Chapter 15
Coherence, Dipole Radiation and Laser

Abstract This chapter is focused mainly on coherence, a vital concept if one wants to go beyond a rudimentary understanding of waves; the notion is rooted in the statistical properties of a wave. We also elucidate how mixing of real physical waves leads to unexpected relationships: for example, it would be disastrous if the members of a choir managed to sing in perfect tune. This is related to the difference between light from a lamp compared to laser light. A computer program is provided for numerical explorations of both temporal and spatial coherence. The discussion of coherence is followed by a conceptual explanation of how electric charges in motion may lead to radiation of electromagnetic waves, and the radiation diagram for a dipole antenna is presented. The chapter concludes with a brief description of the basic principles for generating laser light.

15.1 Coherence, a Qualitative Approach

"Coherence" is a very useful concept when we want to describe how regular a wave is. In modern physics, it is clear that waves may interact very differently in various systems, depending on the statistical features of the waves. It is insufficient to characterize a wave only with amplitude, frequency, wavelength and the spatial volume where the wave is found. Figure 15.1 tries to illustrate this point using pictures of various surface waves on water.

The dictionary meaning of the word *coherence*, "the quality of being logically consistent", may be inferred from its etymology (*co* together + *hæreo* to stick). In physics, however, coherence is defined differently, and they are comparatively recent concepts. They describe important statistical properties. Wikipedia states that "two wave sources are perfectly coherent if they have *a constant phase difference* and *the same frequency*, and *the same waveform*". Additionally, it says: "Coherence describes all properties of the *correlation* between physical quantities *of a single wave, or between several waves or wave packets*" (emphasis added).

A. I. Vistnes, *Physics of Oscillations and Waves*, Undergraduate Texts in Physics,
https://doi.org/10.1007/978-3-319-72314-3_15

Fig. 15.1 Three examples of surface waves on water, illustrating the need for a statistical description of waves in addition to the parameters we have used so far

Thus, coherence is a term related to waves and other time-varying signals. When we say that waves at two points A and B in space are coherent, we mean that there is a certain relationship between the waves passing at points A and B at any time.

We distinguish between temporal and spatial coherence. For **temporal coherence** (or longitudinal coherence), we consider the wave at two points A and B which lie along the direction of the wave motion (see Fig. 15.2). We then check if there is a sustained definite relationship between the wave at point A and the wave at another point B which the same part of the wave passed a little earlier.

For **spatial coherence**, we compare the wave at a location A with the wave at location B, the two points being adjacent to each other, but the direction from A to B is perpendicular to the direction of the wave motion.

For real waves, there is often a high degree of correlation between the waves at A and B if the two points are very close to each other (less than the smallest wavelength found in the wave). On the other hand, it is always true for real waves that the correlation between the waves at A and B becomes exceedingly poor if the distance between A and B is made big enough.

If the distance between A and B must be less than a few wavelengths to get a high degree of correlation, we say that the wave is *incoherent*. If, on the other hand, we can find a high degree of *sustained definite relationship* between the wave at point A and the wave at point B even when the distance between A and B is very many

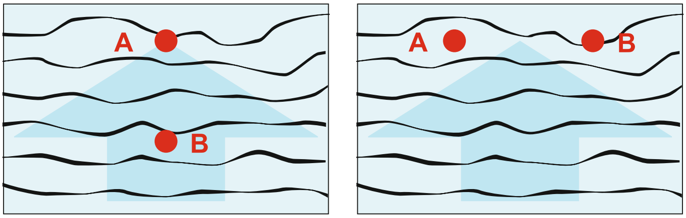

Fig. 15.2 Temporal (also longitudinal) and spatial (also transverse) coherence tell us something about the regularity in waves in the direction along which the wave moves or in the perpendicular direction, respectively. The wave is believed to be two dimensional in this case (e.g. surface waves on water). The black stripes indicate wave peaks, and their thickness indicates the amplitude of the wave at the current location. The wave in this figure is rather irregular

wavelengths, we call the wave *coherent*. However, there is a continuous transition between incoherence and coherence.

The term *coherence length* will be used for the largest distance between A and B for which significant correlation can still be found. We can specify both a temporal and a spatial coherence length. For temporal coherence, we can also operate with *coherence time* which is the time the wave uses in traversing the temporal coherence length.

Since there is always a certain degree of variability and unpredictability in real waves, and randomness can be described by statistics, the degree of coherence may be quantified statistically.

15.1.1 When Is Coherence Important?

Coherence is always important when two or more waves superimpose on each other. Thus, coherence is an important condition for interference. In deriving the intensity distribution at a double slit, we assumed that the wave has the same amplitude at all times in the two slit openings. It is equivalent to saying that the spatial coherence length must be at least as large as the distance between the two slits. In order to have several interference fringes outside the central one, it is also necessary that the temporal coherence length is at least several wavelengths since we add, in that case, one wave with a time-shifted part of the same wave.

An implicit consequence of a long temporal coherence length is that the wave must last for at least as long as the coherence time. This implies that there is a close relationship between temporal coherence length and width of frequency (or wavelength) distributions. The relationship is given by the time-bandwidth product discussed in Chaps. 3 and 5. Thus, narrow linewidth light emitted from atoms will have a long temporal coherence length, while light with a broad frequency distribution will have a short temporal coherence length. Light from the sun has a temporal coherence length of only a few wavelengths. In comparison, a laser may have a temporal coherence length of the order a million wavelengths.

If a source of waves is very small ("point-like", of the order one wavelength or less), the source will radiate waves with circular wavefronts. If the medium is isotropic, there will be more or less perfect circular wavefronts, at least within a sector of radiation. For such a system, the spatial coherence length can be very large, even if the temporal coherence length is short.

Even for a very extended object with many independent sources of radiation, the spatial coherence length can be large if the distance to the object is very large compared to the size of the object. We will return to the point later in this chapter when we discuss a famous experiment performed by Hanbury Brown and Twiss more than 60 years ago. We also often use a so-called pinhole in optics in order to increase spatial coherence length of light, as will be discussed later in this chapter.

15.1.2 Mathematical/Statistical Treatment of Coherence

Waves can be irregular in many different ways. Several methods are worked out to characterize the irregularities. We will limit ourselves to one of the simplest methods based on the calculation of a first-order correlation function.

If we want to characterize a wave, we can either record the amplitude at one point in space as a function of time, or we can record the amplitude at one instant as a function of position in space. In both cases, we acquire real measurements as a row of numbers, an array. If we make measurements at two points, as is indicated in Fig. 15.2, we end up with two sets of numbers, two signals.

There are many ways to compare two such arrays, but for analysing coherence physicists chose many years ago a strategy similar to Fourier transformation. In Fourier transformation, we actually do a correlation analysis between the signal we are studying and a perfect mathematical harmonic function with a given frequency (and we change the frequency to get the entire frequency range). Fourier transformation thus involves calculating the inner product between the signal we analyse and a perfect harmonic function.

When analysing correlation between signals (waves), we also calculate the inner product, but now between the two signals we compare! In a manner of speaking, we use one of the signals as a reference to check how much it resembles the other signal.

If the signal recorded at point A is called $f(t)$ and the signal at B is called $g(t)$, our predecessors have chosen to calculate the correlation between f and g as follows:

$$C = \frac{\int f(t) g(t)\, \mathrm{d}t}{\sqrt{\int f^2(t)\, \mathrm{d}t \ \int g^2(t)\, \mathrm{d}t}} \,. \tag{15.1}$$

The integrations extend over an arbitrary interval of time. If the signals are stationary (in the sense that their statistical character does not change over time) and also ergodic (in which case the statistical information derived by analysing many independent signals will be equivalent to the statistical information derived by following only one sufficiently long-lasting signal), C will approach a well-defined value when the integration time increases.

Defined this way, correlation is simply a number. There will be no function before we find some means of systematically changing how f and g are measured or generated.

15.1.2.1 Autocorrelation

In the left part of Fig. 15.2, we chose to compare waves at two points A and B situated along the direction in which the wave moves. Practical measurements could be carried out if we chose to analyse sound, because very small microphones are available that perturb the wave so little that the signal at A would not be affected by the presence of the microphone at point B.

In other contexts, it is not feasible to place a sensor at B without interfering with the wave that reaches point A.

This is one of the reasons why we often choose, when we analyse temporal coherence of a wave, to use only one detector, for example at point B, and no detector at point A. We assume that the shape of the overall wave pattern (as given by the black lines in Fig. 15.2) does not change much during the time it takes for the wave to move from B to A; the signal in point A will be approximately the same as at point B, only time shifted (because of the wave pattern moving at a given speed).

In such cases, we will have

$$f(t) \approx g(t + \tau)$$

where τ is the time used by the wave (wave pattern) to move from B to A.

The temporal correlation between the wave at A and at B is then given by:

$$C = \frac{\int g(t)g(t+\tau)\,\mathrm{d}t}{\sqrt{\int g^2(t)\,\mathrm{d}t \ \int g^2(t+\tau)\,\mathrm{d}t}} \; . \tag{15.2}$$

If the statistical properties do not change over time, we say that the wave is *stationary*. For such waves

$$\int g^2(t)\,\mathrm{d}t \;\approx\; \int g^2(t+\tau)\,\mathrm{d}t \; .$$

With this viewpoint, we are able to calculate correlations for many different distances between points A and B. In practice, we do this by changing the time shift τ in Eq. (15.2) above. If we let τ vary continuously from zero onwards, the correlations $C(\tau)$ will become what we call *autocorrelation function* for the signal.

Thus, the expression for the autocorrelation function for a signal g becomes:

$$C(\tau) = \frac{\int g(t)g(t+\tau)\,\mathrm{d}t}{\int g^2(t)\,\mathrm{d}t} \,. \tag{15.3}$$

For a digitized signal $g_i \equiv g(t_i)$ for $i = 1, 2, \ldots, N$ (discrete instants), the corresponding expression is:

$$C(j+1) = \frac{\sum_{i=1}^{M} g_i g_{i+j}}{\sum_{i=1}^{M} g_i g_i} \tag{15.4}$$

for $j = 0, \ldots, N - M$.

Note that since we shift the selection of points used for describing the signals at points A and B from the same data string, $M < N$. The largest j value we can calculate for a signal described by N points is $j = N/2$.

The left part of Fig. 15.3 shows an example of an autocorrelation function for a wave. Along the x-axis, we have the time difference as $f(t) = g(t+\tau)$ is offset relative to $g(t)$. If the sampling rate is F_s, the relation between the index j and the time delay τ will be:

$$j = \text{round}(F_s\tau)$$

where "round$(\cdots)$" means the integer nearest to the numerical value of the expression enclosed in the parentheses.

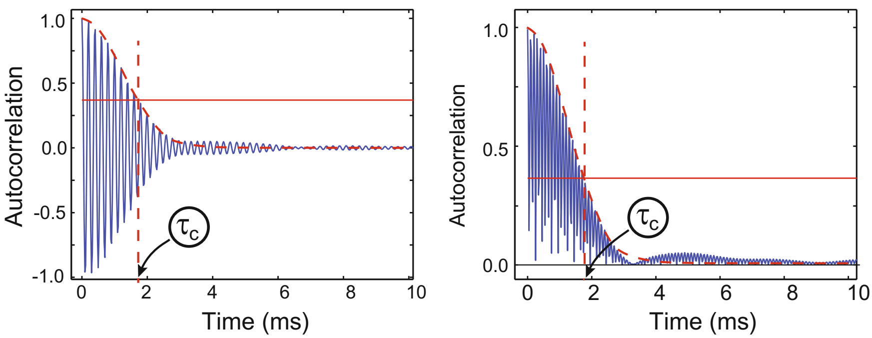

Fig. 15.3 An example of how the autocorrelation function of a wave might look like. To the left, the usual autocorrelation function is plotted, while in the right part the absolute value is plotted. The time shift τ_c that causes the correlation to decrease to $1/e$ of the maximum value is called the coherence time of this wave. (Note: The curve in the right part of this figure seems to never touch the x-axis. The reason for this is that the sampling frequency is not much larger than the signal frequency. We just do not happen to measure values very close to zero.)

We see that the autocorrelation varies from $+1$ to near -1 and oscillates up and down while the amplitude of the variation decreases to zero. Correlation equal 1 (for the standard formula we used here) corresponds to the fact that f and g are identical. They are always at the first point (no shift). Correlation -1 would mean that $f = -g$. For a wave, just say that when f has a wave peak, g will have a valley, and vice versa. In that case, there is still perfect correlation, but the sign has reversed.

When the distance between points A and B in Fig.:15.2 increases, in Fig. 15.3, there is a gradual transition from high correlation (dashed upper envelope curve has a value close to 1) to a smaller and smaller correlation (envelope near zero). After about 1.8 ms, the number has fallen to $1/e$ of max. We say that *the temporal coherence time* τ_c for this signal is 1.8 ms. The corresponding temporal coherence length is $\tau_c v$ where v is the wave phase velocity.

In the right part of Fig. 15.3, we have plotted the absolute value of the autocorrelation function as a function of the displacement τ. We have seen that correlation $+1$ and -1 both correspond to a perfect correlation, only with a change of sign in the amplitude in the latter case. It is therefore best to draw a hypothetical envelope curve touching the peaks in the above plot, when we want to determine the value of τ at which the correlation decreases from 1 to $1/e$.

It should be noted that in practice, the autocorrelation will never vanish when the distance between A and B increases. However, if the average value of g is zero and if the total observation time is much longer than τ_c', the asymptotic value of the autocorrelation ratio will come so close to zero that the remaining variation does not affect the determination of coherence time.

It turns out that the variation in the first part of the autocorrelation function is quite stable if we make more subsequent recording of the signal $g(t)$, but the oscillations around zero when we have passed at least twice the coherence time will change from one data record to the next.

If we take the average of many runs and add the autocorrelations, the first part of the correlations will be added constructively while what we like to call the "noise" around zero will eventually be considerably reduced. The autocorrelation function can then often be written almost:

$$C(\tau) = \cos(\bar{\omega}\tau)\exp[-(\tau/\tau_c)^2] \qquad (15.5)$$

where $\bar{\omega}$ corresponds to the mean of the angular frequencies in the original signal. τ_c is the correlation time and corresponds (with some reservations) to what we also call the "coherence time" for our wave/oscillation. The coherence time is then defined by Eq. (15.5).

If g was a perfect sinusoid, a shift of $\tau = 2\pi n$ (n an integer) would cause $g(t)g(t + \tau)$ in Eq. (15.3) in practice to be a $\sin^2$ function. The autocorrelation function C will then simply be the mean of $\sin^2$ divided by the same value, which is 1.

Similarly, we can show that when τ equals $(2n + 1)\pi$ in the perfect sinus signal g, the calculation will mean the mean of $-\sin^2$ and the answer would be -1. For a quarter wavelength offset relative to the cases we have already mentioned, C will be the mean of a $\sin \times \cos$ expression, which is equal to zero.

The autocorrelation function for a perfect sine will therefore be a periodic function that varies from $+1$ to -1. The autocorrelation function simply gets a cosine shape and will never get an amplitude that decreases to zero as shown in Fig. 15.3. Coherence time would then become infinite.

No real waves have infinite coherence time, but if the time frame we have available for analysis is no more than twice the coherence time long, we cannot determine the coherence time.

15.1.3 Real Physical Signals

We will now take a few real physical signals to show important features when we mix real waves. We have at our disposal three different analytical methods to study the time signal, namely Fourier transformation, wavelet analysis and calculation of the autocorrelation function.

The signals are microphone signals after sampling sound waves from a person who sings "eeeeee" throughout the sampling process. Eight separate audio recordings of the same person have been made. From these recordings, we have constructed three: signal 1 is simply one of the sound recordings. Signal 2 is the sum of two recordings, and signal 3 is the sum of all eight recordings. Signal 2 simulates a data recording of two people who sing "eeeeee" simultaneously, and likewise for signal 3. The signals are shown in Fig. 15.4 together with the analyses we have performed.

Signal from one source
We see in the left-hand column in Fig. 15.4 that the amplitude of the signal (and thus the sound intensity) remains quite similar throughout the period. The frequency range has a width of about 15 Hz. The wavelet diagram shows how the frequency and the amplitude changed during data recording. At the bottom of the figure, the first part of the autocorrelation function is given, and we can estimate the coherence time of the signal to about 0.18 s.

The sound speed in air is about 340 m/s. This means that we can predict correlation in the phase of the signal in hand within a range of about $\Delta L = 340 \times 0.18$ m, that is, about 60 m. This quantity we call the *coherence length* for our wave (more precisely the *temporal coherence length*).

There seems to be nothing special about data recording for one source. Everything is as we expected, but we realize that the instability in frequency (pitch) about the middle of the data recording probably affects the calculated coherence time. We also notice that the autocorrelation function does not settle down to the zero baseline after we have passed the coherence time, which is as expected.

Signal from several sources
We see from the middle column in Fig. 15.4 that the amplitude in the sum of two almost similar waves varies drastically during data recording, although the amplitudes of the individual contributions remained rather stable. The reason for this is

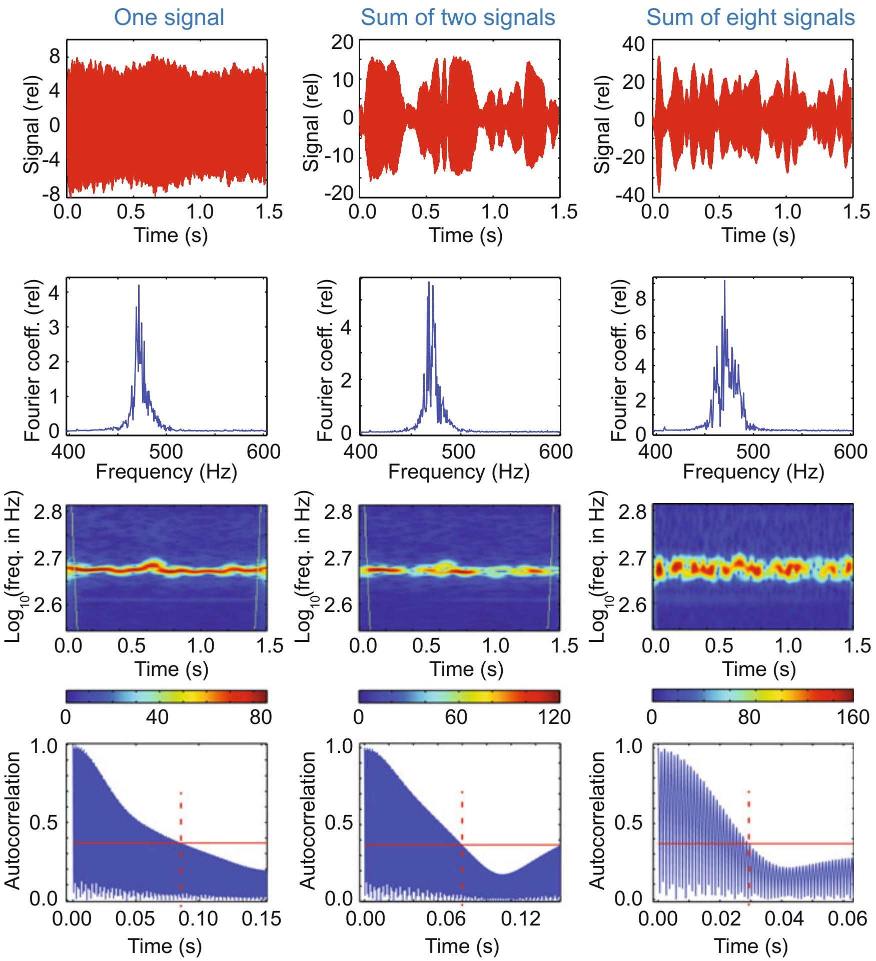

Fig. 15.4 Three examples of real sampled audio signals, with left to right one, two and eight voices at the same time. Frequency analysis is also shown (a little second harmonic component is not included), as well as wavelet analysis and the first part of the autocorrelation function. For the wavelet analysis, $K = 100$ was applied to the first two signals and $K = 32$ to the last

that the sound waves from singer 1 and singer 2 add constructively for some periods, so that the amplitude becomes about twice as large as that for each single wave. At the same time, the sound waves in other periods happen to add destructively, and the total amplitude of the sound waves falls almost to zero.

All waves seem in general to have the following properties:

The amplitude at an arbitrary place in the wave can have many contributions. However, the amplitude has only one value within a sufficiently small volume (lengths in each direction much smaller than the shortest wavelength in the waves that contribute). For example, the sound pressure will only have one value in small volumes when a sound wave passes, the height of the water surface has only one value in any place, and the electric field in the sum of all electromagnetic waves has only one value in each small volume even though many electromagnetic waves contribute.

We can only recognize different contributions to the waves and their origins by examining the pattern in the sum of all the waves and seeing how this pattern evolves in time. Contribution from circular ripples on a water surface impacted by a stone can only be recognized by looking at the rings in the region surrounding the small volume under consideration.

It is the *summed wave* which evolves in time, not every single contribution separately. However, when the physical system behaves linear, we can still describe the evolution of a wave *mathematically* as a sum of several contributions that individually conform to the wave equation. That we can use mathematics in this way should, however, be considered more a happy exception than the rule (because when we deal with nonlinear processes, it does not apply).

When we add more independent waves, there will always be some periods of constructive interference and some periods of destructive interference. The duration of the periods of constructive interference depends very much on the frequency variation in the signals that are added (which is related to the temporal coherence time). We will show more examples of this a little later in the chapter. The effect is manifested particularly well in continuous wavelet analysis with Morlet wavelets.

Besides displaying the characteristic fluctuation in amplitude due to the summation, Fig. 15.4 shows that the frequency spectrum becomes larger as more signals contribute. Each signal has its centre frequency and variation, and the sum of signals therefore gets a wider width than each individual contribution. In our case, the centre frequencies of the eight contributions do not differ by more than 1 Hz (tested separately, data not shown).

An increase in the width of frequency distribution affects also the coherence time. With more contributions, we get a shorter coherence time than with individual signals. However, it is impossible to draw conclusions about relationships between, for example, the width of the frequency spectrum and coherence time, on the basis of these files alone. The statistics are too poor for the task. We will come back to the issue about a little.

The wavelet diagrams in columns 2 and 3 in Fig. 15.4 show essentially the same characteristic features as the amplitude variation in the time domain. However, the frequency is so well defined that sometimes we can see small changes in the dominant frequency in the sum signal as time passes.

Amplitude and spatial distribution of the summed signal
It is worth noting that the average amplitude of the sum of eight similar signals is about 20 on a scale where the average amplitude of one of the signals is about 7. It can easily be shown that $20 \approx 7 \times \sqrt{8}$. With the addition of independent waves of only approximately the same amplitude, frequency and degree of variation over time, the sum of amplitudes is not proportional to the number of signals that contribute, but only roughly proportional to the square root of the number of signals. The intensity (proportional to the amplitude squared) is proportional to the number of signals.

However, if the contributions were close to perfect harmonic signals with the same frequency, amplitude and phase, the sum would have got an amplitude proportional to the number of signals we add, and the intensity proportional to the *square* of the number of signals added. That is what makes the intensity of a laser beam so impressive even if its power is only a few milliwatts.

Let us go back to our singers who sing "eeeeee"s. If the singers had sung exactly and constantly in phase, eight singers would give an intensity equal to 64 times the intensity of each individual. This is quite different from the eight times intensity we found in practice with our real signals. Is this something we can utilize?

Unfortunately, "*There ain't no such thing as a free lunch*". We get nothing for nothing. It may appear that eight hypothetical singers who sing in unison would give *eight times greater* sound intensity than eight singers who do not sing coherently. How, then, can the requirement for energy conservation be satisfied?

The energy ledger will stand up if *spatial* relationships are also taken into account. Eight hypothetical singers who sing coherently will not give eight times larger sound energy everywhere in space, only in those places where the signals from all eight are in phase with one another. At other places in space, where the signals are out of phase, the sound energy could drop to almost zero. That is not what happens with the real singers. In no region of space will there be permanent constructive or destructive interference for these singers. We hear the sound of the real singers everywhere.

If we integrate the sound energy over all space, it will be about the same regardless of whether the singers are singing coherently or incoherently.

These spatial considerations are analogous to the intensity distribution in the interference fringes from many slits. When the number of slits increases, the fringes become narrower and narrower, and the bright streaks become more intense even though the total luminous flux out of the slits is unchanged.

15.2 Finer Details of Coherence

As stated above, there is a close relationship between temporal coherence length and width of frequency (or wavelength) distributions. We also claimed that the spatial coherence length may depend on the size and other features of the source of the waves.

In this sub-chapter, we will explore finer details of coherence which in fact can be very useful to know—useful for experimental physics and for better understanding of some exciting phenomena in physics.

We are now going to analyse signals where we can choose between different widths in the frequency (or wavelength) distributions of the wave. We could avail ourselves of audible noise found in nature or other types of waves with similar characteristics. However, we choose to generate the signals numerically that makes it easy to change the characteristics.

15.2.1 Numerical Model Used

We will make a kind of noise signal, a signal similar to the sound of a large waterfall. The sound may be composed of many sources that act independently of each other (the sound of small and large masses of water hitting rocks and water surface at the bottom of the fall). Since the sound is created in a variety of unrelated processes, we call it "random" or "stochastic". The sound has many frequency components that cover an entire frequency band.

We choose an approach in which we create the frequency spectrum of the signal we wish to work with, and then, we use an inverse Fourier transform to generate the signal in the time domain.

We choose that the frequency spectrum (frequency image) should have many frequency contributions with an optional centre frequency and a Gaussian distribution of nearby frequency components. The width of the frequency distribution must be optional. In order to get a large variation each time we generate a signal, we allow each frequency component to have an arbitrary value between zero and the variance of the Gauss distribution. In addition, we allow the phase of each frequency component to be arbitrary, lying between 0 and 2π.

A Matlab function that generates such arbitrary signal with given centre frequency and given full spread in the frequency distribution is given at the end of this chapter.

In Fig. 15.5, one sees three examples of arbitrary signals generated in this way, along with the analysis of the signals by the same three methods of analysis as before. The same centre frequency (5000 Hz) has been selected for all three signals, but three different full widths (down to $1/e$ of max), namely 50, 500 and 5000 Hz. 2^{16} points are used, and the sampling rate is 44,100 Hz (same as audio on CDs). There are a number of interesting results.

15.2.2 Variegated Wavelet Diagram

It is particularly interesting to see the wavelet diagram for these arbitrary signals. Already in Fig. 15.4 we saw that the sum of several signals led to the amplitude in some time intervals being large, but small in other periods. We got a "clumping" of the signal in time. However, in Fig. 15.4 there was little difference in frequency.

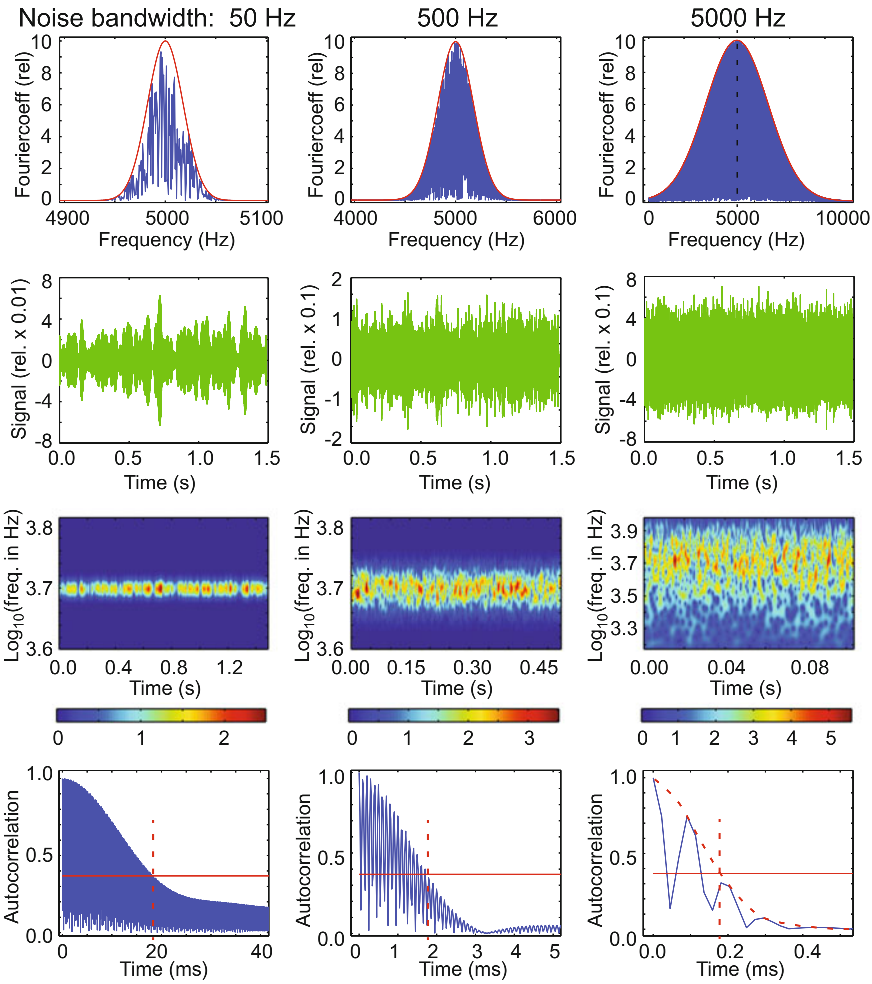

Fig. 15.5 Three examples of synthetic chaotic fluctuations, with a width in the frequency range of, from left to right, 50, 500 and 5000 Hz. The centre frequency is 5000 Hz. Frequency analysis, wavelet analysis and the first part of the autocorrelation function are also shown. For the wavelet analysis, $K = 120$, 36 and 12 were used on the analyses from left to right (almost optimal values). Note the differences in time scales for the three signals in the wavelet diagrams and autocorrelation graphs

We see this much better in Fig. 15.5. The clumping of the signal occurs in a rather chaotic manner both in frequency and time when the width of the frequency distribution is large enough. There are up to thousands of contributions (frequency components) to the final signal (all frequency components within the frequency distribution), and the result is a rather chaotic variegated pattern in the wavelet diagram.

If we scale frequencies and sampling rates in our calculations to visible light, the last column in Fig. 15.5 would be comparable to electromagnetic waves from the sun. In that case, it is tempting to make believe that one is dealing with photons. Each red spot in the wavelet diagram could then be associated with a photon coming at a given time and having a certain frequency. But we know from the way we have generated this signal that the effect is due to a straightforward summation of many random waves with a width in frequency distribution. This is a fingerprint of the sum of many simultaneous independent wave contributions which is quite natural since the light emitted from some parts of the sun arises in a chaotic manner and there is no correlation between light coming from one part of the sun surface with what is coming from other parts.

Note that the time excerpt of the wavelet diagrams is ten times smaller for the right one compared to the left one. It shows that the duration of each red spot (periods with a significant amplitude for the frequency to which the spot corresponds) becomes shorter when the width of the frequency distribution increases. It is difficult to estimate some sort of average duration for the red spots, and the result is partly also dependent on the choice of K value for the wavelet analysis. Nevertheless, we can give a (very rough) estimate of the duration of the spots as follows:

Frequency width (Hz)	Duration of red spots (ms)
50	20–50
500	6–10
5000	1–2

We see that the duration of the spots decreases as the width of the frequency distribution increases. We further note that for the smallest width of the frequency distribution it is not possible to detect that more frequencies occur simultaneously, but when the width of the frequency distribution increases, there are many examples that more than one frequency can be significantly present at the same time.

15.2.2.1 Width of Frequency Distribution Versus Coherence Length

We find the following relation between the width of frequency distribution and coherence length.

Frequency width (Hz)	Coherence time (ms)	Product-of-these
50	18	0.9
500	1.8	0.9
5000	0.18	0.9

The interesting point is that if the width of the frequency distribution is small, it takes a relatively long time between the occurrence of constructive and destructive interference (for fictitious sub-signals with slightly different frequencies). Each time

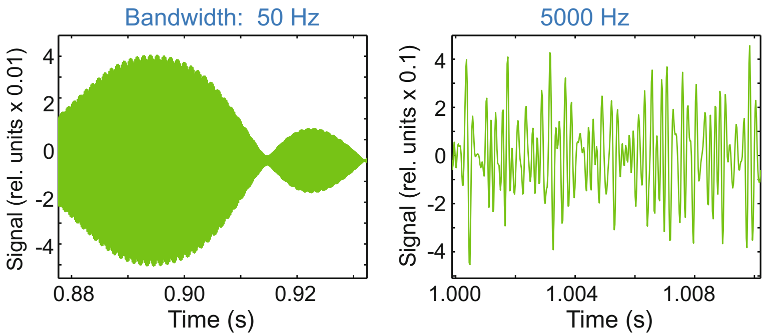

Fig. 15.6 Sections from the signal in the time domain to show that each "bubble" with relatively pure signal is much larger when the frequency width is small than when it is large. Note the difference in time intervals displayed

the signal gets an increased amplitude, it will take a time equal to the inverse of the width in the frequency distribution before the amplitude again decreases towards zero.

Figure 15.6 shows a detail in the time-domain description of a signal generated with widths (in the frequency distribution) of 50 Hz (left) and 5000 Hz (right). The centre frequency is still 5000 Hz. Note that for the smallest frequency band, each bubble takes in the time domain 15–40 ms (more than 100 time periods, individual oscillations do not appear in the plot). With a hundred times greater width in frequency distribution, the width of each bubble (to the extent that it is possible to define such) is only about 0.3–0.5 ms (about two periods). With goodwill, we can say that an increase in the width of the frequency distribution by a factor of hundred led to a 100-fold reduction in the duration of each bubble. This accords with the relationship between width of the frequency distribution and coherence time of the signal.

The result is related to the so-called *Wiener–Khintchine theorem* which states that the Fourier transform of the autocorrelation function of a function is equal to the power spectrum of the function (also called the spectral power density). We do not go into details about this last relationship.

*15.2.3 Sum of Several Random Signals; Spatial Coherence **

So far, when we have discussed random signals (chaotic signals), we have only studied each signal in itself. In calculating coherence time, however, we have in a way compared a chaotic signal with itself, but slightly shifted in time. We saw that coherence length was very small when the width of the frequency distribution was about as large as the centre frequency.

We shall now study *spatial* coherence; that is, we will investigate the correlation between the wave passing point A with the wave passing point B in the right part of

Fig. 15.2. A key for understanding spatial coherence is a spatially distributed source of waves.

This subsection deals with the experiments carried out by Hanbury Brown and Twiss more than 60 years ago. Our treatment explains how we can generate light with considerable spatial coherence by sending a light beam (initially with much less spatial coherence) through a "pinhole" with a few micrometre diameter.

The description is, however, somewhat demanding and is not expected to be treated at bachelor level in physics. Jump to Sect. 15.3 if you want to skip these finer details.

A hypothetical "point source"

Suppose we have "point source" of a wave with an extent less than a wavelength. Assume further that the wave exits from this source with an almost spherical symmetry (at least for the part we are interested in). Assume further that A and B are equally distant from the source. In that case, the wave at A will be equal at each instant to the wave at B. A wave peak will pass A and B at the same time. We can say that A and B are on the same well-defined wavefront, which is part of a spherical surface with the source at the centre. In that case, the spatial coherence is as long as the extent of the part of the sphere where this relationship holds.

This means that we could insert a double slit perpendicular to the wave motion direction and see interference on a screen behind the double slit. If the coherence time is more than a few time periods, we will be able to get more fringes in the interference pattern.

Similarly, we could put a spherical obstacle and demonstrate the existence of Arago's spot (bright point) in the centre of the shadow (image) of the obstacle.

A more realistic source of light

However, for light it is difficult to create a light source with an extent less than the wavelength. It is possible in the so-called quantum dots, but when a filament lamp emits light, we can look at the filament as a mass of independent light sources, each emitting chaotic light signals. If we now study light waves passing two points A and B at the same distance from the incandescent lamp, there will no longer be a good correlation between the waves at the two points. This is because there is a certain distribution of distances between the different parts of the filament and the point A, and a *different* distribution of distances between the same parts of the filament and the point B.

The same reasoning can also be used for light from the surface of a star. Hanbury Brown and Twiss developed an elegant method in 1954 and 1956 that can be used to measure the extent of a star using the properties of chaotic waves (see references at the end of this chapter).

A simple model to point out the essence we want to discuss

In Fig. 15.7, the principle of the Hanbury Brown and Twiss effect is shown. For the sake of simplicity, we have only included two independent sources with the same type of chaotic signal, near each other, and two detectors A and B. The signal into A consists of the sum of waves from sources 1 and 2 travelled equal path lengths.

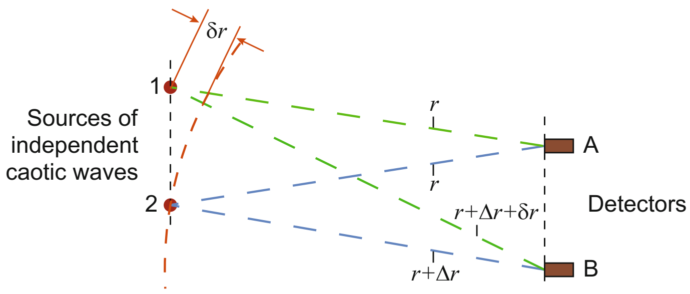

Fig. 15.7 Principle sketch to show the basis of the so-called Hanbury Brown and Twiss effect. See the text for details

The signal into B is also a sum of the signals from the sources 1 and 2, but this time the path length is slightly different. The signal from source 2 is delayed relative to the signal from source 1, which is different to the situation at A.

The signals from the sources 1 and 2 are both chaotic, and their sum is also chaotic. The spots in the wavelet diagram will have short duration and will be distributed chaotically in time and frequency (assuming a large width in the frequency spectrum).

Since the path length are equal between A and source 1 and 2, and equal to the path length between B and source 2, but not between B and source 1, the wavelet diagrams of A and B may be quite different. This means that there may be a bad correlation between the signals into detectors A and B. However, if we placed A and B in the same location, the correlation would be maximum.

Although the distribution of distances is always different at points A and B, if these points do not coincide, the difference may not be demonstrated in practice. If the differences in the distribution of distances of A and B differ by significantly smaller than a wavelength, we cannot expect to see any difference in the waves at points A and B. In that case, there will be a high degree of correlation between the waves in these points. If we make the distance between A and B larger, we will sooner or later get a difference of more than one wavelength in distributions of distances from different parts of the light source to point A and corresponding to B. In that case, we would expect the correlation between the waves at A and B to decrease.

The largest distance between A and B for which we can still get a significant correlation will be called the spatial coherence of the wave at the place under consideration.

If the light source, which consists of independent chaotic parts, has an extent given by an angular diameter θ judged from observer, and the average wavelength is λ, the spatial coherence length a will be in magnitude

$$a = \lambda/\theta$$

Hanbury Brown and Twiss used this relationship in 1956 to calculate the size of the star Sirius which is 8.6 light years away from us. The spatial coherence length of the light from Sirius was about 8 metres here on earth. The angular diameter was estimated at 0.0068 arc seconds which correspond to 3.3×10^{-8} rad.

If we use this relationship and consider a halogen bulb (with a filament of 1 cm extension) as light source and considering the light from this 100 m away, we will be able to find correlation in the signals at points A and B, which are up to about 6 mm apart. In other words, the spatial coherence of this lamp at 100 m distance would be about 6 mm.

For random (chaotic, stochastic) light, we have so far only considered how the spots in a wavelet diagram change, for example, when two chaotic signals are added with and without a time offset. Basing our discussion on the signals with the given frequency and the width of the frequency distribution, we have calculated temporal correlations in the amplitude (first-order correlations). This approach works well when the frequency is less than a few GHz, but we have no detectors that can follow the time variation of the signal for visible light (6×10^{14} at 500 nm). Accordingly, we cannot sample and calculate the autocorrelation function for light waves.

The detectors for light are so-called square law detectors that provide a response proportional to the *square* of the amplitude of the light coming in. The detectors cannot follow the instantaneous intensity, which varies as fast as the underlying sinusoid itself, but provides an integrated intensity over a significantly longer period. This means that light detectors can only detect time variations in intensity in a frequency range below 1 GHz. When Hanbury Brown and Twiss performed their famous experiment about Sirius in 1956, the available bandwidth of detectors and amplifiers was only 38 MHz. How could they follow the much faster changes in the light signals themselves?

The clue is that when we sum up two frequencies and squares the sum, we get the following:

$$(\cos\omega_1 t + \cos\omega_2 t)^2 = 1 + \frac{1}{2}\cos(2\omega_1 t) + \frac{1}{2}\cos(2\omega_2 t) + \cos\left[(\omega_1 + \omega_2)t\right] + \cos\left[(\omega_1 - \omega_2)t\right] .$$

In addition to the constant term, the frequency of three terms is about twice the original, and for the detection of light, they are completely beyond the possibility of detection. The term $\cos\left[(\omega_1 - \omega_2)t\right]$ is, however, a kind of "beat frequency term". For continuous frequency distributions we have worked with, this term will provide contributions from the frequency zero to a frequency that corresponds to the width of the frequency distribution.

Because of this "beat frequency term", Hanbury Brown and Twiss (and everyone else for that matter) could transform the variation in the visible light frequency range to a much lower frequency range. The signal that forms within the "beat frequency range" can in our modelling case also be analysed by wavelet analysis, and we will have similar chaotic patterns there as well. This illustrates that the correlation of signals as shown in Fig. 15.7 can also be studied in cases where detection occurs with "square law detectors" in a completely different frequency range than the original waves.

Spatial coherence in the light from the sun

If we use the same relation for the light from the sun seen from the earth, we find that the spatial coherence length is only about 60 μm. This means that it is impossible to use sunlight directly for double-slit experiments and to detect Arago's spot. What

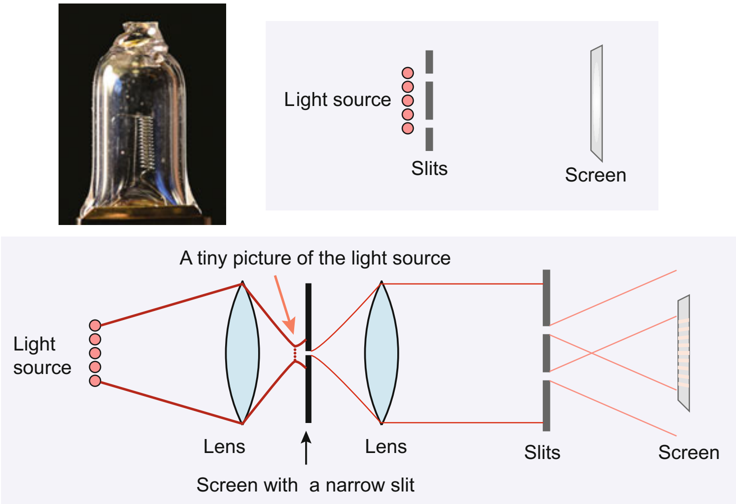

Fig. 15.8 We do not get interference fringes when light from a halogen bulb is sent directly through a double slit. However, if we first send the light from the bulb through a very narrow slit, we can get sufficiently large spatial coherence that interference fringes can be detected. To get a sizable amount of light through the slit and enough light to be able to see the fringes on a screen or on a photographic plate, it is advantageous to use convex lenses (cylindrical) both before and after the first slit

did Young do in 1801 and Arago in about 1820? The secret lies in making the angle of the light source we actually use small enough. We can achieve that by sending light through a so-called pinhole. We can make a pinhole by sticking a very thin needle through an aluminium foil, or we can buy foils with a well-defined pinhole in a holder for about a hundred US dollars. We have a wide choice of the hole sizes, and if we choose, for example, the diameter to be 10 μm, the spatial coherence length of the light passing through the hole will be about 5 cm after the light has travelled 1.0 m from the hole (at 550 nm).

However, the cross-sectional area of a hole with a diameter of 10 μm is exceedingly small. If we want to experiment with a filament lamp or sun as a light source (incoherent light source), we can increase the intensity of the light that passes through the hole by passing the light through a convex lens and positioning the hole exactly where the image of the sun (or filament) is formed. In this case, one captures only a small part of the light from the source, but much more than if one did not use a lens.

It may also be advantageous to use a lens after the hole to prevent the light from spreading too much. This last lens should then be positioned so that the hole is at the focal point of the lens (see Fig. 15.8).

Fig. 15.9 Surface waves on water at some point. Within small patches on the surface, we have a rather "pure" wave. See the text for further discussion

15.3 Demonstration of Coherence

It is no easy matter to get a good understanding of coherence. We therefore choose to include a photograph of surface waves on water to illustrate coherence in an altogether different way.

Within small patches on the surface we have a rather "pure" wave (see Fig. 15.9). Within these patches, it is possible to predict with fair confidence mutual phase relationships in the direction along which the wave is moving (red lines). Within the patches, the phase and amplitude of the wave are approximately constant in a direction that is normal to the propagation direction (yellow lines). The patches are very different in size. In the direction of propagation, the patches vary between two and twelve wavelengths. This means that the temporal coherence length is of the order of 5–7 wavelengths, but this is hardly a sound estimate of size. In a direction normally to the direction of propagation, the yellow lines in this case are on average about as long as the average red line. This means that the spatial coherence length is about as long as the temporal in this case. Perspective conditions, however, make it difficult to specify the width of the patches in terms of uniform waves.

If we consider waves in three dimensions, the "patches" where one sees moderately well-defined waves will be replaced with small volumes where there are fairly well-defined (and almost flat) waves.

However, the patches or volumes with fairly well-defined waves will change in time, which aggravates the complexity even more. One readily appreciates what an enormous statistical challenge it is to describe this dynamic situation, which one often comes across in practice when waves propagate in space.

A small detail in Fig. 15.9 may be worth reminding. At any point, the water surface at a certain moment has a fairly well-defined height. Put another way: The height of the water surface does not have multiple values at the same time! We are so used to thinking that "there are several waves at the same time", but at one and the same place, the local air pressure has only one value at a given moment for sound waves in air, and at one place, the electric field has only one value and only one direction for the sum of all contemporaneous electromagnetic waves at this location.

This is a property well worth pondering over!

15.4 Measurement of Coherence Length for Light

Visible light has such a high frequency that we cannot detect the sinusoidal vibration of the electric field as the wave passes. We cannot use the mathematics mentioned above directly.

However, we can perform an *analogue* calculation of a quantity closely related to the autocorrelation function. This is done by splitting a light beam into two sub-beams by a so-called beam splitter. The two sub-beams are then reunited, but only after one has been made to travel a longer path than the other. When the sub-beams are brought together, their electric fields are added, and so are the magnetic fields, and we deal with the intensity of the sum. We simply consider:

$$\begin{aligned} G(\tau) &= 1/T \int_0^T [f(t) + f(t+\tau)]^2 \mathrm{d}t \\ &= 1/T \int_0^T \left[f^2(t) + 2f(t)f(t+\tau) + f^2(t+\tau)\right] \mathrm{d}t \\ &= 1 + 2/T \int_0^T f(t)f(t+\tau)\mathrm{d}t \ . \end{aligned} \tag{15.6}$$

Here is $f(t)$ to be considered as, for example, the electric field in the beam after it is divided into two and the amplitude is normalized to 1.

This means that we can simply change the path of one sub-beam compared to the other before they combine, and then, we get the autocorrelation function just like the curve in the left part of Fig. 15.3, except that the entire curve is offset by $+1$ so the minimum is zero (intensity cannot be negative). Figure 15.10 shows the principle of a so-called Michelson interferometer commonly used in such measurements. The path difference between the two sub-beams is $\Delta L = 2L1 - 2L2$.

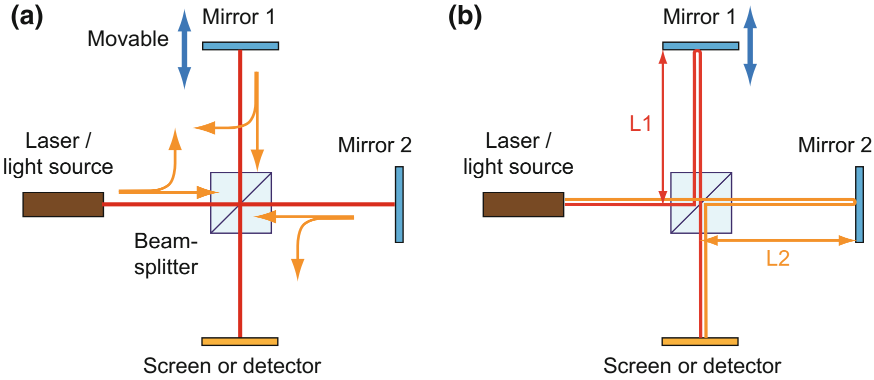

Fig. 15.10 In a Michelson interferometer, a beam of light is brought to a beam splitter. Half of the beam goes to a fixed mirror and is reflected from here, while the other half goes to a movable mirror. Half of the light reflected from the mirrors is sent to a screen or detector where electric fields from the two contributions are added. In the right part of the figure, the light path for the two sub-beams is marked schematically

Light from thermal light sources, such as incandescent lamps, may have a temporal coherence length of only a few wavelengths (that is, just a few microns). Light with such small coherence length is called "incoherent". Light from a good laser can have a coherence length of up to several hundred metres. A laser that costs a few thousand dollars typically has a coherence length of a few centimetres (i.e. in the order of 100,000 wavelengths). Light with long coherence length is called "coherent". There is no sharp boundary between incoherent and coherent light.

Albert Abraham Michelson (1852–1931)
was an eminent experimental physicist. He is perhaps best known for the Michelson–Morley experiment in 1887. Michelson and Morley concluded that their experiment and showed no evidence for the relative motion of the earth and ether. Michelson measured the speed of light with great precision. Furthermore, he developed stellar interferometers, thus measuring the diameter of distant stars, and measuring the distance between star pairs ("binary stars" in English). He received the Nobel Prize in Physics in 1907, the first American to win this distinction.

15.5 Radiation from an Electric Charge

Coherence is linked to the mechanisms of how waves are generated. We have previously discussed mechanisms for producing waves on a string, sound waves and surface waves on water. For electromagnetic waves, we have so far shown that, for

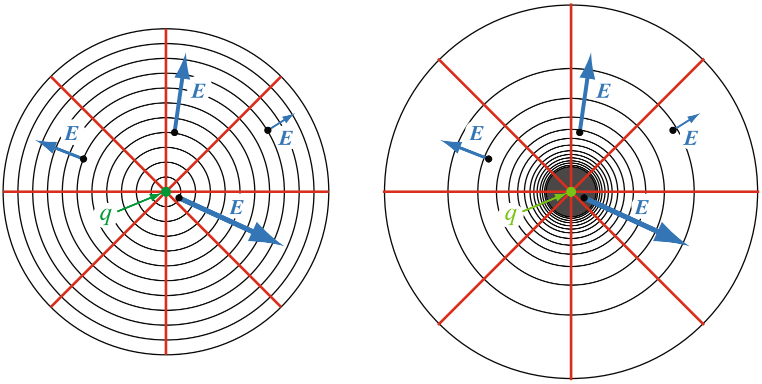

Fig. 15.11 A charge at rest has spherical equipotential surfaces around it (black circles) and radial electric field lines (red lines). The electric field is drawn at four points in the plane. The field is strong close to the charge and decreases with increasing distance. In the left part, the distance between the equipotential surfaces does not reflect the differences in potential between them. In the right part, the potential difference is equal for every set of neighbour surfaces, so that lines close to each other reflect a space with higher electric field than a space where the lines are further apart from each other. However, for point sources the equipotential curves get so close to each other near the charge that the lines overlap. Equipotential lines are therefore often drawn with a somewhat arbitrary selection of electric potentials (as in the left part)

example, plane waves are solutions of Maxwell's equations in the remote field zone in vacuum (at least in the absence of free charges). But what is usually the source or the mechanism behind the generation of electromagnetic waves? We will barely touch this vast field of physics. First, we will see how charge in motion can produce waves, and then, we will look at some of the main features behind the laser.

We can through calculations show that we can make electromagnetic waves in the radio frequency range by sending an alternating current to an antenna. In this case, we have free charges and free currents in action, and Maxwell's equations give us an inhomogeneous second order partial differential equation for the electric field $\vec{E}$ and a corresponding equation for the magnetic field $\vec{H}$. Calculations of this type can be done with finite element methods as mentioned earlier. We do not go into details here.

We choose a "picture and words" presentation instead of a rigorous mathematical treatment, but hope that it will be sufficient to throw light on the key features.

Figure 15.11 shows schematically that a charge q at rest has electric field lines that point radially outward (if q is positive). Equipotential surfaces are spherical shells centred at the charge.

If the charge is moving at a constant speed (left part of Fig. 15.12), the equipotential surfaces according to the theory of relativity will become "squeezed", that is, slightly discus-shaped with the shorter axis in the direction of motion. In a system where the charge is at rest, there is only an electric field. In a system where the charge is in motion, there will be both an electric and a magnetic field. However, when we talk

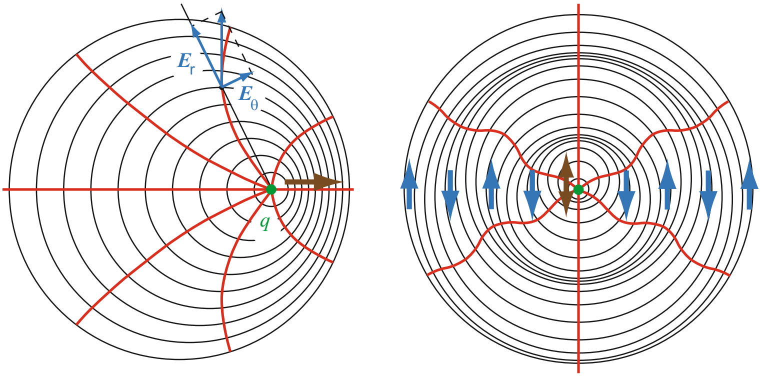

Fig. 15.12 *Left part*: A charge in steady rectilinear motion has spherical equipotential surfaces around it, but only in the sense that the equipotential surface at a certain distance is centred at the charge position at the time d/c earlier, where d is the distance from the equipotential plane and the charge at the earlier time. On account of the relative displacement of the equipotential surfaces, all the electric field lines, except those moving in the same (and opposite) direction as the charge, do not remain purely radial and acquire a tangential component. *Right part*: An oscillatory charge (black arrow) will give equipotential surfaces that are slightly offset relative to each other, as suggested. This causes electrical field lines in tangential direction (*blue arrows*), and these change in principle similarly as in an electromagnetic wave

about generating waves, we must include the so-called *retarded potentials*. We will not go into a more advanced treatment of this topic, but only look at some superficial features.

We assume that *changes* in electrical and magnetic fields move in space with the velocity of light. We do not see a supernova when it happens, but only after light has travelled the enormous distance from the nova to us. What we see today is the supernova as it was for the exact time d/c since, where d is the distance between us and the supernova and c is the velocity of light.

This is also true when we move a charge in space. The field somewhere in space has a distribution corresponding to the location of the charge at the time

$$t' = t - d/c$$

where t is the present time and d is the distance from the charge to the point where the field is being measured at the instant t'.

If we draw equipotential surfaces from a charge moving with constant speed, the planes will have relative positions as indicated in the left part of Fig. 15.12. The effect is greatly exaggerated as the charge would actually have a velocity above half the light velocity as the figure is now drawn.

The electric field is usually given as the gradient of the electrical potential, and we apply this rule also when we use retarded potentials. Then, we get electric field lines that are curved, as shown in the figure.

Suppose that we are at rest and a charged particle is going past us at constant speed. We will first experience an electric (and magnetic) field at our location that has a time development in which the electric field has the same direction as that of the moving charge (for positive charge), and then, we will experience a much stronger field perpendicular to this direction as the charge passes, and end up with a weak field in the opposite direction. This is a "pulse" of electrical (and magnetic) field, and not a wave in the usual sense.

An observer who happens to be following a charge moving at constant speed will describe the electric field as static, and it would appear to him to conform to Fig. 15.11. Such a situation does not qualify for the radiation of energy. In our own reference system, where the charge is in motion, the electric fields will be built up at one place in space, while a completely equivalent depletion of fields takes place somewhere else in space. To be sure, the region that has the highest electromagnetic field energy density will move, in the same way as the charge, but this displacement is of local character, and does not represent energy flowing out of the region around the charge.

To get a wave that extends beyond the vicinity of the charge, we must strive for a situation similar to that of an electromagnetic wave in Chap. 9. Electrical (and magnetic) fields must oscillate and have a direction perpendicular to the direction of wave motion. To get this with our charge in motion, we must have a charge that is subjected to an *acceleration*. For example, the charge can oscillate back and forth in space, preferably in a harmonic motion. The electric field a little away will then oscillate as outlined in the right part of Fig. 15.12. This change in electrical field will have both a radial and a tangential component relative to the radius vector from the charge to the point we consider.

The component in the *radial* direction (when we are at some distance from the charge compared to the amplitude of charge oscillation) will (almost) not change over time. Therefore, this component will (almost) not give rise to any wave that will propagate.

However, the component *perpendicular to the radial direction* (in the plane perpendicular to the charge oscillation direction) will oscillate (almost) as a sinusoid over time. This component could give rise to an electromagnetic wave that spreads into space.

The curvature of the electric field increases with the speed of charge while it oscillates. The time derivative of this again determines how large $\partial E/\partial t$ becomes. These two factors together cause the radiated energy to be proportional to the square of the frequency of oscillation. Therefore, we often say that the radiation is proportional to the *acceleration* of the charge.

Some side remarks:
It may be tempting to think that the electric field from a charge "radiates" outward all the time. Such thoughts could be nourished by the notion of a retarded potential where we think the field in one place is due to the charge where it was a while ago. However, a constantly radiating electric field would soon contradict energy conservation, etc. Fortunately, there is no need to think along these lines. There are *changes* in electrical and magnetic fields that propagate with the light velocity. Before a particular change has spread and reached a given location, it is the field distribution that is rooted in the relationship *before* the change that applies. The electric field from a charge at rest is in equilibrium with itself. It is a solution of Maxwell's equations, and there are no changes in fields and no transport of energy. As soon as movement and particularly acceleration enter, things become different.

15.5.1 Dipole Radiation

An alternative way of generating electromagnetic waves is to use an electrical (or magnetic) dipole that varies in time. This is a very effective way to make waves. We can understand this by considering electrical field distribution from a permanent electrical dipole (see left part of Fig. 15.13). The electric field is directed perpendicularly to the radial direction in the plane normal to the dipole direction.

If we change the polarity of the dipole in a harmonic way, we get an electric field in this equatorial plane that will vary just the way we want it for generating an electromagnetic wave that can propagate in space (electric field perpendicular to the direction of motion). In the direction of the dipole itself (and in the opposite direction), the electric field from the dipole is nearly radially directed and has negligible component across the radial direction. In these two directions, virtually no waves are transmitted.

In the right part of Fig. 15.13 is shown a diagram of electric field distribution near a dipole antenna at a given time. The electric field is the strongest where the field lines are closest. The entire pattern moves outwards with the velocity of light, and new loops form near the antenna twice for each period (direction of the field changes direction in the two systems of loops that form each period in the dipole variation). An animation of the time course (and much other information) is available on Wikipedia under the heading "dipole radiation".

[A remark: The right part of Figs. 15.12 and 15.13 has a certain relation to each other, but is still different. Try to point out differences and similarities.]

It is common to draw *direction diagram* for antennas. A direction diagram indicates the relative temporal intensity of the transmitted waves for different directions

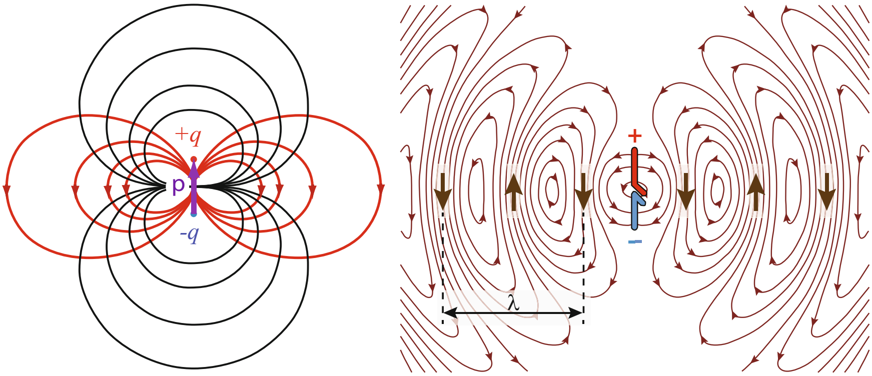

Fig. 15.13 *Left part*: A static electric dipole consists of two identical charges, but with opposite signs, slightly spaced apart. Equipotential surfaces (black, with lobes pointing upward and downward) are drawn as well as electric field lines (red, with belly outwards to the sides). The physical extent of the dipole is greatly exaggerated in relation to the field line pattern. *Right part*: An oscillating electric dipole will create an electric field in the surrounding space as indicated. The field pattern moves outwards at the speed of light

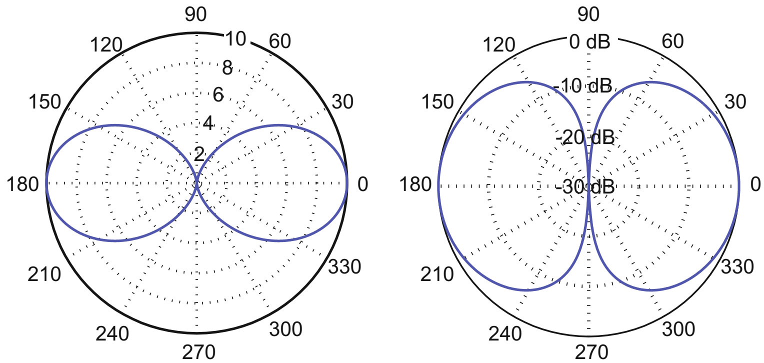

Fig. 15.14 Directional diagram for a single vertical dipole antenna (also called "antenna diagram" or "radiation diagram"). The diagram only applies to a distance from the antenna which is large in relation to the length of the antenna. Linear scale for the intensity in the radial direction is used in the left part of the figure and a logarithmic scale in the right part. Intensities are all relative to the maximum value. The diagram is read as follows (*left part*): the relative intensity at 0° is set to "10". Then, the relative intensity at 30° is about 7.3 and at 60° about 2.8. For the *right part*, see the text

in space, in a so-called polar diagram. Figure 15.14 shows the direction diagram in a vertical plane passing through a single vertical dipole antenna.

The intensity (time-averaged Poynting vector) as a function of angle θ deviation from the meridian plane of the dipole is given by:

$$I(\theta, r) = constant \times f^4 \cos^2(\theta)/r^2$$

where f is the frequency of the alternating current and r the distance from the dipole (has to be many wavelength away to avoid near-field conditions).

A direction diagram can be given with a linear scale in the radial direction (left part of the figure), but most commonly used is a logarithmic scale in the radial direction (right part of the figure).

A remark:
Suppose we create a polar graph with logarithmic scaling of intensity in the radial direction. Since the logarithm of 0 does not exist, a cut-off intensity needs to be chosen when the graph is drawn. In our case, a relative intensity of 1.0 corresponds to the outer radius in the chart, while a relative intensity of 0.001 is chosen at the centre of the diagram. Relative intensities less than 0.001 would be negative on a logarithmic scale and would appear on the opposite side of the chart. To avoid misunderstandings, we remove negative values before plotting.

The Matlab program used to create Fig. 15.14 is given below:

```
function antennaDiagram3
N = 1024;
theta = linspace(-pi/2.0,3.0*pi/2.0,N);     % Angles
costheta = cos(theta);
intensities = costheta.*costheta;
intensities = log10(intensities*1000.0);
for i = 1:N
    if(intensities(i)<0)
        intensities(i)=0;
    end;
end;
polar(theta,intensities);
```

A dipole antenna can in some ways be viewed as a single slit of very small width. The radiation becomes identical in all directions perpendicular to the direction of the dipole. However, if we insert two dipoles next to each other and feed both antennas with identical signal, the radiation diagram will look like a double-slit pattern. By placing many identical antennas in sequence, we get a radiation diagram similar to a single slit (in a "meridian plane" including the complete line of antennas). By inserting reflectors and directors, we can further influence the radiation diagram, and an example is given in Fig. 15.15. There is an antenna diagram for an antenna that is widely used in base stations for mobile telephony. Note that the diagram is radial with decibels in radial direction.

The thinking that lies behind the antenna pattern diagrams is much the same as for the intensity distribution of light after passing one or more slits. The principal motif in the calculations is interference between sufficiently coherent waves. Thus guided, we use the idea of differences in the path lengths and add various contributions with the correct mutual phase.

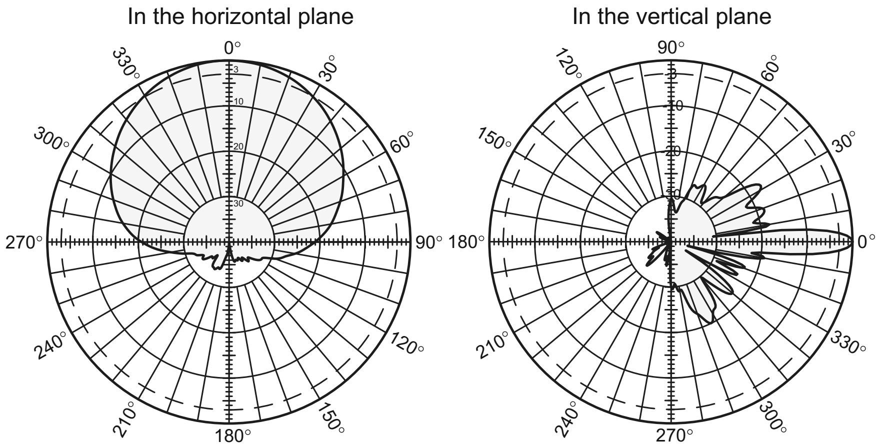

Fig. 15.15 Directional diagrams for a commonly used base station antenna for GSM 900 mobile telephone (Kathrein 80010621). One diagram gives angular distribution in the vertical direction and the other for horizontal direction

15.6 Lasers

One of the most important light sources in science nowadays, lasers, is used in everyday technological appliances, such as CD, DVD players and laser printers. Lasers are also used for cutting metals and other materials, and in medicine, for example, by reshaping the cornea in our eyes, and other operations. Even my dentist has switched to using lasers for "drilling" in the teeth. Some car manufacturers now use lasers as headlights. The range of laser applications is impressive and still increasing!

The word laser is an acronym for Light Amplification by Stimulated Emission of Radiation. Theodore Maiman at Hughes Research Laboratories managed to make the world's first laser (see Fig. 15.16). This happened on 16 May 1960. The laser celebrated its 50th anniversary in 2010. However, there are many physicists who have been involved in the development and utilization of the laser, and it was Charles H. Townes who in 1964 received the Nobel Prize in Physics "for the development of laser principles". Also, other Nobel prizes in physics are fairly closely linked to the laser in one way or another. It is therefore natural that we devote some time to the concepts that lie behind a laser. However, only the main principles are mentioned, and here too we will use a "picture and words" approach.

A laser is based on the so-called *stimulated emission*. Einstein had already shown in 1917 that we could get stimulated emissions from, for example, atoms. By that we mean that we do not have to excite an atom and wait until it finds it convenient to send out light and fall back to its ground state. By shining some light on an excited atom, we can actually *trigger/stimulate* it to return to the ground state. The laser requires a medium that is amenable to stimulation and an arrangement where light produced by

Fig. 15.16 Photograph of the first laser. A flash tube encircles a ruby rod, which is coated with an almost 100% reflective mirror at one end and approximately 95% reflective mirror at the other end. The laser emitted pulsed coherent light. Image courtesy of HRL Laboratories—Malibu California

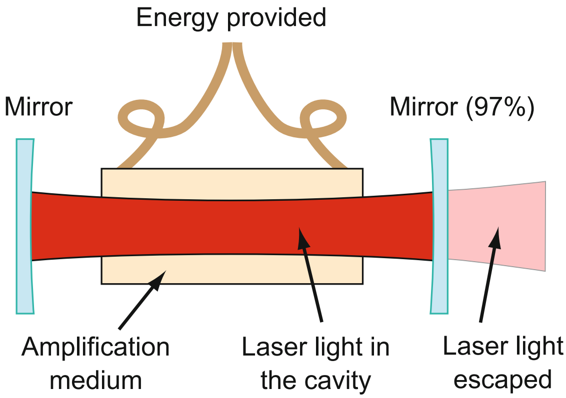

Fig. 15.17 Main constituents of a traditional laser (schematic)

stimulation may lead to the stimulation of even more light from the medium. In this way, we get a positive feedback in the process that makes it almost self-sustained. However, since the energy that is consumed in bringing about the emission of laser light must be compensated for, an external source of energy is necessary if the process is to be sustained.

Figure 15.17 shows the main ingredients in a traditional laser. It contains an amplification medium to which energy can be supplied from an external source. The medium is in an optical cavity ("cavity" or "box") limited by two mirrors. Light that is produced in the medium will initially spread in all directions, but light that hits one mirror tends to be reflected back through the media, hit the mirror on the other side, is reflected once more and sets up a standing wave of electric and magnetic fields in the cavity. The light that is reflected back and forth many, many times can stimulate the medium to give even more light.

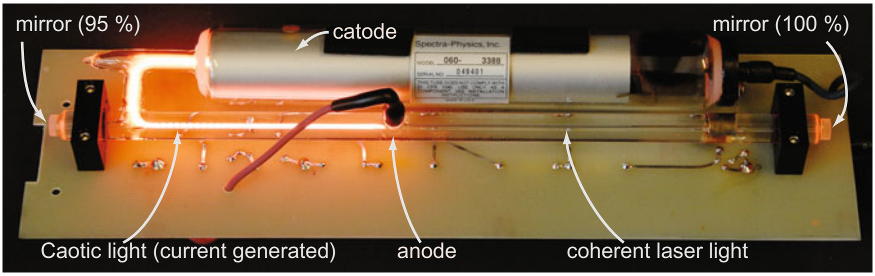

Fig. 15.18 A photograph showing the innards of a standard HeNe type laboratory laser. An electric current is sent through a low-pressure mixture of helium and neon and leads to the emission of chaotic, incoherent light emitted in all directions. Part of this light energy is (at proper conditions) building up in a cavity between two mirrors. The result is a strong, coherent beam of light between the two mirrors. We do not see this beam from the side because the light beam in the cavity is almost perfectly aligned along the axis between the mirrors, and does not exit from the sides. A tiny fraction of this light beam is transmitted through the mirror to the left in this figure

With ordinary sources, the light comes from many atoms or molecules that appear quite independent of each other. Then, we get a state that corresponds to our real singers and our chaotic waves mentioned earlier in this chapter. The light intensity from the light source increases approximately proportionally with the number of atoms/molecules emitting light and the light goes in "all" directions.

The conditions are different in lasers, which will be exemplified through a description of a HeNe laser (see Fig. 15.18). A low-pressure mixture of helium and neon (in a ratio of approximately 10:1) is held in a glass receptacle. Between a cathode and an anode a high voltage is applied, which generates an electric current through the intervening gas mixture. Electric current through the gas volume between the cathode and anode leads to the emission of chaotic, incoherent light emitted in all directions. This process is similar to what is found in a standard fluorescent lamp used in coloured advertisement lighting at night ("neon lights").

Energy of helium atoms (the majority species) excited by the electric current is transferred to the neon atoms during collisions. The excited neon atoms emit light initially through the so-called spontaneous emission in the same manner as the helium atoms.

Part of the light emitting atoms is located in a cavity between two mirrors. Some light will be reflected back and forth between the mirrors as described above, and standing electromagnetic waves may form. The presence of the electromagnetic field with correct wavelength leads to an enhanced probability for an excited atom to emit light, and the process is called "stimulated emission". Light coming from the different atoms through stimulated emission has nearly the same phase and is directed in the same directions as the standing waves between the mirrors. Then, the amplitudes of electrical and magnetic fields will be added directly and the intensity of the light within the beam will be proportional to the *square* of the number of emitting atoms. This leads to increased stimulated emission and thus a positive feedback. The light intensity within the cavity builds up in time, but after few seconds the intensity will reach a plateau which depends among others on the efficiencies to excite He atoms and transfer of energy to Ne atoms.

One of the two mirrors reflects only 95–99% of the light. About 1–5% of the light intensity will be transmitted and is responsible for the laser beam available for use in the laboratory.

It should be mentioned that the energy levels for helium and neon allows us to make lasers with several different wavelengths. Usually, the wavelength of light from a HeNe laser is about 633 nm. Other wavelengths would lead to less efficiency and even light that is not visible for the human eye. Special tricks are used in order to avoid building up standing waves with wavelengths different from about 633 nm.

Since the light waves of a laser are created in a "cavity" (with mirrors at both ends, see Fig. 15.18), the light will form standing waves as mentioned. Then, the frequency will be very precise, in a similar way as the sound of a guitar string attached to both ends is pretty precise. If the distance between mirrors is 30 cm for a HeNe laser wavelength of about 633 nm, there will be about 473,940 wavelengths between the mirrors. The actual line width of the energy transition we use in the neon gas is so wide that sometimes there may be more simultaneous wavelengths in the cavity (the line width is broadened due to collisions with other atoms). With 473,940 wavelengths between the mirrors, the wavelength will be 632.9915 nm, but with one wavelength more or less than this, the wavelength will be 632.9902 and 632.9929 nm, respectively. We are talking about "modes" for the laser light. Mechanical heating of the laser cavity will cause small changes in the distance between the mirrors. In that case, the wavelengths will also change, which will lead to what is called "mode hopping".

In some contexts, lasers are constructed to play an active part in the various modes the laser can operate in. We can then achieve many wavelengths whose mutual difference is almost constant, and the phenomenon is called a "frequency comb". Theodor Hänsch received the Nobel Prize in 2005 for creating a "frequency comb synthesizer" that made it possible for the first time to measure the oscillations in light with extreme precision. The method forms the basis for our most modern atomic clock.

The laser light that escapes from the cavity through the 95–99% reflecting mirror is quite different from light emitted by, for example, an incandescent lamp. If we compare the phase of the electromagnetic wave on a plane normal to the beam, the phase throughout the plane will be almost identical. As stated earlier in this chapter, this is equivalent with a high degree of spatial coherence. This is unlike the light that originates from many atoms which have no definite phase relationship with each other, for example light from a filament lamp. Such light has little spatial coherence, and the wavefront is very uneven.

The stimulated emission in the cavity is fairly stable, and the frequency is so well-defined all the time that we can predict the phase of the laser light beam many wavelengths in the future. Thus, even the temporal coherence is high in laser light.

Since the wavefront is very well defined across the beam, while at the same time we can predict the phase of the laser light long distances along the beam itself and a laser is a far superior light source for interference and diffraction experiments compared to so-called thermal light (“incoherent” light). It also means that a laser beam will maintain a very well-defined shape where diffraction is kept to a minimum. The light in a laser beam is one of the closest we can come to a mathematically idealized wave description in practice. Laser light is therefore sometimes called “classic light”, but such a term confuses more than it instructs.

15.6.1 Population Inversion

While we are explaining lasing action, we cannot completely leave out a detail called population conversion. We will not go into detail since this theme is not so important in our context. Nevertheless, a brief review is given below.

An atom may be in one of the several different energy states. We often draw the energy states schematically as in the left part of Fig. 15.19. The ground state is usually labelled E_0 and the first excited state as E_1. An atom can be exited from the ground to the first excited state by, among other ways, placing it in an electromagnetic field with the frequency $\nu = (E_1 - E_0)/h$ where h is Planck’s constant. An atom in an excited state can fall back to the basic state entirely on its own. The transition may then be accompanied by the emission of light (called spontaneous emission which is a radiative process), or it may be through a so-called nonradiative processes. We can also stimulate the transition with an electromagnetic field with the same frequency as indicated for absorption (a radiative process called stimulated emission).

For the simple system in the left part of Fig. 15.19, energy is stolen from a beam of light to excite the atom at an absorption, while in stimulated emission, energy is released by the atom. There is the same probability of one transition as the other per atom, assuming that the pertinent initial state is occupied. In order to release more light than we insert (as required by a laser), there must be

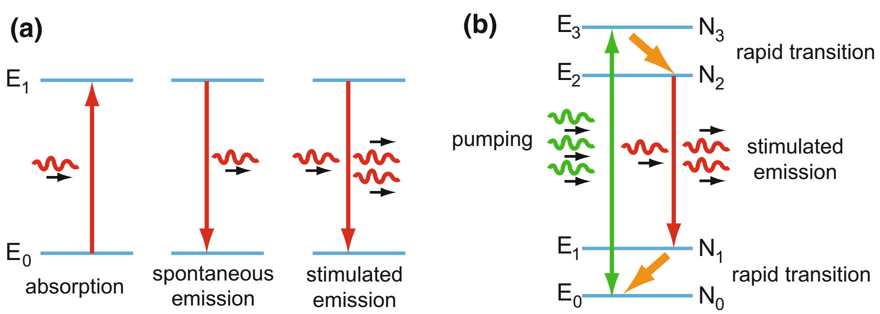

Fig. 15.19 *Left part*: Two energy states in one atom, and schematic transitions between these (very simplified). *Right part*: Population inversion can be achieved by pumping between energy levels other than those involving the emission of laser light. See text for details

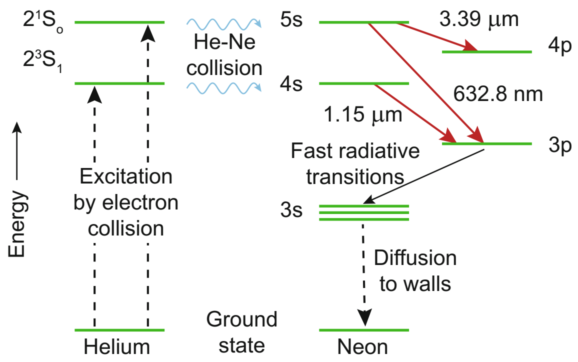

Fig. 15.20 Energy levels which are involved in a common HeNe laser. XuPanda, Wikipedia Commons, CC BY-SA 4.0. Modified from [1]

more atoms in the excited state than in the state to which the atom falls back after emission. The population of energy levels usually follows Boltzmann statistics. Then, there are more atoms in a low energy state than in a higher one. That is, it is usually impossible to form a laser from atoms in a state of thermal equilibrium and few energy transition possibilities available.

The right part of Fig. 15.19 shows a way to get to a higher population in an energy state than in a lower state. The principle is used in neodymium YAG lasers and is based on four energy levels. The atoms are excited, by the use of strong light from some light source, from the ground to the fourth energy level (E_3). The atom then spontaneously falls rapidly to the energy state E_2 through a nonradiative process, but stays here (in a "metastable state"). There is also a fast spontaneous transition from E_1 down to the ground state. However, the transition from E_2 to E_1 is not fast, and after some pumping, there are more atoms in E_2 than in E_1. We have got a population version!

If we now send a (weak) light with the frequency $\nu = (E_2 - E_1)/h$, we will get more emitted light than absorbed light from the atoms, and we may form a laser. However, the intensity of the laser will be limited by how fast we can pump atoms from the initial state to E_3.

For the helium–neon laser, the energy levels involved are a bit more complicated. A sketch (for orientation) is shown in Fig. 15.20. In this case, helium atoms are excited by sending an electrical current through a gas mixture of helium and neon. Electrons with significant speed supply the excitation energy for helium atoms. Helium has two excited levels that are "metastable" so that helium can be in these states quite long before they fall into lower energies. If such an excited helium atom collides with a neon atom, the excited energy can be transferred from helium to neon. The neon atom can then be further deexcited through a transition that gives light at 632.8 nm. It is this red light we recognize from a HeNe laser. The transition from the 3p to the ground state is rapid and involves both radiative and nonradiative processes.

Today, there are many different ways to make a laser. Most people have more and more lasers in their homes, as CD and DVD players use lasers. In addition, many also have a laser pointer. In all these examples, semiconductor laser diodes are used.

The light from such laser diodes is continuous in time. There are also lasers that only give off, in part, very short light pulses. The pulse length can be completely down in the so-called femtosecond area, even down to the so-called attosecond area (uses this term for pulses a bit shorter than 1 fs). The wavelength is not well defined for such short laser pulses!

Relevance for us?

When we went through the generation of electromagnetic waves using an oscillatory charge or oscillating dipole, we based the treatment on Maxwell's equations. The process was described as continuous, and we obtained an electromagnetic wave that lasted as long as the oscillation continued.

When we explained the laser, we used energy levels and jump from one energy state to another. Such an image is based on quantum physics, but only a quantum physics based on energy states where we use perturbation theory to look at probabilities for transitions. How should such energy diagrams as shown in Fig. 15.19 be understood? When a transition first takes place, does it take place immediately or does the transition take some time? A good answer to this question is difficult to get, since there are no consensus on this matter!

We like to draw "photons" like small wave packages (wavy lines) in diagrams like the ones in Fig. 15.19, which implies that there is a small wave when a photon is released from an atom. But how long is this wave? Can we have waves that come out having a phase memory (coherence length) that corresponds to several hundred thousand wavelengths, but which itself has almost no extent?

And how come that electrons at lower frequencies provide a continuous, sustained wave in Maxwell's formalism, but all of a sudden electromagnetic waves with frequencies corresponding to light cannot be described as continuous waves (but as photon particles)?

There are similarities in how quantum mechanics and classic electromagnetism describe the emission of electromagnetic wave/light, but the interpretation of the formalism is quite different.

In spite of the fact that physicists today often refer to light as "photons" which have a slightly fuzzy particle nature, the vast majority of phenomena involving light may be explained by the wave model of light. There are very few phenomena where we must use a particle model.

But what is meant by waves and what is meant by particles after all? Could it be the wave–particle duality and the apparent paradoxes that follow from such an opinion may disappear if we try to get a little more precise in our description?

The problems regarding the wave–particle dualism is intimately connected with philosophy. If we take Niels Bohr's view that the purpose of physics is not to tell how the world *is*, but just to find relations between properties we can measure, it is no problem to switch between a particle and wave description of light. However, if we have a philosophical stance close to realism, the huge differences between a physics behind particle and wave phenomena in nature make the wave–particle dualism unacceptable.

It is now 100 years since the last time physicists shifted from one paradigm about light to another paradigm. Perhaps, the time is now ripe for a new paradigm shift?

15.7 A Matlab Program for Generating Noise in a Gaussian Frequency Band

Actually, the full program code for the Hanbury Brown and Twiss model, is available from this Web page (Matlab version only).

```
function  [xx] = whiteNoiseGauss(Fs,N,fcenter,fullFwidth)

% Parameters: Fs : Sampling frequency, N : # data points
% fcenter, fullFwidth : The frequency spectrum has a gaussian
% distribution with center frequency fcenter and full width
% (1/e) in frequency spectrum equal to fullFwidth

% In the calling program, use for example:
% Fs = 44100;
% N = 2^16;
% fcenter = 400.0;
% fullFwidth = 100.0;

fsigma = fullFwidth/2.0;
y = zeros(N,1);
xy = zeros(N,2);
T = N/Fs;
t = linspace(0,T*(N-1)/N,N);
f = linspace(0,Fs*(N-1)/N,N);
ncenter = floor(N*fcenter/(Fs*(N-1)/N));
nsigma = floor(N*fsigma/(Fs*(N-1)/N));
gauss = exp(-(f-fcenter).*(f-fcenter)/(fsigma*fsigma));
ampl = rand(N,1);
ampl = ampl.*transpose(gauss);
phases = rand(N,1);
phases = phases*2*pi;
y = ampl.*(cos(phases) + i*sin(phases));

% Mirror of lower half (Nhalf+1) to make upper half correct
Nhalf = round(N/2);
for k = 1:Nhalf -1
    y(N-k+1) = conj(y(k+1));
end;
y(Nhalf+1) = real(y(Nhalf+1));
y(1) = 0.0;
% plot(f,abs(y),'-g');  % Plotting as a check if desired
% figure;

xy = ifft(y);
xr = real(xy*400);
xx = xr;

plot(t,xr,'-b'); % Plotting if wanted
hold on;
plot(t,imag(xy),'-r');
xlabel('Time (s)');
ylabel('Sound signal (rel units)');
sound(xr,Fs); % Playing the sound for proper frequencies if wanted
```

15.8 Original and New Work, Hanbury Brown and Twiss

R. Hanbury Brown and R.Q. Twiss. A test of a new type of stellar interferometer on Sirius. Nature 178, 1046 (1956).

R. Hanbury Brown, R.C. Jennison, and M.K.D. Gupta. Apparent angular sizes of discrete radio sources: Observations at Jodrell Bank Observatory, Manchester. Nature 170, 1061 (1952).

R. Hanbury Brown and R.Q. Twiss. Correlation between photons in two coherent beams of light. Nature 177, 27 (1956).

A relatively new article 6:37980—https://doi.org/10.1038/srep37980 "The colored Hanbury BrownTwiss effect" based on a quantum approach by Silva et al. is available for free at Nature.

15.9 Learning Objectives

After working through this chapter, you should be able to:

- Explain what distinguishes a real wave from an idealized simple mathematical description of a wave.
- Explain what is meant by temporal coherence and coherence length, and how they can be determined.
- Explain what is meant by spatial coherence.
- Explain why coherence plays a role in experiments involving interference.
- Explain qualitatively why the line width in a frequency spectrum is related to coherence length.
- Explain why summation of N perfectly coherent waves leads to an intensity of N^2 times that of each wave, while the intensity is only N times that of each wave for the summation of N incoherent waves.
- Explain how energy conservation is fulfilled in the two cases in the previous learning objective.
- Explain why we claim that it would have been disastrous if all members in a choir sung perfectly in tune with each other.
- Explain how we can measure coherence lengths with a Michelson interferometer.
- Explain qualitatively why an oscillating charge leads to the emission of electromagnetic waves.
- Provide a qualitative correlation between a composite radio frequency antenna and the diffraction pattern of light from two or more slits.
- Explain why a laser basically gets a (temporal and longitudinal) coherence length that is far greater than does thermal light from, for example, an incandescent lamp.
- Explain qualitatively why population inversion is important for making a laser work.

15.10 Exercises

Suggested concepts for student active learning activities: Temporal/longitudinal coherence, spatial coherence, autocorrelation function, correlation function, coherence time, coherence length, chaotic signals, intensities after summation of coherent/noncoherent waves, frequency width, pinhole, interferometer, retarded potential, dipole radiation, dipole antenna, antenna diagram, laser, population inversion, stimulated emission, spontaneous emission, cavity, standing waves.

Comprehension/discussion questions

1. Attempt to explain why the signal in the middle of Fig. 15.4 varies so much in amplitude even though the signals we started with had a more even distribution of amplitudes.
2. If you have sung in choirs, you have probably noticed that the volume reached with more and more singers does not increase as much as we might think. A soloist with a good voice is not outsung by a choir of maybe 30 people. Why is not that one voice totally drowned by that of the 30?
3. Will a singer who sings a "dead" tone be expected to have a greater or less coherence time for his/her voice than a singer who has significant "vibrato" in his song?
4. Indicate the advantages and disadvantages of people who sing in a choir having relatively short coherent times for the voice (the sound).
5. In Fig. 15.21, red lines mark the portion of the wave pattern that will pass a double slit as waves are passing by. Do you think you would be able to detect interference fringes for the case on the left and/or the case on the right? Explain what would happen in the two cases.
6. Some believe that when we talk about coherent and incoherent waves, there are two well-defined types of waves. In reality, there is a continuous range all the way from "incoherent" to "coherent". Explain.
7. There are advantages and disadvantages associated with both coherent light and incoherent light. In which contexts would you prefer one and when would you prefer the other? As usual, explain your answers.

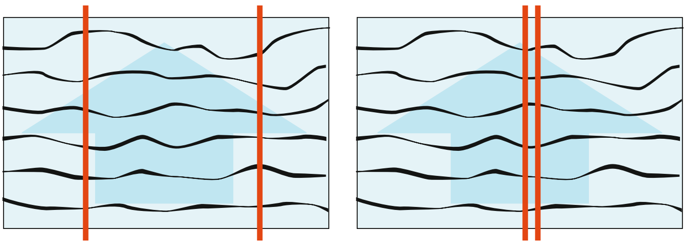

Fig. 15.21 Marking the position of the two apertures in a double slit relative to the wave pattern. The gap between the slits is much larger in the example to the left than in that to the right

8. What will it take to obtain at least five interference fringes after sending light through a double slit when the light we use comes from the sun?
9. Try to describe in your own words what we mean by coherence and coherence time for a wave. What is the difference between spatial coherence and temporal coherence?
10. What is the main idea behind the method of Hanbury Brown and Twiss portrayed in a simplified form in Fig. 15.7?
11. What is the main idea behind "retarded potential"?
12. What is most important for creating an electromagnetic wave from an oscillatory charge or oscillating dipole: the radial component of the electrical field or the tangential component? Explain!
13. Antennas are often designed as electrical dipoles as indicated in Fig. 15.13. Occasionally, however, magnetic dipoles are used as antennas (a simple circular current loop). Sketch the magnetic field around a magnetic dipole, and use this as a basis to argue that the antenna can be expected to be quite effective.
14. In a previous chapter, we found intensity distribution of light diffracted at a single slit. How would intensity as a function of angle from the symmetry axis be expressed in a polar graph (qualitative)? Would the plot have some resemblance to what we find in one of the antenna diagrams in this chapter?
15. Explain how population inversion can be achieved in a four-level system.

Problems

16. Use the program code earlier in the chapter to generate a chaotic signal with centre frequency 5000 Hz and a 3000 Hz width in the frequency distribution. Create a function that can calculate the autocorrelation function. Make a plot. Is the coherence time you arrive at close to that expected? Comment on the results.
17. Use the program code given earlier in the chapter to generate a chaotic signal with centre frequency 5000 Hz and width in the frequency distribution of 3000 Hz. Make a Fourier and wavelet analysis of the signal. Then, calculate the square of the original signal (elementwise squaring). Make a Fourier and wavelet analysis of the squared signal (be sure to include a large enough frequency range in the wavelet analysis). Comment on the results.
18. Determine the coherence time for your own voice. Specifically, the task entails the following steps:
    (a) Create a computer program where you can digitize audio, calculate the autocorrelation function, and plot a selected part (see program snippets in Chap. 5). Use the plot to estimate the approximate coherence time of the signal. If you have created a program for digitizing audio when working with Chap. 5, you may save a lot of time using it here as well.
    (b) In particular, explain how you chose to utilize the data string you received when digitizing in the analysis. Specifically: How did you choose to let the i and j in Eq. (15.4) run in relation to the total data string?
    (c) Determine the approximate coherence time of your own voice when you sing

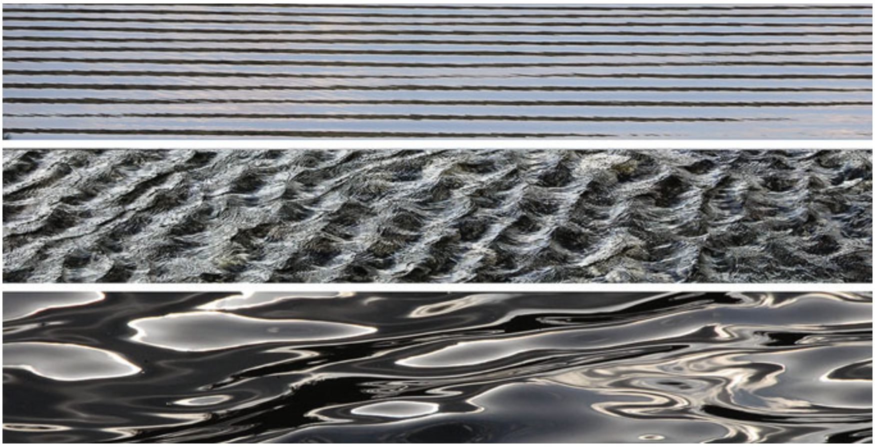

Fig. 15.22 Waves on water surface at some instant

"eeeeee" as evenly as you can. Perform this for 2–3 different pitches. Does the correlation time seem to change much with the pitch?
(d) Perform a wavelet analysis of the signal. Comment on the results.
(e) Digitize another sound and determine the coherence time also for this (Suggested sound: Own voice, same pitch as you used in point c, except that you are now singing "oooooo" instead of "eeeeee". Alternatively: Audio from a piano, guitar or some other musical instrument.). Do you find any interesting differences or similarities compared to what you found in point c?

19. Try to mark, in Fig. 15.22, regions where the waves are relatively well-defined. How big are these regions approximate? And how much of the entire surface have you taken into account in the selection of these areas? Enter lengths in numbers of "approximate wavelengths". [It is helpful to draw a line indicating the approximate direction of waves moving in. To get a correct picture of coherence lengths, the *entire* surface must be included in the statistical processing.]
20. Can you suggest a way to make an interferometer for sound that corresponds to a Michelson light interferometer? (Must be used to measure coherence lengths for sound from, for example, different musical instruments.)
21. Let us look at the antenna diagrams in Figs. 15.15 and 15.23. The radiation pattern in the horizontal plane is approximately what one expects for a single dipole, except that the radiation backward is damped heavily. The radiation diagram in the vertical plane is different and has something in common with the diffraction pattern for a single slit, but with a great deal of deviations. Figure 15.23 shows a photograph of a regular base station for mobile telephony and indicates how the antenna is built up of multiple dipole antennas that have identical radiance.

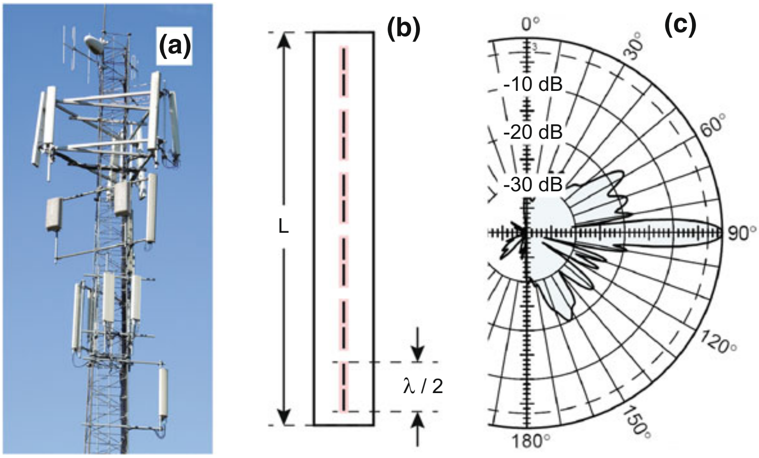

Fig. 15.23 **a** Base station with many different antennas for mobile telephony. **b** Each antenna is assembled by pair of dipole antennas opposite each other (here are six identical antennas drawn). Each dipole antenna is a half wavelength long. **c** Radiation diagram in the vertical plane of the antenna in Fig. 15.15 (Kathrein 80010621). For this antenna $L = 1.4$m, and it is intended for signals at approximately 2 GHz

This basic construction shows that the comparison with diffraction from a slit is not completely groundless. Note, however, that "slit" may be completely confusing since we use the *length* of the antenna as *slit width*.
(a) Use the radiation diagram in Fig. 15.23c as well as the formula of the angle to the first minimum by diffraction from a single slit:

$$\sin\theta = \lambda/a$$

and the information that the antenna is used for approximately 2 GHz signals, to calculate a, the "effective slit width" based on the given formula.
(b) Compare a with the antenna's outer vertical length. By the way, approximately how many dipole antennas (half a wavelength long) would there be room for within L?
(c) There is a mismatch between L and a even if they are of the same order of magnitude. Do you have any ideas about what the disagreement can bump into? Could you test your ideas using numerical calculation?

22. (a) In Figs. 15.15 and 15.23 are given antenna diagrams for a much used base station antenna. In a specific case, such an antenna is on a mast 22 m above the ground. Determine the intensity at ground level at a distance of 30 m from the mast (measured along the ground) relative to the intensity 500 m from the antenna in that direction in the horizontal plane where the intensity is greatest. [Hint 1: The radial direction in the antenna diagram indicates relative intensity in the number of dB for waves that go out in different directions (in a vertical

plane) from the antenna. Hint 2: We can only specify *relative* intensities. In the right part of the figure, the origin indicates that the intensity is 40 dB lower than in the maximum direction (0°).]
(b) Perform the same calculation if instead a single dipole antenna was used (use Fig. 15.14).
(c) Is it favourable that the base station antenna has the intensity profile it has, or would it be advantageous if a single dipole antenna was used instead?

23. (a) Use the "bandwidth theorem" (the classic counterpart of Heisenberg's uncertainty relationship) to determine the width of the frequency spectrum of a laser pulse using 40 fs (femtoseconds) to pass a point in space (image from the laser is shown on the first page of this chapter). The wavelength corresponding to the centre of the frequency spectrum is 810 nm. [The pulses have a near Gaussian envelope curve in both time and frequency.]
(b) How many wavelengths are there in a pulse?
(c) How "long" is the light pulse (measured along the direction of propagation)? [It will get rather messy if you mix the theory of relativity into such calculations, so we recommend that you stick to a nonrelativistic description.]
(d) Repeat the same calculations as above for a 7.7 fs pulse. [This is a laser currently used in Munich in their attempt to create attosecond laser. Interested is referring to a notice on 7 May 2015 in http://www.Photonics.com headed: "Laser Design Brings Attosecond Spectroscopy Closer" (http://www.photonics.com/Article.aspx?AID=57412 accessed 10 May 2015.)
It is interesting to see what the researchers behind this work write:
"*This field of ultrafast physics focuses on phenomena such as electron motions in molecules and atoms, which can take place on attosecond time scales. The ability to generate attosecond laser pulses would effectively permit electron motions to be 'photographed'.*" As you can see, exciting development is taking place in physics today!
24. Search the Internet to get an impression of the status of lasers in the X-ray region. How far has the development come? Which applications will an X-ray laser find?

Reference

1. XuPanda, https://en.wikipedia.org/wiki/Helium-neon_laser#/media/File:HeNe_Laser_Levels.png. Accessed April 2018