

An Interactive Shader for Natural Diffraction Gratings

Bachelorarbeit

der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Michael Single

2014

Leiter der Arbeit:
Prof. Dr. Matthias Zwicker
Institut für Informatik und angewandte Mathematik

Abstract

In nature, animals exhibit structural colors because of the physical interaction of light with the nanostructures of their skin. In his pioneering work, J.Stam developed a reflectance model based on wave optics capturing the effect of diffraction from surface nanostructures. His model is dependent on an accurate estimate of the correlation function using statistical properties of the surface's height field. We propose an adaption of his BRDF model that can handle complex natural gratings directly. Furthermore, we describe a method for interactive rendering of diffraction effects due to interaction of light with biological nanostructures such as those on snake skins. Our method uses discrete height fields of natural gratings acquired by using atomic force microscopy (AFM) as an input and employs Fourier Optics to simulate far-field diffraction. Based on a Taylor Series approximation for the phase shifts at the nanoscale surface, we leverage the precomputation of the discrete Fourier Transformations involved in our model, to achieve interactive rendering speed (about 5-15 fps). We demonstrate results of our approach using surface nanostructures of two snake species, namely the *Elaphe* and the *Xenopeltis* species, when applied to a measured snake geometry. Lastly, we evaluate the quality of our method by comparing its (peak) viewing angles with maximum reflectance for a fixed incident beam with those resulting from the grating equation at different wavelengths. We conclude that our method produces accurate results for complex, natural gratings at interactive speed.

Contents

1	Theoretical Background	1
1.1	Basics in Modelling Light in Computer Graphics	1
1.1.1	Radiometry	1
1.1.2	Spectral Energy	1
1.1.3	Spectral Power	2
1.1.4	Spectral Irradiance	2
1.1.5	Spectral Radiance	2
1.1.6	BRDF	3
1.1.7	Wavespectrum and Colors	4
1.1.8	Colorspace	5
1.1.9	Spectral Rendering	7
1.1.10	Rendering Equation	7
1.2	Wave Theory for Light and Diffraction	8
1.2.1	Basics in Wave Theory	8
1.2.2	Wave Interference	9
1.2.3	Wave Coherence	11
1.2.4	Huygen's Principle	12
1.2.5	Waves Diffraction	12
1.3	Stam's BRDF formulation	14
2	Derivations	18
2.1	Problem Statement and Challenges	18
2.2	Approximate a FT by a DFT	19
2.2.1	Reproduce FT by DTFT	19
2.2.2	Spatial Coherence and Windowing	21
2.2.3	Reproduce DTFT by DFT	22
2.3	Adaption of Stam's BRDF for Discrete Height Fields	24
2.3.1	Reflected Radiance of Stam's BRDF	24
2.3.2	Relative Reflectance	25
2.4	Optimization using Taylor Series	27
2.5	Spectral Rendering using DFTs	29
2.6	An Alternative Approach	30
2.6.1	PQ factors	30
2.6.2	Sinc Interpolation in Frequency Domain	32

3 Results	34
3.1 BRDF maps	34
3.2 Rendering Surface Geometries	45
3.2.1 A Comparison with Stam	50
3.2.2 Merge with above subsec	52
Bibliography	54

Chapter 1

Theoretical Background

1.1 Basics in Modelling Light in Computer Graphics

1.1.1 Radiometry

One purpose of Computer Graphics is to simulate the interaction of light with a surface and how a real-world observer, such as a human eye, will perceive this. These visual sensations of an eye are modelled relying on a virtual camera which captures the emitted light from the surface. The physical basis to measure such reflected light is studied under radiometry which deals with measures on the electromagnetic radiation transferred from a source to a receiver.

Fundamentally, light is a form of energy propagation, consisting of a large collection of photons, whereat each photon can be considered as a quantum of light that has a position, a direction of propagation and a wavelength λ . A photon travels at a certain speed $v = c/n$, that depends only the speed of light c and the refractive index n through which it propagates. Its frequency is defined by $f = v/\lambda$ and its carried amount of energy q , measured in the SI unit Joule, is given by $q = hf = h\nu/\lambda n$ where h is the Plank's constant. The total energy of a large collection of photons is hence $Q = \sum_i q_i$.

1.1.2 Spectral Energy

It is important to understand that the human eye is not equally sensitive to all wavelength of the spectrum of light and therefore responds differently to specific wavelengths. Remember that our goal is to model the human visual perception. This is why we consider the energy distribution of a light spectrum rather than considering the total energy of a photon collection since then we could weight the distribution according to the human visual system. So the question we want to answer is: How is the energy distributed across wavelengths of light?

One idea is to make an energy histogram from a given photon collection. For this we have to order all photons by their associated wavelength, discretize the wavelength spectrum, count all photons which then will fall in same wavelength-interval, and then, finally, normalize each interval by the total energy Q . This will give us a histogram which tells us the spectral energy Q_λ for a given discrete λ interval and thus models the so called spectral energy distribution¹.

¹Intensive quantities can be thought of as density functions that tell the density of an extensive quantity at an infinitesimal by a small interval or a point.

1.1.3 Spectral Power

Rendering an image in Computer Graphics corresponds to capturing the color sensation of an illuminated, target scene at a certain point in time. Each color is associated with a wavelength and is directly related to a certain amount of energy. In order to determine the color of a to-be-rendered pixel of an image, we have to first get a sense of how much light (in terms of energy) passes through the area which the pixel corresponds to. We begin by considering the flow of energy $\Phi = \frac{\Delta Q}{\Delta t}$ transferred through this area over a unit period of time. This allows us to measure the energy flow through a pixel during a certain amount of time. In general, power is the estimated rate of energy production for light sources and corresponds to the flux. It is measured in the unit Watts, denoted by Q . Since power is a rate over time, it is well defined even when energy production is varying over time. As with Spectral Energy for rendering, we are really interested in the spectral power $\Phi_\lambda = \frac{\Delta \Phi}{\Delta \lambda}$, measured in Watts per nanometer.

1.1.4 Spectral Irradiance

Before we can tell how much light is reflected from a given point on a surface towards the viewing direction of an observer, we first have to know how much light arrives at this point. Since in general a point has no length, area or even volume associated, let us instead consider an infinitesimal area ΔA around a such a point. Then, we can ask ourself how much light falls in such a small area. Further while observing this process over a short period of time, the measured quantity gives us the the spectral irradiance E as illustrated in figure 1.1. Summarized, this quantity tells us how much spectral power is incident on a surface per unit area and mathematically is equal:

$$E_\lambda = \frac{\Phi_\lambda}{\Delta A} \quad (1.1)$$

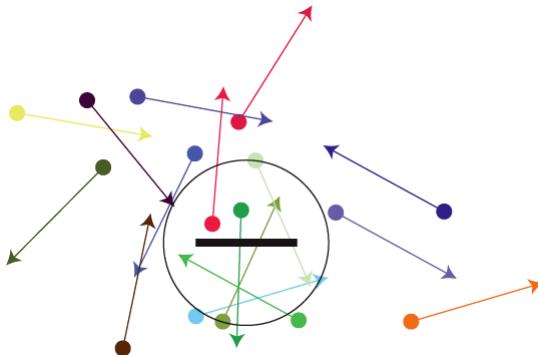
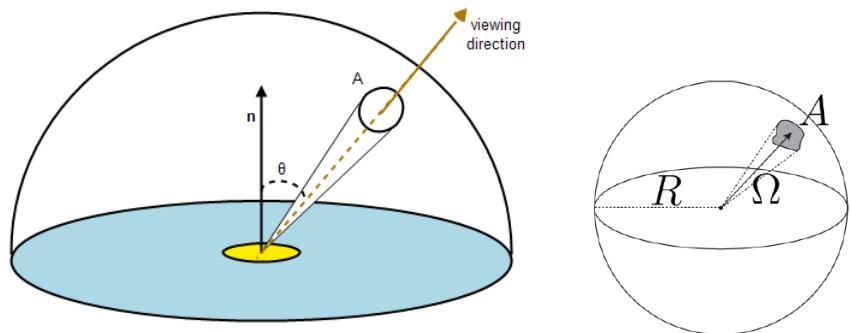


Figure 1.1: Irradiance is the summed up radiance over all directions. The black border is representing a surface element.

1.1.5 Spectral Radiance

When rendering an image we have to determine the color of each pixel of the image. Although irradiance tells us how much light is arriving at a point and gets reflected, it tells us nothing about the power distribution across different directions. The direction of the element is important because the human eye may perceive the brightness of an illuminated objects differently when looking from direction.



(a) Radiances is the density of photons per area per solid angle

(b) Solid angle is the area of a surface patch on a sphere with radius R which is spanned by a set of directions

This concept is described by the radiometric quantity radiance. Basically, radiance is the measure of light energy passing through or emitted from a small area around a point on a surface towards a given direction during a short period in time. More formally this is the spectral power emerging from an arbitrary point (an infinitesimal area around this point) and falls within a given solid angle (see figure² 1.2(b)) in a specific direction (usually towards the observer) as shown in figure 1.2(a). Formally, this leads us to the following mathematical formalism:

$$L_\lambda(\omega) = \frac{d^2\Phi_\lambda}{dAd\Omega} \approx \frac{\Phi_\lambda}{\Omega A} \quad (1.2)$$

where L is the observed spectral radiance in $\text{Wm}^{-2}\text{sr}^{-1}$ in direction ω , Φ_λ is the total power emitted, θ is the angle between the surface normal and the specified direction, A is the area of the surface and Ω is the solid angle subtended by the observation or measurement.

It is useful to distinguish between radiance incident at a point on a surface and excitant from that point. Terms for these concepts sometimes used in the graphics literature are surface radiance L_r for the radiance *reflected* from a surface and field radiance L_i for the radiance *incident* at a surface.

1.1.6 BRDF

In order to render the colorization of an observed object, a natural question in Computer Graphics is what portion of the incident light a viewer will receive after reflection, after when he looks at an illuminated object. For any given surface which is illuminated from a certain direction ω_i , we can ask ourselves how much light is reflected from any point on this surface towards a viewing direction ω_r . This is where the Bidirectional Reflectance Distribution Function (BRDF) comes into play, which is a radiometric quantity telling us how much light is reflected at an opaque surface. Mathematically speaking, the BRDF is the ratio of the reflected radiance pointing in the direction ω_r to the incident irradiance coming from the inverse direction of ω_i as illustrated in figure 1.2. Hence the BRDF is a four dimensional function defined by four angles θ_i , ϕ_i , θ_r and ϕ_r .

Formally, a BRDF for any given wavelength λ is defined as:

²Modified from a figure in Computer Graphics class 2012 in chapter *Colors*

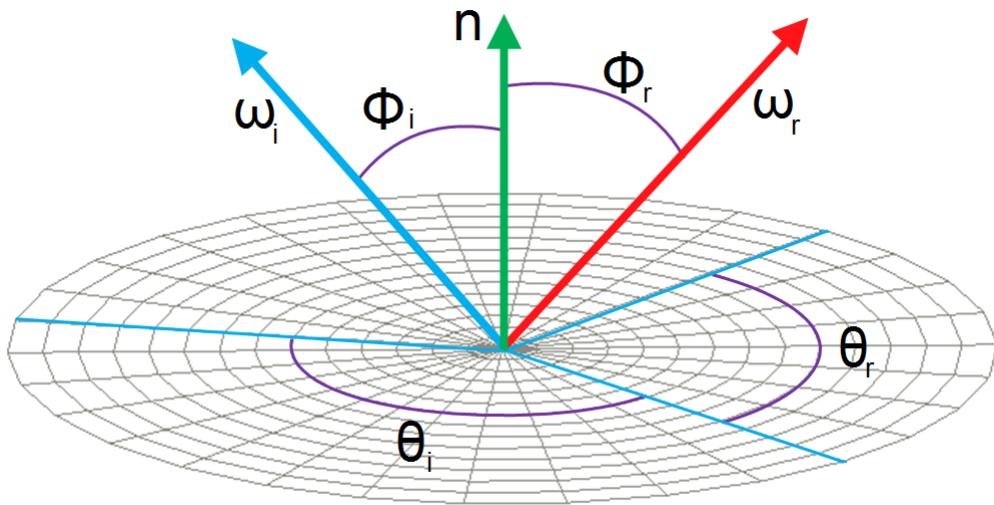


Figure 1.2: Illustration of the BRDF model, where ω_i is pointing to the light source and the viewing direction is denoted by ω_r . Both unit direction vectors are defined w.r.t to a surface normal \mathbf{n} for every point on the surface.

$$\begin{aligned} BRDF_\lambda(\omega_i, \omega_r) &= \frac{dL_r(\omega_r)}{dE_i(\omega_i)} \\ &= \frac{dL_r(\omega_r)}{L_i(\omega_i)\cos(\theta_i)d\omega_i} \end{aligned} \quad (1.3)$$

Where L_r is the reflected spectral radiance, E_i is the incident spectral irradiance and θ_i is the angle between ω_i and the surface normal \mathbf{n} . Also, L_i is the incident spectral radiance.

1.1.7 Wavespectrum and Colors

In order to see how crucial the role of human vision plays, let us consider the following definition of color by *Wyszeckiu and Siles*³ stating that "*Color is the aspect of visual perception by which an observer may distinguish differences between two structure-free fields of view of the same size and shape such as may be caused by differences in the spectral composition of the radiant energy concerned in the observation*". Therefore, similarly like the humans' perceived sensation of smell and taste, color vision is just another individual sense of perception giving us the ability to distinguish between different frequency distributions of light which we experienced as different colors.

³mentioned in Computer Graphics Fundamentals Book from the year 2000

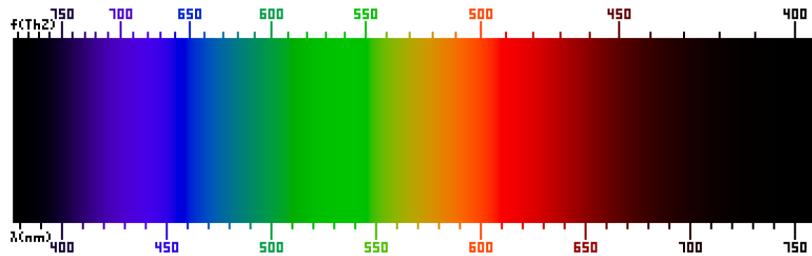


Figure 1.3: Frequency (top) and wavelength (bottom) of colors of the visible light spectrum⁴.

In general, an eye consists of photoreceptor cells which are responsible for providing ability of color-perception. A schematic of an eye is illustrated in figure 1.4. Basically, there are two specialized types of photoreceptor cells, cone cells which are responsible for color vision and rod cells, which allow an eye to perceive different brightness levels.

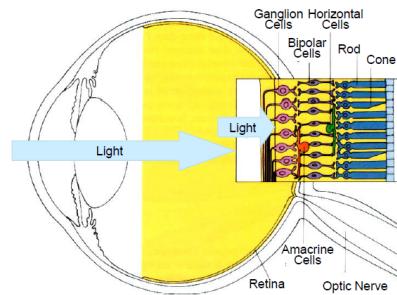


Figure 1.4: Schematic⁵ of photoreceptor cells, cones and rods, in a human eye

A human eye is made of three different types of cone cells, having their peak sensitivity in different wavelength ranges. More precisely, there are cone cells most sensitive to short wavelengths between 420nm and 440nm , those which are most sensitive in the middle range between 530nm and 550nm and those which have their peak in the long range, from 560nm to 580nm . Therefore, any color sensation in human color perception can therefore be described by just three parameters, corresponding to the levels of stimulus of these three types of cone cells.

1.1.8 Colorspace

In order to render accurate images of how a human observer sees its world, a mathematical model of the human color perception is required. Remember that color sensation is due to a visual stimulus processed by cone cells in an eye. A human eye contains three different types of cone cells. Therefore, one possible approach is to describe each kind of these cone cells with a function mapping each wavelength to a certain sensitivity. In the early 1920 from a series of experiments, the so called CIE RGB color space was derived. This space describes the response of cone cells of an average human individual, the so called standard observer. Basically, a statistically sufficiently large number of probands were exposed to different target light colors expressed by their wavelength. The task of each proband was to reproduce these target colors by mixing three given primary colors, red, green and blue light. The strength of each primary color could be manually adjusted by

⁴Similar figure like used in computer graphics class 2012 in chapter colors

⁵image of illustration has been taken from http://en.wikipedia.org/wiki/Bidirectional_reflectance_distribution_function

setting their relative sensitivity. Those adjustment weights have been measured, aggregated and averaged among all probands for each primary color. This model describes each color as a triple of three real valued numbers⁶, the so called tristimulus values. Summarized, these experiments provided certain weights of primary colors in order to match a color at a certain wavelength according to the average human color perception. However, some of these weights could have a negative value.

The disadvantage of the CIE RGB colorspace is that some of its color weights are negative. Thus, scientist derived the CIE XYZ colorspace which has no negative color matching functions but is still additive⁷. Figure 1.5 visualizes the matching functions of the CIE XYZ space. Another property of the CIE XYZ space is that its Y component is representing the luminance of the corresponding color. Usually, the CIE XYZ space is used as a reference colorspace to define colorspace transformations.

Pragmatically speaking, color spaces describe the range of colors a camera can see, a printer can print or a monitor can display. Thus, formally we can define it as a mapping from a range of physically produced colors from mixed light to a standard objective description of color sensations registered in the eye of an observer in terms of tristimulus values.

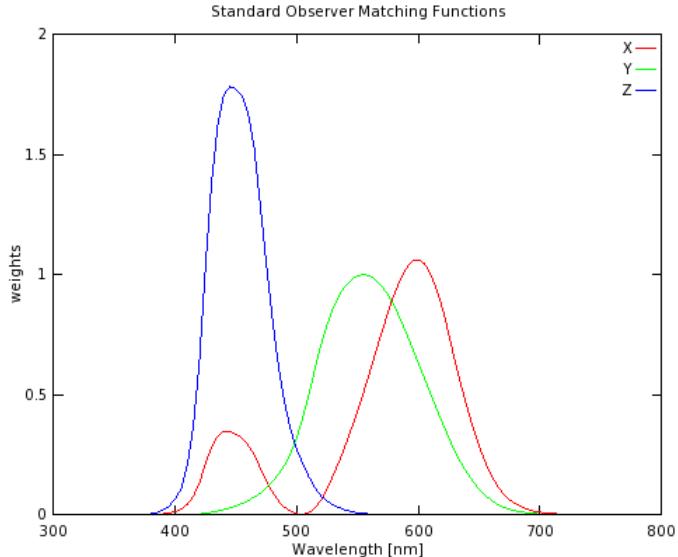


Figure 1.5: Plots of our color matching functions we used for rendering

Interpolating all measured tristimuli values gives us three basis functions, the CIE color matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$. In figure 1.5 are the numerical description of the chromatic response of the observer. They can be thought of as the spectral sensitivity curves of three linear light de-

⁶note that there are negative color weights possible in the CIE RGB colorspace. This is why some human perceived color sensations could not be reconstructed using just an additive color model (adding three positively weighted primary values). Therefore, a proband was also allowed to move one of the primary colors to the target color and instead was supposed to reproduce this new color mix using the two remaining primaries (subtractive model). The value of the selected, moved primary was then interpreted as being negative weighted in an additive color model.

⁷Remember, the property of an additive colorspace is that any color can be represented as a weighted sum of matching functions of that color space.

tectors yielding the CIE tristimulus values X, Y and Z.

The tristimulus values for a color with a spectral power distribution $I(\lambda)$, are given in terms of the standard observer by:

$$\begin{aligned} X &= \int_{\Lambda} I(\lambda) \bar{x}(\lambda) d\lambda \\ Y &= \int_{\Lambda} I(\lambda) \bar{y}(\lambda) d\lambda \\ Z &= \int_{\Lambda} I(\lambda) \bar{z}(\lambda) d\lambda \end{aligned} \quad (1.4)$$

where λ is the wavelength of the equivalent monochromatic light spectrum $\Lambda = [380nm, 780nm]$. Note that it is not possible to build a display that corresponds to the CIE XYZ colorspace. For this reason it is necessary to design other color spaces, which are physically realizable, efficiently encoded, perceptually uniform and have an intuitive color specification. There are simple conversions between XYZ color spaces to other color space - such as the RGB colorspace - described as linear transformations.

1.1.9 Spectral Rendering

When rendering an image, most of the time we are using colors described in a certain RGB color space. However, a RGB colorspace results from a colorspace transformation of the tristimulus values, which themselves are inherent to the human visual system. Therefore, many physical phenomenon are poorly modelled when they rely on RGB colors for rendering. Using only RGB colors for rendering is like assuming that a given light source emits light of only three particular wavelengths. But in reality this is rarely the case. Spectral rendering refers to using certain wavelength spectrum, for e.g. the human visible light spectrum, instead of simply using only the range of RGB values in order to render an illuminated scene. This captures the physical reality of specific light sources way more accurately. Keep in mind that even when we make use of a spectral rendering approach, we have to convert the final spectra to RGB color values when we want to display an image on an actual display.

1.1.10 Rendering Equation

As discussed previously, colors are associated to radiance. Since we are starting with Stam's BRDF⁸ formulation but want to perform a simulation rendering structural colors, we have to reformulate this BRDF equation such that we will end up with an identity of the reflected spectral radiance. This is where the rendering equation comes into play. Let us assume we are given an incoming light source directional at a solid angle ω_i and θ_i is its angle of incidence and that ω_r is the solid angle for the viewing direction. Further let λ denote the wavelength⁹ and Ω is the hemisphere of integration for the incoming light. Then, we are able to formulate a $BRDF_{\lambda}$ by using its definition 1.3:

⁸Remember that a BRDF is the portion of a incident light source reflected off a given surface towards a specified viewing direction.

⁹Notice that, to keep our terms simple, we have dropped all λ subscripts for spectral radiance quantities.

$$\begin{aligned}
f_r(\omega_i, \omega_r) &= \frac{dL_r(\omega_r)}{L_i(\omega_i)\cos(\theta_i)d\omega_i} \\
\Rightarrow f_r(\omega_i, \omega_r)L_i(\omega_i)\cos(\theta_i)d\omega_i &= dL_r(\omega_r) \\
\Rightarrow \int_{\Omega} f_r(\omega_i, \omega_r)L_i(\omega_i)\cos(\theta_i)d\omega_i &= \int_{\Omega} dL_r(\omega_r) \\
L_r(\omega_r) &\Rightarrow \int_{\Omega} f_r(\omega_i, \omega_r)L_i(\omega_i)\cos(\theta_i)d\omega_i
\end{aligned} \tag{1.5}$$

The last equation is the so called rendering equation . We assume that our incident light is a directional, unpolarized light source like sunlight and therefore its radiance is given as

$$L_{\lambda}(\omega) = I(\lambda)\delta(\omega - \omega_i) \tag{1.6}$$

where $I(\lambda)$ is the intensity of the relative spectral power for the wavelength λ . By plugging the identity in equation 1.6 into our current rendering equation 1.5, we will get:

$$\begin{aligned}
L_{\lambda}(\omega_r) &= \int_{\Omega} BRDF_{\lambda}(\omega_i, \omega_r)L_{\lambda}(\omega_i)\cos(\theta_i)d\omega_i \\
&= BRDF_{\lambda}(\omega_i, \omega_r)I(\lambda)\cos(\theta_i)
\end{aligned} \tag{1.7}$$

where $L_{\lambda}(\omega_i)$ is the incident radiance and $L_{\lambda}(\omega_r)$ is the radiance reflected by the given surface. Note that the integral in equation 1.7 vanishes since $\delta(\omega - \omega_i)$ is only equal one if and only if $\omega = \omega_i$.

1.2 Wave Theory for Light and Diffraction

1.2.1 Basics in Wave Theory

In order to prepare the reader for relevant concepts in physics which are used later for derivations and reasonings within this thesis, I am going to provide a quick introduction to the basics of wave theory and related concepts. In physics, a wave describes a disturbance that travels from one location to another through a certain medium. The disturbance temporarily displaces the particles in the medium from their rest position which results in an energy transport along the medium during wave propagation. Usually, when talking about waves we are actually referring to a complex valued function which is a solution to the so called *wave equation* which is modelling how the wave disturbance proceeds in space during time.

There are two types of waves, (a) mechanical waves which deform their mediums during propagation like sound waves and (b) electromagnetic waves consisting of periodic oscillations of an electromagnetic field, such as light. As illustrated in figure 1.6, there are several properties someone can use and apply in order to compare and distinguish different waves:

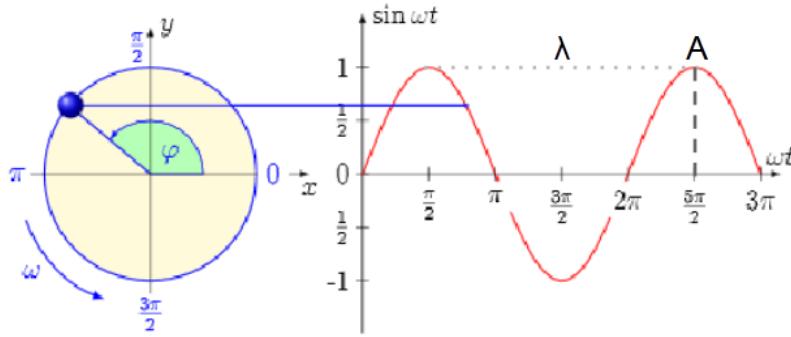


Figure 1.6: Simplified, one dimensionally real valued wave function¹⁰, giving an idea about some important wave properties. We denote the crest of a wave as the highest point relative to the equilibrium line (zero height along time axis) and similarly the trough as the lowest point.

Wavelength: It is usually denoted by λ and is a measure for the spatial distance from one point to another until the shape of a wave repeats

Amplitude: It is denoted by A and there are two possible interpretations: Firstly, it is a measure of the height from the equilibrium point to the highest point of a crest on the wave or the lowest point of a trough. This means that the amplitude can be positive or negative. However, usually, someone is just interested in the absolute value of an amplitude, i.e. the magnitude of a wave. For light waves it is a relative measure of intensity or brightness to other light waves of the same wavelength. And secondly, it can be interpreted as a measure how much energy a wave carries whereat the greater the absolute amplitude value, the bigger the amount of energy being carried.

Frequency: Is a measure of the number of waves which are passing through a particular point in the propagation medium during one unit of time and is denoted by f .

Phase: It is denoted by φ . It describes either the offset of initial position of a wave or the relative displacement between or among waves having the same frequency. Two waves with same frequency are said to be in phase if they have the same phase. This means that they line up everywhere. As a remark, we denote by ω the angular frequency which is equal $2\pi f$.

A geometrical property of waves is their wavefront. This is either a surface or line along the path of wave propagation on which the disturbance at every point has the same phase. Basically, a wavefront can have any kind of shape although three prominent types of wavefronts are: spherical-, cylindrical- and plane wavefronts. If a point in an isotropic medium is sending out waves in three dimensions, then the corresponding wavefronts are spheres, centered on the source point. Hence spherical wavefront is the result of a spherical wave, also denoted as a wavelet. Note that for electromagnetic waves, the phase is a position of a point in time on a wavefront cycle (motion of wave over a whole wavelength) whereat a complete cycle is defined as being equal to 360° .

1.2.2 Wave Interference

Next, after having seen that a wave is simply a traveling disturbance along a medium, having some special properties, someone could ask what happens when there are several waves traveling on the

¹⁰Image source: <http://neutrino.ethz.ch/Vorlesung/FS2013/index.php/vorlesungsskript>

same medium. Especially, we are interested in how these waves will interact with each other. In physics the term interference denotes the interaction of waves when they encounter each other at a point along their propagation medium. At each point where two waves superpose, their total displacement is the sum of the displacements of each individual wave at that point. Then, the resulting wave is having a greater or lower amplitude than each separate wave and we can interpret the interference as the addition operates on waves. Two extreme scenarios are illustrated in figure 1.7 for waves with same frequency and equal amplitude. There are basically three variants of interferences which can occur, depending on how crest and troughs of the waves are matched up:

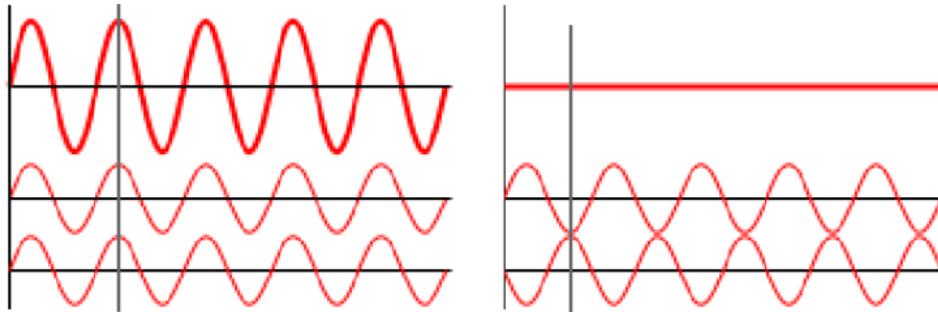


Figure 1.7: Interference scenarios¹¹ when two waves meet: On the left hand-side, there is constructive interference and on the right hand-side there is destructive interference illustrated.

- A crest of a wave meets a crest of another wave and similarly a trough meets a trough of another wave. This scenario is denoted as constructive interference and occurs at any location along the medium where the two interfering waves have a displacement in the same direction. This is like saying that the phase difference between the waves is a multiple of 2π . Then the resulting amplitude at that point is being much larger than the amplitude of an individual wave. For two waves with an equal amplitude interfering constructively, the resulting amplitude is twice as large as the amplitude of an individual wave.
- A crest of a wave meets a trough of another wave and vice versa. This scenario is denoted as destructive interference and occurs at any location along the medium where the two interfering waves have a displacement in the opposite direction. This is like saying that the phase difference between the waves is an odd multiple of π . Then the waves completely cancel each other out at any point they superimpose.
- If the phase difference between two waves is intermediate between the first two scenarios, then the magnitude of the displacement lies between the minimal and maximal values which we could get from constructive interference.

Keep in mind that when two or more waves interfere with each other, the resulting wave will have a different frequency. For a wave, having a different frequency also means having a different wavelength. Therefore, this directly implies that a light of a different color, than its source waves have, is emitted.

¹¹Image source: [http://en.wikipedia.org/wiki/Interference_\(wave_propagation\)](http://en.wikipedia.org/wiki/Interference_(wave_propagation))

1.2.3 Wave Coherence

When considering waves which are traveling on a shared medium along the same direction, we could examine how their phase difference is changing over time. Formulating the change in their relative phase as a function of time will provide us a quantitative measure of the synchronization between those two waves, the so called wave coherence. In order to better understand this concept, let us consider a perfectly mathematical sine wave and a second wave which is a phase-shifted replica of the first one. A property of these mathematical waves is that they keep their shape over an infinity amount of time (i.e. propagated wavelengths). In our scenario, both waves are traveling along the same direction on the same medium, as illustrated in figure 1.8.

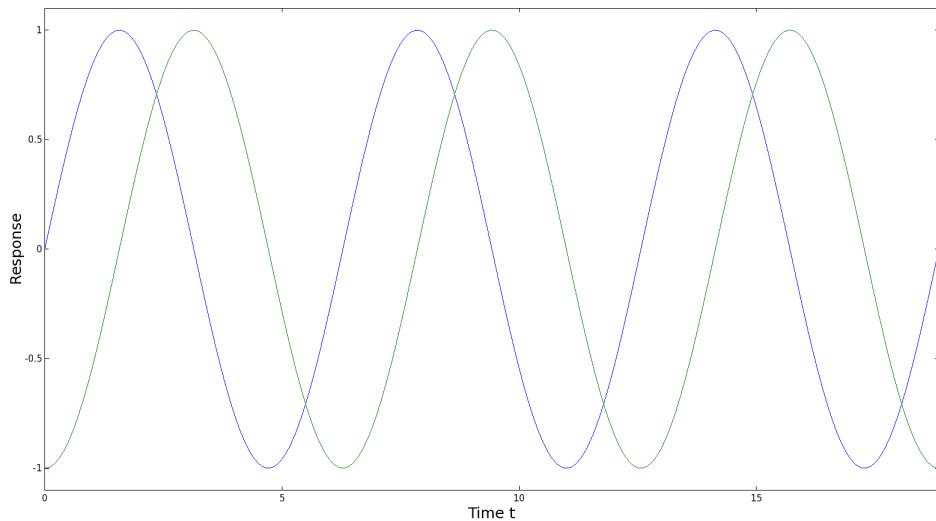


Figure 1.8: Two mathematical sine waves which are perfectly coherent which means that their phase difference is constant for every point in time.

Taking the difference between these two sine waves always yields a constant number. Therefore, those two waves are said to be coherent and hence perfectly synchronous over time. Notice that this scenario is completely artificial since in nature there are no mathematical sine waves. Rather, the phase difference is then a function of time $p(t)$. The more coherent two waves are, the slower this function will change over time. In fact, two waves are said to be coherent if they are of the same frequency, are temporally in phase or have the same amplitude at every point in time. Thus two waves are coherent if they are generated at the same time, having the same frequency, amplitude, and phase. Conversely, waves are considered incoherent or also asynchronous if they have no stable phase difference. This means $p(t)$ is heavily varying over time. Coherence describes the effect of whether waves will tend to interfere with each other constructively or destructively at a certain point in time and space. Thus this is a property of waves that enables stationary interference. The more correlated two waves are, the higher their degree of coherence is. In physics coherence between waves is quantified by the cross-correlation function, which basically predicts the value of a second wave using the value of the first one. There are two basic coherence classifications:

- Spatial coherence is dealing with the question of what is the range of distance between two points in space in the span of a wave for which there is occurring a significant effect of stationary interference when averaged over time. This is formally answered by considering the

correlation between waves at different point in space. The range of distance with significant coherence is also denoted as the coherence area.

- Temporal coherence examines how well two waves which are observed at two different moments in time correlate with each other. Thus it may be used for predicting how well a wave interferes temporally with itself. Mathematically, this kind of coherence is computed by measuring the correlation between the value of the wave and the delayed version of itself. The coherence time denotes the maximum time delay for which the waves are coherent. The distance a wave has traveled during the coherence time is denoted as their temporal coherence length.

1.2.4 Huygen's Principle

Besides the phases and the amplitudes of waves, their propagation directly affects the interaction between different waves and how they could interfere with each other. This is why it makes sense to formulate a model which allows us to predict the position of a moving wavefront and how it moves in space. This is where *Huygen's Principle* comes into play. It states that any point of a wavefront may be regarded as a point source that emits spherical wavelets in every direction. Within the same propagation medium, these wavelets travel at the same speed as their source wavefront. The position of the new wavefront results by superimposing all of these emitted wavelets. Geometrically, the surface that is tangential to the secondary waves can be used in order to determine the future position of the wavefront. Therefore, the new wavefront encloses all emitted wavelets. Figure 1.9 visualizes Huygen's principle for a wavefront reflected off from a plane surface.

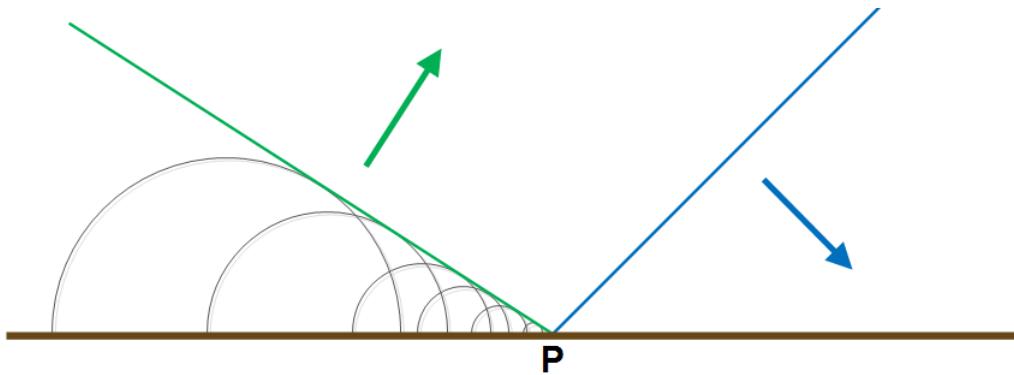


Figure 1.9: A moving wavefront (blue) encounters an obstacle (a surface shown in brown colors) and produces a new wavefront (green) as a result of superposition of all secondary wavelets.

1.2.5 Waves Diffraction

Revisiting Hugen's Principle we know that each point on a wavefront can be considered as a source of a spherical wavelet which propagates in every direction. But what exactly happens when a wave's propagation direction is only partially occluded by an object? What will be the outcome on applying Huygen's Principle to this case? An example scenario for this case is shown in figure 1.10.

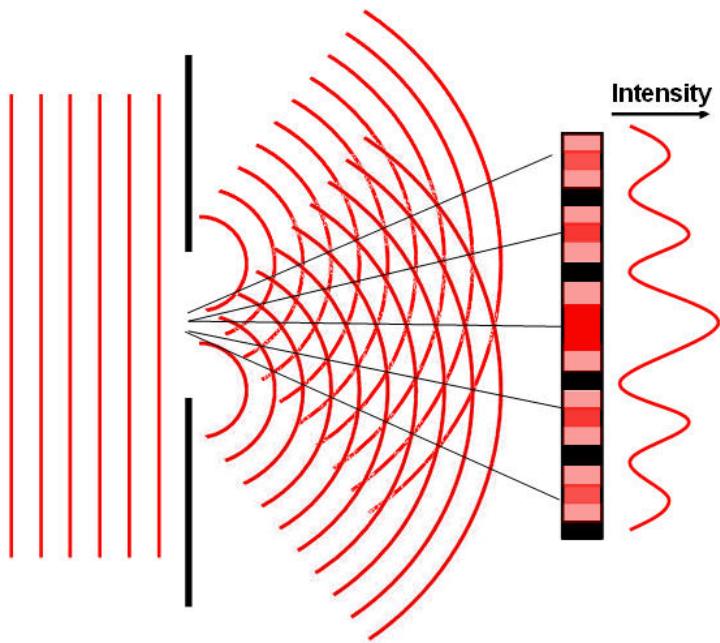


Figure 1.10: Illustration¹² of a diffraction scenario in which a plane wavefront passes through a surface with a certain width and how the wave will be bent, also showing the intensity of the resulting wave along a straight line in its path.

Whenever a propagating wavefront is partially occluded by an obstacle, the wave is not only moving in the direction along its propagation, but is also bent around the edges of the obstacle. In physics, this phenomenon is called diffraction. Waves are diffracted due to interference which occurs among all wavelets when applying Huygen's Principle for the case when a wavefront hits an obstacle. Generally, the effect of diffraction is most pronounced for the waves whose wavelength is roughly similar in size to the dimension of the occluding object. Conversely, if the wavelength is much smaller in size, then there is almost no wave diffraction perceivable at a far off distance. This relationship between the strength of wave diffraction and the wavelength is conceptually illustrated in figure 1.11 when a wave is transmitted through an opening in a surface. A reflective example for diffraction provided in figure 1.9.

¹²Image source:http://cronodon.com/images/Single_slit_diffraction_2b.jpg

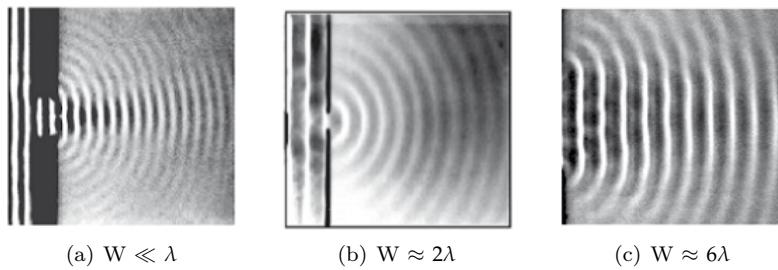


Figure 1.11: Illustration¹³ of how diffraction changes when a wave with wavelength λ propagates through a slit of width equal W .

In everyday's life, we can see the direct outcome of the effect of wave diffraction in form of structural colors. There are examples from nature such as the iridescent colors on various snake skins as well as artificial examples such as the colorful patterns notable when having a close look at an illuminated compact disc. All these examples comprise a surface made of highly regular nanostructures which diffract an incident light significantly. Such a nanostructure which exhibits a certain degree of regularity is also denoted as a diffraction grating. Further information about diffraction gratings can be found in section ??.

1.3 Stam's BRDF formulation

The theoretical foundation of this thesis is based on the pioneering work of J.Stam who derived a BRDF formulation to model the effect of far field diffraction for various analytical anisotropic surfaces, relying on the so called scalar wave theory of diffraction for which a wave is assumed to be a complex valued scalar. It's noteworthy that Stam's BRDF formulation does not take into account the polarization of the incident light. Fortunately, light sources like sunlight and light bulbs are unpolarized. The principal behind J. Stam's approach is illustrated in figure 1.12.

¹³Image taken from: <http://neutrino.ethz.ch/Vorlesung/FS2013/index.php/vorlesungsskript>, chapter 9, figure 9.14

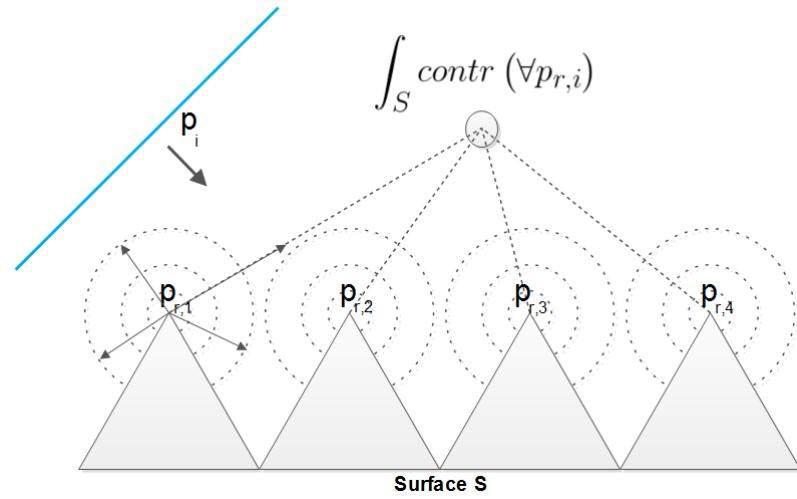


Figure 1.12: Illustration of secondary wavelets reflected off a surface. An integration over all secondary sources resulting from an incident wave according to Huygen's principle will give us an identity for the total contribution at a certain point in space.

An incident wave p_i from a light source encounters a surface representing a diffraction grating. According to Huygen's Principle, at any point i on the grating at which the incident wave meets the grating a secondary, spherical wavelet $p_{r,i}$ will be emitted. A viewer, indicated by a gray circle in the figure, will perceive the superimposed contribution of all wavelets along the surface S (in the figure indicated by a integration symbol), which will directly follow the laws of wave interference. Therefore the resulting color which an observer sees is the final radiance at that point which reflects from stationary interference of all emitted secondary wavelets and per due to Huygen's principle.

A further assumption in Stam's Paper is, that the emanated waves from the source are stationary, which implies the wave is a superposition of independent monochromatic waves. This further implies that each wave is associated with a definite wavelength λ . Directional light sources such as sunlight fulfill this fact and since we are using these kinds of light sources for our simulations, Stam's model can be used for our modelling purposes.

The main model is the formulate a BRDF as the Fourier Transform applied on the given height field, representing a surface like shown in figure 1.13. The classes of surfaces his model is able to support either exhibit a very regular structure or may be considered as a superposition of bumps forming a periodic like structure. Therefore, the surfaces he is dealing with can either be modelled by probabilistic distributions or have a direct analytical representation. Both cases allow him to derive an analytical solution for his BRDF model.

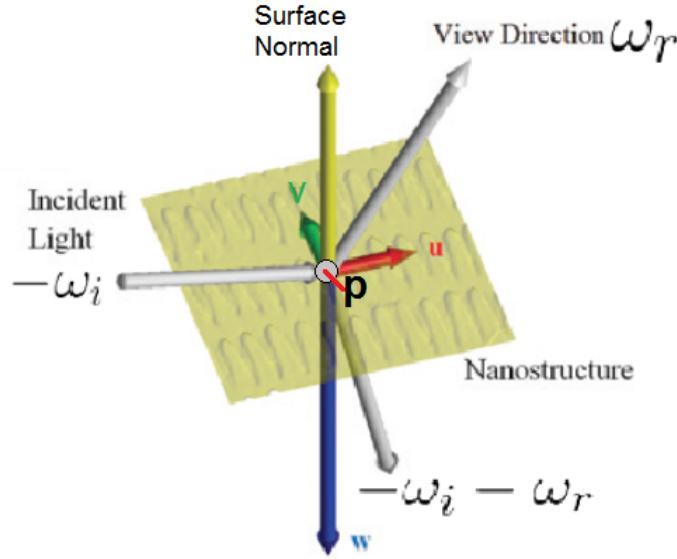


Figure 1.13: Illustration¹⁴ of geometrical setup of Stam's approach where ω_i is a direction, pointing towards the light source, ω_r points towards the camera, n is the surface normal, (u, v, w) are the components of the vector $-\omega_i - \omega_r$.

The direction vector of the secondary wavelet can be computed by taking the difference between the incident and viewing direction like shown in equation 1.8:

$$(u, v, w) = -\omega_i - \omega_r \quad (1.8)$$

These coordinates will later be used in order to compute the total contribution of all secondary sources used in Stam's BRDF in equation 1.11. For simplification, let us introduce an auxiliary function Φ defined in equation 1.9, which models the phase of a wave from the given height field.

$$\Phi(x, y) = \frac{2\pi}{\lambda} wh(x, y) \quad (1.9)$$

Then, any secondary wavelet p which is emitted off from the given surface is equal:

$$p(x, y) = e^{i\Phi(x, y)} \quad (1.10)$$

using the idea presented for figure 1.12 and performing all mathematical steps shown in the appendix ??, will lead us to the final BRDF representation, modelling the total contribution of all secondary sources reflected off the the provided surface h in the direction ω_r :

$$BRDF_\lambda(\omega_i, \omega_r) = \frac{k^2 F^2 G}{4\pi^2 A w^2} \langle |P(ku, kv)|^2 \rangle \quad (1.11)$$

where F denotes the Fresnel coefficient and G is the so called geometry term¹⁵ which is equal to:

$$G = \frac{(1 + \omega_i \cdot \omega_r)^2}{\cos(\theta_i) \cos(\theta_r)} \quad (1.12)$$

¹⁴Modified image which originally has been taken from D.S. Dhillon et. al. poster[DD14].

¹⁵The geometric terms expresses the correction factor to perform an integration over an area instead over a surface. For further information, please have a look at http://en.wikipedia.org/wiki/Surface_integral, and read the definition about *surface element*

One last word about the Fourier transform terms that Stam uses in his derivation: Conventionally, following the definitions of the Fourier Transformation, we are dealing with the inverse Fourier Transformation. However, especially in electrical engineering, it is quite common to define this inverse Fourier transformation by the Fourier Transformation. The reason behind this lies in the fact that we simply could substitute the minus sign as in the following equation 1.13:

$$\begin{aligned}\mathcal{F}_{FT}\{f\}(w) &= \int_{\mathbb{R}^n} f(x)e^{-iwt}dt \\ &= \int_{\mathbb{R}^n} f(x)e^{i\hat{w}t}dt \\ &= \mathcal{F}_{FT}^{-1}\{f\}(w)\end{aligned}\tag{1.13}$$

where \hat{w} is equal $-w$.

The height fields we are dealing with in this work are, however, natural gratings containing a complex shaped nano-structure and hence far from being very regularly aligned. The reason why Stam's approach in its current form is not suitable for our purpose is twofold: First his approach does not capture the complexity of natural gratings accurately well enough when relying on his statistical approaches and secondly it is way too slow in order to be usable for interactive rendering since his BRDF needs an evaluation of a Fourier Transform for every directional changing.

Chapter 2

Derivations

2.1 Problem Statement and Challenges

The goal of this thesis is to perform a physically accurate and interactive simulation of structural color production as shown in figure 2.2, which we can see whenever light is diffracted from a natural grating. For this purpose we need the following input data (see figure 2.1):

- A mesh representing a snake surface¹ as shown in figure 2.1(a).
- A natural diffraction grating represented as a height field, its maximum height and its pixel width².
- A vector field which describes how the given nanostructure patch is oriented on the surface (see figure 2.1(c)).

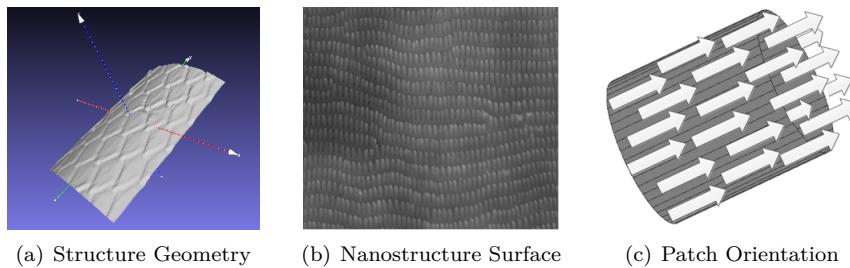


Figure 2.1: Input for our simulation

We want to rely on the integral equation 1.11 derived by J. Stam in his paper [Sta99] about diffraction shaders. This equation formulates a BRDF modelling the effect of diffraction under the assumption that a given grating can either be formulated as an analytical function or its structure is simple enough being modelled relying on statistical methods. These assumptions guarantee that 1.11 has an explicit solution. However, the complexity of a biological nanostructure cannot

¹In our simulation it is an actual reconstruction of a real snake skin. These measurements are provided by the Laboratory of Artificial and Natural Evolution at Geneva. See their website: www.lanevol.org.

²Since the nanostructure is stored as a grayscale image, we need a scale telling us what length and height one pixel corresponds to in this provided image.

sufficiently and accurately modelled simply using statistical methods. This is why interactive computation at high resolution becomes a hard task, since we cannot evaluate the given integral equation on the fly. Therefore, we have to adapt Stam's equation such that we are able to perform interactive rendering using explicitly provided height fields at interactive rates.

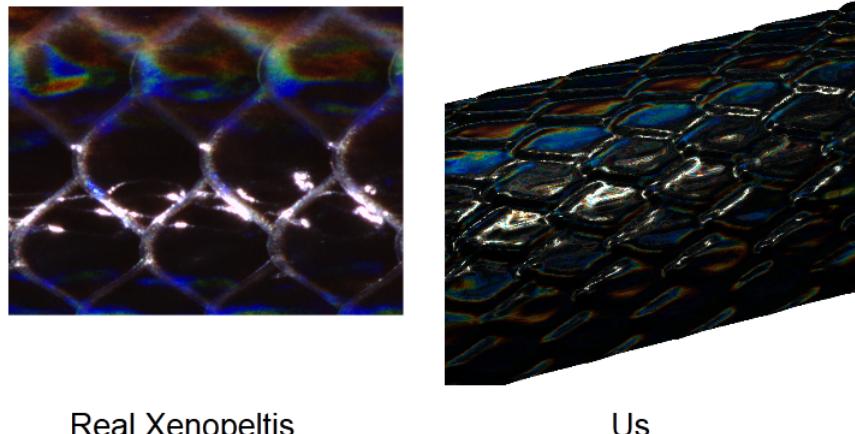


Figure 2.2: Output: Rendered Structural Colors

2.2 Approximate a FT by a DFT

2.2.1 Reproduce FT by DTFT

In the previous section, we have found an identity for the reflected spectral radiance $L_\lambda(\omega_r)$ when using Stam's BRDF for a given input height field. However, the derived expression in equation 2.10 requires to evaluate the Fourier Transform of our height field³ for every direction. In this section we explain how to approximate the FT by the DTFT and apply it to our previous derivations. Figure 2.3 graphically shows how to obtain the DTFT from the FT for a one dimensional signal⁴

The first step is to uniformly discretize the given signal since computers are working finite, discrete arithmetic. We rely on the Nyquist–Shannon sampling theorem tells us how dense we have to sample a given signal $s(x)$ such that can be reconstructed its sampled version $\hat{s}[n]$ ⁵. In particular, a sampled version according to the Nyquist–Shannon sampling theorem will have the same Fourier transform as its original signal when it has a limited bandwidth. The sampling theorem states that if f_{max} denotes the highest frequency of $s(x)$, then, it has to be sampled by a rate of f_s with $2f_{max} \leq f_s$ in order to be reconstructable. By convention $T = \frac{1}{f_s}$ represent the interval length between two samples.

³actually it requires the computation of the inverse Fourier Transform of a transformed version of the given height field, the function $p(x,y)$ defined in equation 1.10.

⁴For our case we are dealing with a two dimensional, spatial signal, the given height field. Nevertheless, without any constraints of generality, the explained approach applies to multi dimensional problems.

⁵Images of function plots taken from http://en.wikipedia.org/wiki/Discrete_Fourier_transform and are modified.

⁶ n denotes the number of samples.

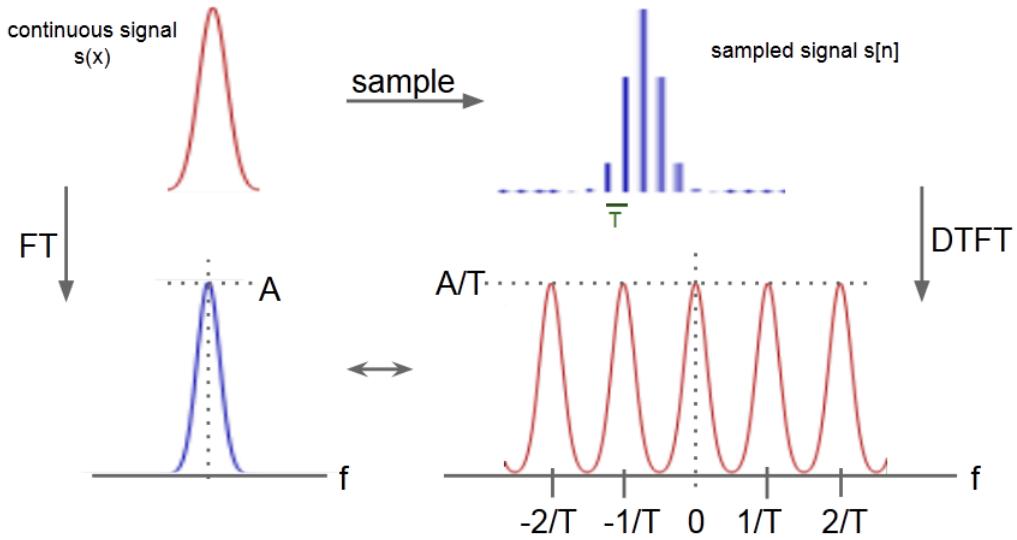


Figure 2.3: Illustration of how to approximate the analytical Fourier Transform (FT)⁵ of a given continuous signal by a Discrete Time Fourier Transform (DTFT). The DTFT applied on a band-limited, discretized signal yields a continuous, periodic response in frequency space.

Next, we apply the Fourier transformation operator on the discretized signal \hat{s} which gives us the following expression:

$$\begin{aligned}
 \mathcal{F}_{FT}\{\hat{s}\}(w) &= \int_{\mathbb{R}} \hat{s}[n] e^{-iwx} dx \\
 &= \int_{\mathbb{R}} \text{mask}(x) s(x) e^{-iwx} dx \\
 &= T \sum_{x=-\infty}^{\infty} \hat{s}[x] e^{-iwx} \\
 &= T \mathcal{F}_{DTFT}\{s\}(w)
 \end{aligned} \tag{2.1}$$

Equation 2.1 tells us that if \hat{s} is sufficiently sampled, then its DTFT corresponds to the FT of $s(x)$. Notice that the resulting DTFT from the sampled signal has a height of $\frac{A}{T}$ where A is the height of the FT of s and thus is a scaled version of the FT.

For a given height field h , let us compute Stam's auxiliary function p defined as in equation 1.10. For the remainder of this thesis we introduce the following definition:

$$P_{dtft} \equiv \mathcal{F}_{DTFT}\{p\} \tag{2.2}$$

Therefore P_{dtft} denotes the DTFT of a transformed version of our height field h ⁷.

⁷By transformed height field we mean $p(x,y) = e^{i\frac{2\pi}{\lambda}wh(x,y)}$ which we get, when we plug h into equation 1.9 and this expression again plug into equation 1.10.

2.2.2 Spatial Coherence and Windowing

Before we can derive a final expression in order to approximate a FT by a DFT, we first have to revisit the concept of coherence introduced in section 1.2.3 of chapter 2. Previously we have seen that wave-theory tells us what is the total contribution of all secondary sources which allows us to say what is the reflected spectral radiance at a certain point in space. This is related to stationary interference which itself depends on the coherence property of the emitted secondary wave sources. The ability for two points in space, t_1 and t_2 , to interfere in the extend of a wave when being averages over time is the so called spatial coherence. The spatial distance between such two points over which there is significant interference is limited by the quantity coherence area. For filtered sunlight on earth this is equal to $65\mu m^8$.

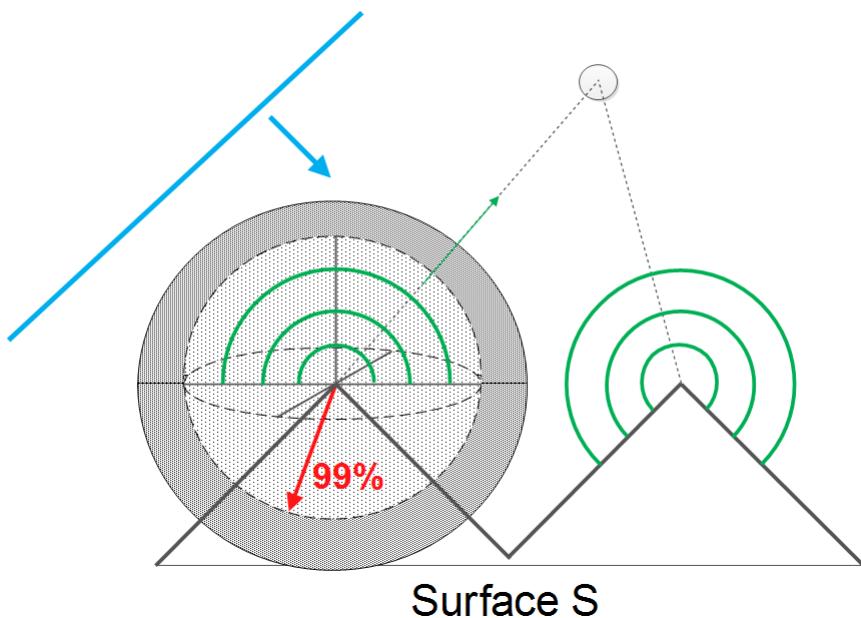


Figure 2.4: A plane wave encounters a surface. According to Huygens principle, secondary wavelets are emitted off from this surface. The resulting wave at a certain point in space (here indicated by a gray circle) depends on the interference among all waves encountering at this position. The amount of significant interference is directly affected by the spatial coherence property of all the wavelets.

Figure 2.4 illustrates the concept of spatial coherence. A wavefront (blue line) encounters a surface. Due to Huygen's principle, secondary wavelets are emitted off from the surface. The reflected radiance at a certain point in space, e.g. at a viewer's eye position (denoted by the gray circle), is a result of interference among all wavelets at that point. This interference is directly affected by the spatial coherence property of all the emitted wavelets.

In physics spatial coherence is predicted by the cross correlation between t_1 and t_2 and usually modelled by a Gaussian Random Process. For any such Gaussian Process we can use a spatial gaussian window $g(x)$ which is equal:

⁸A proof for this number can be looked up in the book Optical Coherence and Quantum Optics[LM95] on page 153 and 154.

$$g(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{x^2}{2\sigma^2}} \quad (2.3)$$

We have chosen standard deviation σ_s of the window such that it fulfills the equation $4\sigma_s = 65\mu m$. This is equivalent to saying that we want to predict about 99.99%⁹ of the resulting spatial coherence interference effects in our model by a cross correlation function.

By applying the Fourier transformation to the spatial window we get the corresponding window in frequency space as:

$$G(f) = e^{-\frac{f^2}{2\sigma_f^2}} \quad (2.4)$$

Notice that this frequency space window has a standard deviation σ_f equal to $\frac{1}{2\pi\sigma_s}$. Those two windows, the spatial- and the frequency space window, will be used in the next section in order to approximate the DTFT by the DFT by a windowing approach.

2.2.3 Reproduce DTFT by DFT

In this section we explain how and under what assumptions the DTFT of a discretized signal¹⁰ can be approximated by a DFT. The whole idea how to reproduce the DTFT by DFT is schematically illustrated in figure 2.5.

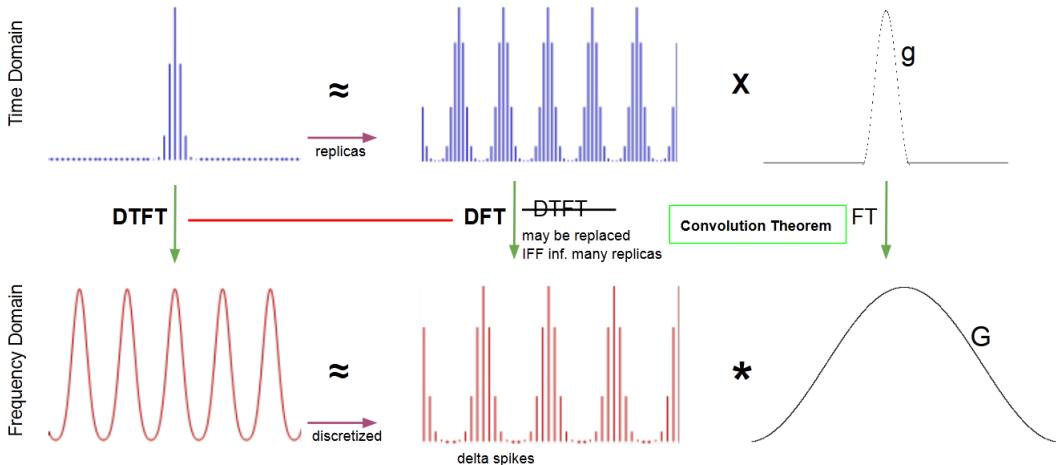


Figure 2.5: Illustration of how to approximate the DTFT¹¹ by the DFT relying on the Convolution Theorem, using a gaussian window function.

Lets say, we are given a spatial, band-limited and discretized one dimensional signal \hat{s} . Our goal is to approximate this spatial signal in a way such that when taking the DFT of this approximated signal, it will yield the same response as taking the DTFT of the original sampled \hat{s} . For

⁹Standard deviation values from confidence intervals table of normal distribution provided by Wolfram MathWorld <http://mathworld.wolfram.com/StandardDeviation.html>.

¹⁰E.g. a sampled signal like already presented in figure 2.3

¹¹Images of function plots taken from http://en.wikipedia.org/wiki/Discrete_Fourier_transform and are modified. Note that the scales in the graphic are not appropriate.

for this purpose we will use the previously introduced concept of gaussian windows and the so called Convolution Theorem which is a fundamental property of all Fourier transformations.

The Convolution Theorem states that the Fourier transformation of a product of two functions, f and g , is equal to convolving the Fourier Transformations of each individual function. Mathematically, this statement corresponds to equation 2.5:

$$\mathcal{F}\{f \cdot g\} = \mathcal{F}\{f\} * \mathcal{F}\{g\} \quad (2.5)$$

The principal issue is how to approximate our given signal \hat{s} . Therefore, let us consider another signal s_N^* which is the N times replicated version of \hat{s} (blue signal at center top in figure).

Remember that, in general, the wave magnitude at a certain point in space is the result of interference among all wavelength meeting at that position. In our scenario, the source of those signals are emitted secondary wavelets. The interference strength between wavelets is related to their spatial coherence. Windowing the signal by a gaussian window g is akin to modelling the effects of spatial coherence interference on the surface. From the previous section 2.2.2 we know that we can use gaussian window like in equation 2.3 in order to approximate such spatial signals interference effects.

Using this insight, we can approximate \hat{s} by taking the product of s_N^* with a gaussian window g . This fact is illustrated in the first row of figure 2.3. So, what will the DTFT of this approximation yield? We already know that the DTFT of \hat{s} is a continuous, periodic signal, since \hat{s} is band-limited. Thus, taking the DTFT of this found approximation should give us approximatively the same continuous, periodic signal.

This is where the convolution theorem comes into play: Applying the DTFT to the product of s_N^* and g is the same as convolving the DTFT of s_N^* by DTFT of g . From equation 2.4 we already know that the DTFT of g is just another gaussian, denoted by G . On the other hand the DTFT of s_N^* yields a continuous, periodic signal. The higher the value of N , the sharper the signal gets (denoted by delta spiked) and the closer it converges toward to the DFT. This is why the DFT is the limit of a DTFT applied on periodic and discrete signals. Therefore, for a large number of N we can replace the DTFT by the DFT operator when applied on s_N^* .

Lastly, we see that the DTFT of \hat{s} is approximately the same like convolving a gaussian window by the DFT of s_N^* . This also makes sense, since convolving a discrete, periodic signal (DFT of s_N^*) by a continuous window function G yields a continuous, periodic function.

In general, for every non-windowed signal, we cannot compute its DTFT ?? numerically due to finite computer arithmetic and hence working with the DFT is our only option. Furthermore, there are numerically fast algorithms in order to compute the DFT values of a function, the Fast Fourier Transformation (FFT). The DFT ?? of a discrete height field is equal to the DTFT of an infinitely periodic function consisting of replicas of the same height field. Now, let a spatial gaussian window g having a standard deviation for which $4\sigma_g$ is equal $65\mu m$. Then, from before, it follows:

$$\mathcal{F}_{dtft}\{\mathbf{s}\} \equiv \mathcal{F}_{dft}\{\mathbf{s}\} * G(\sigma_f) \quad (2.6)$$

Therefore we can deduce the following expression:

$$\begin{aligned}
\mathcal{F}_{dft}\{\mathbf{t}\}(u, v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{dft}\{\mathbf{t}\}(w_u, w_v) \phi(u - w_u, v - w_v) dw_u dw_v \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_i \sum_j F_{dft}\{\mathbf{t}\}(w_u, w_v) \\
&\quad \delta(w_u - w_i, w_v - w_j) \phi(u - w_u, v - w_v) dw_u dw_v \\
&= \sum_i \sum_j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{dft}\{\mathbf{t}\}(w_u, w_v) \\
&\quad \delta(w_u - w_i, w_v - w_j) \phi(u - w_u, v - w_v) dw_u dw_v \\
&= \sum_i \sum_j F_{dft}\{\mathbf{t}\}(w_u, w_v) \phi(u - w_u, v - w_v)
\end{aligned} \tag{2.7}$$

where

$$\phi(x, y) = \pi e^{-\frac{x^2+y^2}{2x_f^2}} \tag{2.8}$$

2.3 Adaption of Stam's BRDF for Discrete Height Fields

2.3.1 Reflected Radiance of Stam's BRDF

We are going to use Stam's main derivation (1.11) for the $BRDF(\omega_i, \omega_r)$ in 1.7 by applying the fact that the wavenumber is equal $k = \frac{2\pi}{\lambda}$:

$$\begin{aligned}
BRDF(\omega_i, \omega_r) &= \frac{k^2 F^2 G}{4\pi^2 A w^2} \langle |P(ku, kv)|^2 \rangle \\
&= \frac{4\pi^2 F^2 G}{4\pi^2 A \lambda^2 w^2} \langle |P(ku, kv)|^2 \rangle \\
&= \frac{F^2 G}{A \lambda^2 w^2} \left\langle \left| P\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle
\end{aligned} \tag{2.9}$$

Going back to equation 1.7 and plugging equation 2.9 into it, using the definition of equation 1.12 and the equation ?? for ω we will get the following:

$$\begin{aligned}
L_\lambda(\omega_r) &= \frac{F^2 (1 + \omega_i \cdot \omega_r)^2}{A \lambda^2 \cos(\theta_i) \cos(\theta_r) \omega^2} \left\langle \left| P\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle \cos(\theta_i) I(\lambda) \\
&= I(\lambda) \frac{F^2 (1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A \omega^2 \cos(\theta_r)} \left\langle \left| P\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle
\end{aligned} \tag{2.10}$$

Note that the Fresnel term F is actually a function of (w_i, w_r) , but in order to keep the equations simple, we omitted its arguments.

So far we just plugged Stam's BRDF identity into the rendering equation and hence have not significantly deviated from his formulation. Keep in mind that P denotes the Fourier transform

of the provided height field which depends on the viewing and incidence light direction. Thus this Fourier Transform has to be recomputed for every direction which will slow down the whole computation quite a lot¹². One particular strategy to solve this issue is to approximate P by the Discrete Fourier Transform (DFT)¹³ and separate its computation such that terms for many directions can be precomputed and then later retrieved by look ups. The approximation of P happens in two steps: First we approximate the Fourier Transform by the Discrete Time Fourier Transform (DTFT) and then, afterwards, we approximate the DTFT by the DFT. For further about basics of signal processing and Fourier Transformations please consult the appendix ??.

Using the insight gained by equation 2.1 allows us to further simplify equation 2.10:

$$\begin{aligned} L_\lambda(\omega_r) &= I(\lambda) \frac{F^2(1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A w^2 \cos(\theta_r)} \left\langle \left| P\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle \\ &= I(\lambda) \frac{F^2(1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A w^2 \cos(\theta_r)} \left\langle \left| T^2 P_{dtft}\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle \end{aligned} \quad (2.11)$$

Where P_{dtft} is a substitute for $\mathcal{F}_{DTFT}\{s\}(w)$. Furthermore T the sampling distance for the discretization of $p(x, y)$ assuming equal and uniform sampling in both dimensions x and y .

2.3.2 Relative Reflectance

In this section we are going to explain how to scale our BRDF formulation such that all of its possible output values are mapped into the range $[0, 1]$. Such a relative reflectance formulation will ease our life for later rendering purposes since usually color values are within the range $[0, 1]$, too. Furthermore, this will allow us to properly blend the resulting illumination caused by diffraction with a texture map.

Let us examine what $L_\lambda(\omega_r)$ will be for a purely specular surface, for which $\omega_r = \omega_0 = \omega_i$ such that $\omega_0 = (0, 0, 1)$. For this specular reflection case, the corresponding radiance will be denoted as $L_\lambda^{spec}(\omega_0)$. We define the relative reflected radiance for our problem 2.10 by simply taking the fraction between $L_\lambda(\omega_r)$ and $L_\lambda^{spec}(\omega_0)$ which is denoted by:

$$\rho_\lambda(\omega_i, \omega_r) = \frac{L_\lambda(\omega_r)}{L_\lambda^{spec}(\omega_0)} \quad (2.12)$$

Notice that the third component w from the vector in equation 1.8 is squared equal to $(\cos(\theta_i) + \cos(\theta_r))^2$ ¹⁴. But first, let us derive the following expression:

¹²Even a fast variant of computation the Fourier Transform has a runtime complexity of $O(N \log N)$ where N is the number of sample.

¹³See appendix ?? for further information about different kinds of Fourier transformations.

¹⁴Consult section ?? in the appendix

$$\begin{aligned}
L_{\lambda}^{spec}(\omega_0) &= I(\lambda) \frac{F(\omega_0, \omega_0)^2 (1 + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix})^2}{\lambda^2 A (\cos(0) + \cos(0))^2 \cos(0)} \langle \left| T_0^2 P_{dtft}(0, 0) \right|^2 \rangle \\
&= I(\lambda) \frac{F(\omega_0, \omega_0)^2 (1+1)^2}{\lambda^2 A (1+1)^2 1} \left| T_0^2 N_{sample} \right|^2 \\
&= I(\lambda) \frac{F(\omega_0, \omega_0)^2}{\lambda^2 A} \left| T_0^2 N_{sample} \right|^2
\end{aligned} \tag{2.13}$$

Where $N_{samples}$ is the number of samples of the DTFT $\langle \cdot \rangle$. Thus, we can plug our last derived expression 2.13 into the definition for the relative reflectance radiance 2.12 in the direction w_r and we get:

$$\begin{aligned}
\rho_{\lambda}(\omega_i, \omega_r) &= \frac{L_{\lambda}(\omega_r)}{L_{\lambda}^{spec}(\omega_0)} \\
&= \frac{I(\lambda) \frac{F(\omega_i, \omega_r)^2 (1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r)} \langle \left| T_0^2 P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}) \right|^2 \rangle}{I(\lambda) \frac{F(\omega_0, \omega_0)^2}{\lambda^2 A} \left| T_0^2 N_{sample} \right|^2} \\
&= \frac{F^2(\omega_i, \omega_r) (1 + \omega_i \cdot \omega_r)^2}{F^2(\omega_0, \omega_0) (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r)} \langle \left| \frac{P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda})}{N_{samples}} \right|^2 \rangle
\end{aligned} \tag{2.14}$$

For simplification and better readability, let us define the following gain-factor:

$$C(\omega_i, \omega_r) = \frac{F^2(\omega_i, \omega_r) (1 + \omega_i \cdot \omega_r)^2}{F^2(\omega_0, \omega_0) (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r) N_{samples}^2} \tag{2.15}$$

Using equation 2.15, we get the following expression for the relative reflectance radiance from equation 2.14:

$$\rho_{\lambda}(\omega_i, \omega_r) = C(\omega_i, \omega_r) \langle \left| P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}) \right|^2 \rangle \tag{2.16}$$

Using the previous definition for the relative reflectance radiance equation 2.12:

$$\rho_{\lambda}(\omega_i, \omega_r) = \frac{L_{\lambda}(\omega_r)}{L_{\lambda}^{spec}(\omega_0)}$$

which we can rearrange to the expression:

$$L_{\lambda}(\omega_r) = \rho_{\lambda}(\omega_i, \omega_r) L_{\lambda}^{spec}(\omega_0) \tag{2.17}$$

Let us choose $L_{\lambda}^{spec}(\omega_0) = S(\lambda)$ such that it has the same profile as the relative spectral power distribution of CIE Standard Illuminant D65 (discussed in ??). Furthermore, when integrating over λ for a specular surface, we should get CIE_{XYZ} values corresponding to the white point for D65. The corresponding tristimulus values using CIE colormatching functions 1.4 for the CIE_{XYZ} values look like:

$$\begin{aligned} X &= \int_{\lambda} L_{\lambda}(\omega_r) \bar{x}(\lambda) d\lambda \\ Y &= \int_{\lambda} L_{\lambda}(\omega_r) \bar{y}(\lambda) d\lambda \\ Z &= \int_{\lambda} L_{\lambda}(\omega_r) \bar{z}(\lambda) d\lambda \end{aligned} \quad (2.18)$$

where \bar{x} , \bar{y} , \bar{z} are the color matching functions. Combining our last finding from equation 2.17 for $L_{\lambda}(\omega_r)$ with the definition of the tristimulus values from equation 2.18, allows us to derive a formula for computing the colors values using Stam's BRDF formula relying on the rendering equation 1.5. Without any loss of generality it suffices to derive an explicit expression for just one tristimulus term, for example Y , the luminance:

$$\begin{aligned} Y &= \int_{\lambda} L_{\lambda}(\omega_r) \bar{y}(\lambda) d\lambda \\ &= \int_{\lambda} \rho_{\lambda}(\omega_i, \omega_r) L_{\lambda}^{spec}(\omega_0) \bar{y}(\lambda) d\lambda \\ &= \int_{\lambda} \rho_{\lambda}(\omega_i, \omega_r) S(\lambda) \bar{y}(\lambda) d\lambda \\ &= \int_{\lambda} C(\omega_i, \omega_r) \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S(\lambda) \bar{y}(\lambda) d\lambda \\ &= C(\omega_i, \omega_r) \int_{\lambda} \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S(\lambda) \bar{y}(\lambda) d\lambda \\ &= C(\omega_i, \omega_r) \int_{\lambda} \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S_y(\lambda) d\lambda \end{aligned} \quad (2.19)$$

Where we used the definition $S_y(\lambda) \bar{y}(\lambda)$ in the last step.

2.4 Optimization using Taylor Series

Our final goal is to render structural colors resulting by the effect of wave diffraction. So far, we have derived an expression which can be used for rendering. Nevertheless, our current equation 2.19 used for computing structural colors, cannot directly be used for interactive rendering, since P_{dtft} had to be recomputed for every change in any direction¹⁵.

In this section, we will address this issue and deliver an approximation for P_{dtft} defined in equation 2.2. This approximation will allow us to separate P_{dtft} in a certain way such that some computational expensive terms can be precomputed. The main idea is to formulate P_{dtft} as a series expansion relying on the definition of Taylor Series, as defined in equation ???. Further, we will provide an error bound for our approximation approach for a given number of terms. Lastly, we will substitute our approximation into our BRDF formula from equation 2.19.

Let us consider $p(x, y) = e^{ikwh(x, y)}$ form Stam's Paper 1.3 where $h(x, y)$ is a given height field and $k = \frac{2\pi}{\lambda}$ denotes the wavenumber of wavelength λ . For any complex number t the power series expansion of the exponential function is equal to:

¹⁵According to changes in viewing- or incident light direction.

$$e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{t^n}{n!} \quad (2.20)$$

Now, when we use the exponent¹⁶ of $p(x, y)$ as an input argument for equation 2.20 we get:

$$\begin{aligned} e^t &= e^{ikwh} \\ &= 1 + (ikwh) + \frac{1}{2!}(ikwh)^2 + \frac{1}{3!}(ikwh)^3 + \dots \\ &= \sum_{n=0}^{\infty} \frac{(ikwh)^n}{n!}. \end{aligned} \quad (2.21)$$

where i is the imaginary unit for complex numbers. For simplification, in the remainder of this section we omitted the arguments of h . Equation 2.21 gives us an expression for an exponential series expansion for the exponent of $p(x, y)$. Please note that the above Taylor series is convergent for any complex valued number. Therefore the equation 2.21 is equal to

$$p(x, y) = \sum_{n=0}^{\infty} \frac{(ikwh(x, y))^n}{n!} \quad (2.22)$$

and thus gives us a series representation of $p(x, y)$. Next, calculating the Fourier transformation \mathcal{F} of equation 2.22 gives us the identity:

$$\begin{aligned} \mathcal{F}\{p\} &\equiv \mathcal{F}\left\{\sum_{n=0}^{\infty} \frac{(ikwh)^n}{n!}\right\} \\ &\equiv \sum_{n=0}^{\infty} \mathcal{F}\left\{\frac{(ikwh)^n}{n!}\right\} \\ &\equiv \sum_{n=0}^{\infty} \frac{(ikw)^n}{n!} \mathcal{F}\{h^n\} \end{aligned} \quad (2.23)$$

Where we have exploited the fact that the Fourier transformation is a linear operator. Therefore, in equation 2.23, we have shown that the Fourier transformation of a series is equal to the sum of the Fourier transformation, applied on each individual series term. Reusing the identifier P ¹⁷ in order to determine the Fourier transformation of p from equation 1.10 equation 2.23 then correspond to:

$$P(\alpha, \beta) = \sum_{n=0}^{\infty} \frac{(ikw)^n}{n!} \mathcal{F}\{h^n\}(\alpha, \beta) \quad (2.24)$$

Up to now we have found a infinity series representation for P_{dtft} . Next we are going to look for an upper bound $N \in \mathbb{N}$ such that

$$\tilde{P}_N(\alpha, \beta) := \sum_{n=0}^N \frac{(ikwh)^n}{n!} \mathcal{F}\{h^n\}(\alpha, \beta) \approx P(\alpha, \beta) \quad (2.25)$$

¹⁶This exponent is a complex valued function, equal to $ikwh(x, y)$.

¹⁷This identifier P may be subscripted by $dtft$ which will denote the DTFT variant of P .

\tilde{P}_N is a good approximation of P , i.e. their absolute difference is small¹⁸. But first, the following two facts would have to be proven¹⁹:

1. Show that there exist such an $N \in \mathbb{N}$ s.t. the approximation of equation 2.25 holds true.
2. Find a value for N s.t. this approximation is below a certain error bound, e.g. close to machine precision ϵ .

Our facts are valid and proven (see appendix ??) implies that there actually exists such an N . Thus, we can make use of the taylor series approximation from equation 2.25 and use it for approximating P_{dft} . This idea allows us to adapt equation 2.19, which is used for computing the structural colors of our BRDF model, in a numerically fast way. Finally the equation for the luminance is equal to:

$$\begin{aligned} Y &= C(w_i, w_r) \int_{\lambda} \left\langle \left| P_{dft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S_y(\lambda) d\lambda \\ &= C(w_i, w_r) \int_{\lambda} \left\langle \sum_{n=0}^N \frac{(wk)^n}{n!} \mathcal{F}\{i^n h^n\} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right\rangle^2 S_y(\lambda) d\lambda \end{aligned} \quad (2.26)$$

Notice that equation 2.26 is constrained by N and hence is an approximation of equation 2.19. Furthermore, it is possible to separate out all the Fourier Terms in the summation and precompute them. This is why the approach in equation 2.26 is fast in order to compute structural color values according to our BRDF model.

2.5 Spectral Rendering using DFTs

In this section we describe how our final model for rendering structural colors due to diffraction will look like. For this purpose we use all our previous findings and plug them together to one big equation. For a given height field h representing the surface of a grating, we want to compute the resulting color due to light diffracted on that grating. For rendering we rely on the CIE_{XYZ} colorspace. For given direction vectors w_i and w_r as shown in figure 1.13 the DFT of the height field is equal to:

$$DFT_n\{h\}(u, v) = F_{dft}\{i^n h^n\} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \quad (2.27)$$

From section 2.2.3 we know that we can reproduce a FT by applying a Gaussian window on the DFT from equation 2.27. This windowing approach gives us:

$$W_n(u, v) = \sum_{(r,s) \in \mathcal{N}_1(u,v)} |DFT_n\{h\}(u - w_r, v - w_s)|^2 \phi(u - w_r, v - w_s) \quad (2.28)$$

where $\phi(x, y) = \pi e^{-\frac{x^2+y^2}{2\sigma_f^2}}$ is the Gaussian window from equation 2.2.3 and $\mathcal{N}_1(u, v)$ denotes an the k-ring neighborhood around (u, v) with k as a parameter.

¹⁸Mathematically speaking, this statement correspond to $||\tilde{P}_N - P|| \leq \epsilon$, where $\epsilon > 0$ is a small number.

¹⁹Please have a look in section ?? in the appendix

In section 2.4 we derived equation 2.26 which tells us how to compute the CIE_{XYZ} color value of a particular color channel using our relative reflectance BRDF model from section 2.3.2. Plugging all these findings together and using our windowing approach, listed in equation 2.28, Then $\forall(u, v, w)$ like (defined in equation 1.8), our final expression for computing structural colors due to diffraction, using all our previous derivations, will be equal:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = C(\omega_i, \omega_r) \int_{\Lambda} \sum_{n=0}^N \frac{(2\pi w)^n}{\lambda^n n!} W_n(u, v) \begin{pmatrix} S_x(\lambda) \\ S_y(\lambda) \\ S_z(\lambda) \end{pmatrix} d\lambda \quad (2.29)$$

Where $C(\omega_i, \omega_r)$ is the defined in equation 2.15. Note that equation 2.29 integrates over a given wavelength spectrum, denoted by Λ . Usually, this Λ is equal to $[\lambda_{min}, \lambda_{max}]$ where $\lambda_{min} = 380nm$ and $\lambda_{max} = 780nm$.

2.6 An Alternative Approach

2.6.1 PQ factors

In this section we are presenting an alternative approach to the previous Gaussian window approach described in section 2.2.3 in order to solve the issue of working with the $DTFT$ instead of DFT . We assume, that a given surface S is covered by a number of replicas of a provided representative surface patch f . In a simplified, one dimensional scenario, mathematically speaking, f is assumed to be a repetitive function, i.e. $\forall x \in \mathbb{R} : S(x) = S(x + nT)$, where T is its fundamental period and $n \in \mathbb{N}_0$. Thus, the surfaces can be written formally as:

$$S(x) = \sum_{n=0}^N f(x + nT) \quad (2.30)$$

What we are looking for is an identity for the Fourier transform²⁰ of our surface S , required in order to simplify the (X, Y, Z) colors from 2.26:

$$\begin{aligned} \mathcal{F}\{S\}(w) &= \int f(x) e^{iwx} dx \\ &= \int_{-\infty}^{\infty} \sum_{n=0}^N f(x + nT) e^{iwx} dx \\ &= \sum_{n=0}^N \int_{-\infty}^{\infty} f(x + nT) e^{iwx} dx \end{aligned} \quad (2.31)$$

Next, apply the following substitution $x + nT = y$ which will lead us to:

$$\begin{aligned} x &= y - nT \\ dx &= dy \end{aligned} \quad (2.32)$$

²⁰Remember that we are using the definition of Fourier Transform used in electrical engineering where \mathcal{F} actually corresponds to the inverse Fourier Transform.

Plugging this substitution back into equation 2.31 we will get:

$$\begin{aligned}
\mathcal{F}\{S\}(w) &= \sum_{n=0}^N \int_{-\infty}^{\infty} f(x + nT) e^{iwx} dx \\
&= \sum_{n=0}^N \int_{-\infty}^{\infty} f(y) e^{iw(y-nT)} dy \\
&= \sum_{n=0}^N e^{-iwnT} \int_{-\infty}^{\infty} f(y) e^{iwy} dy \\
&= \sum_{n=0}^N e^{-iwnT} \mathcal{F}\{f\}(w) \\
&= \mathcal{F}\{f\}(w) \sum_{n=0}^N e^{-iwnT}
\end{aligned} \tag{2.33}$$

We used the fact that the exponential term e^{-iwnT} is a constant factor when integrating along dy and the identity for the Fourier Transform of the function f . Next, let us examine the series $\sum_{n=0}^N e^{-iwnT}$ closer:

$$\begin{aligned}
\sum_{n=0}^N e^{-iwnT} &= \sum_{n=0}^N (e^{-iwT})^n \\
&= \frac{1 - e^{iwT(N+1)}}{1 - e^{-iwT}}
\end{aligned} \tag{2.34}$$

We recognize the geometric series identity for the left-hand-side of equation 2.34. Mainly relying on trigonometric identities, equation 2.33 can be further simplified to:

$$\mathcal{F}\{S\}(w) = (p + iq)\mathcal{F}\{f\}(w) \tag{2.35}$$

where p and q are defined as:

$$\begin{aligned}
p &= \frac{1}{2} + \frac{1}{2} \left(\frac{\cos(wTN) - \cos(wT(N+1))}{1 - \cos(wT)} \right) \\
q &= \frac{\sin(wT(N+1)) - \sin(wTN) - \sin(wT)}{2(1 - \cos(wT))}
\end{aligned} \tag{2.36}$$

Please notice, all derivation steps can be found in the appendix in section ??.

Now lets consider our actual problem description. Given a patch of a nano-scaled surface snake shed represented as a two dimensional height field $h(x, y)$. We once again assume that this provided patch is representing the whole surface S of our geometry by some number of replicas of itself. Therefore, $S(x, y) = \sum_{n=0}^N h(x + nT_1, y + mT_2)$, assuming that the given height field has the dimensions T_1 by T_2 . In order to derive an identity for the two dimensional Fourier transformation of S we can similarly proceed like we did to derive equation 2.35.

$$\mathcal{F}\{S\}(w_1, w_2) = (p + iq)\mathcal{F}_{DTFT}\{h\}(w_1, w_2) \quad (2.37)$$

Note that a detailed derivation of equation 2.37 can be found in the appendix in section ?? and we have defined :

$$\begin{aligned} p &:= (p_1 p_2 - q_1 q_2) \\ q &:= (p_1 p_2 + q_1 q_2) \end{aligned} \quad (2.38)$$

For the identity of equation 2.37 we made use of Green's integration rule which allowed us to split the double integral to the product of two single integrations. Also, we used the definition of the 2-dimensional inverse Fourier transform of the height field function. We applied a similar substitution like we did in 2.32, but this time twice, once for x_1 and once for x_2 separately. The last step in equation 2.37, substituting with p and q in equation ?? will be useful later in the implementation. The insight should be, that the product of two complex numbers is again a complex number. We will have to compute the absolute value of $\mathcal{F}\{S\}(w_1, w_2)$ which will then be equal $(p^2 + q^2)^{\frac{1}{2}} |\mathcal{F}\{h\}(w_1, w_2)|$

2.6.2 Sinc Interpolation in Frequency Domain

In section 2.6.1 we derived an alternative approach to the gaussian window approach described in section 2.2.3 in order to approximate our height field. We assume that our height field is a superposition of periodically aligned substructures (i.e. finger structures). This so called PQ approach allows us to integrate over one period of a substructure in our height field, instead of iterating over the whole domain. Nevertheless, this main finding, described in equation 2.37, is using the DTFT. Thus, since our original height field is supposed to be a continuous-time band-limited function we can reconstruct it by applying a sinc-interpolation.

In general, for a sinc-interpolation, we are interested in recovering an original analog signal $x(t)$ from its samples. Therefore, for a given sequence of real numbers $x[n]$, representing a digital signal, its correspond continuous function is:

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \text{sinc}\left(\frac{t - nT}{T}\right) \quad (2.39)$$

which has the Fourier transformation $X(f)$ whose non-zero values are confined to the region $|f| \leq \frac{1}{2T} = B$. When $x[n]$ represents time samples at interval T of a continuous function, then the quantity $f_s = \frac{1}{T}$ is known as its sample rate and $\frac{f_s}{2}$ denotes the Nyquist frequency. The sampling Theorem states that when a function has a Bandlimit B less than the Nyquist frequency, then $x(t)$ is a perfect reconstruction of the original function. Figure 2.6 illustrates a reconstruction of a 1d signal relying on a sinc-interpolation.

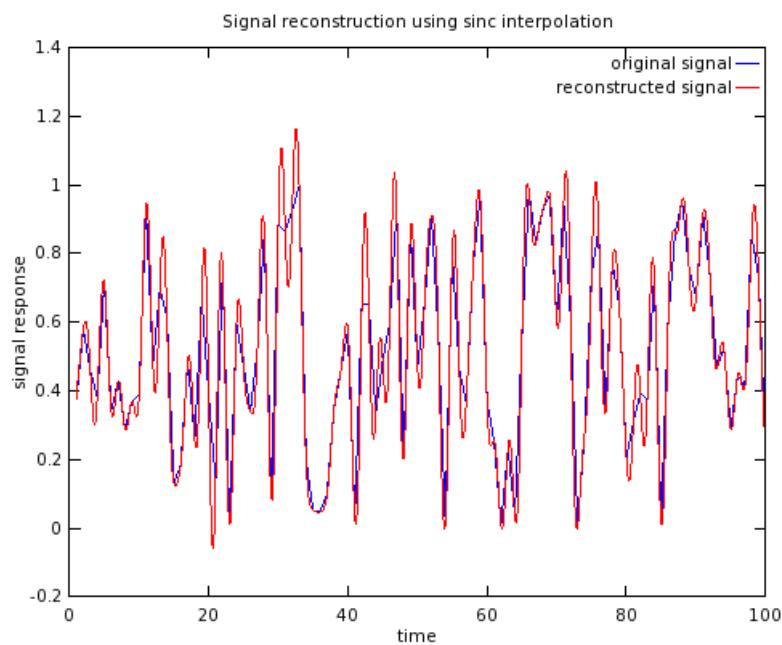


Figure 2.6: Comparison between a given random one dimensional input signal $s(t)$ and its sinc interpolation $\hat{s}(t)$. Notice that for the interpolation there were $N = 100$ samples from the original signal provided.

Chapter 3

Results

In this chapter we examine the rendered output results of our implementation of our BRDF models applied to different input patches such as Blaze grating or Elaphe ?? and Xenopeltis ?? snake nano-scaled surface sheds. We are discussing and comparing both, their BRDF maps 3.1 and the corresponding renderings on a snake geometry like shown in section 3.2 for various input parameters. Last we also show a real experimental image showing the effect of diffraction for similar parameters like we have.

3.1 BRDF maps

A BRDF map shows a shader's output for all possible viewing directions for a given, fixed, incident light direction. We assume that each viewing direction is expressed in spherical coordinates (See appendix ??) (θ_v, ϕ_v) and is represented in the map at point

$$(x, y) = (\sin(\theta_v)\cos(\phi_v), \sin(\theta_v)\sin(\phi_v)) \quad (3.1)$$

with its origin at the map center. The light direction for normal incidence (θ_i, ϕ_i) has been fixed to $(0, 0)$ for our rendered results.

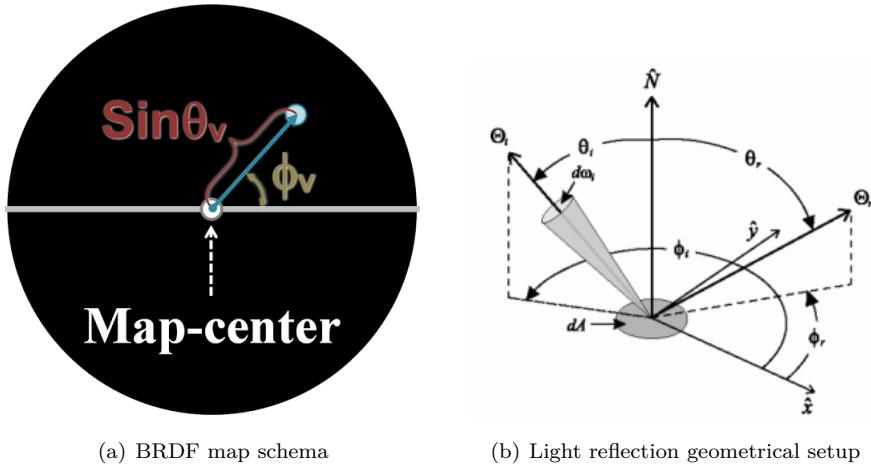


Figure 3.1: BRDF maps¹ for different patches: $\Theta = (\theta_i, \phi_i)$ is the direction of light propagation

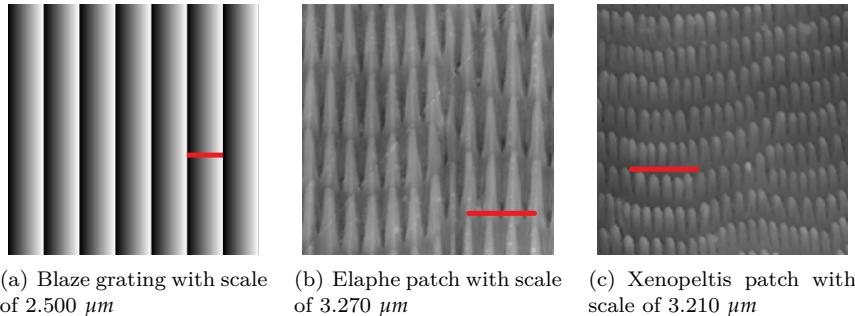


Figure 3.2: Cutouts of our nano-scaled surface gratings used for rendering within our shader with a scale indicator (red line) for each patch. Note that for rendering, we use larger patches.

Figure 3.3 shows the BRDF maps of the full lambda space sampling approach (**FLSS** like introduced in section ??) as described in section ?? applied on different nanoscale surface gratings as shown in figure 3.2. In Subfigure 3.3(a) we see the BRDF map for the Blazed grating, showing high relative brightness for its first order diffraction, i.e. for the Blazed gratings most of the diffracted spectral energy lies in its first order. Notice that the surface of blazed grating is forming a step structure for which the angle between the step normal and the grating normal is denoted by *blaze angle*. Every blazed grating is manufactured in the Littrow² configuration. This means that the blaze angle is chosen such that the diffraction angle and incidence angle are identical. Thus it a blazed grating it has a maximal efficiency for the wavelength of the used light. Higher diffraction modes are still perceivable (second and higher diffraction orders) but with a much lower relative

¹image source of figure:

- 3.1(a): Taken from D.S.Dhillon's Paper [D.S14]
- 3.1(b): Taken from <http://math.nist.gov/~FHunt/appearance/brdf.html>

²For further information please see http://en.wikipedia.org/wiki/Blazed_grating.

brightness. The asymmetry of the pattern is due to the asymmetric geometry of the grating 3.2(a).

The finger-like structures contained in the Elaphe surface grating 3.2(b) are quite regularly aligned and hence diffraction occurs along the horizontal axis for the BRDF map as shown in figure 3.3(b). The reason for not seeing any strong diffraction color contribution along other directions in the BRDF map is due to the fact that these ‘nano-fingers’ overlap across layers and thus do not exhibit any well-formed periodicity along finger direction.

For Xenopeltis surface grating 3.2(c), we observe diffraction along many different, almost vertical directions in the BRDF map 3.3(c) since the layers of the finger-like structures do not overlap and are shifted significantly along their length but still exhibit some local consistency. A similar argument holds true for diffraction across locally periodic finger patches with slightly different orientations.

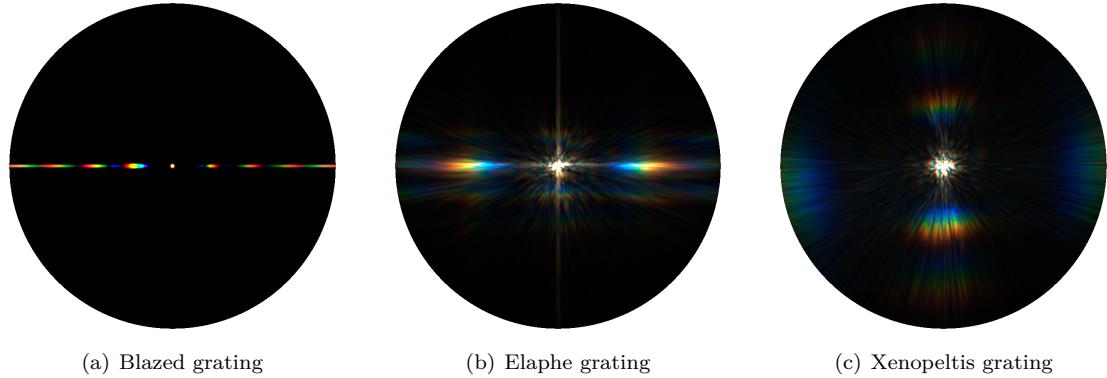


Figure 3.3: BRDF maps for different patches

Figure 3.4 shows BRDF maps of all our BRDF models applied on the Blaze grating. Figure 3.4(a) shows the FLSS shading approach result for our blazed grating and it is used in order to compare with our other rendering approaches.

Figure 3.4(b) shows the BRDF map for the NMM approach, introduced in section ??, which is close to the FLSS approach, verified in section ??, just like in the case of corresponding evaluation in figure ???. Nevertheless there is a small, noticeable difference: For the NMM approach we see a white, circular spot around the map center. Nevertheless, apart from this white spot, the NMM approach resembles the FLSS approach. The reason for this differences is due to the fact that the NMM approach treats the center of a BRDF map as a special case, like described in section ???. Technically, every location around a small ϵ -circumference from the map center gets white color assigned.

Figure 3.4(c) shows the BRDF map for the PQ approach which relies on sinc-interpolation. The PQ BRDF map and the FLSS results are visual alike. Compared to the evaluation plots in figure ???, the BRDF maps even persuade more. Compared to FLSS, one difference we notice is that the first order of diffraction is a little spread for the PQ approach. Without³ applying a sinc-interpolation, this spreading effect would be even strengthened.

³Note that if we do not perform a sinc-interpolation this would correspond to apply a linear interpolation instead.

Last, let us consider figure 3.4(d) which shows the BRDF map produced by using Nvidia Gem's implementation [JG04] of Stam's BRDF model when constraining the y-axis of the BRDF map. This corresponds to a 1d diffraction grating, along the x-axis. This model only uses the spacing d of a given grating. It also always produces highly symmetric results.

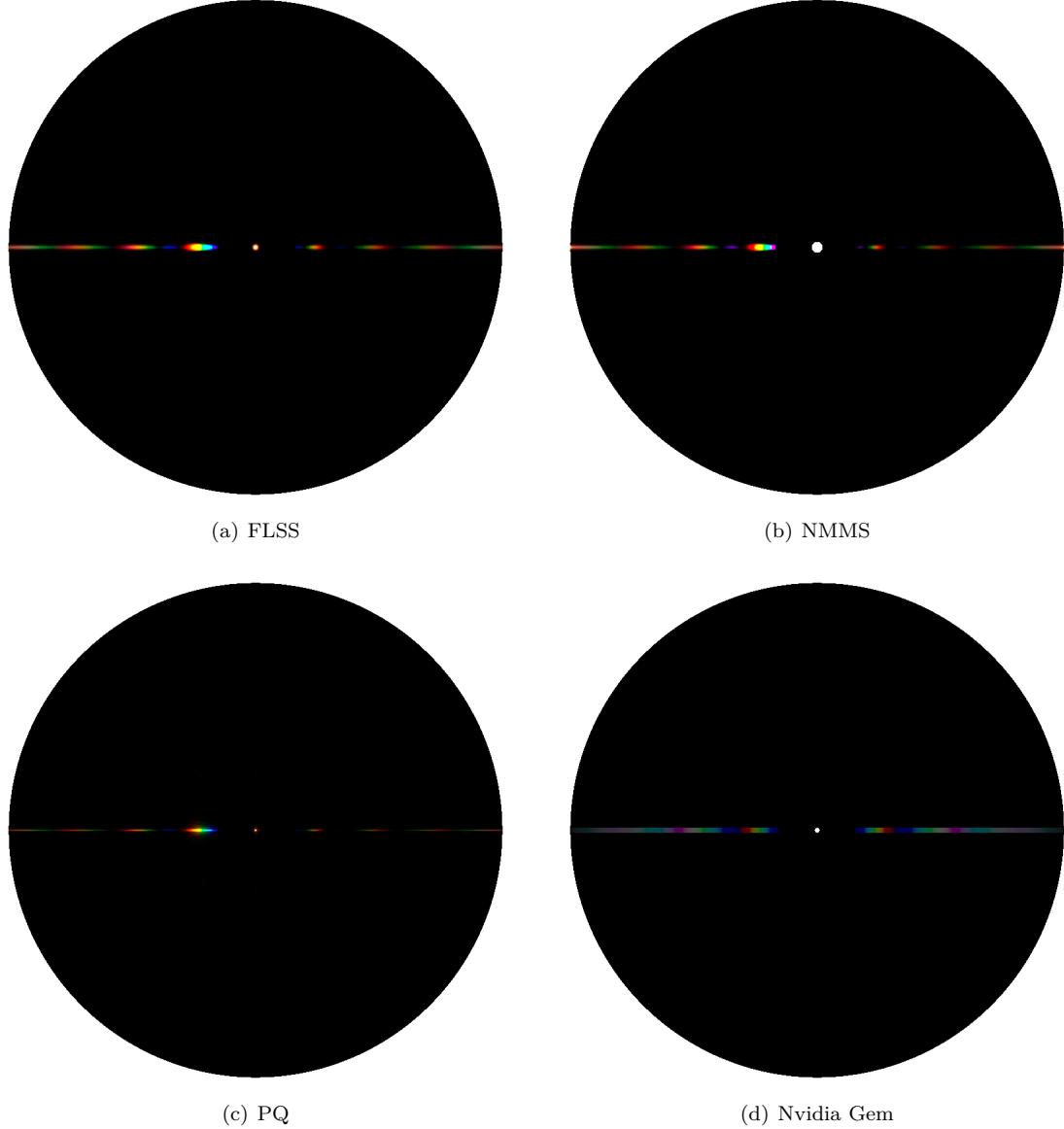


Figure 3.4: BRDF maps for Blazed grating comparing our different rendering approaches

Figure 3.5 and figure 3.6 show the BRDF maps for different wavelength step sizes used in the fragment shader for the FLSS approach applied on the blazed grating and the Elaphe snake shed, respectively. Within our fragment shaders the outermost loop iterates over the range $[380nm, 780nm]$

for a given step size λ_{step} to integrate over the wavelength spectrum. Having bigger step sizes implies having fewer λ -samples which will reduce the overall runtime of a shader but, it will also introduce artifacts and therefore lower the overall shading quality. For Elaphe surface grating, artifacts are perceivable when $\lambda_{step} \leq 10nm$. Results produced by using $5nm$ step sizes do not differ from those produced by using $\lambda_{step} = 1nm$. This allows us to set λ_{step} at $5nm$. For a Blazed grating we may chose even bigger step sized without losing any rendering quality(see figure 3.5).

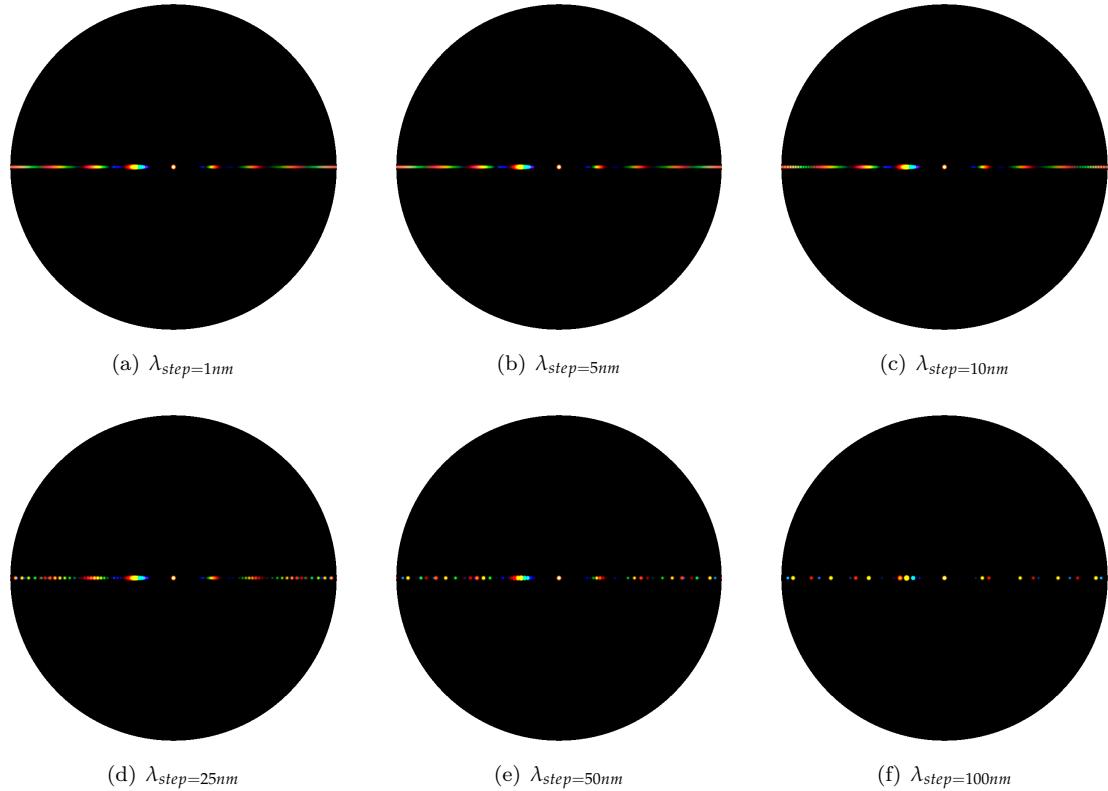


Figure 3.5: Blazed grating at $2.5\mu m$: Different λ step sizes

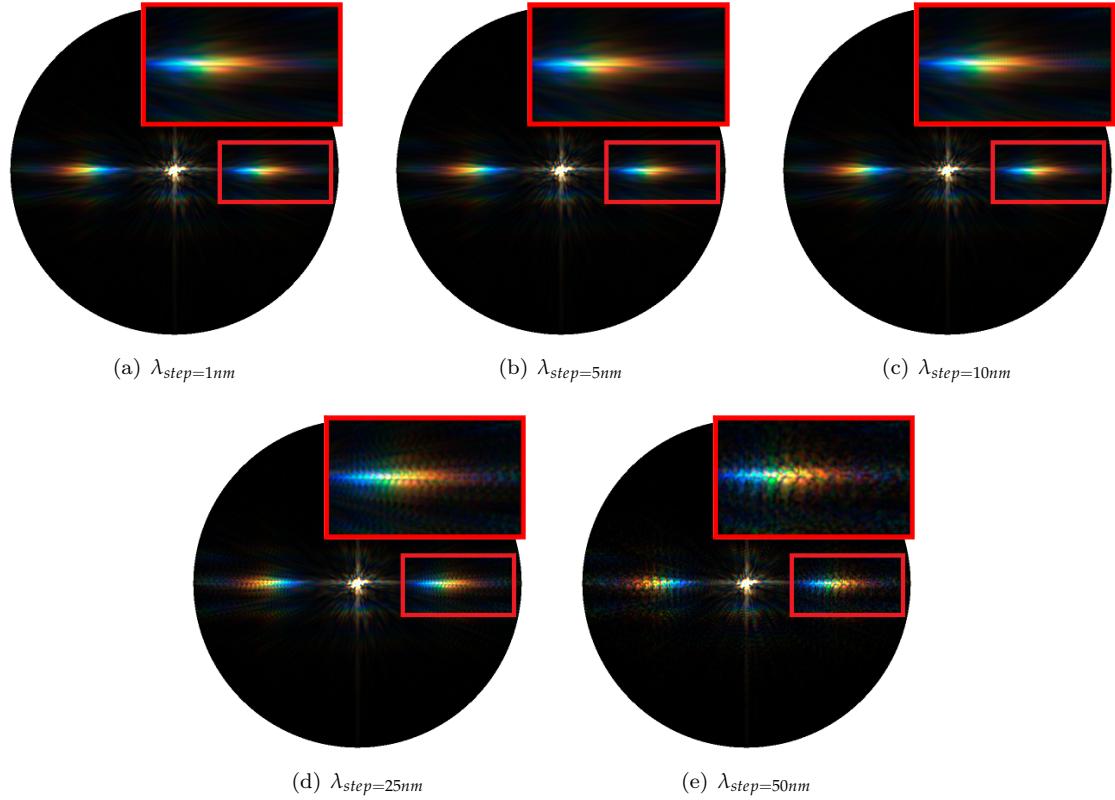


Figure 3.6: Elaphe grating at $65\mu m$: Different λ step sizes

The figures 3.7, 3.8, 3.9 show a comparison of the BRDF maps produced by the FLSS approach (on the left) and the PQ shading approach (on the right) applied on all our patches. For Blazed grating, as already mentioned, we notice that both approaches, FLSS and PQ, resemble each other. We also notice that for PQ map, the first order diffraction color contribution is spread. For the Elaphe and Xenopeltis grating we notice similar shaped BRDF patterns, even when the angle of light varies, but nevertheless, they also contain some artifacts.

In general, a Blazed Grating is manufactured in a way that a large fraction of the incident light is diffracted preferentially into the first order. Therefore, most of the energy in its BRDF map lies in the first order of diffraction at its blaze angle. This implies that largest portion of the color contribution, visible on the corresponding BRDF map, lies at that angle. In figure 3.7, in contrast to the results produced by the FLSS approach, we see color fringes at the first order modes in the BRDF map produced by our PQ approach. This implies that the PQ approach does not produce reliable results which also affirms our evaluation plots shown in figure ???. A similar argumentation hold true for the PQ approach, when we do not apply we a sinc-interpolation like shown in figure 3.7(c).

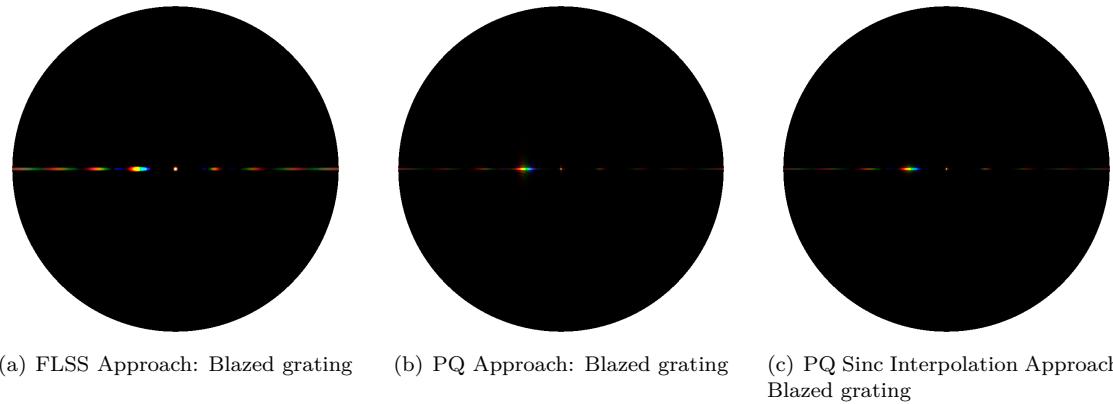


Figure 3.7: A comparison between the PQ- and the FLSS approach applied on an Blazed grating.

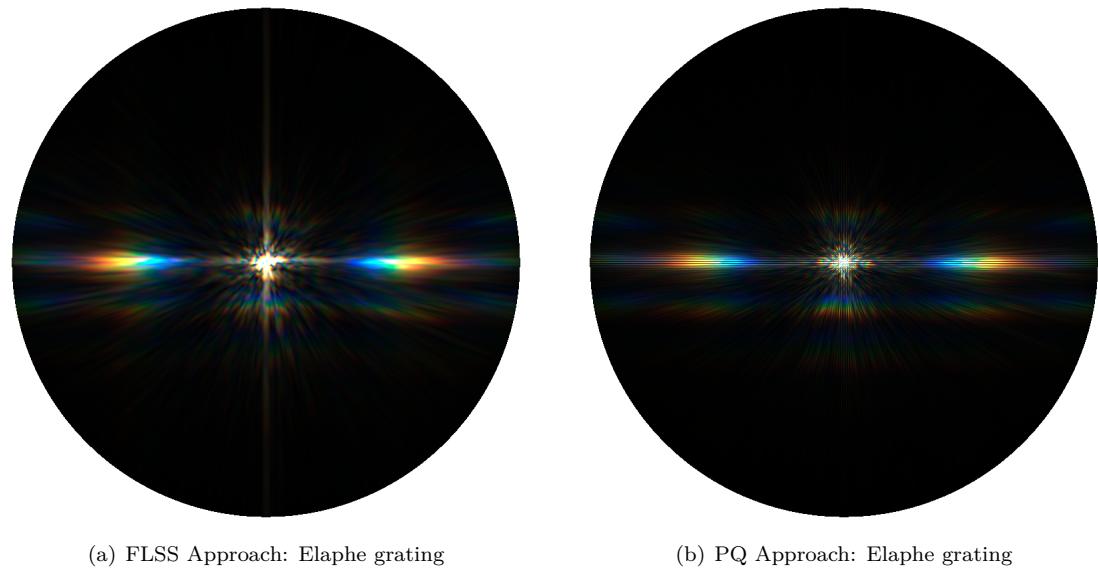


Figure 3.8: A comparison between the PQ- and the FLSS approach applied on an Elaphe grating.

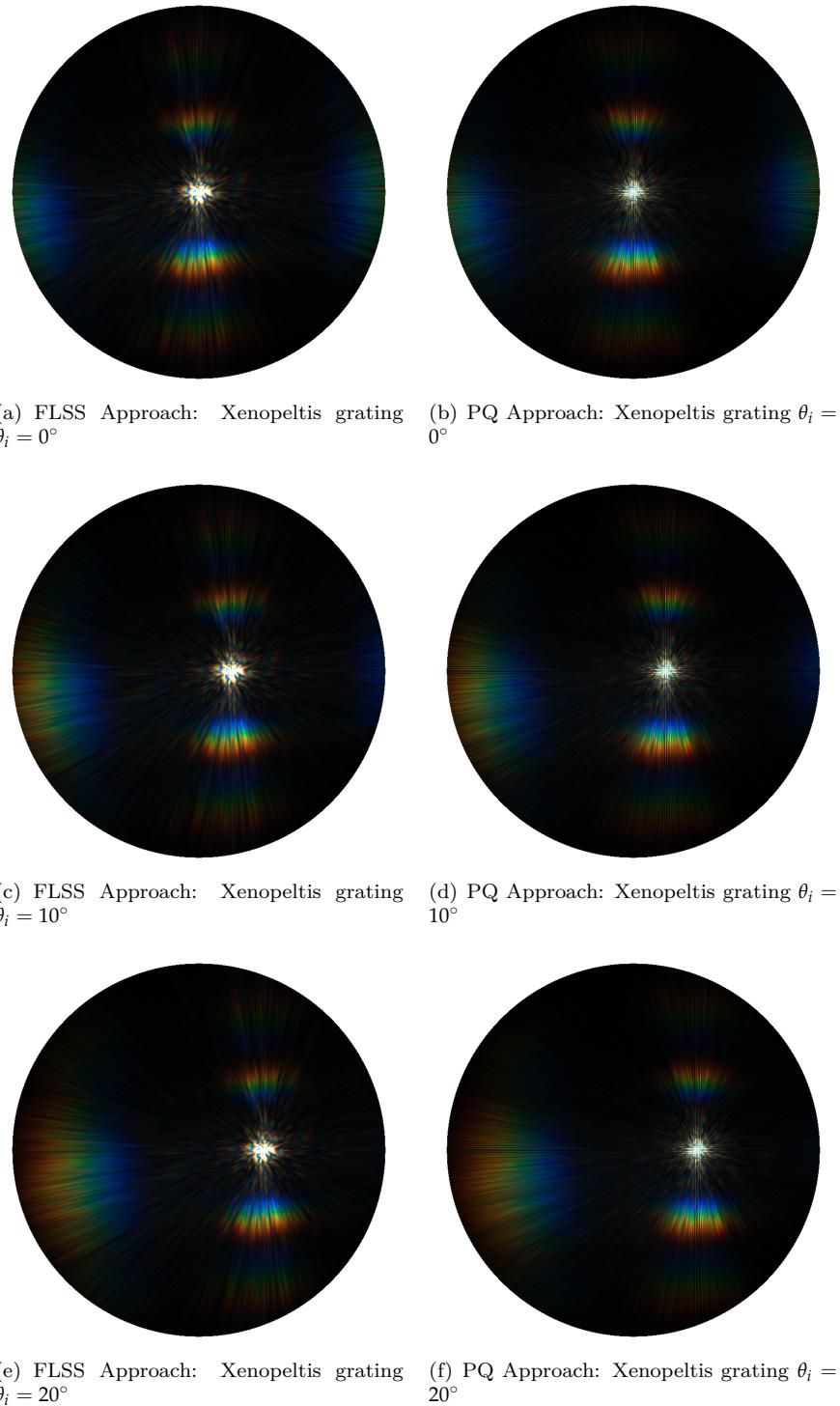


Figure 3.9: A comparison between the PQ- and the FLSS approach applied on an Xenopeltis grating.

Figure 3.10 shows BRDF maps for the full lambda sampling approach applied to the Blazed grating, while varying the value for the spatial variance σ_s . This akin to changing the coherence length for the incident light. The lower the coherence length, the fewer interacting grating periods produce blurred diffraction bands for different λ which overlap to produce poorly resolved colors.

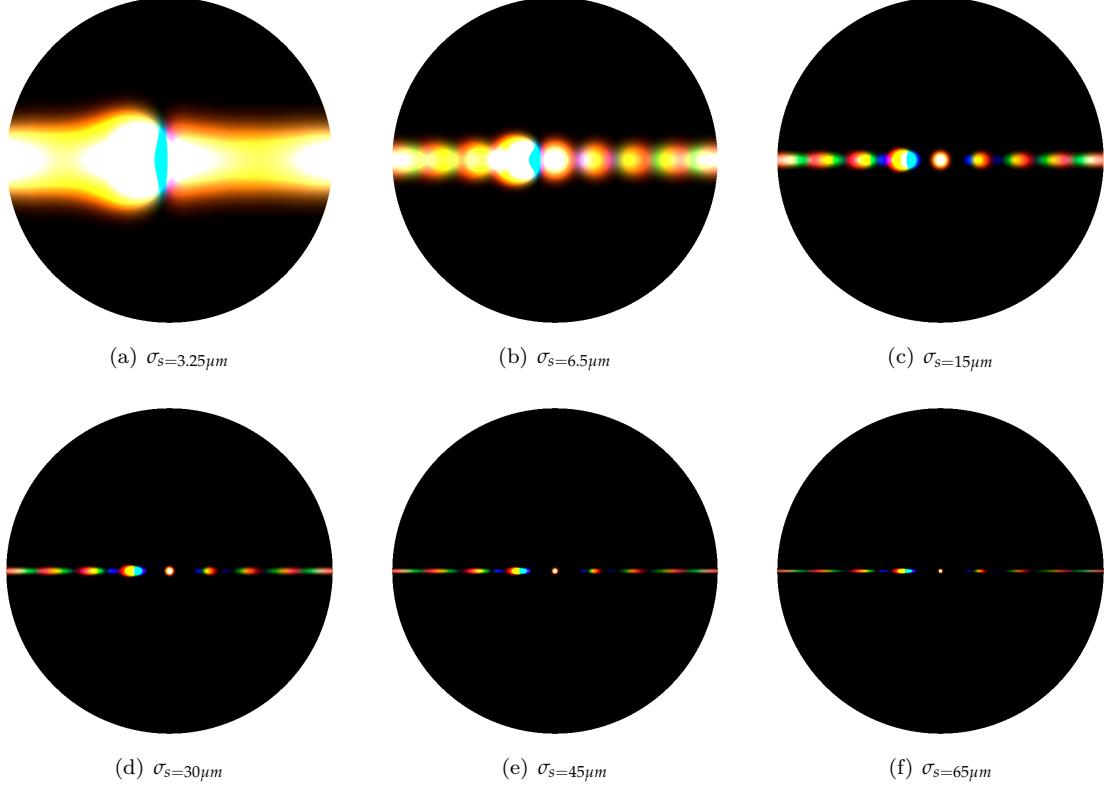


Figure 3.10: Blazed grating with periodicity of $2.5\mu m$: Different σ_s

Figures 3.11 and 3.12 show the BRDF maps the reference-, FLSS approach using different values for N in the taylor series approximation. For both input patches we clearly visually observe the convergence of the taylor series for higher values of N . We visually observe convergence of the Taylor series for all our patches a very large value of N^4 .

Like discussed in section 2.4 there exists a certain value of N for which our approach converges. For all our shading approaches, applied on our gratings, we visually observed a convergence of their BRDF maps when using $N \geq 39$ DFT terms. Furthermore, for a Blazed grating it satisfies to use only $N \geq 7$ - and for an Elaphe grating only $N \geq 9$ DFT terms. Notice, that these numbers of required DFT terms were empirically determined by trial and error strategy.

However, by making use of taylor error term estimates, like introduced in the appendix section ??, we can derive an upper bound for N . Since this computation is dependent on many aspects, such as on the grating spacing, the pixel-width correspondence, the used lambda space for sam-

⁴Using N equal to 39 lead to visual convergence for all our used gratings.

pling, it is usually simpler to determine empirically actual values for N .

In algorithm ?? we compute the DFT terms of a provided height field h raised to the power of the imaginary number i times an integer, i.e. we evaluate the expression $DFT(h)^n \cdot i^n$. Since we multiply our height field by i^n and then apply the DFT operator, basically, there exist four possible convergence images, each having its own convergence radius.

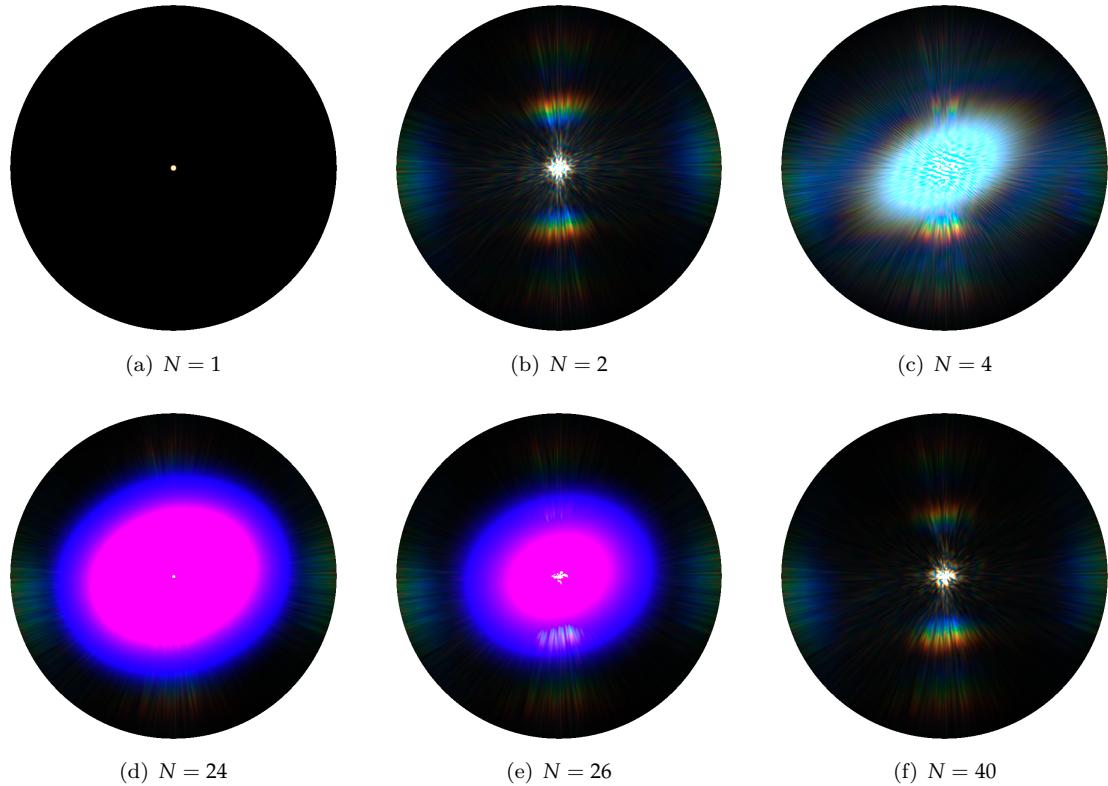


Figure 3.11: Blazed grating at $65\mu m$: N Taylor Iterations

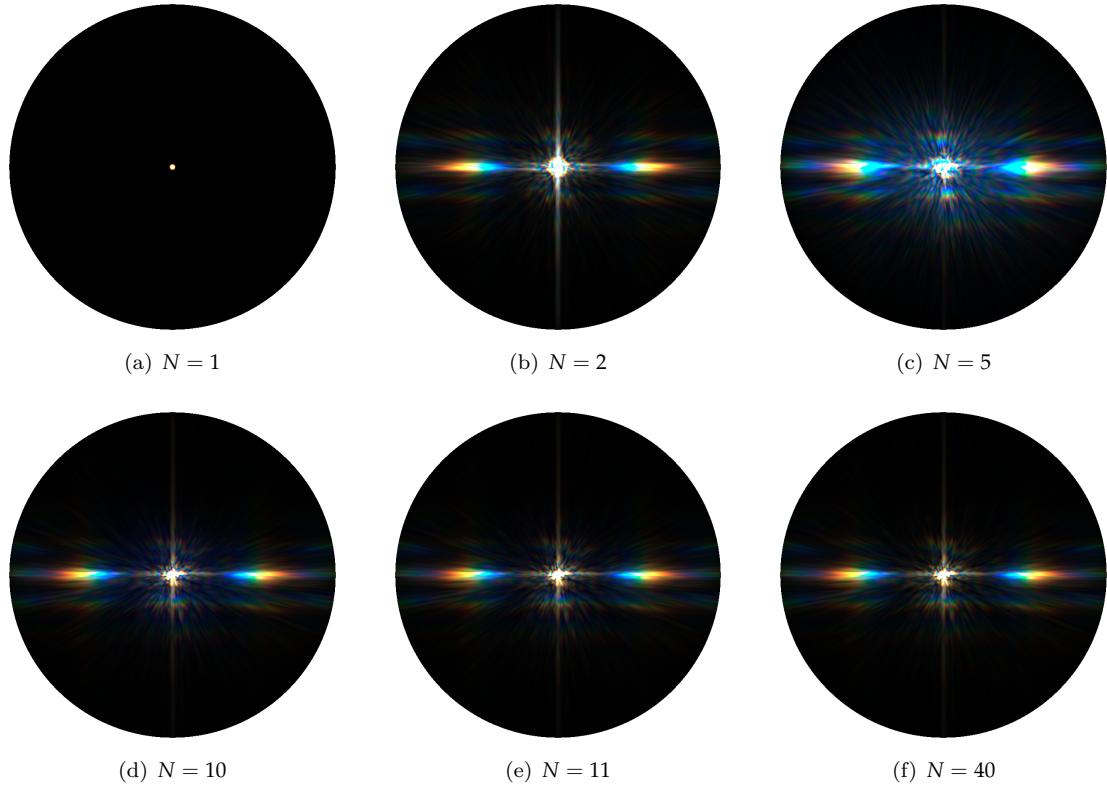
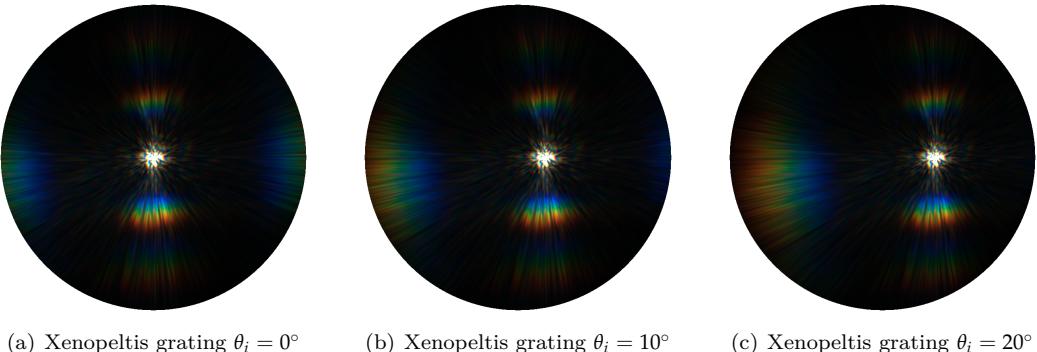
Figure 3.12: Elaphe grating at $65\mu m$: N Taylor Iterations

Figure 3.13 shows the BRDF maps of the FLSS approach applied on the Xenopeltis snake shed, using different θ_i incident angles. When slightly moving the incident angle θ_i , we can observe how the BRDF map changes. For higher values of θ_i we start seeing diffraction color contribution on the right side of the BRDF map.

Figure 3.13: BRDF maps for Xenopeltis grating: different θ_i angles

3.2 Rendering Surface Geometries

In this section we are going to present our actual renderings simulating the effect of diffraction caused when a directional light source encounters different nano-scaled surfaces on a given curved snake mesh. We will see that diffraction colors change dramatically with changes in light direction, surface normals and viewing direction, which is typical for diffraction colors observed in nature. For rendering we are going to rely on our FLSS approach. Unfortunately, this approach is rather slow and can barely be considered as being interactively performing. Nevertheless, we have introduced some optimizations in order to make it interactive.

All rendered results shown in this section are produced by the FLSS approach since we have proven its validity in figure ???. Therefore, we can trust its renderings and may consider them as being accurate. Furthermore, as support of our evaluation plots regarding the NMM approach, that are shown in figure ???, we also show results produced by the NMM approach.

The Laboratory of Artificial and Natural Evolution in Geneva provided us by a triangular mesh of a snake. This mesh was produced by a 3d scan of a Elaphe snake species and consists of 11696 vertices and 22950 faces. Note that, for all our renderings, we used this snake mesh.

Among all the snake species under consideration, their macroscopic geometry is highly similar. Only the geometry of their nano-structures varies and is responsible for a snake's iridescence. Thus, we can use the same snake surface model to render diffraction for different species. Table 3.1 lists the system specifications of the machine I used in order to produce the rendered images.

Processor	Intel i7 CPU 970 @ 3.20 GHz (12 CPUs)
Memory	12288 MB RAM
Graphics Card	GeForce GTX 770
Graphics Clock	1150 MHz
Graphics Memory	4096 MB
Graphics Memory Bandwidth	230.4 GB/s MHz

Table 3.1: Hardware specifications of the machine used to render snake surface. Statistics are provided using the tool *NVIDIA Geforce Experience*.

Figure 3.14 shows renderings produced by the FLSS approach applied on our snake mesh for different, given input patches. Due do the fact that a Blazed grating has its maximum intensity for a certain direction and the geometry of the snake mesh is curved i.e. is non-flat, we can expect rather less diffraction color contribution like shown in figure 3.14(b).

In contrast, For both the renderings, we see colorful patterns on the skin of our snake species, Elaphe and Xenopeltis, due to the effect of diffraction. We see much less colorful patterns for Elaphe like shown in figure 3.14(b) than for Xenopeltis like shown in figure 3.14(c). This is consistent with the observations in the real world as shown in figure ???. As observable figure 3.2(b), the substructures (the finger like structures) in the height field of a Elaphe snake skin are not very regularly aligned along the y-axis. This is why the Elaphe species is less iridescent than the other specie. The Xenopeltis snake has a brownish body with no pigmentation, which makes the iridescence more spectacular than on Elaphe.

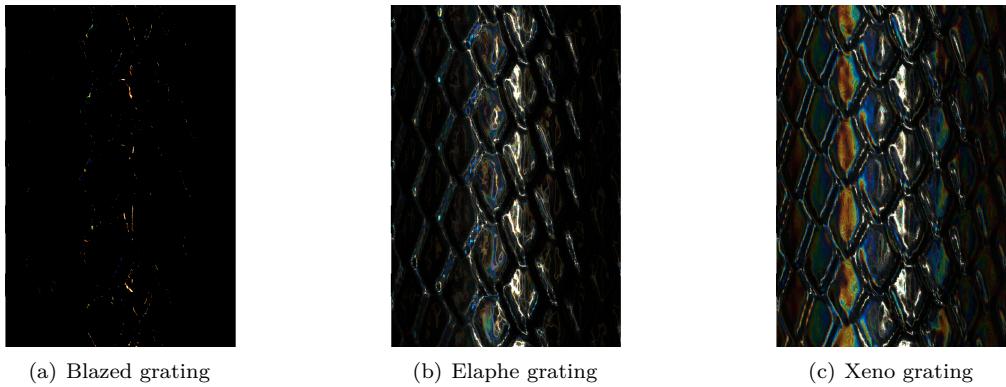


Figure 3.14: Diffraction of different snake skin gratings rendered on a snake geometry

Figure 3.15 shows a set of subfigures for rendering the effect of diffraction produced by the FLSS approach (used as our reference approach), applied on our snake mesh using the Elaphe nano structure. Figure 3.15(b) shows the final diffraction color contribution result with texture-blending. We only see little diffraction color contribution in this subfigure which resembles quite well to the reality as shown in figure ???. In subfigure 3.15(d) we see the light cone in order to show the direction of the light source besides the rendered results. Subfigure 3.15(e) is a sample Fourier image of Elpahe's nano-scale surface structure 3.15(d).

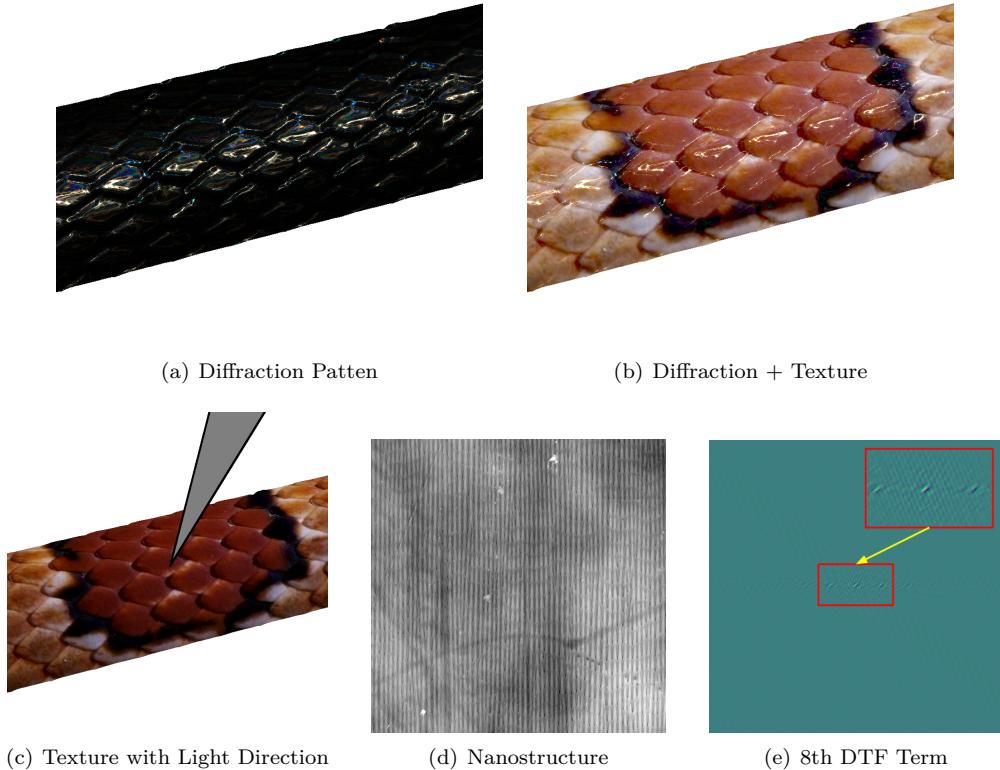


Figure 3.15: Diffraction for Elaphe snake skin produced by our reference approach.

Figure 3.16 shows a set of subfigures for the effect of diffraction for the Xenopeltis snake surface. For Xenopeltis we see quite a lot color contribution due the phenomenon of diffraction like shown in figure 3.16(c). Comparing this to a real image ?? we notice much resemblance regarding the reflectance strength and colorful pattern.

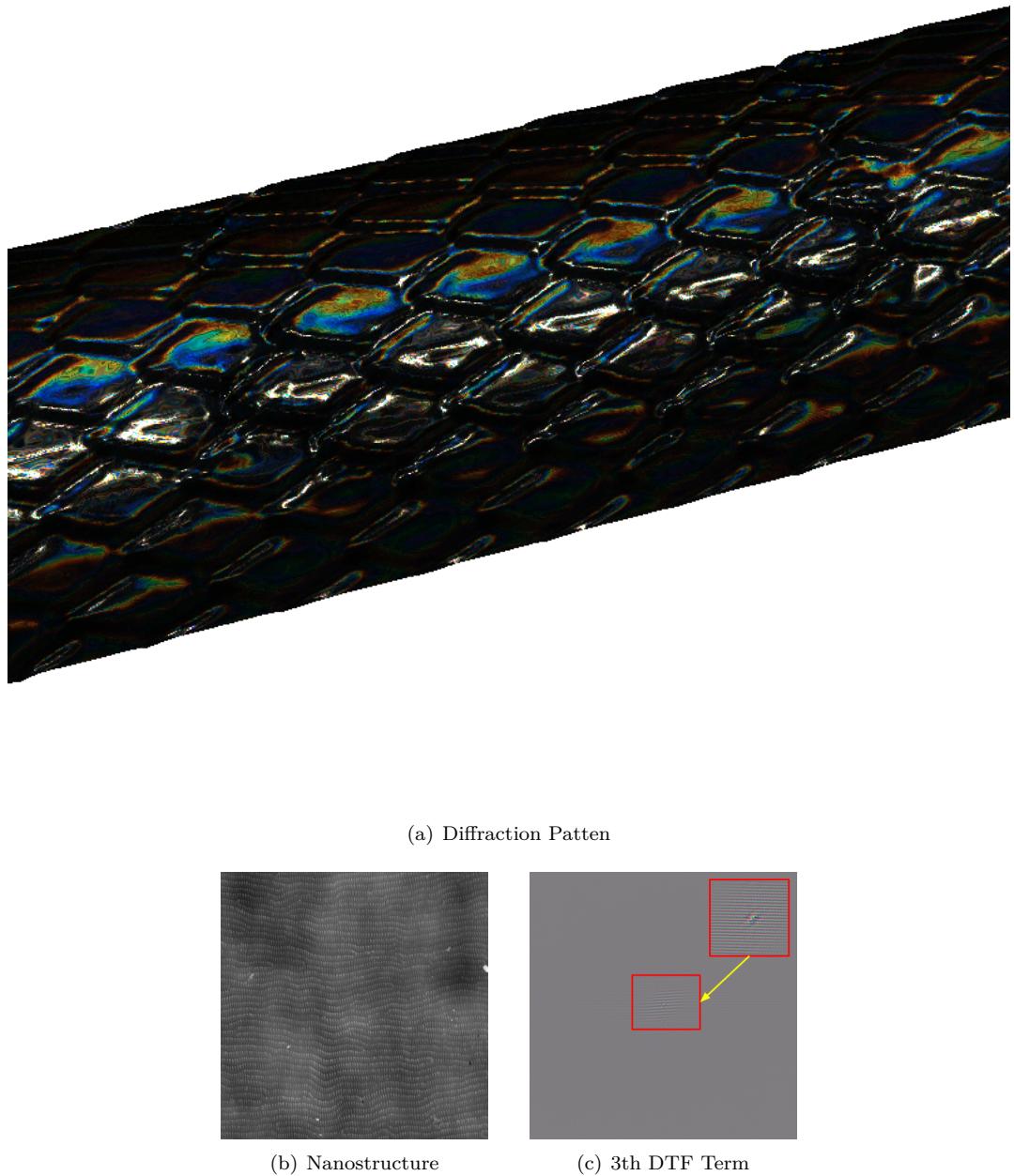


Figure 3.16: Diffraction for Xenopeltis snake skin produced by our reference approach.

Figure 3.17 shows the diffraction pattern for Elaphe snake shed at different zoom levels for fixed incident light and viewing direction. We changed the zoom-levels by adjusting the field of view angle of our camera. For each image in this figure, the one to its right side is a five times zoomed-in version of the region within its red box. The close up perspectives exhibit complex and colorful diffraction patterns.

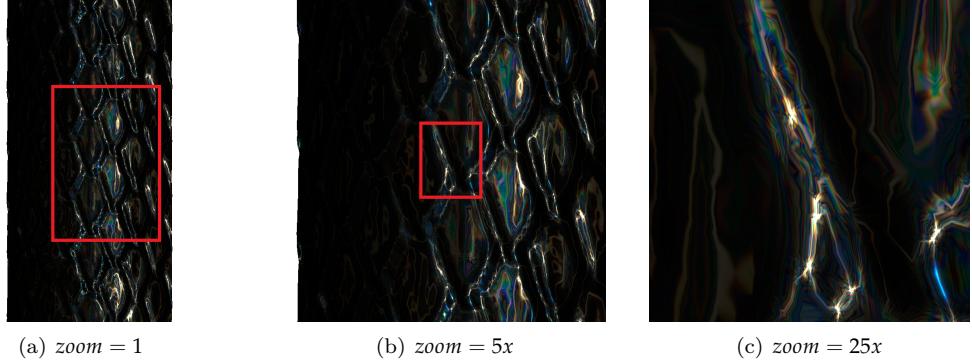


Figure 3.17: Diffraction on Elaphe snake skin grating: Different camera zoom levels by varying the field of view.

Figure 3.18 shows how the diffraction pattern changes when the incident light direction is moved slightly. This Figure gives us an impression what kind of complex, perspective-dependent pattern the diffraction phenomenon produces.

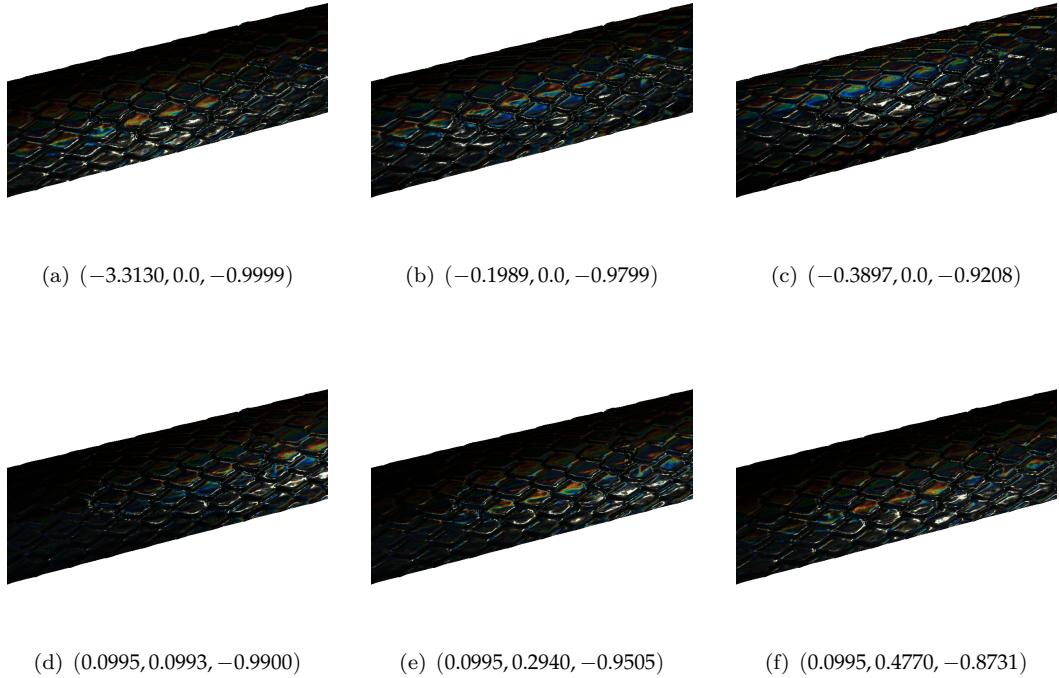


Figure 3.18: Diffraction on Elaphe snake skin grating: Different light directions

3.2.1 A Comparison with Stam

In the following a brief comparison between Stam's⁵ and our final approach using two different kinds of gratings, a synthetic, regularly aligned grating and a natural, complex structured grating. These gratings are shown in figure 3.19.

⁵A reference implementation of Stam's Diffraction Shader[Sta99] is provided by Nvidia's GPU Gems at http://http.developer.nvidia.com/GPUGems/gpugems_08.html

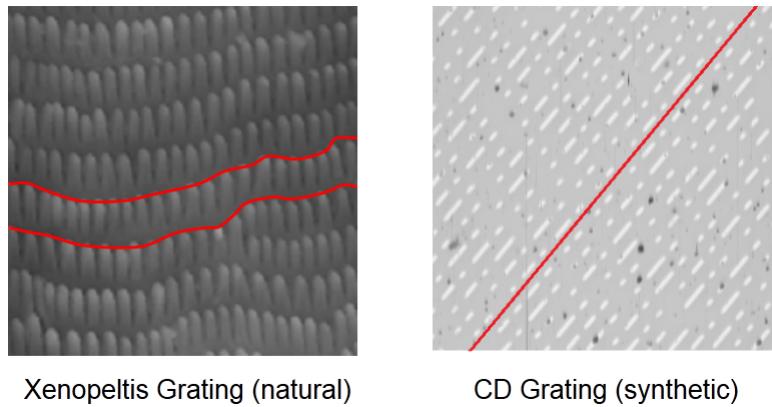


Figure 3.19: Alignment of nano-structures in diffraction gratings. On the left a complex, natural grating of the Elaphe snake species and on the right a synthetic, very regularly aligned grating of a CD.

Figure 3.20 shows an example of a case where Stam’s approach performs well. Considering the red-line in the figure we notice that the nano-scaled structures of a compact disc are very regularly aligned along the surface. Tracks of a CD are uniformly spaced and bumps along a track are distributed according to a poisson distribution⁶. All angles of the diffraction pattern look the same as in the images produced by our approach.

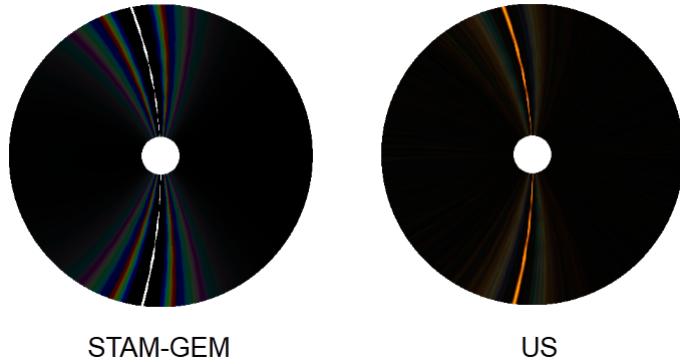


Figure 3.20: Comparison of our approach against a reference implementation of Stam’s method provided by Nvidia Gem. For synthetic diffraction gratings, which have a very regular structure, Stam’s approach is doing well. All angles of the diffraction pattern look the same as in the images produced by our approach.

Figure 3.21 shows an attempt to reproduce real structural colors on the skin of the Xenopeltis snake using our method and comparing it to Stam’s approach. Even the results of Stam might look appropriate. However, there are some notable differences such as missing colors close to specular regions, or such as the color distribution which is rather discrete in Stam’s approach. Furthermore, Stam’s approach has at some places color contribution, where it should not have.

⁶See http://en.wikipedia.org/wiki/Poisson_distribution

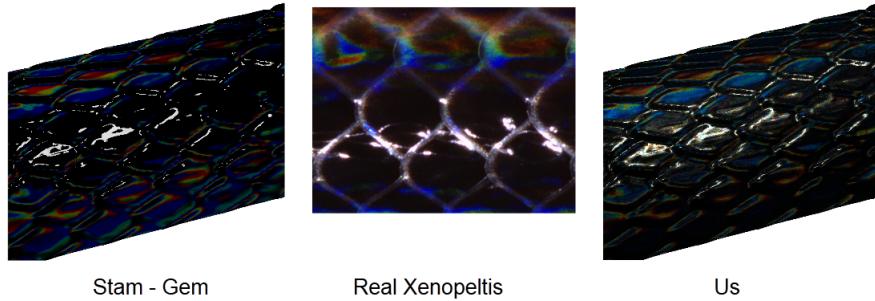


Figure 3.21: Comparison of our approach(on the right) against a reference implementation of Stam’s method(on the left) provided by Nvidia Gem. We attempt to reproduce a real Xenopeltis skin coloration(center). For natural diffraction gratings, which have a rather complex structure, Stam’s approach is qualitatively different from the real image.

In the next chapter we are going to adapt Stam’s BRDF model such that it will be able to handle the kind of surfaces we are dealing with and even will have a runtime complexity which permits interactive rendering.

3.2.2 Merge with above subsec

Figure 3.22 shows a photo of an experimental setup for demonstrating the effect of diffraction using a Elaphe snake grating. The exact parameters for the experimental setup are unknown. Nevertheless this image gives us an impression of how close our model is to the reality comparing it with our simulated results since we notice similar diffraction patterns for our simulated results using an Elaphe snake shed.

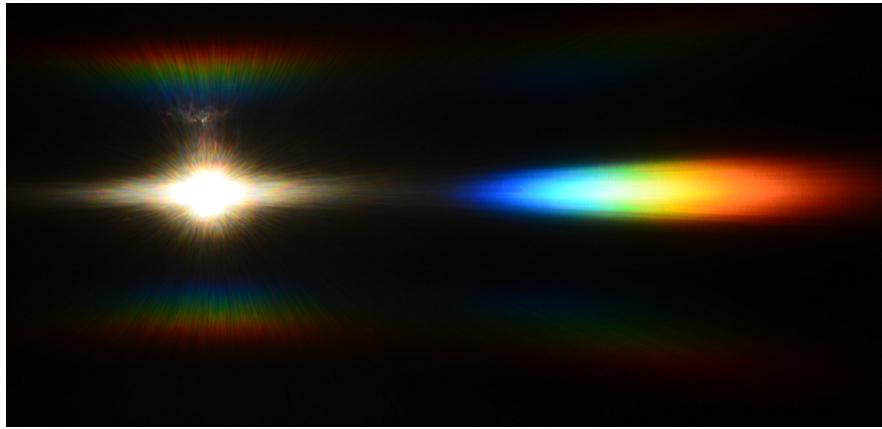
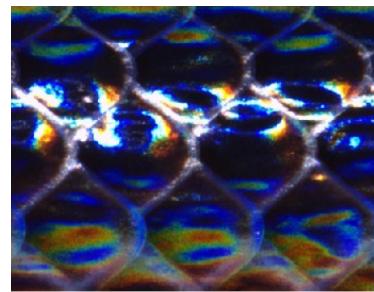
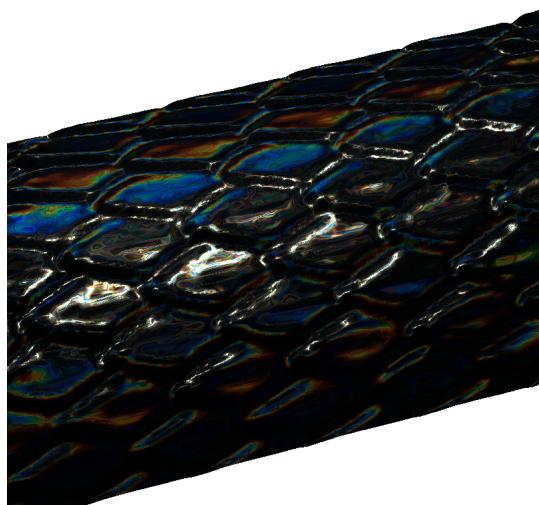


Figure 3.22: Diffraction Elaphe: experimental setup

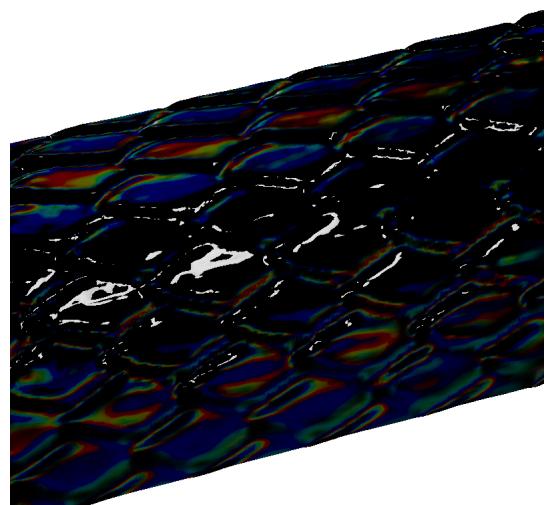
Finally, tried to reproduce a real Xenopeltis image for a unknown light viewing direction. For this purpose I used our FLSSS rendering approach and compared its results against Nvidia Gem’s implementation. The results are shown in figure 3.23.



(a) Xenopeltis Image



(b) FLSS



(c) Stam

Figure 3.23: Attempt to reproduce a real world image: FLSS- vs Stam’s Approach

Bibliography

- [Bar07] BARTSCH, Hans-Jochen: *Taschenbuch Mathematischer Formeln*. 21th edition. HASNER, 2007. – ISBN 978-3-8348-1232-2
- [CT12] CUYPERS T., et a.: Reflectance Model for Diffraction. In: *ACM Trans. Graph.* 31, 5 (2012), September
- [DD14] D.S. DHILLON, et a.: Interactive Diffraction from Biological Nanostructures. In: *EUROGRAPHICS 2014/ M. Paulin and C. Dachsbacher* (2014), January
- [D.S14] D.S.DHILLON, M.Single I.Gaponenko M.C. Milinkovitch M. J.Teyssier: Interactive Diffraction from Biological Nanostructures. In: *Submitted at Computer Graphics Forum* (2014)
- [For11] FORSTER, Otto: *Analysis 3*. 6th edition. VIEWEG+TEUBNER, 2011. – ISBN 978-3-8348-1232-2
- [I.N14] I.NEWTON: *Opticks, reprinted*. CreateSpace Independent Publishing Platform, 2014. – ISBN 978-1499151312
- [JG04] JUAN GUARDADO, NVIDIA: Simulating Diffraction. In: *GPU Gems* (2004). <https://developer.nvidia.com/content/gpu-gems-chapter-8-simulating-diffraction>
- [LM95] LEONARD MANDEL, Emil W.: *Optical Coherence and Quantum Optics*. Cambridge University Press, 1995. – ISBN 978-0521417112
- [MT10] MATIN T.R., et a.: Correlating Nanostructures with Function: Structurnal Colors on the Wings of a Malaysian Bee. (2010), August
- [PAT09] PAUL A. TIPLER, Gene M.: *Physik für Wissenschaftler und Ingenieure*. 6th edition. Spektrum Verlag, 2009. – ISBN 978-3-8274-1945-3
- [PS09] P. SHIRLEY, S. M.: *Fundamentals of Computer Graphics*. 3rd edition. A K Peters, Ltd, 2009. – ISBN 978-1-56881-469-8
- [R.H12] R.HOOKE: *Micrographia, reprinted*. CreateSpace Independent Publishing Platform, 2012. – ISBN 978-1470079031
- [RW11] R. WRIGHT, et a.: *OpenGL SuperBible*. 5th edition. Addison-Wesley, 2011. – ISBN 978-0-32-171261-5
- [Sta99] STAM, J.: Diffraction Shaders. In: *SIGGRPAH 99 Conference Proceedings* (1999), August
- [T.Y07] T.YOUNG: *A course of lectures on natural philosophy and the mechanical arts Volume 1 and 2*. Johnson, 1807, 1807