FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT • DEPARTMENT INFORMATIK

Lehrstuhl für Informatik 10 (Systemsimulation)

Optimization of Mirrorshapes in Optically Pumped Solar Lasers Using Ray Tracing Simulation Techniques

Matthias König

Master's Thesis

Optimization of Mirrorshapes in Optically Pumped Solar Lasers Using Ray Tracing Simulation Techniques

Matthias König

Master's Thesis

Aufgabensteller:	Prof. Dr. C. Pflaum
Betreuer:	
Bearbeitungszeitraum:	1.11.2021 – 2.5.2022

Abstract

This work showcases the application of ray tracing techniques for the calculation of absorption profiles in optically pumped solar lasers. It aims at using a lightweight and fast physically based raytracer combined with a biobjective mesh adaptive direct search algorithm to optimize total power absorption and to minimize variance across the crystal. An exemplatory setup of a side pumped Nd:Yag solar laser was simulated, optimized and the resulting beam quality evaluated.

Erklärung:

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Der Universität Erlangen-Nürnberg, vertreten durch den Lehrstuhl für Systemsimulation (Informatik 10), wird für Zwecke der Forschung und Lehre ein einfaches, kostenloses, zeitlich und örtlich unbeschränktes Nutzungsrecht an den Arbeitsergebnissen der Master's Thesis einschließlich etwaiger Schutzrechte und Urheberrechte eingeräumt.

Erlangen, den 31. März 2022 .....

Contents

1 Introduction

2 Methodology

3 Lasers

3.1 Solar Lasers

4 Raytracing Framework

With the advent of cheap processors and increasingly powerful consumer hardware, ray tracing has become more popular in recent years. For the purpose of global illumination in video games and image processing, more advanced techniques have been continuously developed and improved. In optical design ray tracing is used to analyse the imaging quality of optical systems or as in this work other illumination properties can be simulated. The need for fast refresh rates in video games and the requirement of modelling more complex physical phenomena in optical design have led to tracing and sampling techniques that reduce the computational expense dramatically with minimal loss of accuracy. Focused on the specific problems of laser design, these improvements make it possible to get physically accurate results in an acceptable amount of computational time in an iterative context.

As in optical design systems are mostly rotationally symmetrical, the framework is meant to be used in a two dimensional setup and calculated quantities, e.g. absorbed power in a medium converted to three dimensional values after a simulation step. This significantly reduces the amount of rays needed to avoid undersampling effects and to produce stable results across multiple simulation runs. Intersection tests also require less computation and objects in the scene require less fundamental shapes to test a ray against. The resulting performance gains makes it possible to run the simulation thousands of times in an iterative process to optimize some parameters in the optical setup even on consumer grade hardware. The objects in a scene are preproccessed to group fundamental shapes into leaves of a quadtree to reduce the amount of shapes a ray has to be tested against even further. To achieve the satisfied accuracy and to reduce noise the appropiate sampling strategies have to be used for a given problem. The most important techniques are provided including uniform sampling, stratified and importance sampling.

The framework was designed to provide a simple yet powerful interface for the user and was implemented in C++17. It provides the necessary data structures and algorithms for a fast raytracing solution. The sampling techniques are implemented in specialized classes of abstract interfaces. They can also be used by the user to implement custom techniques. The framework extensively relies on lambda functions to be provided by the user and thus naturally is customizable, although some preset functions are also provided. Because the calculations in the framework are so similar to applications in graphics software the OpenGL Mathematics header only library GLM [1] was used as an underlying maths library. GLM is based on the OpenGL Shading Language (GLSL) and so in a potential later step the framework could be ported to work on graphics cards providing that the data structures are changed to be accessable from a GPU. As in the specific problem in this work the tracing of each ray has side effects on the scene and on itself, i.e. the absorbed power of each ray has to be accumulated, it was decided to focus more on single core performance first and leave the parallel execution and execution on GPUs for a later point. Furthermore IO utilities for simulations are provided for Comma Separated Values (CSV) files and structured output for the commonly used Visualization Toolkit (VTK) [2].

In the following chapters the applied ray tracing techniques explained in detail. Firstly the basics of raytracing, i.e. intersection tests of fundamental shapes and objects and reflection and refraction effects are shown. Then an applied method of subdivision for the performance optimization of the raytracer is explained. Lastly some methods of sampling are shown before the structure of the developed framework is presented with code samples. Here the usage of classes is demonstrated and it is shown how specific objects are defined. In particular the objects which are relevant for the simulation of laser cavities and which are used in the example setup are shown.

4.1 Raytracing Basics

Rays are represented as a parametric line from a ray origin o in direction d. The parameter t goes is in the interval $[0, \infty)$ and represents the closeness of the ray to the origin. The mathematical representation therefore is given as

$$\mathbf{r}(t) = \mathbf{o} + t\mathbf{d} \tag{1}$$

After the ray is generated it is tested against intersections with the scene. Here the smallest $t > 0$ of all the intersections with objects has to be found. The question if a ray intersects an object can usually only be answered for simple fundamental shapes, e.g. lines, circles, axis aligned bounding boxes (AABBs) in 2D or planes, triangles, spheres, etc. in 3D. Therefore objects are normally comprised of a collection of fundamental shapes and an intersection occurs if one of the fundamental shapes is intersected. Naturally, an object can be intersected multiple times by the same ray and so the results have to be searched for the smallest t. Each fundamental shape should be represented in a parametrised form so the intersection test can be represented as a system of equations. The two fundamental shapes used in this work are 2D lines and axis aligned bounding boxes (AABBs).

Lines are represented by two points $\mathbf{a}$ and $\mathbf{b}$. So the intersection problem can be written as a ray-ray intersection as follows:

Find $\alpha \in [0, 1]$ and $t \in [0, \infty)]$ s.t.

$$\mathbf{a} + \alpha(\mathbf{b} - \mathbf{a}) = \mathbf{o} + t\mathbf{d} \tag{2}$$

If such a combination of α and t exists, we have an intersection. As we are in 2D there are two equations for two unknowns and the system always has a solution. The solution can then be checked, s.t. the values are in the right intervals. A small mathematical trick is to define a 2D cross product which is basically just the z component of a 3D cross product if the two input vectors $\mathbf{p}$ and $\mathbf{q}$ were parallel to the xy plane:

$$\mathbf{p} \times \mathbf{q} = p_x \cdot q_y - p_y \cdot q_x \in \mathbb{R} \tag{3}$$

Observe that same as the 3D cross product, the 2D version becomes 0 when you cross a vector with itself. If one now crosses Eq. (2) with $\mathbf{d}$ on both sides the intersection equation becomes:

$$\mathbf{a} \times \mathbf{d} + \alpha(\mathbf{b} - \mathbf{a}) \times \mathbf{d} = \mathbf{o} \times \mathbf{d} \tag{4}$$

So t has been eliminated from the equation and we can solve Eq. (4) for α:

$$\alpha = \frac{(\mathbf{a} - \mathbf{o}) \times \mathbf{d}}{\mathbf{d} \times (\mathbf{b} - \mathbf{a})} \tag{5}$$

If α satisfies the condition, we continue analogously for t by crossing Eq. (2) with $\mathbf{b}-\mathbf{a}$. The resulting t is then checked against the condition and a normal at the intersection point is calculated. The intersected rays can be seen in in Figure 1.

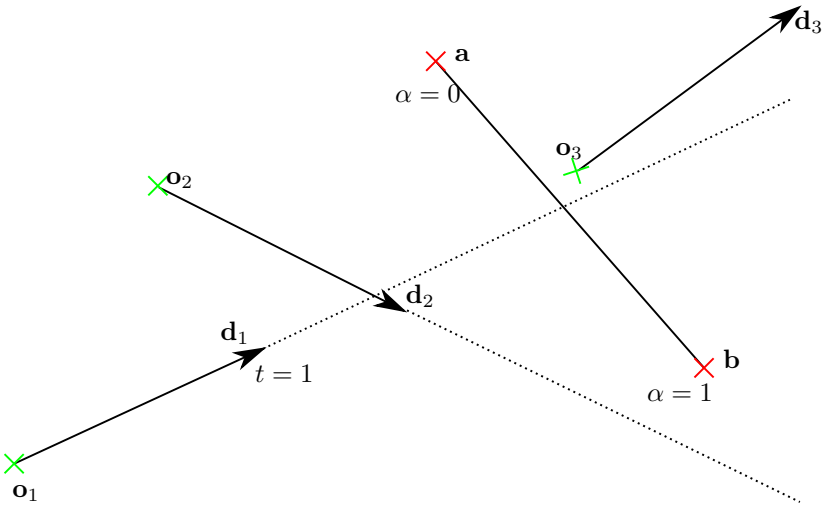

Figure 1: Ray-line intersection of two rays. The line is specified by the points $\mathbf{a}$ and $\mathbf{b}$ and the rays are defined by the origins $\mathbf{o}_i$ and directions $\mathbf{d}_i$. Ray $(\mathbf{o}_1, \mathbf{d}_1)$ satisfies the conditions $t \geq 0$ and $0 \leq \alpha \leq 1$ and therefore causes an intersection, ray $(\mathbf{o}_2, \mathbf{d}_2)$ dissatisfies the α condition and ray $(\mathbf{o}_3, \mathbf{d}_3)$ does not satisfy the t condition.

Another important shape to intersect are AABBs. They are rectangles aligned with the axis of the coordinate system so they require minimal memory space and intersection tests are as simple as possible. They most often used to surround complex objects or parts of it to reduce the amount of intersection tests. First the AABB of the object is tested and only if there is an intersection the actual fundamental shapes inside the AABB are tested. A 2D AABB is defined by two points $\mathbf{b}_{min}$ and $\mathbf{b}_{max}$ which represent the lower left and upper right corner of the rectangle. The intersection test is done by comparing the values of t at each of the axis aligned lines defining the box. The t values for the x axis aligned lines can be calculated as shown in Algorithm 1.

Algorithm 1: Intersection test for a AABB $(\mathbf{b}_{min}, \mathbf{b}_{max})$ with ray $(\mathbf{o}, \mathbf{d})$

$t_{x1} = \frac{b_{min\,x} - o_x}{d_x}$;
$t_{x2} = \frac{b_{max\,x} - o_x}{d_x}$;
$t_{min} = \min(t_{x1}, t_{x2})$;
$t_{max} = \max(t_{x1}, t_{x2})$;
$t_{y1} = \frac{b_{min\,y} - o_y}{d_y}$;
$t_{y2} = \frac{b_{max\,y} - o_y}{d_y}$;
$t_{min} = \max(t_{min}, \min(t_{y1}, t_{y2}))$;
$t_{max} = \min(t_{max}, \max(t_{x1}, t_{x2}))$;
if $t_{min} \geq 0$ *and* $t_{min} \leq t_{max}$ **then**
  AABB was hit!
end

If the conditions $t_{min} \leq t_{max}$ and $t_{min} \geq 0$ hold there is an intersection. This process is better understood visually and is illustrated in Figure 2. If normals are needed they can be easily calculated since there are only four possibilities depending on which side of the box is intersected first.

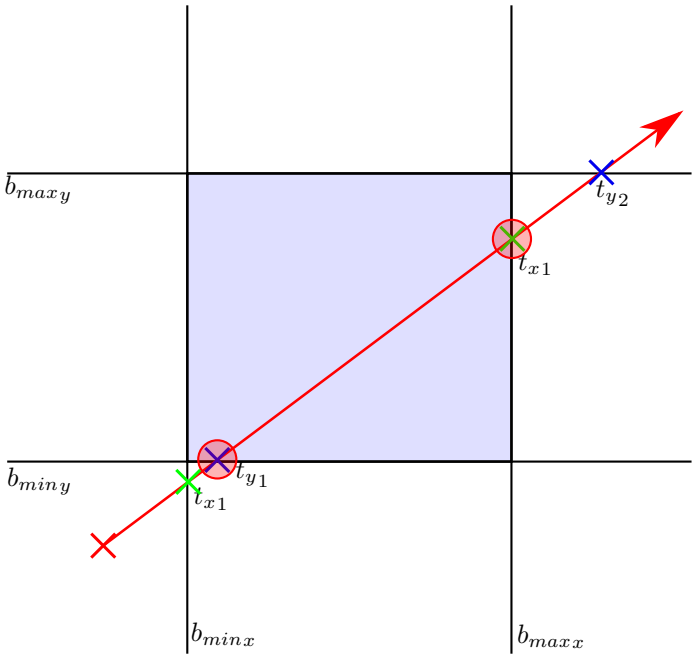

Figure 2: Ray-AABB intersection, where first the intersection points with the x axis (marked in green) and y axis (marked in blue) are calculated. Each of the values t_{i1}, t_{i2} are then split into the minimum and the maximum of the two. Both maximums are then compared and the minimum is chosen as the final t_{max}. Analogously the minimums are compared and the maximum is chosen as t_{min} (marked with red circles). Thus there is an intersection with an entry point $\mathbf{o} + t_{min}\mathbf{d}$ and an exit point $\mathbf{o} + t_{max}\mathbf{d}$ and the normals can be calculated depending on which sides the points reside.

The AABB intersection becomes really handy once one wants to use them for ray tracing acceleration techniques, as they can easily be constructed to surround a cloud of points and then be used as a spacial subdivider in a tree structure. This is described in detail in the next chapter. Other shapes like circles or ellipses are intersected in a similar way but as they are not used in the example below the intersection process is not explained here.

Once an intersection takes place, the ray will be either reflected, terminated or refraction occurs depending on the desired material of the object. Total reflection is only dependent on the incident angle θ_i to the normal of the surface at the hitpoint. Then the reflection angle θ_r is given by Eq. (6).

$$\theta_r = -\theta_i \tag{6}$$

A new ray is then generated at the hitpoint pointing in the direction given by θ_r. Due to limited floating point precision, it is required that the origin of the new ray is shifted by a small ϵ towards the reflection direction in order to make sure the ray is originated at the correct side of the material. Since the reflection is total the entire amount of power of the incident ray is transferred to the reflected ray.

When hitting a material that is transmissible for light the ray will be refracted at the boundary between the two media. The effects of matter on a light beam are described by Snell's law and the Fresnel equations. The ray is split into a reflected and a transmitted ray. The direction of the transmitted ray is governed by Snell's law in Eq. (7) which depends on the indices of refraction of the two media n_i and n_t.

$$n_i \sin(\theta_i) = n_t \sin(\theta_t) \tag{7}$$

Naturally, the reflected ray is still reflected as given in Eq. (6). The transmitted and reflected power can be calculated with the transmission- and reflection rates given by Fresnel's equations. These are dependent on the orientation of the polarization of the incident ray (perpendicular or parallel) to the surface.

$$R_\perp = \frac{\sin^2(\theta_1 - \theta_2)}{\sin^2(\theta_1 + \theta_2)} \quad R_\parallel = \frac{\tan^2(\theta_1 - \theta_2)}{\tan^2(\theta_1 + \theta_2)} \quad T_\perp = 1 - R_\perp \quad T_\parallel = 1 - R_\parallel \tag{8}$$

For unpolarized light the total rates are just given by the average.

$$R_{total} = \frac{R_\perp + R_\parallel}{2} \quad T_{total} = \frac{T_\perp + T_\parallel}{2} \tag{9}$$

4.2 Raytracing Acceleration

As the raytracer is later intended to be used in an iterative optimization algorithm, it is of vital importance that unnecessary computational cost is avoided. For a raytracer this can be achieved in a number of ways. The first and simplest way is to simply reorder the objects in a scene by a heuristic that describes the likelihood of on object to be the first object hit by the majority of the rays. Of course, this only works well if rays are shot into a scene from a dominant direction. Another way would be to subdivide the entire 2D scene with quadtrees and try to fill each branch of the tree with an equal amount of objects or shapes.

Similarly one can also subdivide an object itself and sort the fundamental shapes comprising that object into a quadtree. This method was chosen in this work as there are a limited amount of objects in the scene with the objects possibly being quite complex. Once the fundamental shapes of an object are known, they can be sorted into a quadtree of AABBs of a chosen depth. The outermost AABB is the root of the tree with four children, each encompassing the shapes inside their quarter of space as tightly as possible. This is recursively done until the desired depth is hit. The intersection test of an object then can be done by hitting the root AABB of the tree and then stepping through its children via breadth first search. Each AABB child the ray hits, is persued further and the ones the ray doesn't intersect are ignored. If a leaf has been hit all the shapes inside are then tested for intersection. Finally the t values of all the intersections are compared and the minimum and the maximum chosen as entry and exit points. The advantage of this is that AABB intersection tests are done really fast and a large number of fundamental shape intersection tests are avoided. An illustration of subdivision of a mirror comprised of line segments is given in Figure 3 and a the intersecting algorithm for a single object is givin in Algorithm 2.

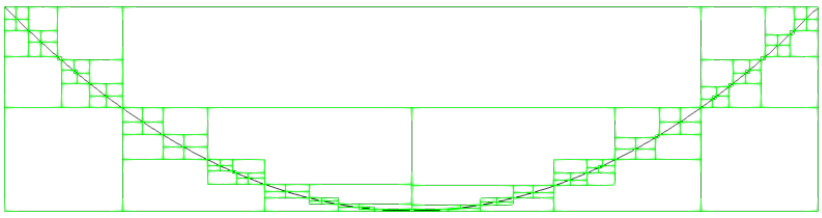

Figure 3: A parabolic mirror comprised of line segments subdivided by a quadtree of AABBs with depth 4. Note that the AABBs are encompassing their contained line segments as tightly as possible.

Algorithm 2: Intersection test for a single object subdivided by a quadtree

```
IntersectionResult objectResult;
objectResult.tEnter = MaxFloatingPoint;
objectResult.tLeave = MinFloatingPoint;
Queue treeQueue;
treeQueue.push(object.root);
while !treeQueue.empty() do
    tree = treeQueue.front();
    IntersectionResult aabbResult = tree.aabb.intersect(ray);
    if aabbResult.hit then
        for shapes in tree.shapes do
            IntersectionResult shapeResult = shape.intersect(ray);
            if shapeResult.hit then
                set objectResult appropiately;
            end
        end
        treeQueue.push(tree.children);
    end
    treeQueue.pop();
end
```

4.3 Sampling Techniques

The accuracy and performance of ray tracing simulations are heavily dependent upon using the correct sampling techniques. One could sample values on a uniform grid or equally spaced intervals. The problem with this is that there is no randomness or irregularity causing structured aliasing errors in most applications. Random sampling however always relies on some sort of random number generation. These are usually pseudo random numbers generated according to some distribution with the generation engine initialized with a seed. Usually when one samples according to some scheme only values in $[0, 1]$ are allowed. The returned sample is then later scaled to the desired range depending on the usecase. It is also to be noted that calls to a sampler must be ensured to be as effiecient as possible as a large number of calls will be made during the simulation.

The simplest sampling scheme is uniform sampling. It returns values uniformly and can be implemented right on top of the random number engine of the used system. The advantage of uniform sampling is that it produces close to random samples without the need for additional logic and therefore performance losses. The disadvantage is the irregular density of samples within the interval. There can be areas with a lot of samples and large gaps between. So in scenarios where there needs to be a more regular distribution of samples uniform sampling is not optimal.

For this reason another sampling technique called stratified uniform sampling exists. Here the domain is split into N equally spaced intervals and the uniform sampling occurs within each interval. This ensures that there is some amount of regularity while still keeping the randomness of uniform sampling. An application of stratified sampling would be the definition of a light source in the simulation. The direction or origin of the rays the light source emitts can be sampled according to stratified sampling to ensure a smooth illumination of the scene. The difference between uniform and stratified uniform sampling can be observed in Figure 4.

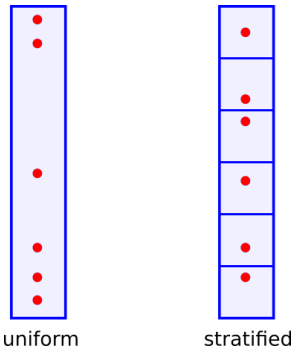

Figure 4: Uniform sampling of a 1D interval (left) vs. stratified sampling (right). Observe the large gap between samples on the left whereas the samples on the right are spaced out more equally.

A more advanced technique is to sample values where they are contributing the most. Suppose one wants to approximate some function f over a domain $[0, 1]$. The criterium that has to be met in order to calculate some quantity Q is given by any arbitrary integral in the domain. With uniform sampling of $x \in [x_1, x_2]$ by the sequence (x_i) and $i = 1...n$ the integral is approximated by Eq. (10).

$$Q = \int_{x_1}^{x_2} f(x)dx \approx \frac{1}{n} \sum_{i=1...n} f(x_i) \tag{10}$$

Now it is also possible to sample according to some other distribution. One just has to know the probability density function (pdf) p to know how likely it is that a sample x_i is generated. Probability density functions are nonnegative across the domain and their integral over the domain is always 1. This corresponds to the probability that a sampled value is in the interval $[x_1, x_2]$ which is of course the case when all possible values are within that interval. The more accurately the pdf p matches f the more $[x_1, x_2]$ is sampled at the points where the function f has a large contribution to the integral. This can significantly reduce the amount of samples needed to get an accurate approximation for the integral Q. The integral is then approximated by Eq. (11).

$$Q = \int_{x_1}^{x_2} f(x)dx \approx \frac{1}{n} \sum_{i=1...n} \frac{f(x_i)}{p(x_i)} \tag{11}$$

Observe that now the values for f are weighted by the likelihood p. If it is likely that a sample x_i is generated - larger p - then the contribution is worth less and vice versa. Of course this requires some preprocessing in order to generate the distribution P from a pdf p. Usually one wants to generate p from f but there are also cases where one might choose another function to deduce the pdf and subsequently the distribution.

One method of building such a distribution is to first integrate f in the domain $[x_1, x_2]$ and normalize the function so the integral is guaranteed to be 1. Then to sample according to the resulting pdf one discretizes the pdf to a finite amount of equidistant intervals $p_i = [x_{1i}, x_{2i}]$. For an interval p_j the values $p(x_{2i})$ for all $i < j$ are then summed up progressively assigned to that interval. These resulting rectangles are then stacked so that each rectangle represents an interval in x and the corresponding values of the sum of the pdf. Then a sample $\xi \in [0, 1]$ is drawn from a uniform sampler and the corresponding interval in which the value ξ lies in is searched in the list of intervals using binary search. Once the interval is found we interpolate linearly between the lower boundary x_{1i} and x_{2i} depending on ξ. Then the resulting $x \in [x_{1i}, x_{2i}]$ is returned along the value of the pdf $p(x)$ as a sample. This process approximates the distribution P and is called the inversion method. It is better understood when visualized as in Figure 5.

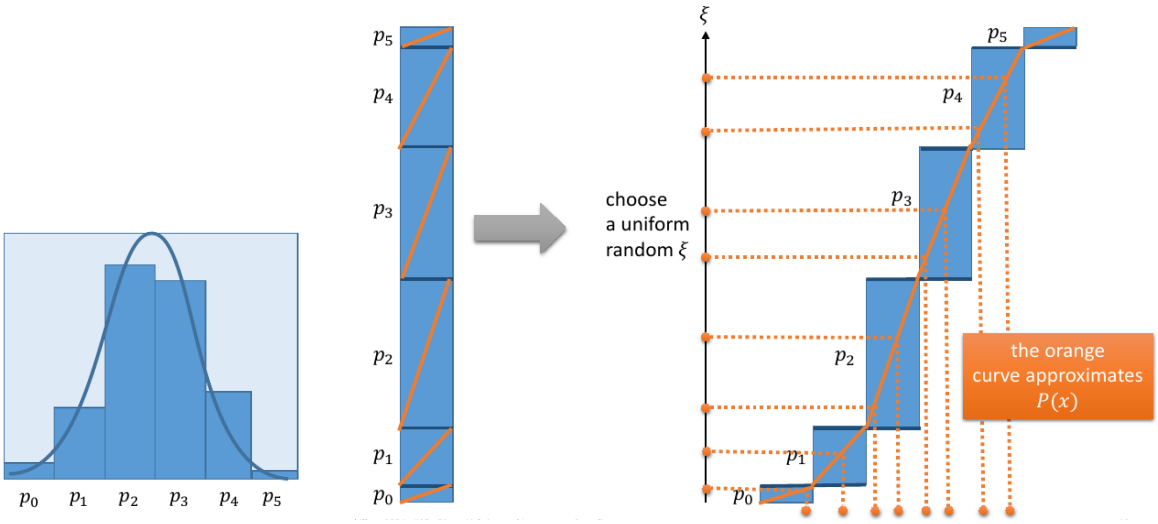

Figure 5: Visualization of inversion method to sample according the pdf on the left. The intevals p_i are summed up and stacked. Then a $\xi \in [0, 1]$ is uniformly sampled and the corresponding interval is searched. The returned value is interpolated linearly according to the value of $/xi$ within the rectangle.

References

[1] Opengl mathematics (glm). https://glm.g-truc.net/0.9.4/api/index.html. [Online; accessed 28-March-2022].

[2] Will Schroeder, Ken Martin, and Bill Lorensen. *The Visualization Toolkit (4th ed.)*. Kitware, 2006.