atom that did not participate in the scattering process is still in the ground state. Due to the availability of "which-path" information no fringes appear in this situation.

We finally consider a *four*-level atom. After the emission of the photon γ and the decay to the intermediate state we pump into the fourth level which decays rapidly to the ground state by emitting a second photon ϕ. The intensity pattern of γ-photons measured in the far field now conditioned on the detection of ϕ-photons displays oscillations since we have erased the information about the path of the excitation.

The quantum eraser brings out clearly that the disappearance of the interference fringes is not related to uncontrollable disturbances of the phases of the atoms, or the field, but rather originates from the correlations established between the internal states of the atom and the field.

Double-Slit with Two Cavities

This shift of paradigm is also emphasized by the gedanken experiment shown in
▶ Fig. 19.10 combining the wave nature of matter, that is, atom optics with cavity
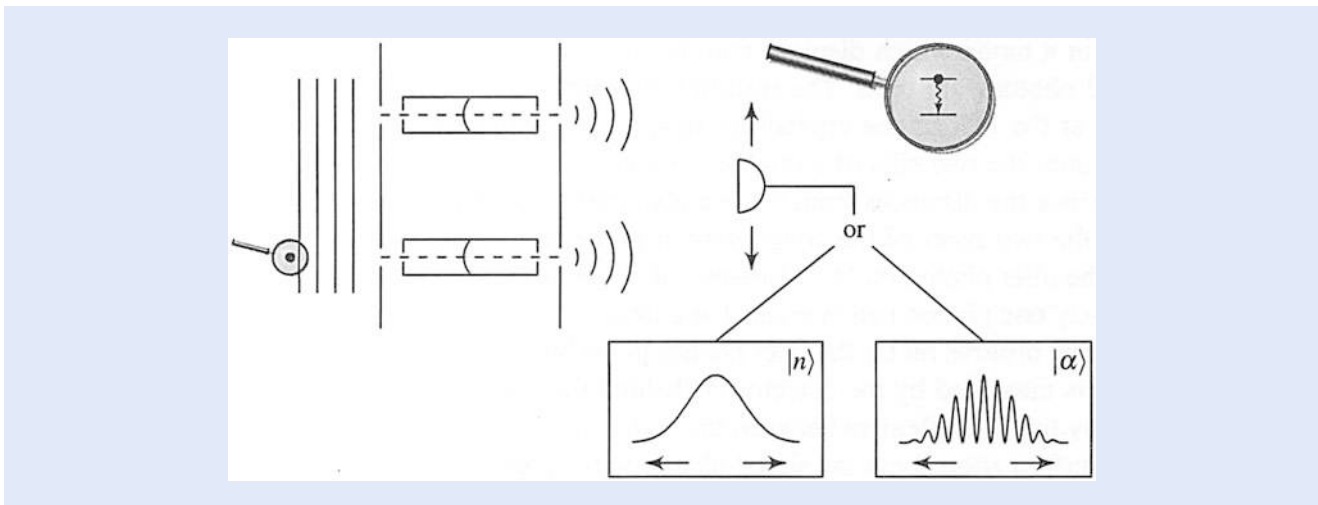
▶ **Fig. 19.10** "Which-path" information encoded in the quantum states of two *cavity fields*. A single two-level atom initially in its excited state and with a center-of-mass motion in a coherent superposition of passing the upper and the lower cavities interacts with the single-mode fields indicated here by their lowest modes in such a way as to deposit with certainty its internal excitation. When both fields are in a number state $|n\rangle$ one of them must change by one quantum, that is go from $|n\rangle$ to $|n+1\rangle$ and in this way create "which-path" information. As a consequence, no interference occurs in the center-of-mass motion in the far field of the second double-slit. However, when both fields are in a coherent state $|\alpha\rangle$ of large amplitude one quantum does not make a difference. In this case the path of the atom cannot be reconstructed and fringes emerge

quantum electrodynamics. Here a single two-level atom whose transverse center-of-mass motion is in a superposition of two locations, prepared, for example, by a double-slit, passes two high Q-cavities whose mode maxima are aligned with these slits. The atom is initially in the excited state and the experimental parameters are chosen such that the atom must deposit its internal excitation in one of the two cavities. Hence, the manifestation of interference in the transverse center-of-mass motion depends on the initial state of the cavities.

Indeed, when both are in a number state $|n\rangle$ the photon placed by the atom in one of the cavities increases the corresponding photon number n by one unit. We can reconstruct by this change the path the atom has taken, and since number states corresponding to different photon numbers such as $|n\rangle$ and $|n+1\rangle$ are orthogonal no interference fringes occur.

However, for a coherent state $|\alpha\rangle$ of large average number $\overline{n}$ of photons, that is $1 \ll \overline{n} \equiv |\alpha|^2$, this change by one photon is negligible and the two coherent states corresponding to $\overline{n}$ and $\overline{n}+1$ are *not* orthogonal. As a consequence, in this case the path is unknown to us and interference fringes emerge.

Orthogonality of field states instead of uncontrollable phase changes as the eraser of interferences—this statement serves as the one-sentence-summary of this version of "which-path" detectors. It is interesting to note that this insight has led to a lively controversy. Indeed, the group of the late Daniel F. Walls has repeatedly emphasized that there is still room for an interpretation in terms of phase disturbances. However, the seminal experiment based on atom interferometry by the group of Gerhard Rempe has tilted the scale towards the notion of orthogonality.

## 19.6   A Two-Photon Double-Slit Experiment

We now briefly highlight the key features of a recent experiment on wave-particle dualism using entangled photons performed in the group of Ralf Menzel. On first sight their results seem to indicate a break-down of the principle of complementarity. However, a closer analysis of the groups of atoms creating the registered photons leading to "which-path" information and interference reveals that they are different in the two arrangements. As a consequence, this experiment constitutes another impressive verification of this corner stone of quantum theory.

19.6.1   A Violation of the Principle of Complementarity?

◘ Figure 19.11 highlights the essential components of the experiment by Menzel et al. Here, a laser in a mode which displays two distinct maxima pumps a crystal and two entangled photons are born. The resulting two pairs consisting of signal and idler photons at the exit of the crystal are imaged with the help of a polarizing beam splitter onto the two slits of a double-slit and onto a single-photon counting detector D_1. Since the distances from the beam splitter to D_1 and to the double-slit are identical, the two spots of the signal photon in the two slits correspond to the two spots of the idler photon on D_1. Moreover, the low intensity of the pump beam ensures that only one photon pair is created at a time.

When we now observe on D_1 the idler photon in the left (right) intensity spot the signal photon is measured by the detector D_2 behind the upper (lower) slit. In this way we employ the entanglement between the signal and the idler photon to obtain "which-path" information about the signal photon without ever touching it.

However, we can also observe interference fringes while at the same time we gain "which-path" information when we scan D_2 in the *far-field* region of the
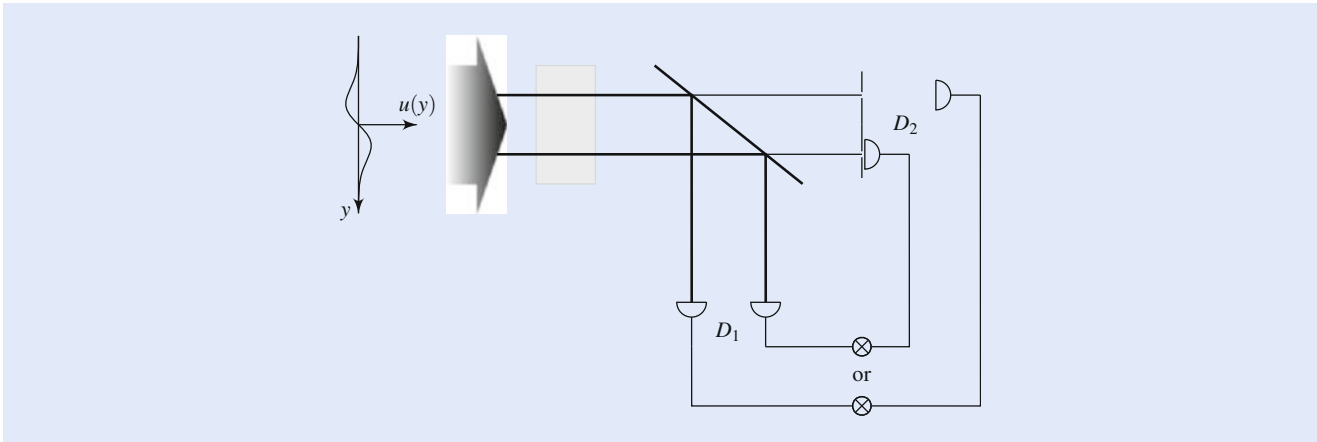
Fig. 19.11 Essential ingredients of the Menzel et al. experiment aimed at observing interference while obtaining "which-path" information in a double-slit arrangement. We pump a nonlinear crystal (*gray area*) with a light beam of transverse mode function $u = u(y)$ with two distinct intensity maxima and a node between them. The correlated photon pair consisting of the signal and the idler photon and emerging from the crystal in two distinct spots is divided by a polarizing beam splitter. The signal photon passes a double-slit and is detected on detector D_2 which is either in the near field, or the far field. The widths and separation of the two slits are adjusted to match the two intensity spots on the end of the crystal. The idler photon is measured on the detector D_1 which is arranged in a way as to ensure that the distances from the beam splitter to D_1 are identical to the ones to the two slits. In this way the idler photon which is entangled with the signal photon allows us to obtain "which-path" information about the latter without ever touching it. Indeed, D_1 can be positioned on the left or right spot as indicated in the figure. We always perform measurements on the signal photon in coincidence with the idler photon

double-slit along the vertical direction and measure in coincidence with the idler photons detected on D_1 in the left spot. In this case we still find interference fringes.

The appearance of fringes in coincidence is surprising since the idler photon provides us with "which-path"-information about the signal photon which still displays interference. In contrast, the principle of complementarity seems to suggest no interference.

19.6.2 Different Atoms Yield "Which-Path" or Interference

We now discuss these rather counter-intuitive results of the Menzel et al. experiment using the model summarized in ◘ Fig. 19.12. In particular, we explain the appearance of an interference pattern in the far field in coincidence with the "which-path" information contained in the correlations of the near field. However, we emphasize that there is no violation of the principle of complementarity.

In order to understand the origin of the observed near-field coincidence measurements, and in particular, the perfect correlations between the signal and the idler photon created by the subsequent emissions from a three-level atom we assume for the sake of simplicity that both detectors are in the plane parallel to the screen and consider a vanishing time delay between the clicks on D_1 and D_2. Hence, the atoms emitting the two photons must be in the plane located between the two detectors. When both detectors are in the same intensity maximum of the double-hump structure this plane is at the center of the maximum as well. As a result, we find simultaneous clicks in the two detectors.

However, when they are in different maxima the only atoms satisfying the coincidence condition are half-in-between the detectors, that is, at the node of the mode as indicated in the top of ◘ Fig. 19.12. Therefore, the only atoms that could cause the appropriate clicks are not excited. From these perfect correlations between idler and signal photon on the left detector-upper slit or right detector-
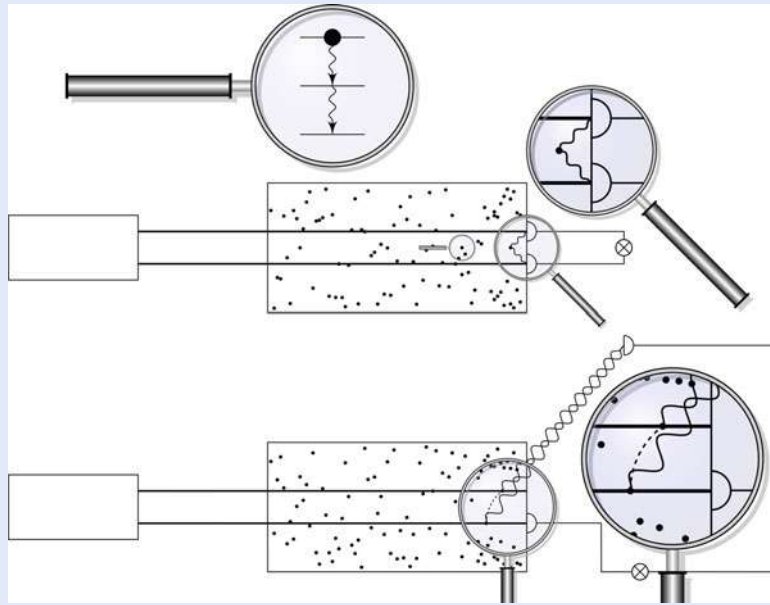
Fig. 19.12 Elementary model to explain the correlations observed in the Menzel et al. experiment. We consider a frozen gas of three-level atoms which are excited by a single photon in a mode consisting of a coherent superposition of two light pencils. When the two detectors are both located at the end of the cell (*near field*) two clicks at the same time registering the signal and idler photon can only result from atoms that are on the symmetry line half-way between the two pencils (*top*). Due to the special mode function these atoms are not excited and the probability for this event to occur vanishes. However, when one detector is in the near field and the other in the far field (*bottom*) the idler photon triggers the detector in the near field. Due to the double-hump structure of the mode function, this click could have come from any atom excited by the two pencils with identical separations from the detector (*dashed line*). As a result, we observe interference fringes in the far field, in complete accordance with the scattering situation of two-level atoms discussed in ◘ Fig. 19.9

lower slit we deduce our "which-path" information. We get this knowledge about the signal photon from the idler photon.

Next we consider the measurements which involve the detector D_1 in the near field and D_2 in the far field. Here the idler photon which triggers D_1 was emitted by one atom anywhere in the light pencils. However, we cannot distinguish between two atoms with identical separations from D_1 as indicated by the bottom of ◘ Fig. 19.12. As a result, we find the interference of the signal photon emitted from two indistinguishable atoms very much in the spirit of ◘ Fig. 19.9.

19.6.3 No Contradiction But Confirmation

We are now in a position to summarize our main results. Three features were crucial for obtaining the on first-sight surprising results of the Menzel et al. experiment: (1) A special mode function of the electromagnetic field consisting of a coherent superposition of two maxima, (2) the entanglement between two photons, and (3) a joint measurement of both of them.

When both detectors are in the near field there are no counts at the same time since the only atoms that could have caused such a result sit at a node of the field and cannot be excited. When one detector is in the near field and one in the far field two indistinguishable atoms lead to an interference signal in the joint count statistics. Since the two experiments correspond to two different arrangements and different atoms are involved there is no contradiction to the principle of complementarity.

19.7 More Questions than Ever

At the center of our essay was the double-slit experiment symbolizing the wave-particle dualism of quantum theory. We have provided a historical perspective and at the end have addressed a rather counter-intuitive experiment. Although there exists a straight-forward explanation many questions associated with this outcome and avenues for further research offer themselves. Here we only allude to one.

Is it possible to perform the Menzel et al. experiment with atoms and would we find the same result? On first sight the answer is "yes" since we can create entangled atoms in complete analogy to photons. But is the measurement process the same?

This question is closely related to the discomfort of W. E. Lamb with the concept of the photon. Whereas in the case of the atoms a description based on single-particle quantum mechanics suffices the photon experiment requires quantum field theory. In the latter the measurement does not reduce the mode function but annihilates the quantum in the mode and leaves the mode intact. However, in the case of the atom the measurement leads to a localization of the particle due to the reduction of the wave packet.

We have started our article by a quote from J. A. Wheeler. Therefore, it is appropriate to close it with the following summary from the same lecture which emphasizes again the influence of the measurement and the role of the observer:

» "Are billions upon billions of acts of observer—participancy the foundations of everything? We are about as far as we can be today from knowing enough about the deeper machinery of the universe to answer this question. Increasing knowledge about detail has brought an increasing ignorance about plan. The very fact that we can ask such a strange question shows how uncertain we are about the deeper foundations of the quantum and its ultimate implications."

Although we have made impressive progress in our understanding of the photon since Wheeler's lecture we are still far away from being able to say that we have discovered "'the plan" or with the words of G. Stein that we "can't understand what made it look strange in the first place."

Further Reading

1. For the historical aspects mentioned in this article we refer to Crease RP (2003) The prism and the pendulum: The ten most beautiful experiments in science. Random House, New York

   Wheeler JA, Zurek WH (1984) Quantum Theory and Measurement. Princeton University Press, Princeton

   Scully RJ, Scully MO (2010) The Demon and the Quantum. Wiley, Berlin
2. The quotes of J. A. Wheeler can be found in Wheeler JA (1994) At Home in the Universe. AIP Press
3. A discussion of the concept of the photon with many historical remarks is presented in Mack H, Schleich WP (2003) A photon viewed from Wigner phase space. Optics and Photonics News Trends 3:29–35
4. An excellent introduction to quantum optics emphasizing the measurement process is provided by Scully MO, Zubairy MS (1997) Quantum Optics. Cambridge University Press, Cambridge
5. For an in-depth discussion of the fascinating facettes of the quantum eraser proposed by Scully MO, Drühl K (1982) Quantum eraser: A proposed photon correlation experiment concerning observation and "delayed choice" in quantum mechanics. Phys Rev A 25:2208–2213

   and further developed by Scully MO, Englert BG, Walther H (1991) Quantum optical tests of complementarity. Nature 351:111–116

   see for example Aharonov Y, Zubairy MS (2005) Time and the quantum: erasing the past and impacting the future. Science 307(5711):875–879
6. An up-to-date review on delayed-choice experiments can be found in Ma X, Kofler J, Zeilinger, A (2016) Delayed-choice gedanken experiments and their realizations. Rev Mod Phys 88,015005
7. The two-photon version of the double-slit experiment is discussed in Menzel R, Puhlmann D, Heuer A, Schleich WP (2012) Wave-particle dualism and complementarity unraveled by a different mode. Proc Natl Acad Sci 109(24):9314–9319; Bolduc E, Leach J, Miatto FM, Leuchs G, Boyd RW (2014) Fair sampling perspective on an apparent violation of duality. Proc Natl Acad Sci 111(34):12337–12341

19