Grating Spectrometer

GRATING.TEX KB 20020119

KLAUS BETZLER[1], FACHBEREICH PHYSIK, UNIVERSITÄT OSNABRÜCK

This short lecture note recalls some of the well-known properties of spectrometers. As an addition to textbooks, it may present some help to students working with such instruments. It is neither intended as a substitute for textbooks in optics nor as a comprehensive overview over the field of spectroscopy.

1 Grating Mathematics

1.1 Transfer Function

As a general rule, the diffraction transfer function of an array of regularly ordered elements is the product of the interference function of the regular structure with the diffraction function of a single element. Applied to a grating, the transfer function is the product of the interference function of the one-dimensional regular structure of equidistant points with the diffraction function of a single groove. We will here restrict our considerations on the interference function. Furthermore, we limit the calculations to a spectrum of a single sharp line (fixed wavevector k_i).

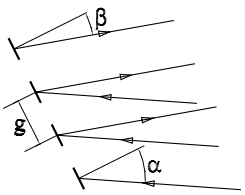

Figure 1: Grating geometry, g is the distance between two adjacent elements, α and β are the angles of incident and reflected beam measured versus the normal on the grating plane.

The grating is illuminated by a plane wave of frequency ω_i and wavevector k_i. In large enough distance after passing the grating we again can assume plane waves. The wave generated by one element n can be written as

$$E_{n,i} = E^0_{n,i} \cdot \exp(j\omega_i t - jk_i r - j\varphi_{n,i}) \tag{1}$$

where $E^0_{n,i}$ is the respective amplitude and $\varphi_{n,i}$ is the phase difference to the first element. This phase difference is the product of the geometrical path difference Δ_n and the wavevector k_i, where the path difference depends on the geometry (see Fig. 1):

$$\varphi_{n,i} = \varphi_{n,i}(g,\alpha,\beta) = k_i \Delta_n(g,\alpha,\beta) = k_i n \delta(g,\alpha,\beta) \quad . \tag{2}$$

[1]KLAUS.BETZLER@UOS.DE

Here, $\delta(g, \alpha, \beta)$ is the path difference between the waves from two adjacent grooves

$$\delta = \delta(g, \alpha, \beta) = g \cdot (\sin\alpha + \sin\beta) \quad . \tag{3}$$

The amplitude of the resulting wave is the sum over all m elements

$$E_i = \sum_{n=1}^{m} E_{n,i} = \exp(j\omega_i t - jk_i r) \cdot \sum_{n=1}^{m} E_{n,i}^0 \exp(-j\varphi_{n,i}) \quad . \tag{4}$$

The complex amplitude of this plane wave as a function of the deflection angle is given by

$$E_i^0 = E_i^0(\delta) = \sum_{n=1}^{m} E_{n,i}^0 \exp(-j\varphi_{n,i}) = \sum_{n=1}^{m} E_{n,i}^0 \exp(-jk_i n\delta) \tag{5}$$

E_i^0 thus is a function of the path difference δ, i. e. of the grating geometry. Eq.5 in principle is a Fourier transform between k-space and real space.

From Eq. 5 the well-known $\frac{\sin(mx)}{\sin(x)}$ formula can be derived using the fact that $\exp(ax) = (\exp(x))^a$.

Putting $E_{n,i}^0 = \text{const.} = 1$ and $A = \exp(-jk_i\delta)$, Eq. 5 transforms to

$$E_i^0(\delta) = \sum_{n=1}^{m} A_i(\delta)^n = \frac{1}{A} \cdot \frac{1 - A^m}{1 - A} \quad . \tag{6}$$

Furthermore

$$\frac{1 - A^m}{1 - A} = \frac{A^{m/2}}{A^{1/2}} \cdot \frac{A^{m/2} - A^{-m/2}}{A^{1/2} - A^{-1/2}} \tag{7}$$

and

$$A^{m/2} - A^{-m/2} = 2j \sin(m/2 \cdot k_i\delta) \quad , \tag{8}$$

$$A^{1/2} - A^{-1/2} = 2j \sin(1/2 \cdot k_i\delta) \quad . \tag{9}$$

Thus

$$E_i^0 \propto \frac{\sin(m/2 \cdot k_i\delta)}{\sin(1/2 \cdot k_i\delta)} \quad . \tag{10}$$

This function has major peaks where the denominator is zero, i. e. at angles where $\delta = N \cdot \lambda$, N is called the order of the interference. Minor peaks occur at peaks of the numerator, i. e. where the argument of the numerator is an odd multiple of $\pi/2$.

To plot it, one usually calculates the absolute value or the intensity from the complex amplitude. Fig. 2 shows two such typical transfer functions for gratings with $m = 15$ and $m = 45$, respectively.

Fig. 2 shows that the halfwidths of the major peaks are inversely proportional to the number of grooves in the grating. Here, the ratio of the two halfwidths is $3 : 1$, corresponding to groove numbers of 15 and 45, respectively. Yet another

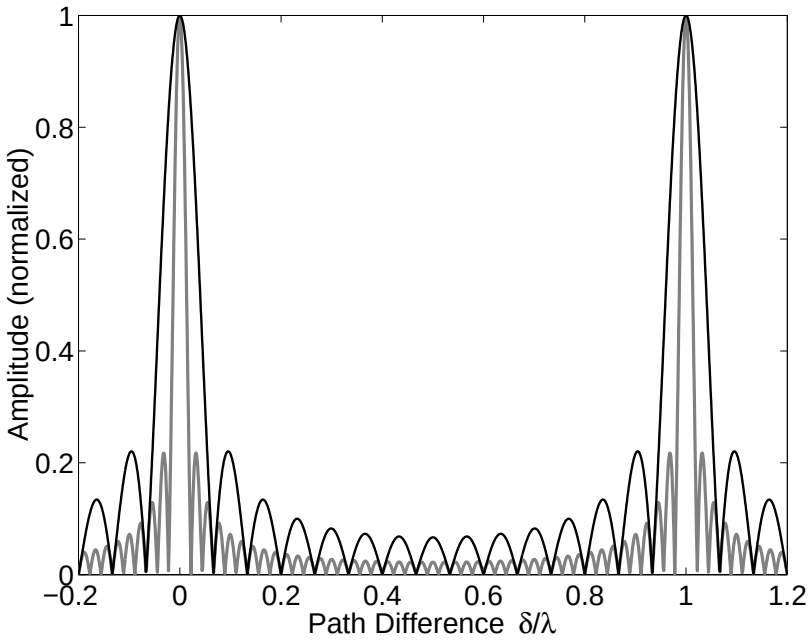

Figure 2: Transfer function of a grating with 15 lines (black) and with 45 lines (gray).

interesting fact can be observed in Fig. 2: The amplitude of the first minor peak stays constant albeit the mean amplitude of the minor peaks decreases – as expected – with increasing groove number. The explanation is obvious, near the major peaks the amplitude function behaves like $(\sin(x))/x$, independent of the number of lines of the grating. The position of the first minor peak is $x = 3\pi/2$, its normalized amplitude thus $2/(3\pi) = 0.2122$.

1.2 Resolution

For defining the resolution of an optical instrument, as a rule of thumb one assumes that two elements can be just resolved when one maximum coincides with the first adjacent minimum of the other element. For the transfer function of a grating this situation is visualized in Fig. 3.

Zeroth, first, and second order are shown, in the first order the two lines just can be resolved.

According to Eq. 10 the peaks of the two lines (wavevectors k_1 and k_2) lie at

$$\frac{1}{2}k_{1,2}\delta_{1,2} = N\pi \tag{11}$$

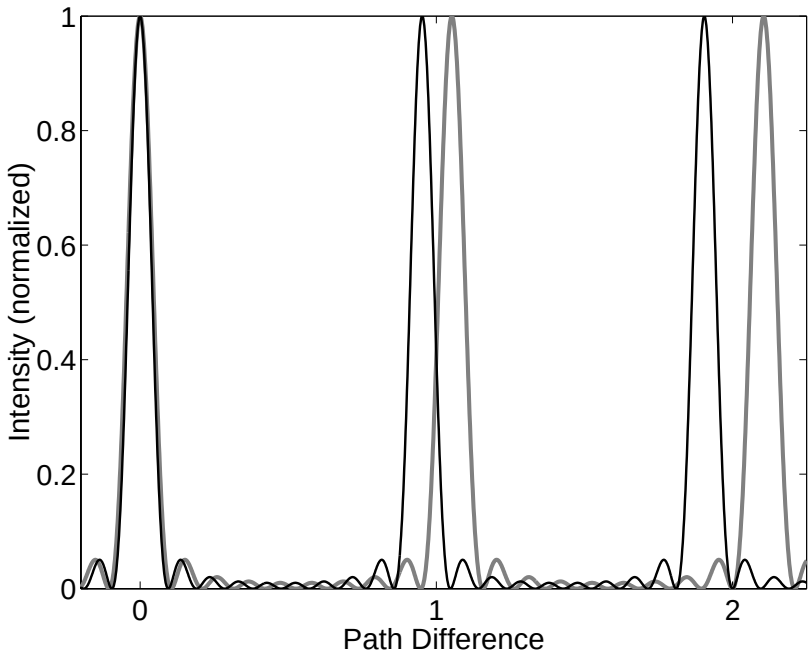

Figure 3: Transfer functions (intensities) of a grating with 10 grooves for two different wavelengths.

where N is the interference order. The first minima of the lines at

$$\frac{m}{2}k_{1,2}\delta_{2,1} = mN\pi \pm \pi \tag{12}$$

have to coincide with the respective maxima. In that way the rule-of-thumb resolution limit is defined. To get it, we e. g. may take the maximum of line one and the minimum of line two, eliminate δ_1 using Eqs. 11 and 12, furthermore substitute k for k_1 and $k + \Delta k$ for k_2. This results in

$$mN\frac{k + \Delta k}{k} = mN + 1 \tag{13}$$

or finally in

$$\frac{k}{\Delta k} = mN \quad , \tag{14}$$

the well-known formula for the resolution limit of a grating, number of grooves times interference order.

It should be noted that only the product is important, thus with a 600 lines/mm grating one might achieve a better resolution as with a 1200 lines/mm grating of the same size, when one can use third order instead of first. Yet at higher orders the spectral working range is reduced accordingly.

The number of grooves m can be written as the quotient of grating width B and grating constant g (distance between adjacent grooves). Moreover, for usual high

resolution applications of gratings in optics, the interference order N is chosen such that $g/N \approx \lambda$, the wavelength used. Putting this to Eq. 14 and using wavenumbers $\tilde{\nu} = 1/\lambda$ instead of wavevectors $k = 2\pi/\lambda$ one arrives at the simplest formula for the resolution

$$\Delta\tilde{\nu} = 1/B \quad . \tag{15}$$

This formula – more or less – is valid for the resolution of all optical instruments when one takes B as the maximum path difference imposed by the instrument to the light under investigation. For a grating this maximum path difference equals approximately the width of the grating.

1.3 Blazing, Echelette Grating

To improve the efficiency of a grating the grooves can be structured in a suitable way. One possibility to achieve this is to incline the individual reflective elements of the grating structure by an angle γ as shown in Fig. 4 (*echelette* grating).

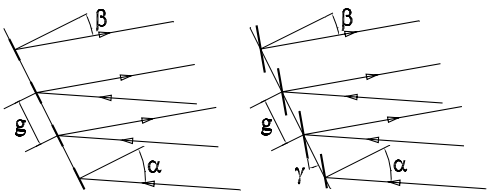

Figure 4: Conventional grating (left) and echelette grating (right). Note that the geometry parameters for the interference function are defined in the same way in each structure.

This angle γ is called the *blaze* angle of the grating. The corresponding wavelength (*blaze* wavelength) is defined as (Eq. 3 for $\alpha = \beta = \gamma$)

$$\lambda_{\mathrm{blaze}} = \delta(g, \alpha = \beta = \gamma) = 2 \cdot g \cdot \sin\gamma \quad . \tag{16}$$

This *blaze* wavelength doesn't shift much for $\alpha \neq \beta$. With $\alpha = \gamma + \epsilon$ and $\beta = \gamma - \epsilon$ (reflection formula) λ_{blaze} shifts to

$$\lambda_{\mathrm{blaze}}(\epsilon) = g \cdot [\sin(\gamma + \epsilon) + \sin(\gamma - \epsilon)] = 2 \cdot g \cdot \sin\gamma \cdot \cos\epsilon \quad , \tag{17}$$

i. e. to only slightly shorter wavelengths, as in conventional spectrometers ϵ usually is a rather small angle.

From the basic blazing scheme as drawn in Fig. 4 for an ideal *echelette* grating one could expect a comparably sharp blazing behaviour of such a grating. Yet one has to take into account the diffraction function of such a narrow element which corresponds to the diffraction function of a single slit. As the width of one element is in the same order of magnitude as the *blaze* wavelength itself, the behaviour broadens considerably.

1.4 Diamond Mask (Apodization)

The relatively intense minor peaks at the side of the major peaks – discussed at the end of subsection 1.1 – usually are no problem. Yet they may affect measurements where high resolution is necessary.

Such expressed ‘side lobes’ are a well known fact in Fourier transform. To reduce it, one usually has to apply a *windowing* or *apodization* function. These functions reduce the amplitudes towards the edges of the function to be Fourier transformed. Thus the influence of the step-like edges on the transform is minimized. The simplest of these functions is a triangle or – what can be regarded as equivalent – a diamond. Fig. 5 shows the effect of such a diamond shaped apodization function on the spectrum. In spectrometers this apodization function is imposed mechanically by a diamond shaped mask on the grating.

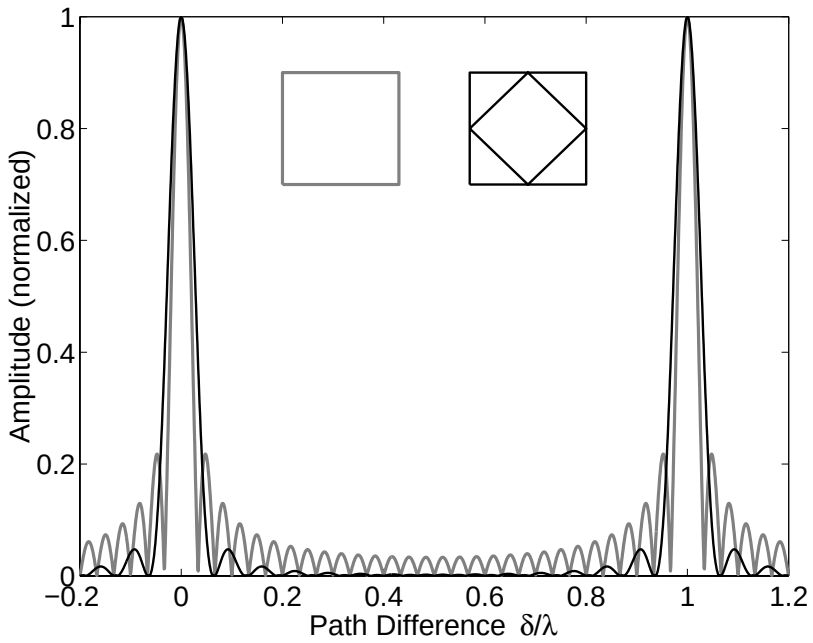

Figure 5: Effect of the ‘diamond mask’ on the side lobes of the transfer function: gray - without, black - with diamond mask.

Even better results can be achieved using smoother functions than the triangle. ‘Smoother’ here means that not only the function value but also the first derivative of the function approaches zero at the borders and that there is no step in the first derivative. A simple function obeying these conditions is the squared sine function. Using it, the results shown in Fig. 6 are achieved.

As has been shown, the application of such apodization masks improves the problem of the side lobes considerably. But their use also decreases the light throughput of an instrument. Both masks discussed here reduce the operating area of a grating

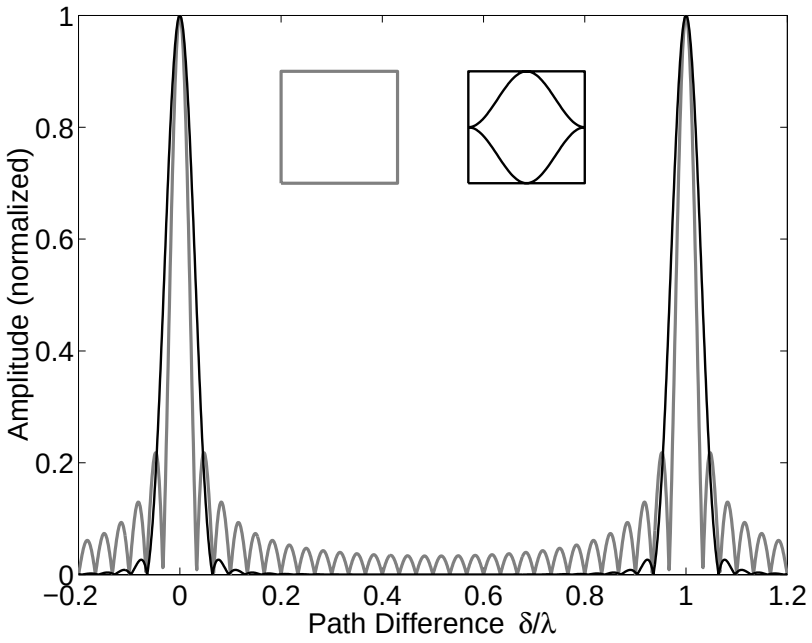

Figure 6: Effect of a sinesquare shaped mask on the side lobes of the transfer function: gray - without, black - with cosine mask.

by a factor of 2 which also affects the amplitude function by approximately this factor. The halfwidth is slightly increased, thus ending up in an intensity loss of around 2. Using the method in both parts of a double spectrometer squares the loss. Therefore one should use the diamond masks only where it's necessary or where it's tolerable with regard to the intensity.

2 Spectrometer

To build a Spectrometer, one has to add imaging optics and slits to the grating. One important parameter to be included is the focus length of the imaging optics.

2.1 Angular Dispersion

The angular dispersion of a grating is the mapping between wavelength variation and the corresponding variation in the deflection angle. Taking the derivative of Eq. 2

$$\frac{d}{d\beta}\{\ \lambda(\beta) = g \cdot \sin\alpha + g \cdot \sin\beta\ \} \tag{18}$$

yields the mapping function wanted

$$\frac{d\lambda}{d\beta} = g \cdot \cos\beta \quad . \tag{19}$$

2.2 Lateral Dispersion

While the angular dispersion still is a property of the grating, the lateral dispersion is a property of the complete instrument, the spectrometer. This property denotes how different wavelengths are laterally dispersed in the exit focal plane of a spectrometer. It is derived from the angular dispersion by multiplying the angle difference $d\beta$ with the focus length f of the optics

$$\frac{d\lambda}{dx} = \frac{d\lambda}{f \cdot d\beta} = \frac{g}{f} \cdot \cos\beta \quad . \tag{20}$$

Eq. 20 is the reason for using large focus lengths, i. e. large spectrometers.

2.3 The 'Real' Spectrometer

To get a sense for the magnitude of typical properties we calculate the data for the Spex Spectrometer used in the Raman spectroscopy experiment. The parameters of the grid used in the spectrometer are 1800 lines/mm and a size of approximately 111 mm (that makes the calculations simple). The focal length of the mirrors used as imaging optics is 0.85 m.

For the resolution limit of the grating we get (we are working in first order)

$$\frac{\tilde{\nu}}{\Delta\tilde{\nu}} = N \cdot m = 1 \times 1800 \times 111 = 200000 \quad . \tag{21}$$

The grating period g is 1/1800 mm = 555 nm. Thus – to make things easy – we will be working at a wavelength of about 555 nm. This yields the typical working angle of the grating

$$\alpha \approx \beta \approx \arcsin(0.5) = 30^\circ \quad . \tag{22}$$

Using Eq. 20 we get the lateral dispersion

$$\frac{d\lambda}{dx} = \frac{g}{f} \cdot \cos\beta = 0.57 \text{ nm/mm} \quad . \tag{23}$$

From the lateral dispersion we can calculate the *bandwidth* by multiplying it with the slit width used. Thus for a slit width of 1 mm the bandwidth is 0.57 nm.

The minimum useful slitwidth is reached when the *bandwidth* equals the resolution limit of the grating. This can be calculated using Eq. 21 and 23:

$$\Delta x = \frac{\lambda}{N \cdot m} \cdot \frac{f}{g \cdot \cos\beta} \quad . \tag{24}$$

The resulting minimum slitwidth for $\lambda = 555$ nm is approximately 5 μm.

2.4 Focus

When adjusting a spectrometer, one of the last adjustments done is that of the focus. To get optimal resolution, the entrance slit has to be exactly imaged to the plane of the exit slit. To get a sense for the accuracy necessary, it is useful to consider the geometrical conditions. We assume that we can tolerate a misadjustment which produces an image equal to the minimum useful slitwidth instead of a sharp line. From the geometry in Fig. 7 we deduce

$$d : b = f : B \tag{25}$$

which yields $d \approx 40$ μm. As two mirrors are moved in parallel, an accuracy of 20 μm in the mechanical adjustment has to be reached.

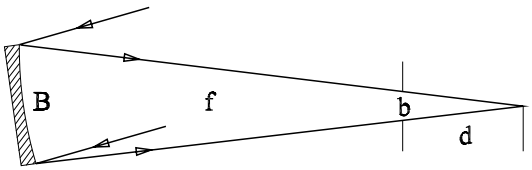

Figure 7: Imaging geometry at the exit slit.

2.5 Diffraction at Apertures

The resolution and throughput of an optical instrument might be limited by diffraction effects at limiting apertures. In a spectrometer, two apertures should be taken into account, the optics, i. e. grating or mirror, and the slits.

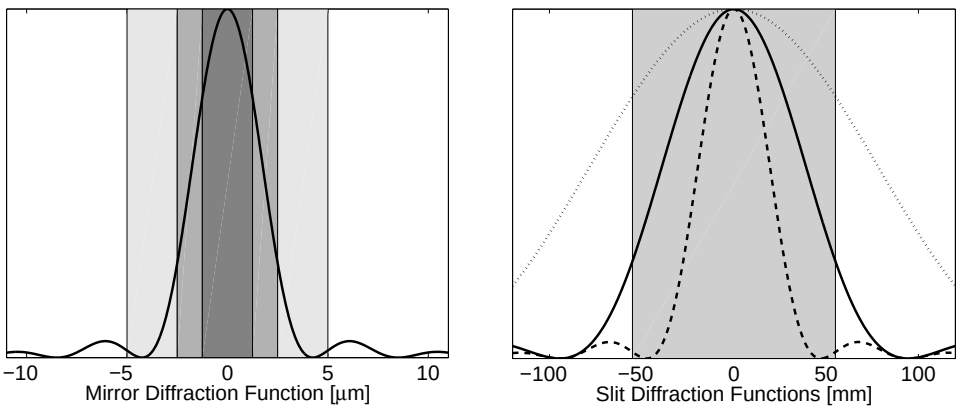

Figure 8: Diffraction at spectrometer apertures. Left: Intensity distribution at the slit position due to diffraction at the imaging optics. For comparison three slit widths are sketched, 2.5, 5, 10 μm. Right: Slit diffraction functions for three different slit widths, dotted – 2.5, full – 5, dashed – 10 μm. The shaded area represents the width of the optics (mirror or grating).

The width of the diffraction maxima should not exceed the width of the complementary apertures to get optimal resolution and throughput. Fig.8 shows the calculated intensity functions. In the left picture the diffraction function from the mirror at the slit position is shown, for comparison three different slit widths are shaded: 2.5, 5, and 10 μm, respectively. The right picture shows the diffraction functions for just these slit widths at the mirror positions, the mirror width is shaded. Both pictures of course are complementary to each other due to the same mathematics. As can be seen, again 5 μm turns out to be the smallest useful width.

3 Annex: MATLAB Programming

Most of the calculation and the function drawing is done using MATLAB, as an example the script for Fig. 6 is shown here.

```
apod = inline('sin(phi).*sin(phi)','phi');
PH = linspace(-pi/2.5,2*pi+pi/2.5,1601);
M = 30;
N = [1:M];
[ph,n] = meshgrid(PH,N);
E = exp(i*n.*ph);
Total = abs(sum(E,1))/M;
plot (PH/2/pi,Total,'Color',0.5*[1,1,1],'Linewidth',2);
hold on;
E = apod(n*pi/(M+1)).*exp(i*n.*ph);
Total = abs(sum(E,1));
Total = Total/max(Total);
plot (PH/2/pi,Total,'Color',0*[1,1,1],'Linewidth',1.5);
hx=XLabel('Path Difference  \delta/\lambda');
hy=YLabel('Amplitude (normalized)');
set([gca,hx,hy],'Fontunits','normalized','Fontsize',0.05);
% now draw the insets
x = [0.2,0.43,0.43,0.2,0.2];
y = [0.7,0.7,0.9,0.9,0.7];
plot (x,y,'Color',0.5*[1,1,1],'Linewidth',2);
x = x+0.37;
plot (x,y,'Color',0*[1,1,1],'Linewidth',1.5);
x = linspace(0.57,0.8,100);
phi = (x-0.57)*pi/0.23;
y = [0.8+0.1*apod(phi);0.8-0.1*apod(phi)];
plot (x,y,'Color',0*[1,1,1],'Linewidth',1.5);
hold off;
```