Importance Sampling Spherical Harmonics

Wojciech Jarosz[1,2]
Nathan A. Carr[2]
Henrik Wann Jensen[1]

[1]University of California, San Diego
[2]Adobe Systems Incorporated

April 2, 2009

Thursday, 6 September 12

Spherical Harmonic Sampling

In this paper we present the first practical method for importance sampling functions represented as spherical harmonics.
* this video shows our method sampling a spherical harmonic function in real time
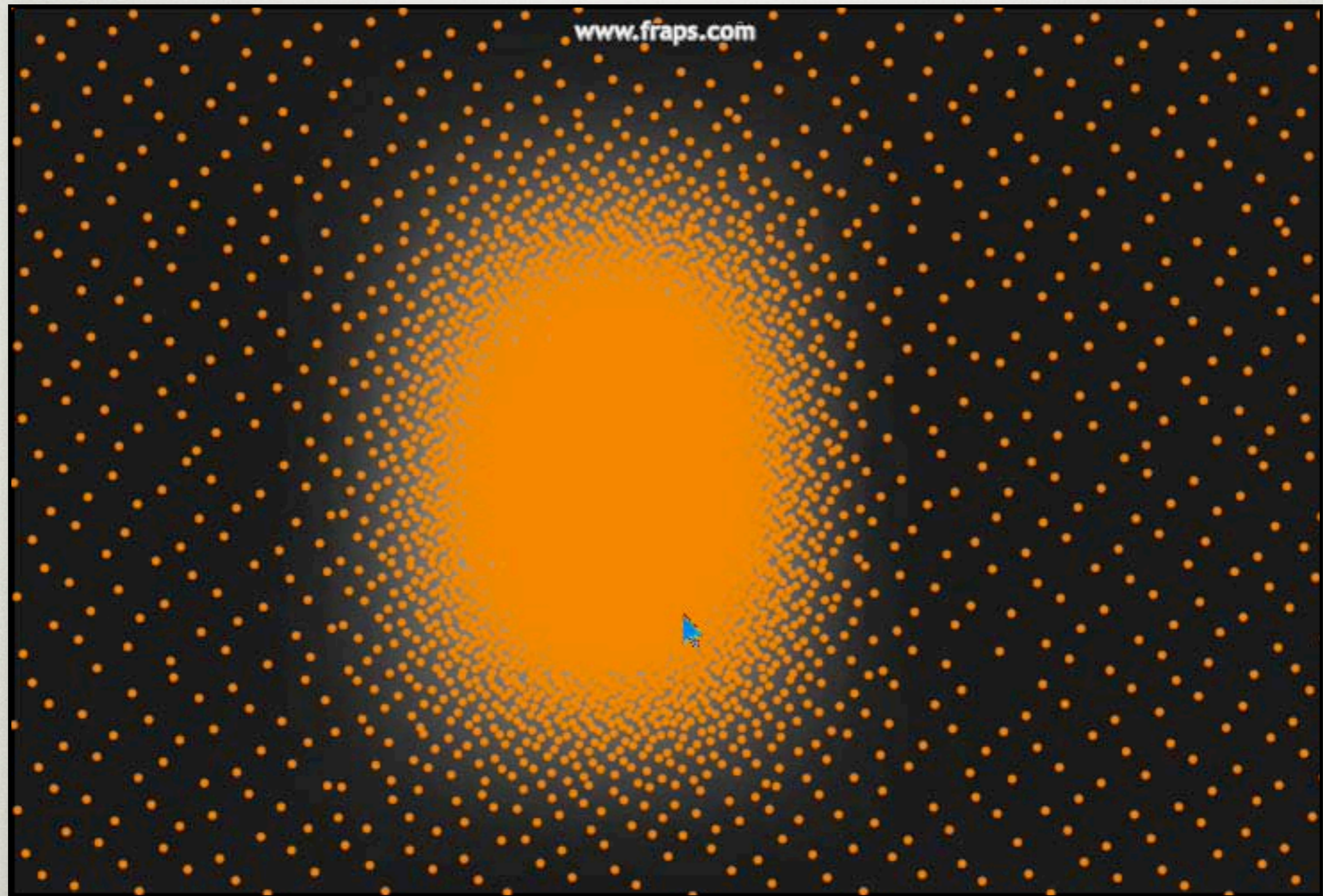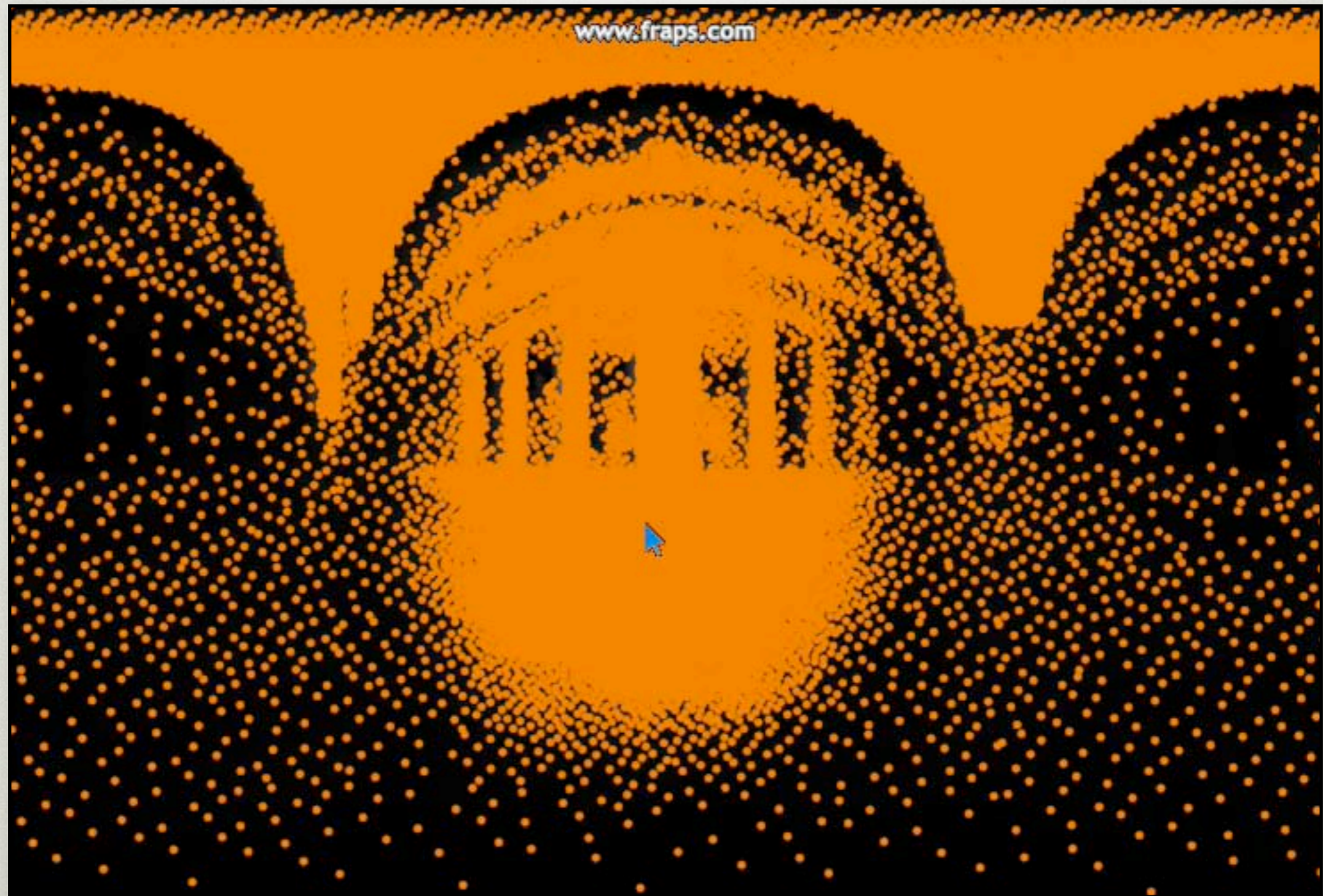
Spherical Harmonic Sampling

Thursday, 6 September 12

In this paper we present the first practical method for importance sampling functions represented as spherical harmonics.
* this video shows our method sampling a spherical harmonic function in real time

SH x Haar Product Sampling

* Our method can also be used to sample the product of a spherical harmonic function and a haar wavelet function.
* In this video we show the product of a BRDF represented using spherical harmonics, and an environment map stored using haar wavelets.
* Since spherical harmonics support efficient rotation, we are able to rotate the BRDF on-the-fly while sampling the product.
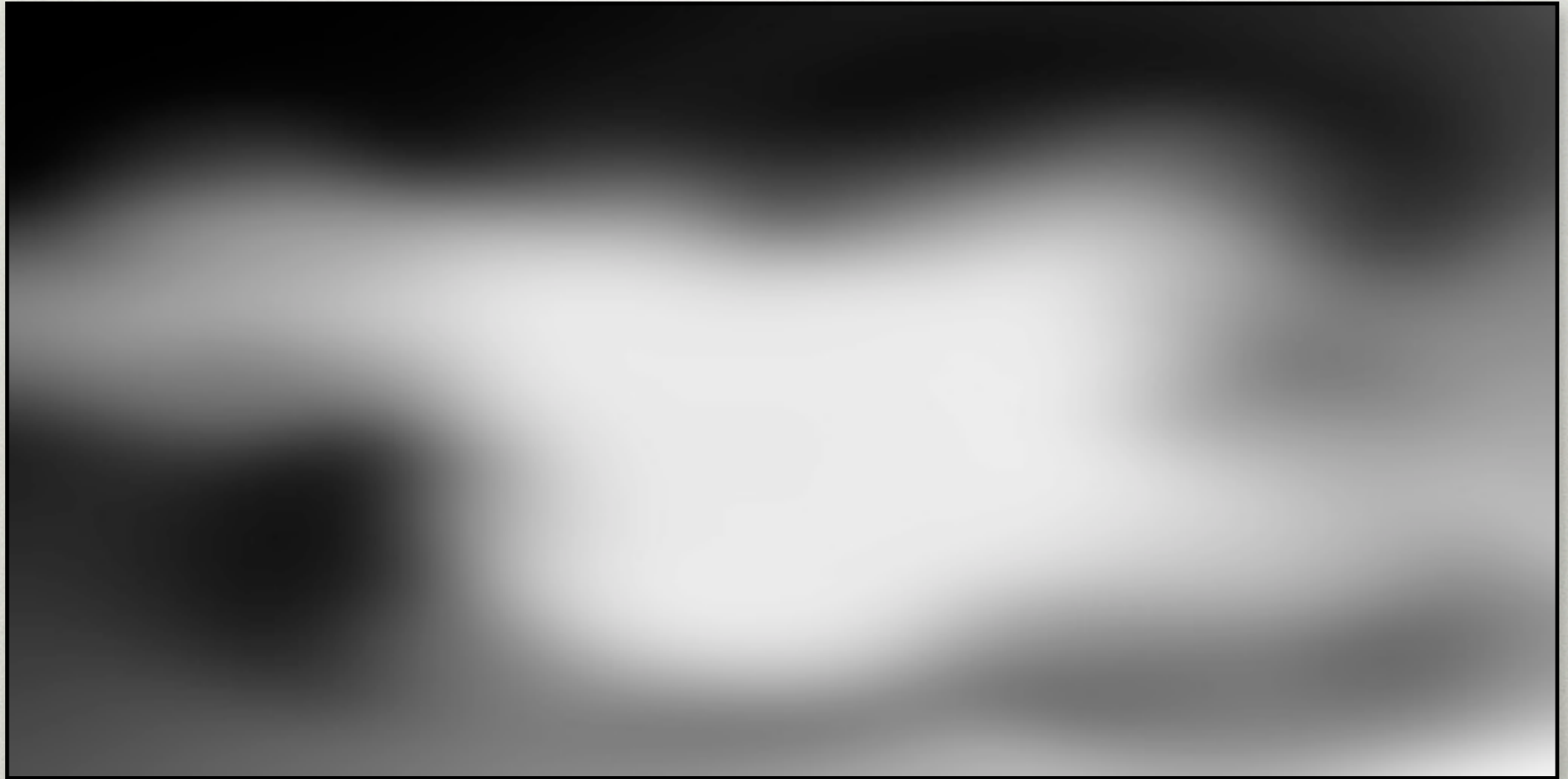
Outline

- What is it?

- Why is it desirable?

- How do we do it?

- How well does it work?

Outline

- ## What is it?

- Why is it desirable?

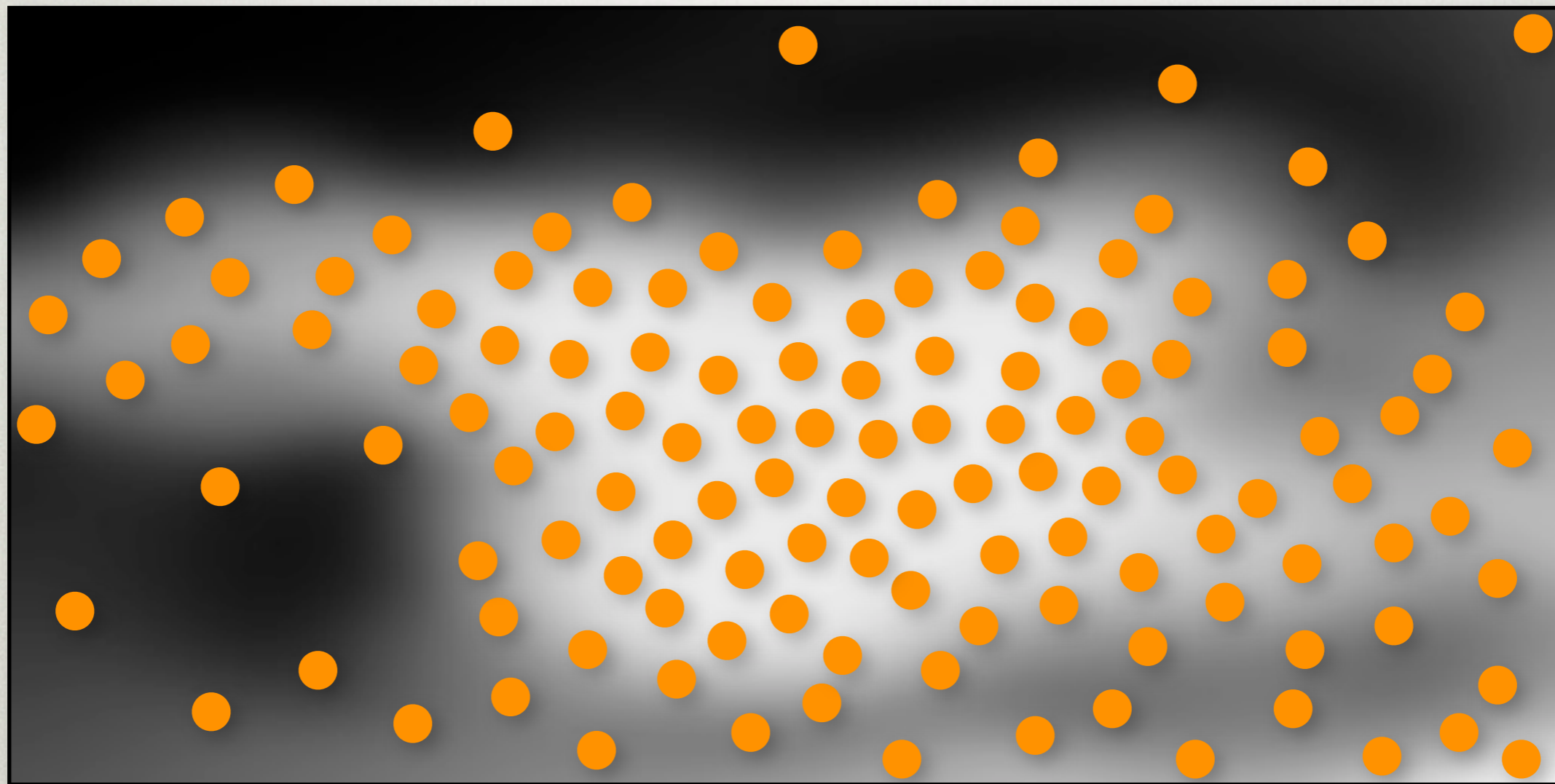- How do we do it?

- How well does it work?

Importance Sampling

* Given some function defined over an arbitrary domain, importance sampling deals with the task of distributing random samples according to this function.

* This means we want to generate more samples where the function is high, and few samples were the function is low

* In this paper, the domain of the function is the sphere, and so we can think of this, for instance, as an image in spherical coordinates

IMPORTANCE SAMPLING

* Given some function defined over an arbitrary domain, importance sampling deals with the task of distributing random samples according to this function.
* This means we want to generate more samples where the function is high, and few samples were the function is low
* In this paper, the domain of the function is the sphere, and so we can think of this, for instance, as an image in spherical coordinates
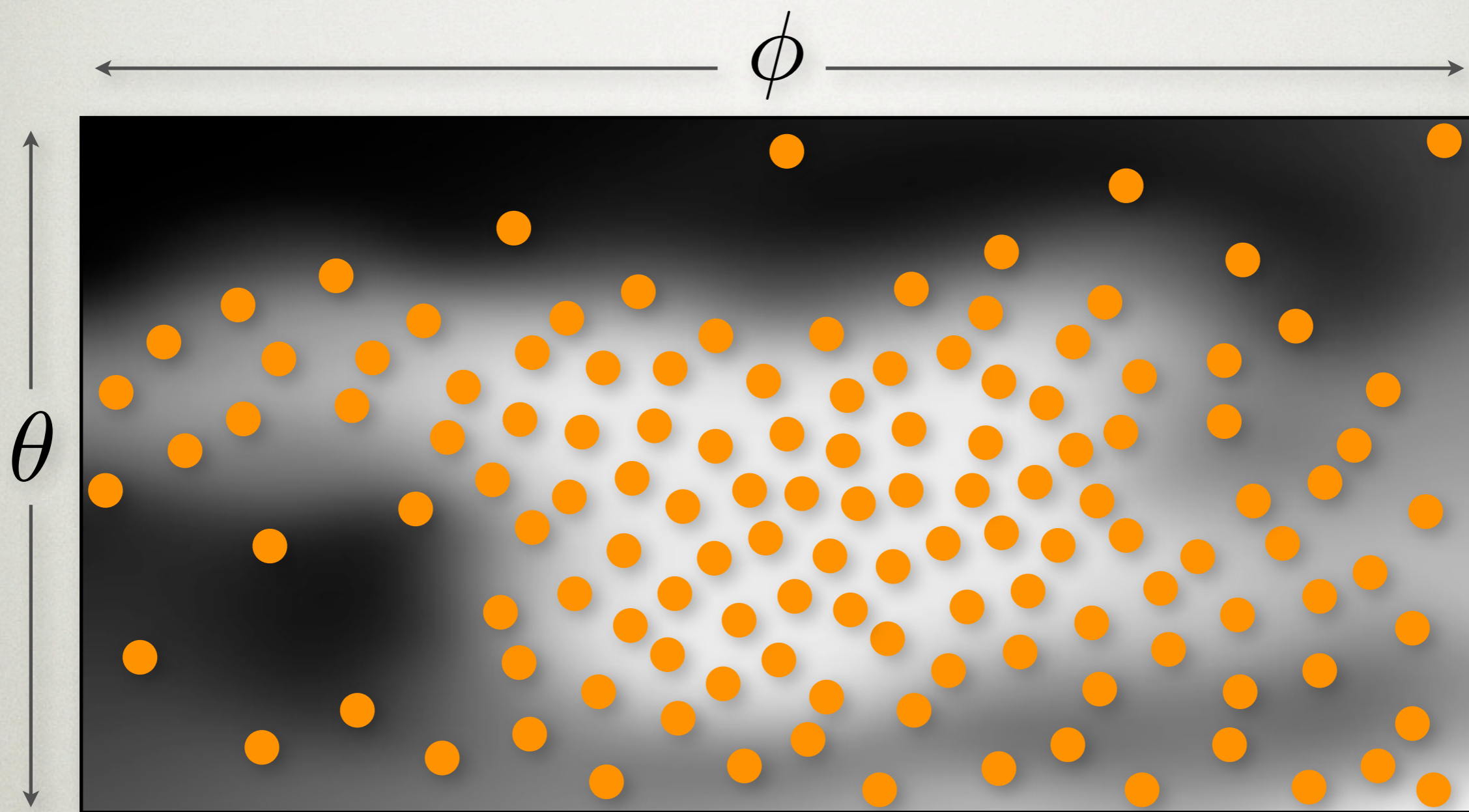
Importance Sampling

$$pdf(\vec{\omega}) = pdf(\theta, \phi) \propto f(\vec{\omega})$$

* Given some function defined over an arbitrary domain, importance sampling deals with the task of distributing random samples according to this function.
* This means we want to generate more samples where the function is high, and few samples were the function is low
* In this paper, the domain of the function is the sphere, and so we can think of this, for instance, as an image in spherical coordinates

Importance Sampling

- There are many applications for importance sampling.

- For rendering:

  - Variance reduction in Monte Carlo integration

$$\int f(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{pdf(x_i)}$$

* In rendering, importance sampling is a method for reducing variance in Monte Carlo integration.
* To reduce variance in a Monte Carlo estimator, we want the term inside the summation (click) to be as close to a constant as possible.
* This is achieved by choosing random samples from a pdf which is as similar to the integrand, f, as possible.

Importance Sampling

- ## There are many applications for importance sampling.

- ## For rendering:

  - ## Variance reduction in Monte Carlo integration

$$\int f(x)\ dx \approx \frac{1}{N} \sum_{i=1}^{N} \boxed{\frac{f(x_i)}{pdf(x_i)}}$$

* In rendering, importance sampling is a method for reducing variance in Monte Carlo integration.

* To reduce variance in a Monte Carlo estimator, we want the term inside the summation (click) to be as close to a constant as possible.

* This is achieved by choosing random samples from a pdf which is as similar to the integrand, f, as possible.

Importance Sampling

$$\int f(x)\ dx \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{pdf(x_i)}$$

Thursday, 6 September 12

* In rendering, importance sampling is a method for reducing variance in Monte Carlo integration.
* To reduce variance in a Monte Carlo estimator, we want the term inside the summation (click) to be as close to a constant as possible.
* This is achieved by choosing random samples from a pdf which is as similar to the integrand, f, as possible.

Monte Carlo

$$\int f(x)\, dx \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{pdf(x_i)}$$

* In the context of rendering, the integrand for computing illumination is the product of three functions, fr, the cosine-weighted brdf, L, the incident lighting, and V the visibility function

Monte Carlo

$$\int f(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} \boxed{\frac{f(x_i)}{pdf(x_i)}}$$

when computing the illumination integral:

$$L(\vec{\omega}') = \int_{\Omega} f_r(\vec{\omega}, \vec{\omega}')\,L(\vec{\omega})\,V(\vec{\omega})\,d\vec{\omega}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} \boxed{\frac{f_r(\vec{\omega}_i, \vec{\omega}')\,L(\vec{\omega}_i)\,V(\vec{\omega})}{pdf(\vec{\omega}_i)}}$$

* In the context of rendering, the integrand for computing illumination is the product of three functions, fr, the cosine-weighted brdf, L, the incident lighting, and V the visibility function

PREVIOUS WORK

$$L(\vec{\omega}') \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f_r(\vec{\omega}_i, \vec{\omega}') \, L(\vec{\omega}_i) \, V(\vec{\omega})}{pdf(\vec{\omega}_i)}$$

* There has been a lot of previous work on importance sampling in the context of rendering. These previous approaches can be categorized according to which portion of the illumination integral they attempt to importance sampling.
* Early approaches focused on importance sampling the BRDF function. Most of these methods create an analytic warping method which inverts the BRDF for sampling purposes. This can be very effective if the BRDF is fairly specular, however it provides little benefit for diffuse BRDFs.
* In the presence of complex distant illumination (represented using environment maps) it is possible to create a pdf which matches the incident lighting. This is very effective if your lighting is complex and your material are diffuse, but is not as effective if you have glossy or specular materials.
* However, optimally, we wish to create a pdf which is proportional to the whole integrand. This is what product sampling techniques try to do. And there have been a number of techniques for this, including our work from SIGGRAPH 2005 on wavelet importance sampling.

Previous Work

$$L(\vec{\omega}') \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f_r(\vec{\omega}_i, \vec{\omega}') \, L(\vec{\omega}_i) \, V(\vec{\omega})}{pdf(\vec{\omega}_i)}$$
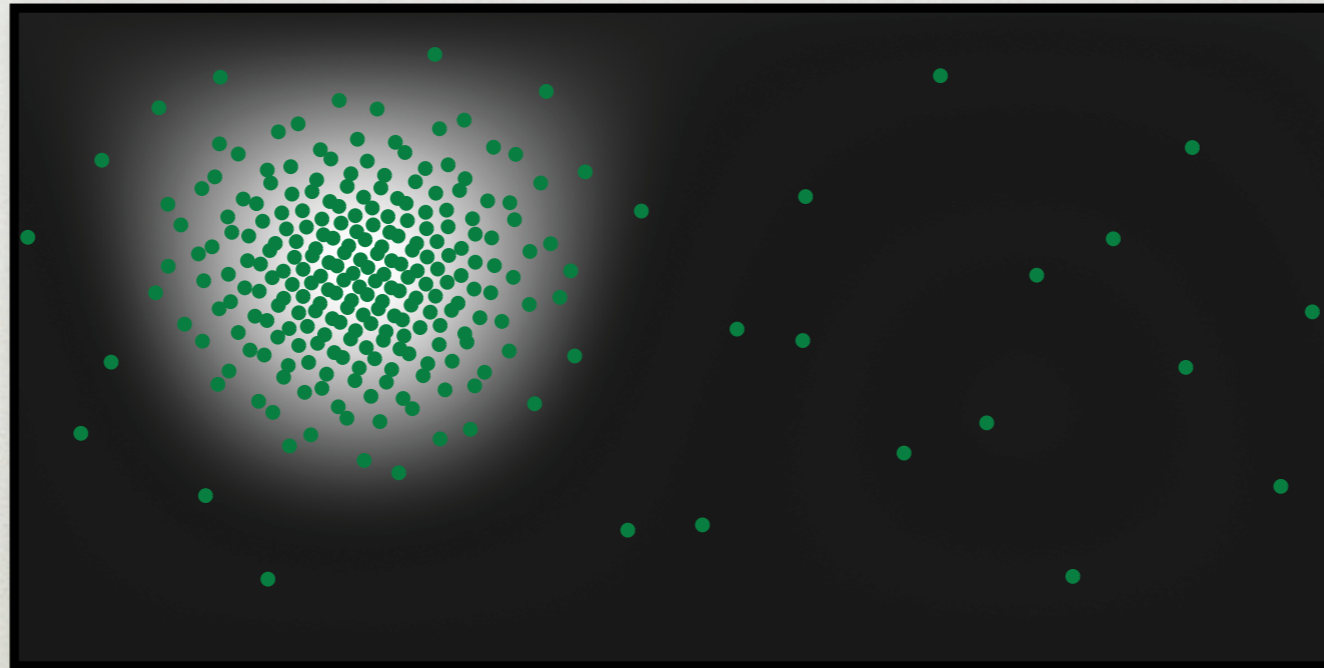
* There has been a lot of previous work on importance sampling in the context of rendering. These previous approaches can be categorized according to which portion of the illumination integral they attempt to importance sampling.
* Early approaches focused on importance sampling the BRDF function. Most of these methods create an analytic warping method which inverts the BRDF for sampling purposes. This can be very effective if the BRDF is fairly specular, however it provides little benefit for diffuse BRDFs.
* In the presence of complex distant illumination (represented using environment maps) it is possible to create a pdf which matches the incident lighting. This is very effective if your lighting is complex and your material are diffuse, but is not as effective if you have glossy or specular materials.
* However, optimally, we wish to create a pdf which is proportional to the whole integrand. This is what product sampling techniques try to do. And there have been a number of techniques for this, including our work from SIGGRAPH 2005 on wavelet importance sampling.

- BRDF importance sampling

Shirley 91, Ward 92, Lafortune 97, Lalonde 97, Claustres et al. 03,04, Matusik 03, Lawrence et al. 04, and many more.

$$L(\vec{\omega}') \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\boxed{f_r(\vec{\omega}_i, \vec{\omega}')} L(\vec{\omega}_i) V(\vec{\omega})}{pdf(\vec{\omega}_i)} \qquad pdf(\vec{\omega}_i) \propto \boxed{f_r(\vec{\omega}_i, \vec{\omega}')}$$

* There has been a lot of previous work on importance sampling in the context of rendering. These previous approaches can be categorized according to which portion of the illumination integral they attempt to importance sampling.
* Early approaches focused on importance sampling the BRDF function. Most of these methods create an analytic warping method which inverts the BRDF for sampling purposes. This can be very effective if the BRDF is fairly specular, however it provides little benefit for diffuse BRDFs.
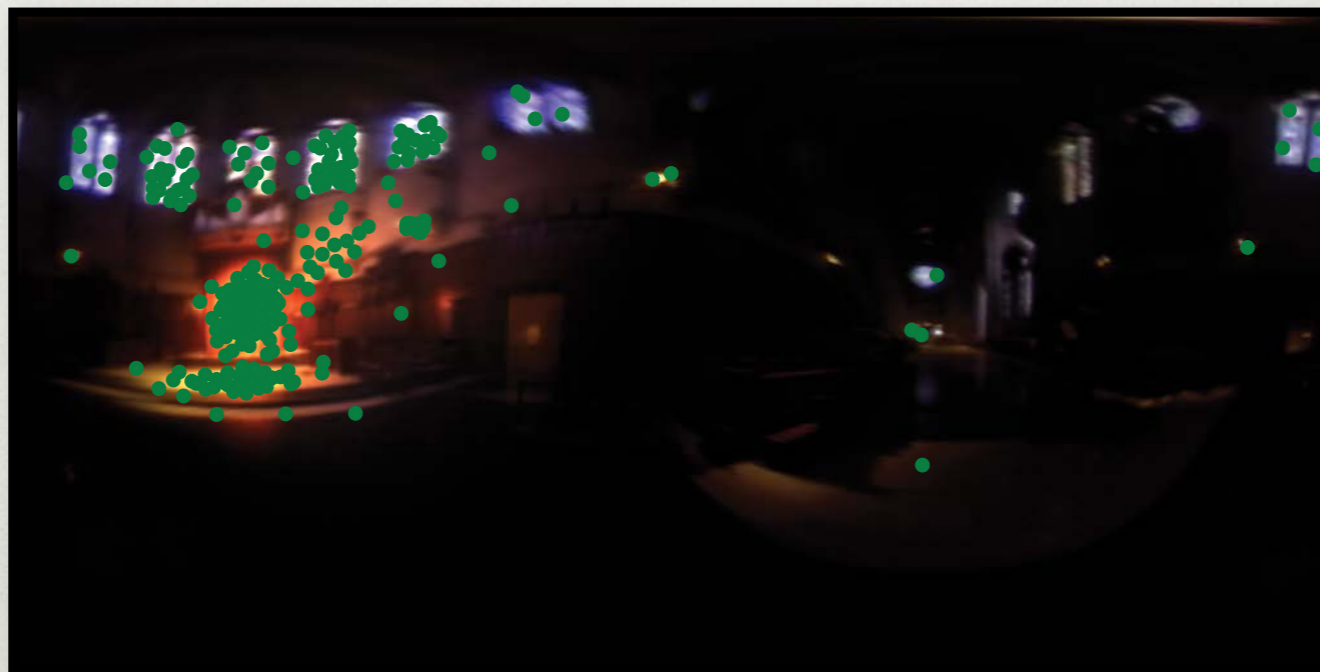* In the presence of complex distant illumination (represented using environment maps) it is possible to create a pdf which matches the incident lighting. This is very effective if your lighting is complex and your material are diffuse, but is not as effective if you have glossy or specular materials.
* However, optimally, we wish to create a pdf which is proportional to the whole integrand. This is what product sampling techniques try to do. And there have been a number of techniques for this, including our work from SIGGRAPH 2005 on wavelet importance sampling.

Previous Work

- Environment importance sampling

Cohen and Debevec 01, Agarwal et al. 03, Kollig and Keller 03, and Ostromoukhov 04.

$$L(\vec{\omega}') \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f_r(\vec{\omega}_i, \vec{\omega}') \boxed{L(\vec{\omega}_i)} V(\vec{\omega})}{pdf(\vec{\omega}_i)} \qquad pdf(\vec{\omega}_i) \propto \boxed{L(\vec{\omega}_i)}$$

* There has been a lot of previous work on importance sampling in the context of rendering. These previous approaches can be categorized according to which portion of the illumination integral they attempt to importance sampling.
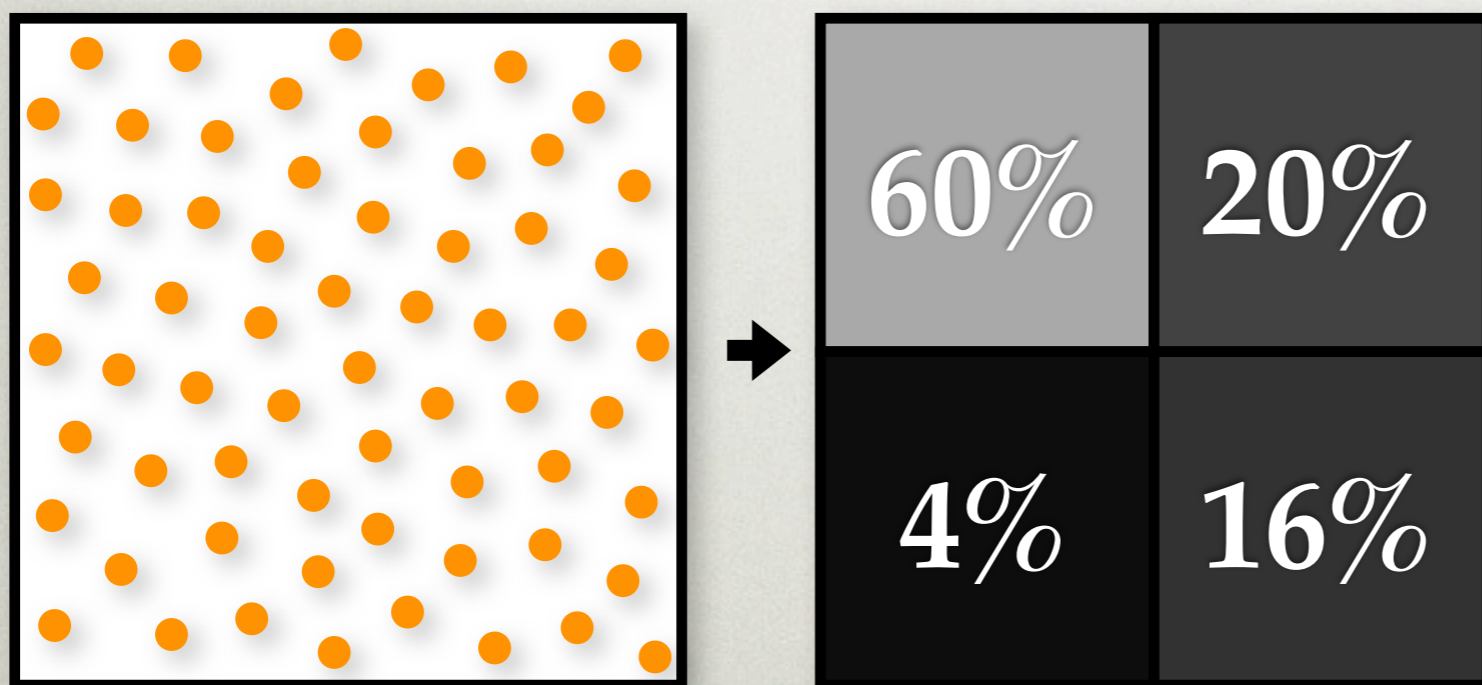* Early approaches focused on importance sampling the BRDF function. Most of these methods create an analytic warping method which inverts the BRDF for sampling purposes. This can be very effective if the BRDF is fairly specular, however it provides little benefit for diffuse BRDFs.
* In the presence of complex distant illumination (represented using environment maps) it is possible to create a pdf which matches the incident lighting. This is very effective if your lighting is complex and your material are diffuse, but is not as effective if you have glossy or specular materials.
* However, optimally, we wish to create a pdf which is proportional to the whole integrand. This is what product sampling techniques try to do. And there have been a number of techniques for this, including our work from SIGGRAPH 2005 on wavelet importance sampling.

Previous Work

- Product importance sampling

Burke et al. 05, Talbot et al. 05, Clarberg et al. 05, Cline et al. 06, Clarberg and Akenine-Moller 08.

$$L(\vec{\omega}') \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f_r(\vec{\omega}_i, \vec{\omega}') \, L(\vec{\omega}_i) \, V(\vec{\omega})}{pdf(\vec{\omega}_i)} \qquad pdf(\vec{\omega}_i) \propto f_r(\vec{\omega}_i, \vec{\omega}') L(\vec{\omega}_i)$$

* There has been a lot of previous work on importance sampling in the context of rendering. These previous approaches can be categorized according to which portion of the illumination integral they attempt to importance sampling.
* Early approaches focused on importance sampling the BRDF function. Most of these methods create an analytic warping method which inverts the BRDF for sampling purposes. This can be very effective if the BRDF is fairly specular, however it provides little benefit for diffuse BRDFs.
* In the presence of complex distant illumination (represented using environment maps) it is possible to create a pdf which matches the incident lighting. This is very effective if your lighting is complex and your material are diffuse, but is not as effective if you have glossy or specular materials.
* However, optimally, we wish to create a pdf which is proportional to both the lighting and the BRDF. This is what product sampling techniques try to do. And there have been a number of techniques for this, including our work from SIGGRAPH 2005 on wavelet importance sampling. This paper is most closely related to this work, so I'll describe this technique in a bit more detail.
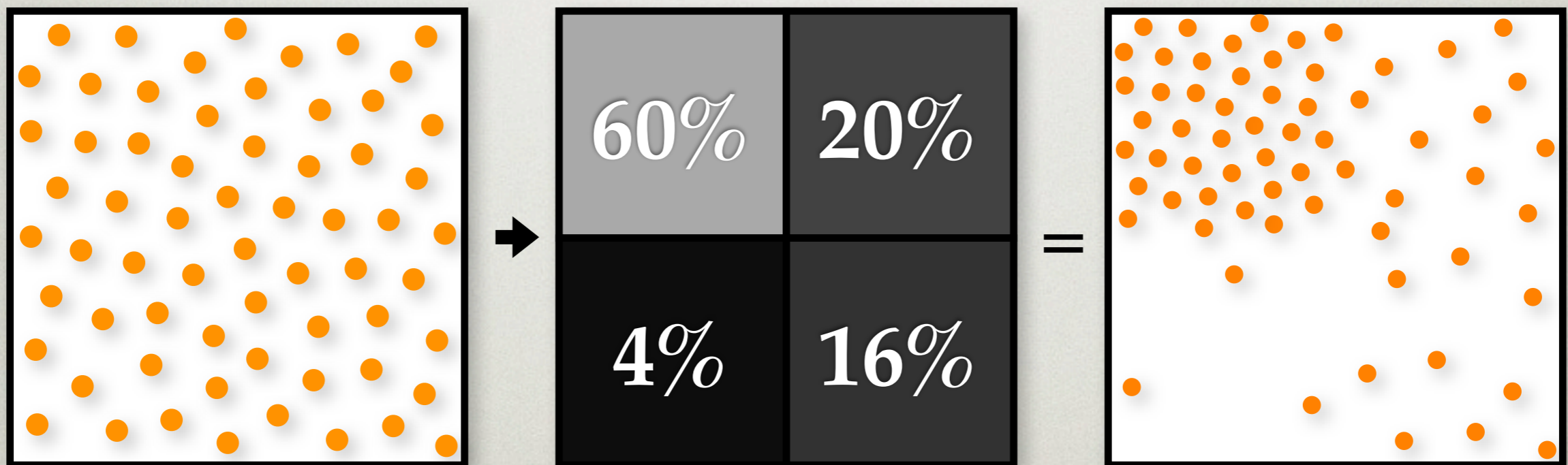
Clarberg et al. 05

Thursday, 6 September 12

* WIS uses a technique called hierarchical sample warping to perform importance sampling
* Given an input point set, and a hierarchical representation of a function
* (click), either as a Haar wavelet or as a mipmap,
* (click), hierarchical sample warping recursively warp the points to match the importance function at each resolution level.
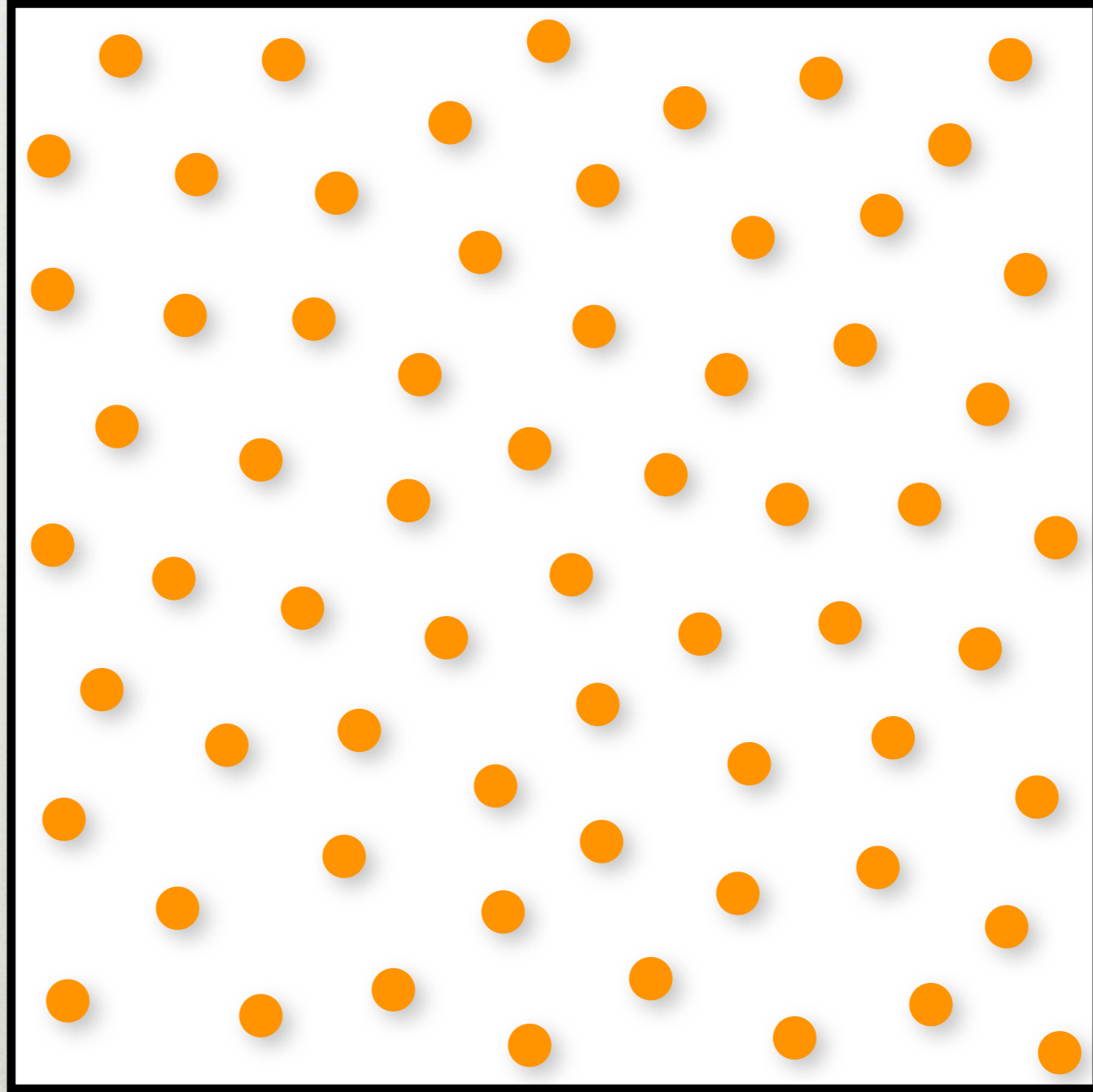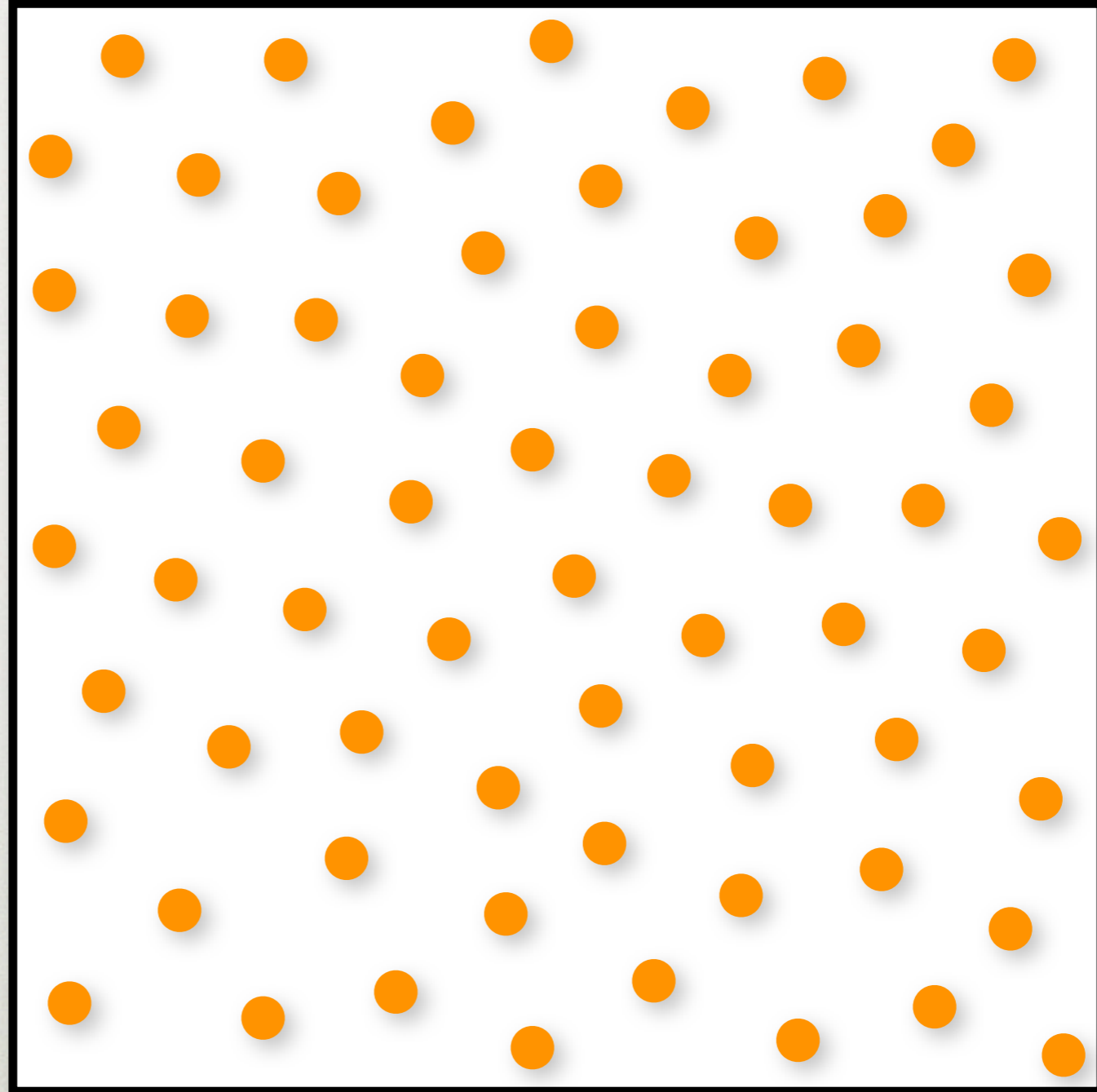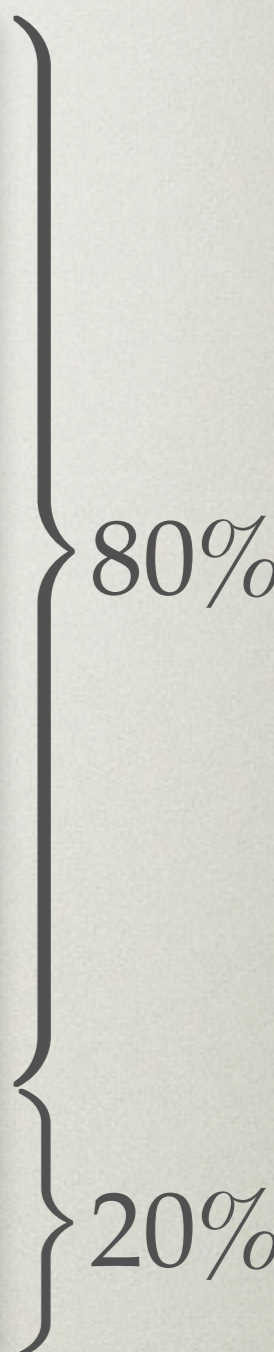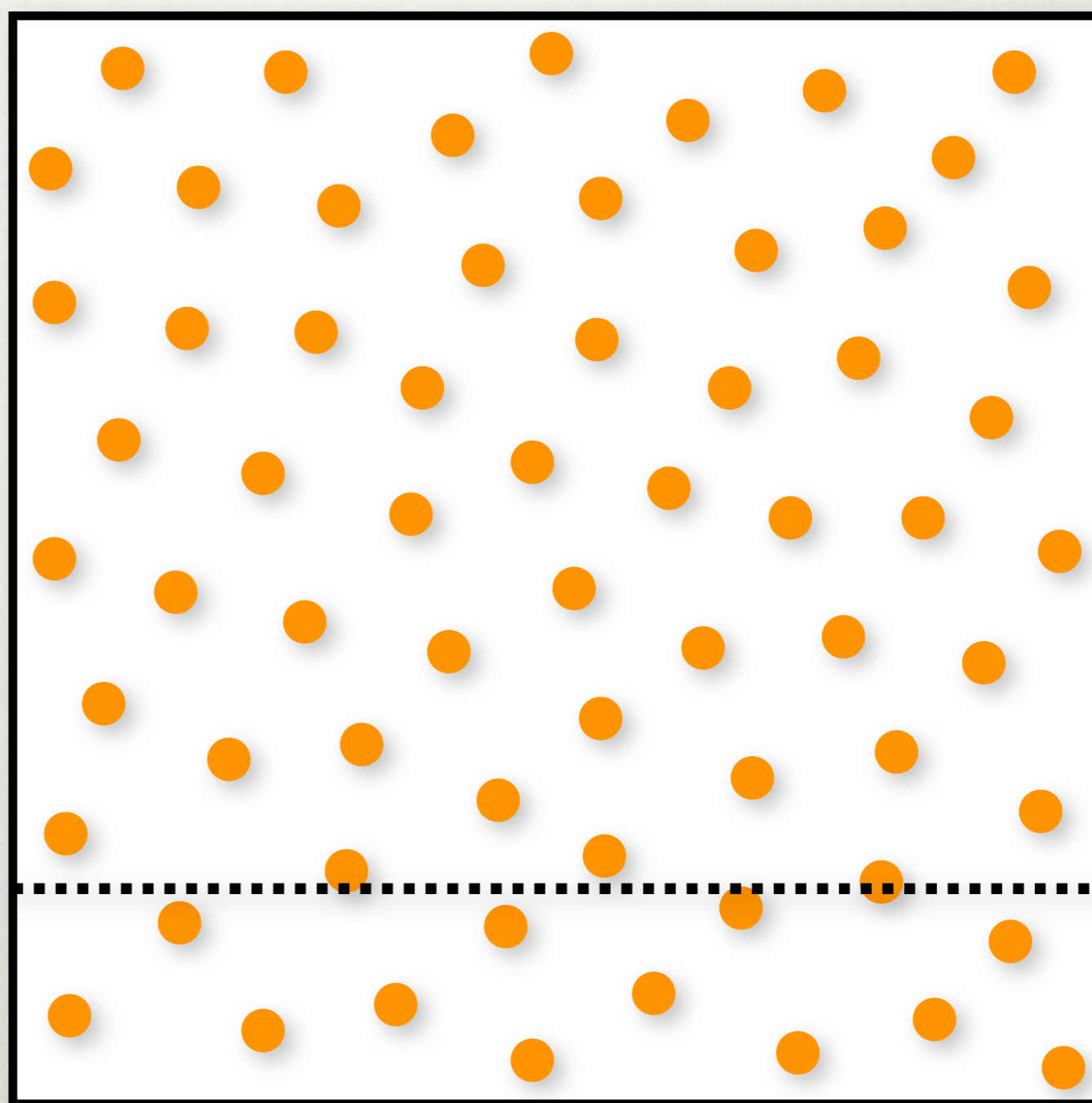
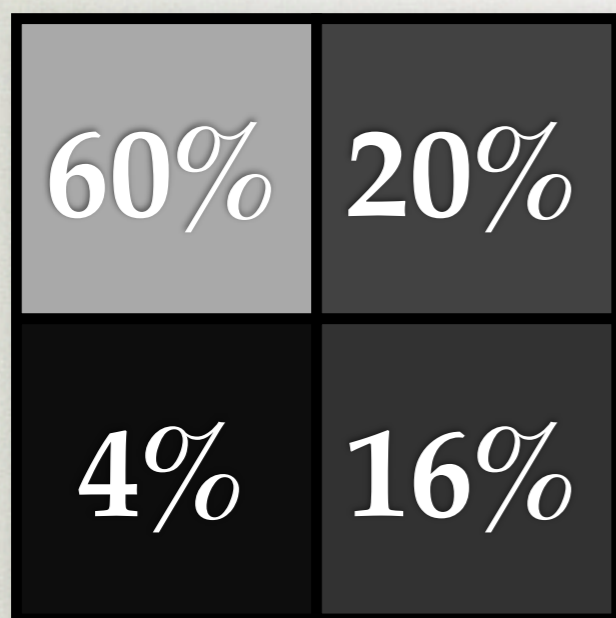# HIERARCHICAL SAMPLE WARPING (HSW)

- Given an input point set and a hierarchical representation of an importance function:
  - recursively warp point set to match the importance function

60%	20%
4%	16%

Clarberg et al. 05

Thursday, 6 September 12

* WIS uses a technique called hierarchical sample warping to perform importance sampling
* Given an input point set, and a hierarchical representation of a function
* (click), either as a Haar wavelet or as a mipmap,
* (click), hierarchical sample warping recursively warp the points to match the importance function at each resolution level.
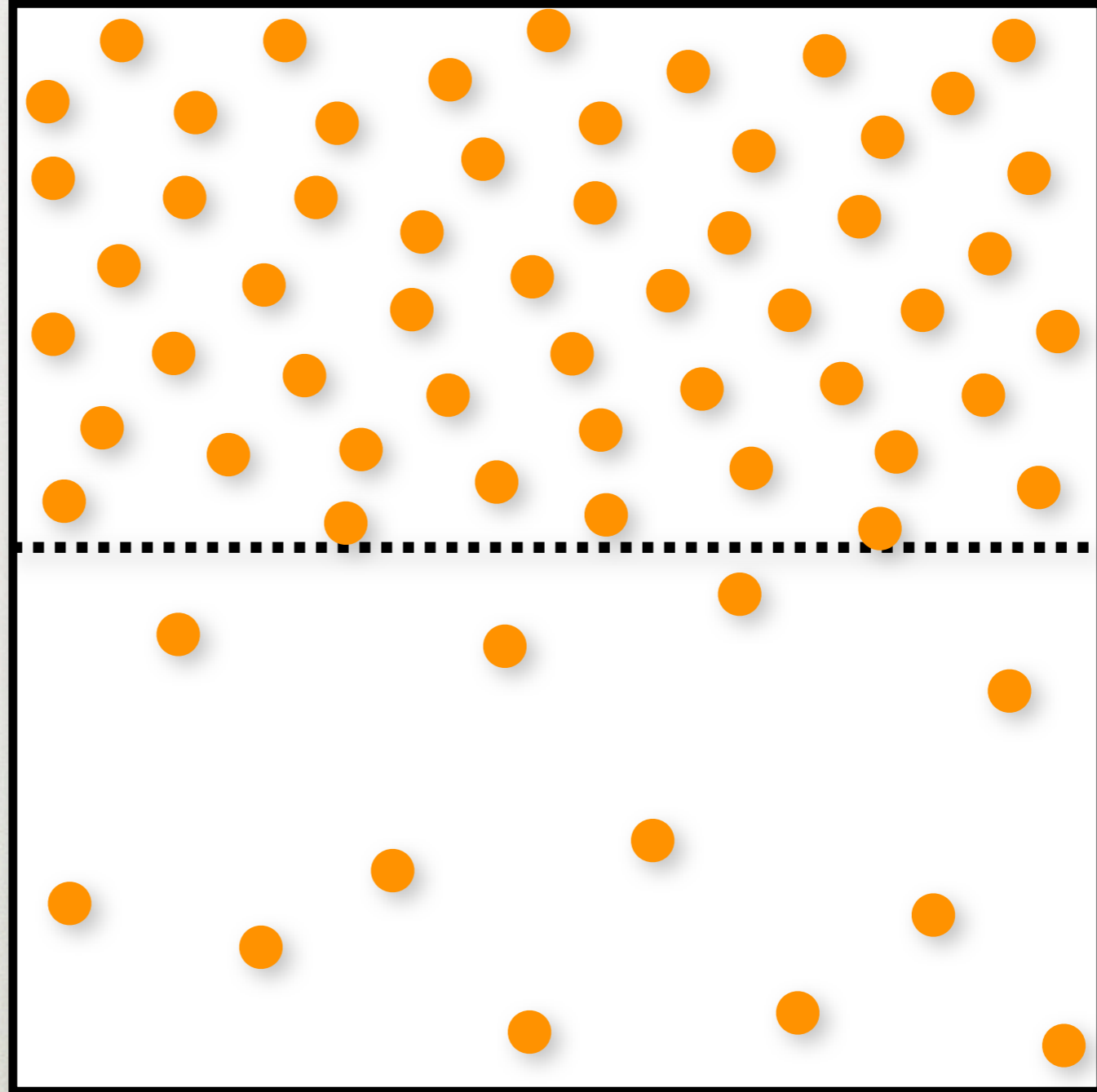
Clarberg et al. 05

Thursday, 6 September 12

* this procedure operates recursively on the levels of a mipmap or Haar wavelet.
* From each level of a mipmap we can determine the total amount of energy contained within each of the 4 child quadrants of the function.
* This process starts at the root.
* First the marginal probabilities for each row are computed (in this case 80% for the top row, and 20% for the bottom row).
* 80% of the samples should be placed in the top half of the importance function.
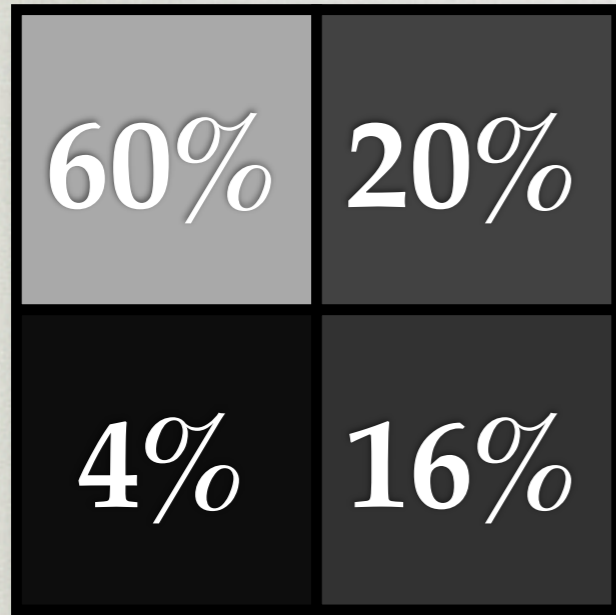* This is accomplished by splitting the point set, and warping the points to fit back into the two rows.

Clarberg et al. 05

Thursday, 6 September 12

* this procedure operates recursively on the levels of a mipmap or Haar wavelet.
* From each level of a mipmap we can determine the total amount of energy contained within each of the 4 child quadrants of the function.
* This process starts at the root.
* First the marginal probabilities for each row are computed (in this case 80% for the top row, and 20% for the bottom row).
* 80% of the samples should be placed in the top half of the importance function.
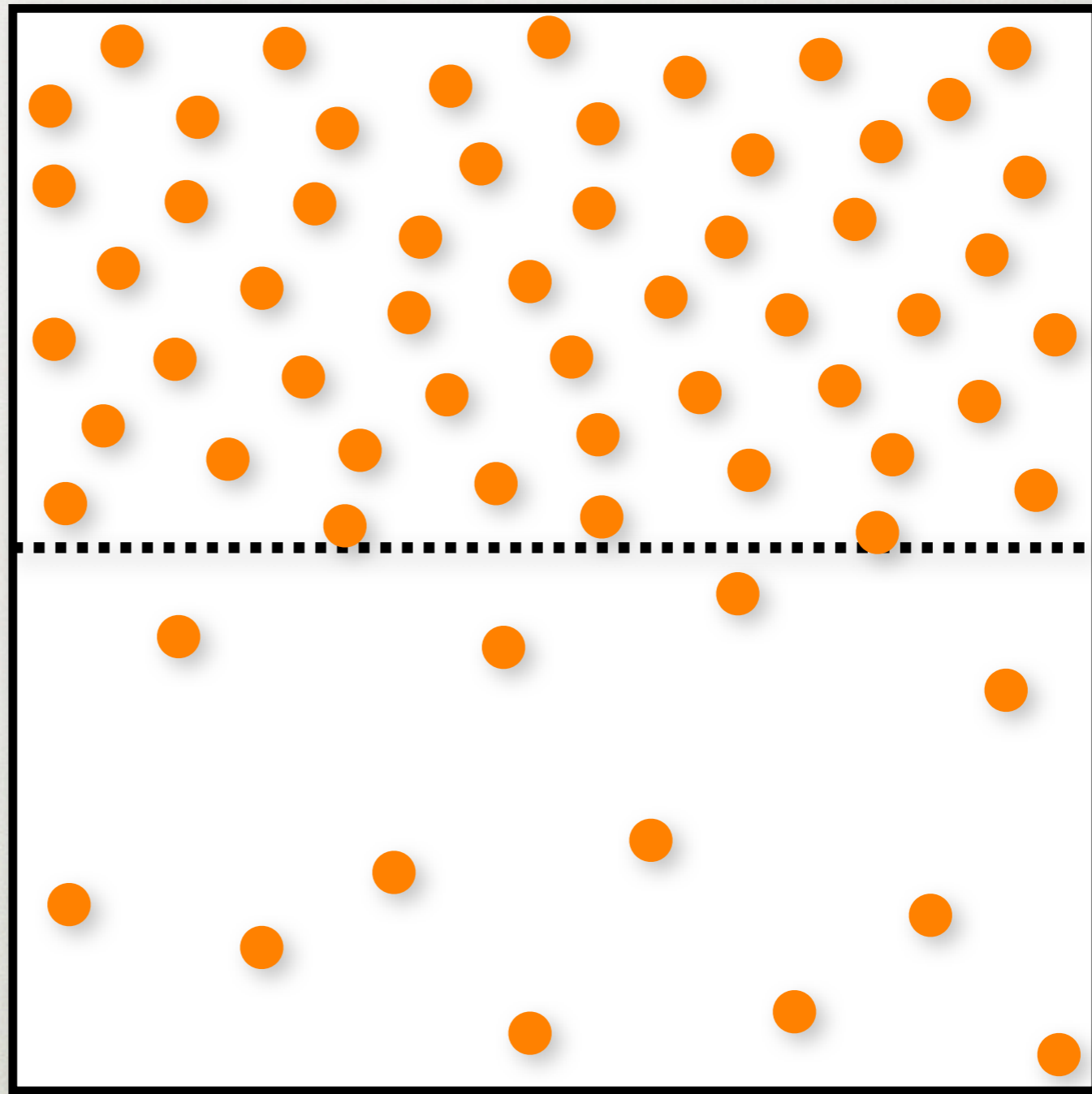* This is accomplished by splitting the point set, and warping the points to fit back into the two rows.

Clarberg et al. 05

* this procedure operates recursively on the levels of a mipmap or Haar wavelet.
* From each level of a mipmap we can determine the total amount of energy contained within each of the 4 child quadrants of the function.
* This process starts at the root.
* First the marginal probabilities for each row are computed (in this case 80% for the top row, and 20% for the bottom row).
* 80% of the samples should be placed in the top half of the importance function.
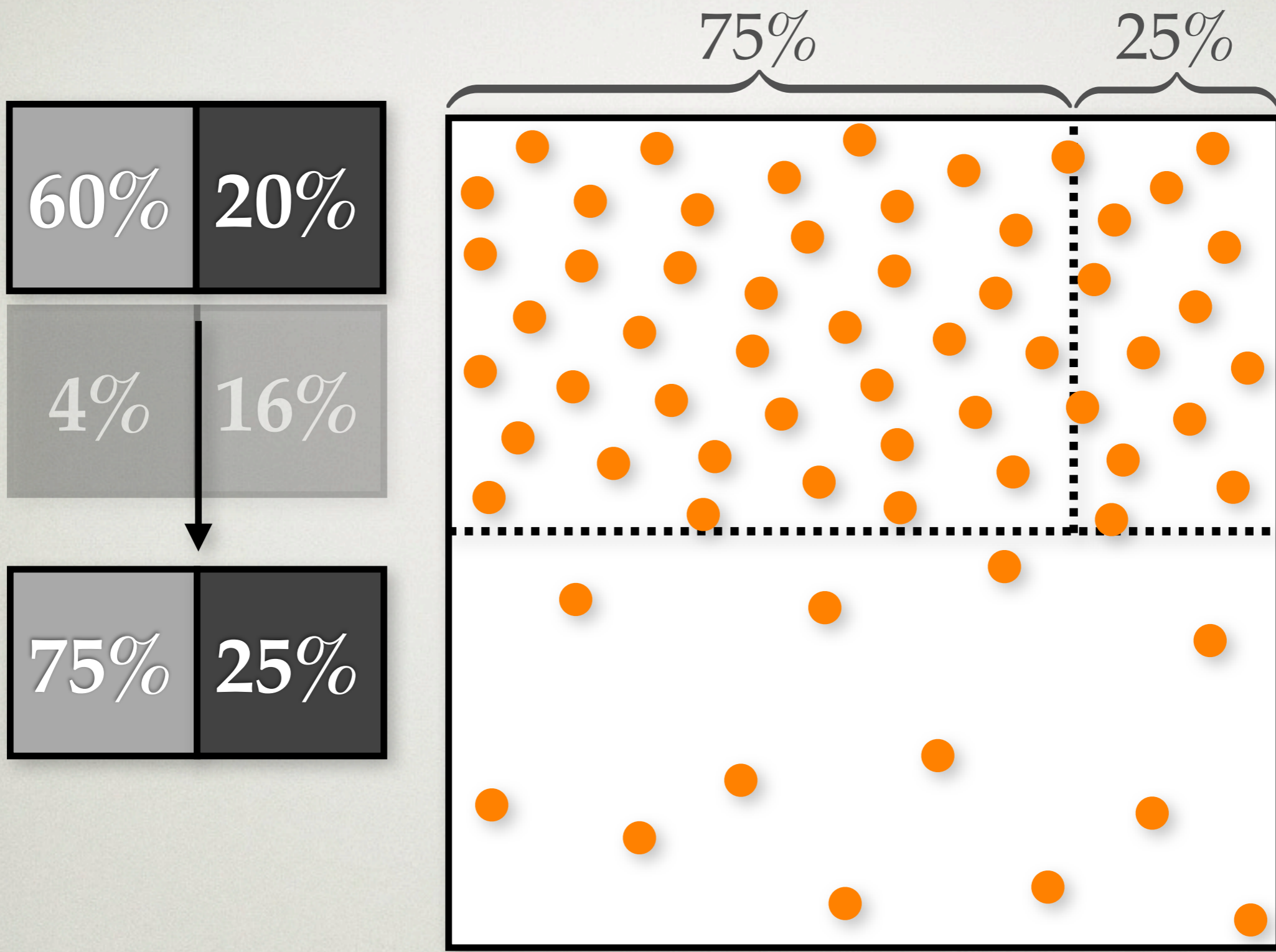* This is accomplished by splitting the point set, and warping the points to fit back into the two rows.

Clarberg et al. 05

Thursday, 6 September 12

* this procedure operates recursively on the levels of a mipmap or Haar wavelet.
* From each level of a mipmap we can determine the total amount of energy contained within each of the 4 child quadrants of the function.
* This process starts at the root.
* First the marginal probabilities for each row are computed (in this case 80% for the top row, and 20% for the bottom row).
* 80% of the samples should be placed in the top half of the importance function.
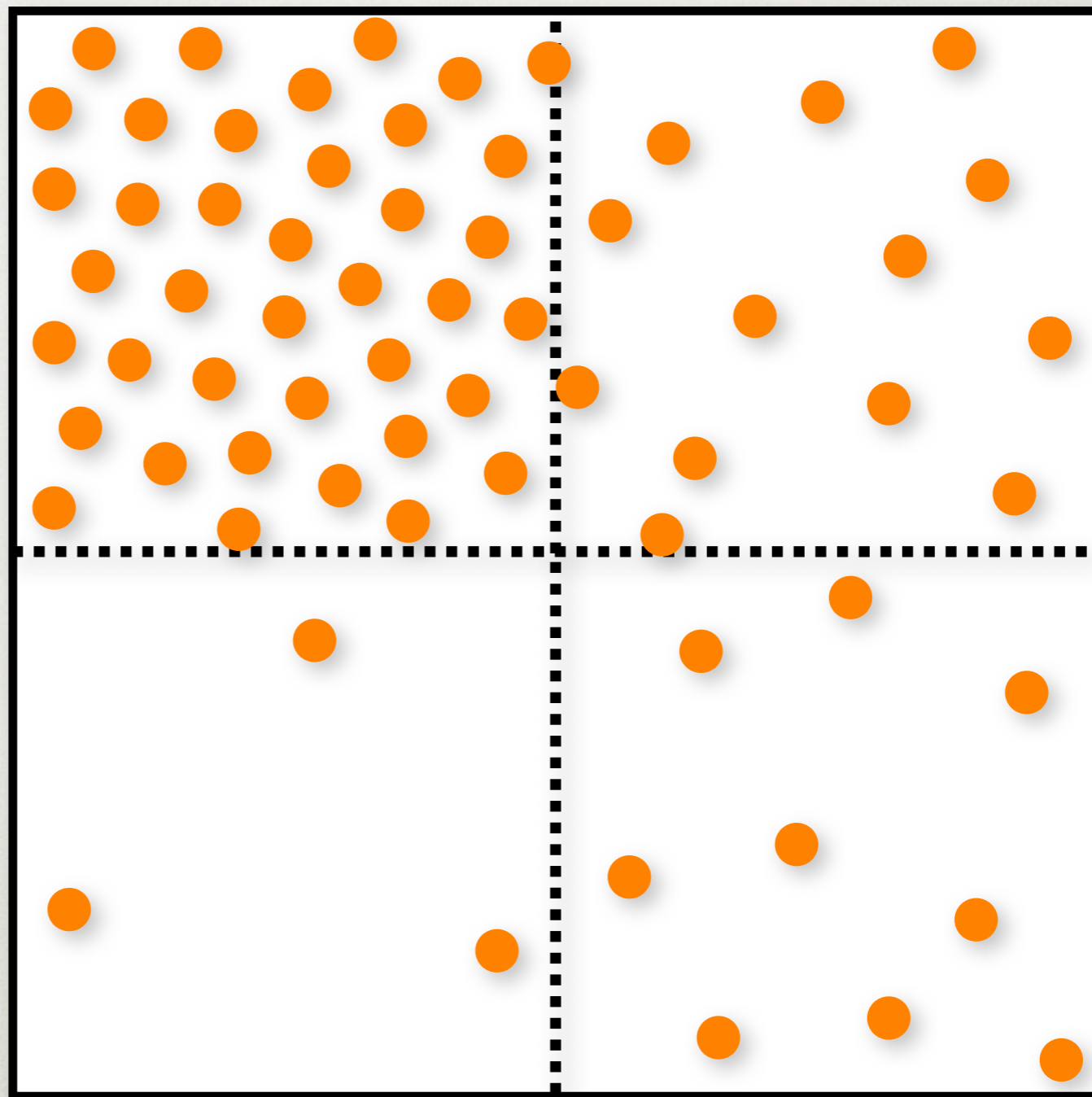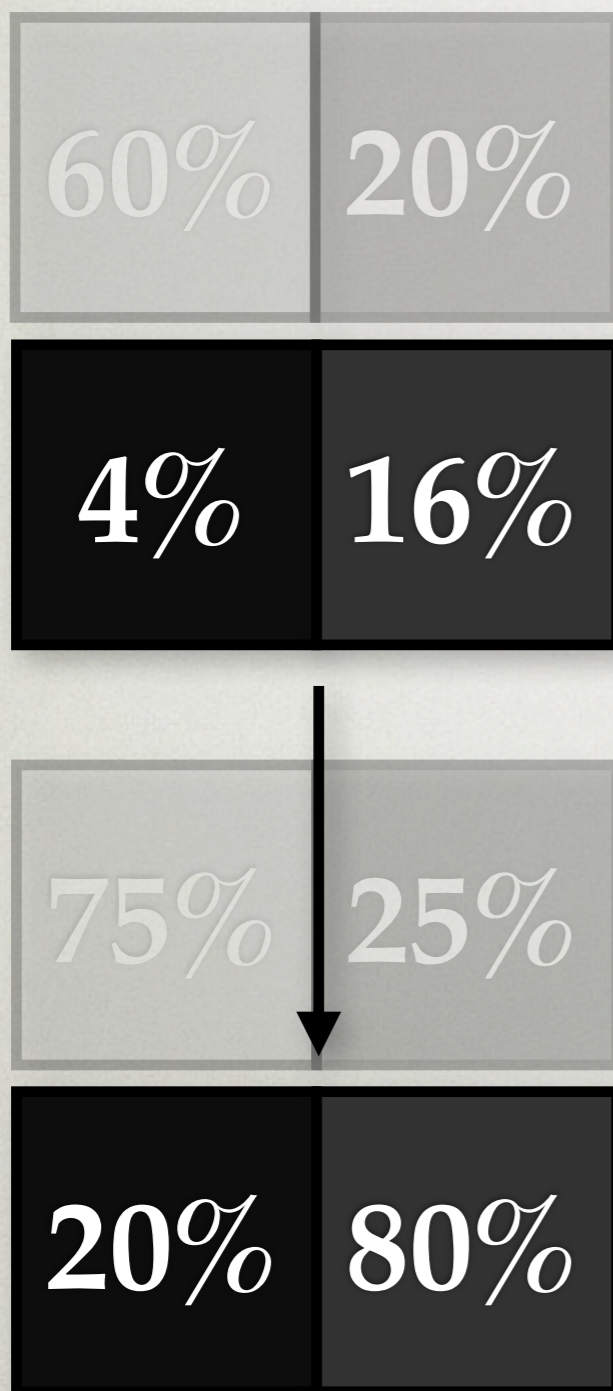* This is accomplished by splitting the point set, and warping the points to fit back into the two rows.

Clarberg et al. 05

* this procedure operates recursively on the levels of a mipmap or Haar wavelet.
* From each level of a mipmap we can determine the total amount of energy contained within each of the 4 child quadrants of the function.
* This process starts at the root.
* First the marginal probabilities for each row are computed (in this case 80% for the top row, and 20% for the bottom row).
* 80% of the samples should be placed in the top half of the importance function.
* This is accomplished by splitting the point set, and warping the points to fit back into the two rows.

Clarberg et al. 05

* The same procedure is executed independently on each row.
* In the top row, 75% of the energy is in the left cell, so the samples are split according to this ratio
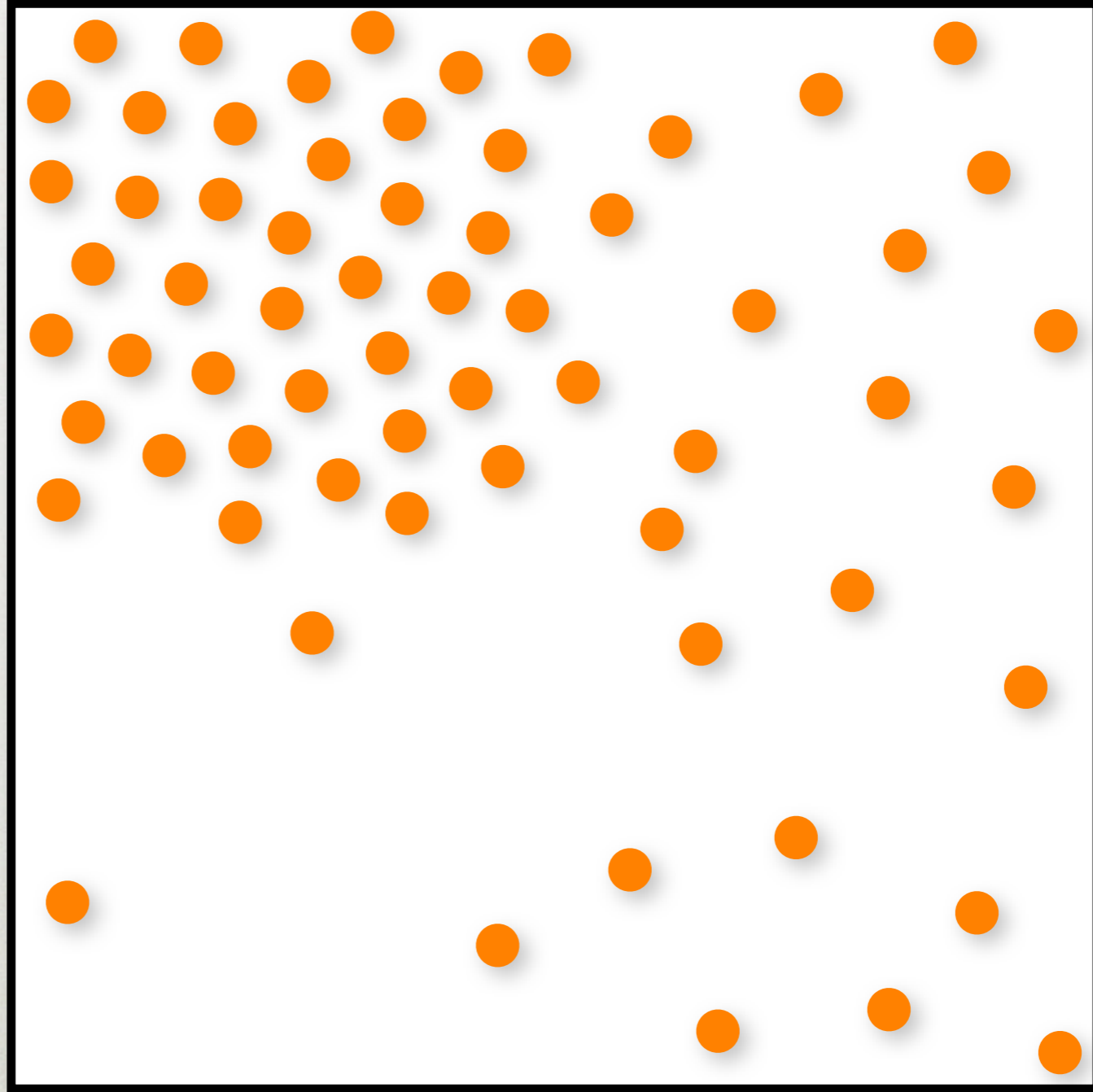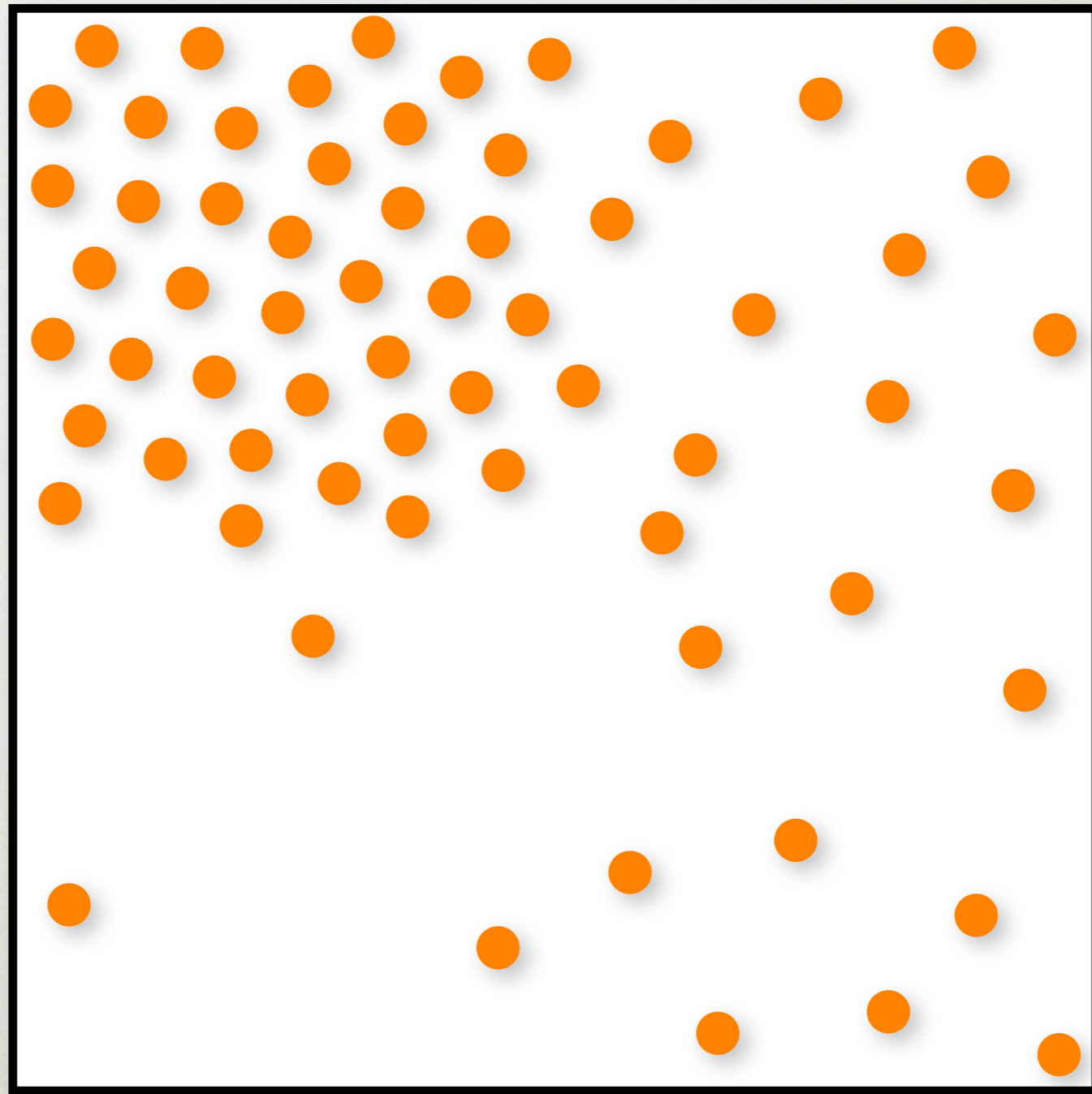* This process is also applied on the bottom row.

Clarberg et al. 05

Thursday, 6 September 12

* The same procedure is executed independently on each row.
* In the top row, 75% of the energy is in the left cell, so the samples are split according to this ratio
* This process is also applied on the bottom row.

Thursday, 6 September 12

* The same procedure is executed independently on each row.
* In the top row, 75% of the energy is in the left cell, so the samples are split according to this ratio
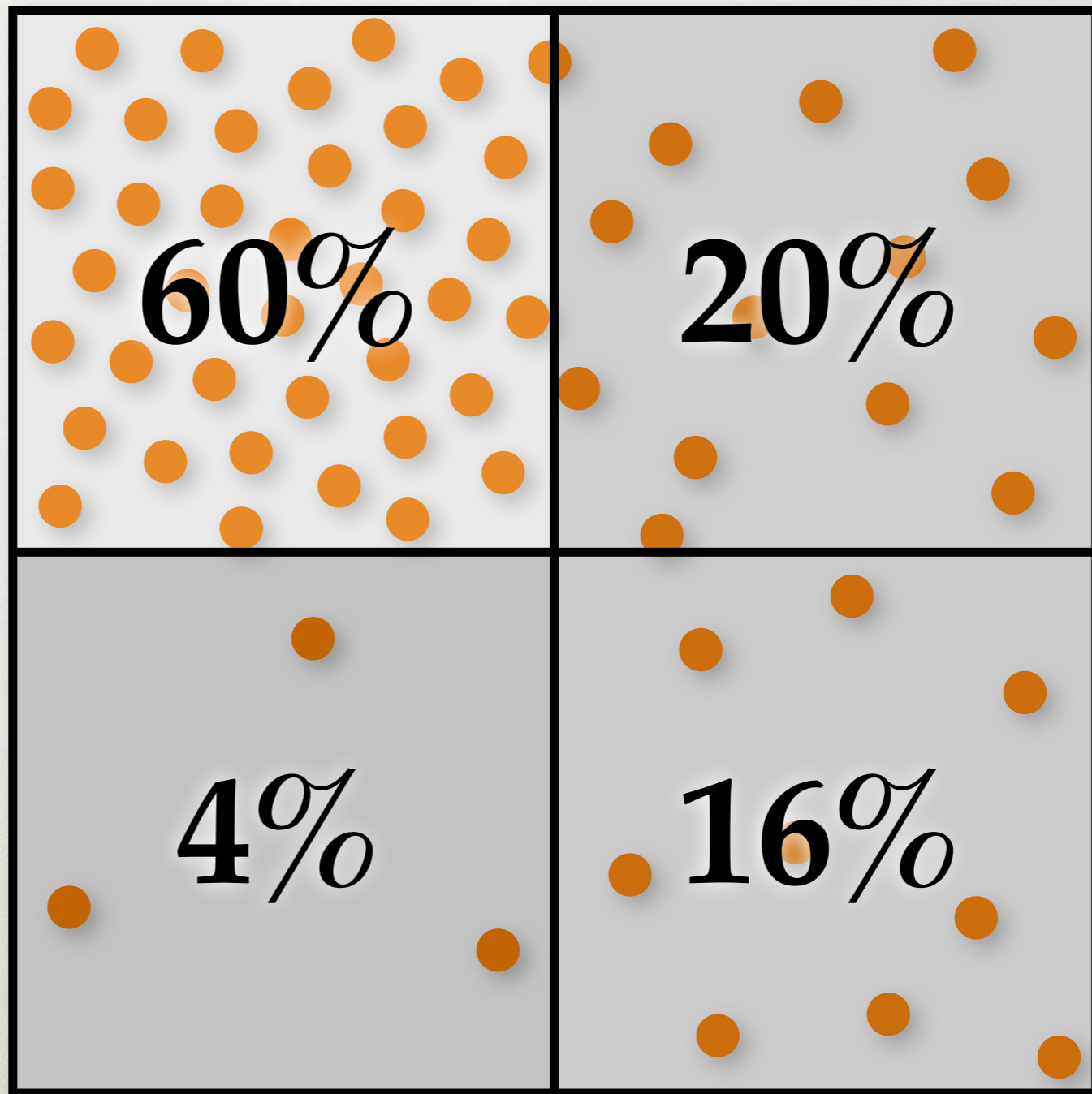* This process is also applied on the bottom row.

Clarberg et al. 05

* This process results in a point set that is distributed according to a 2x2 importance function (click), and within each quadrant the density is uniform. (click)
* The same procedure is applied recursively to the next level of the mipmap within each quadrant. When finished, the density of samples is proportional to the intensity of the hierarchical function.
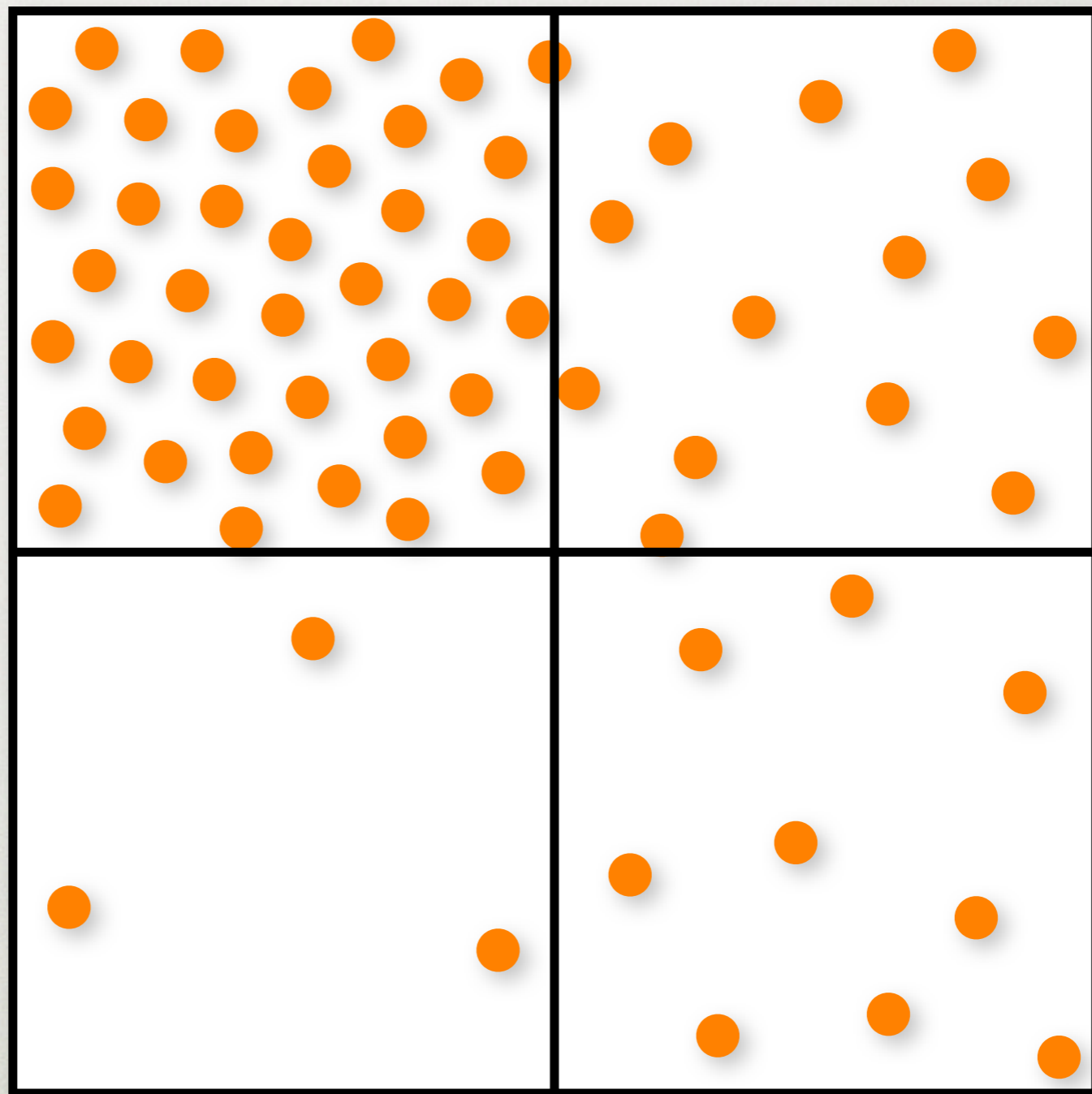*

Clarberg et al. 05

* This process results in a point set that is distributed according to a 2x2 importance function (click), and within each quadrant the density is uniform. (click)
* The same procedure is applied recursively to the next level of the mipmap within each quadrant. When finished, the density of samples is proportional to the intensity of the hierarchical function.
*

Clarberg et al. 05

* This process results in a point set that is distributed according to a 2x2 importance function (click), and within each quadrant the density is uniform. (click)
* The same procedure is applied recursively to the next level of the mipmap within each quadrant. When finished, the density of samples is proportional to the intensity of the hierarchical function.
*

repeat on each quadrant recursively

Clarberg et al. 05

Thursday, 6 September 12

* This process results in a point set that is distributed according to a 2x2 importance function (click), and within each quadrant the density is uniform. (click)
* The same procedure is applied recursively to the next level of the mipmap within each quadrant. When finished, the density of samples is proportional to the intensity of the hierarchical function.
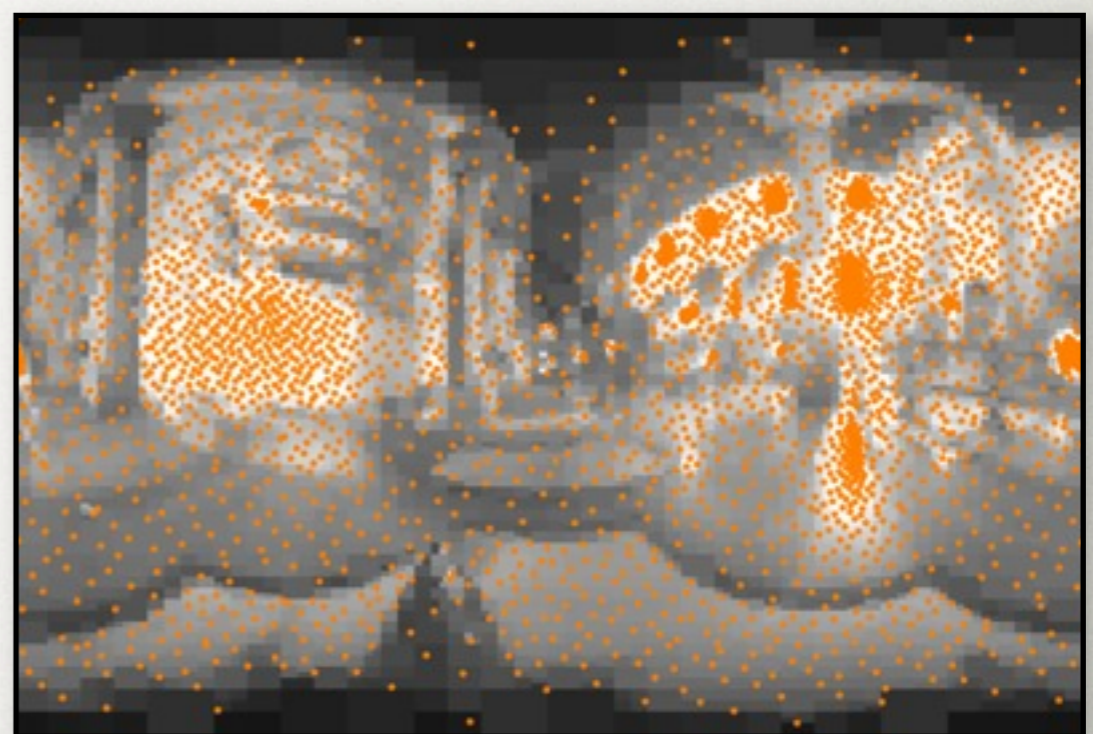*

Wavelet Importance Sampling
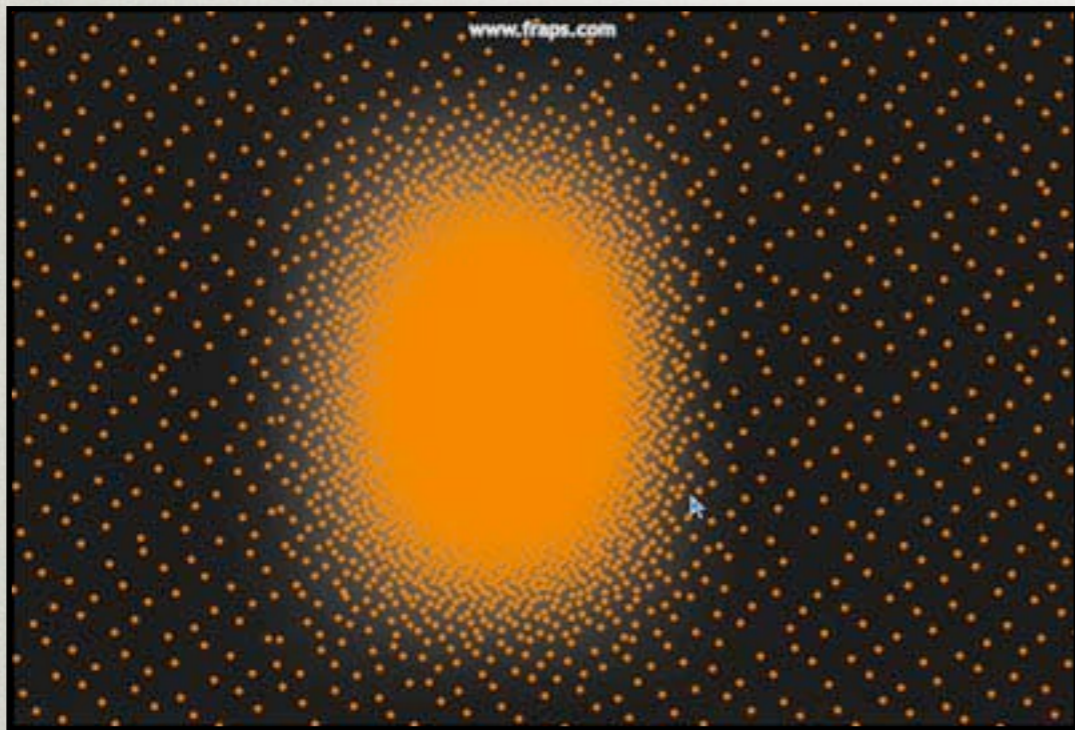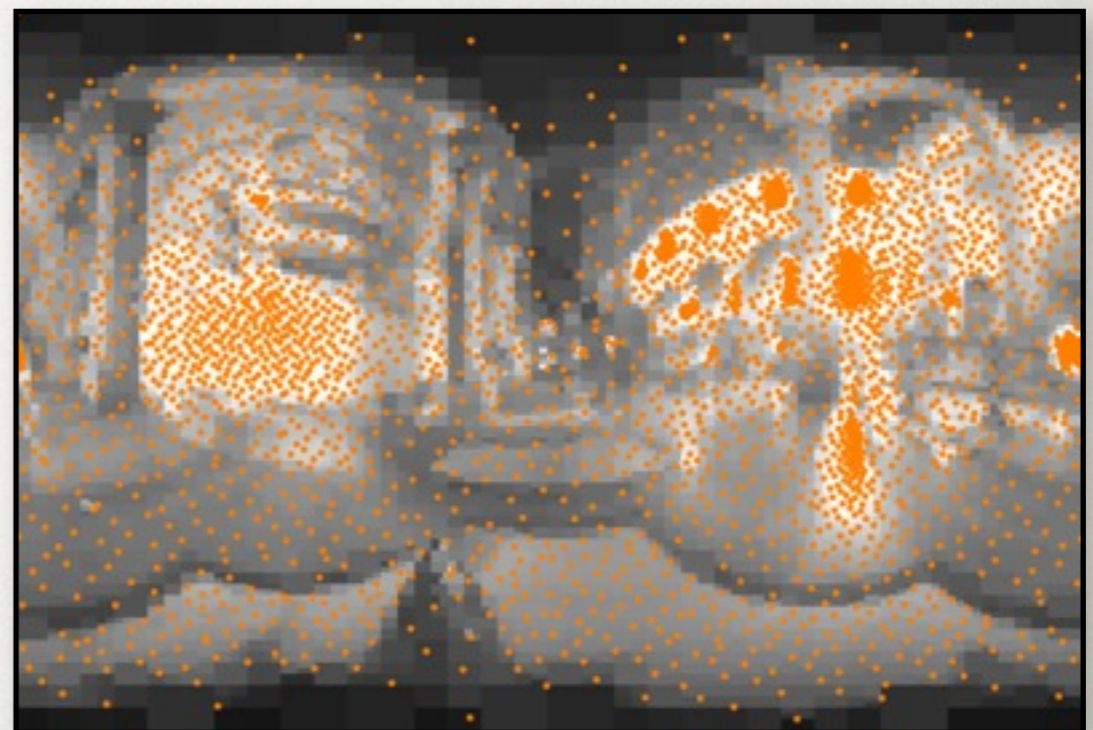
BRDF

Lighting

* With WIS we can sample the BRDF or the lighting if we store these functions as wavelets
* However, one of the big benefits of WIS is that we can compute the product of these two functions on-the-fly, obtain the result as a wavelet, and perform HSW on the product directly
* Pros: fast, samples product, high quality
* Cons: no rotation

Wavelet Importance Sampling

BRDF

×

Lighting

=

BRDF x Lighting

Clarberg et al. 05

* With WIS we can sample the BRDF or the lighting if we store these functions as wavelets
* However, one of the big benefits of WIS is that we can compute the product of these two functions on-the-fly, obtain the result as a wavelet, and perform HSW on the product directly
* Pros: fast, samples product, high quality
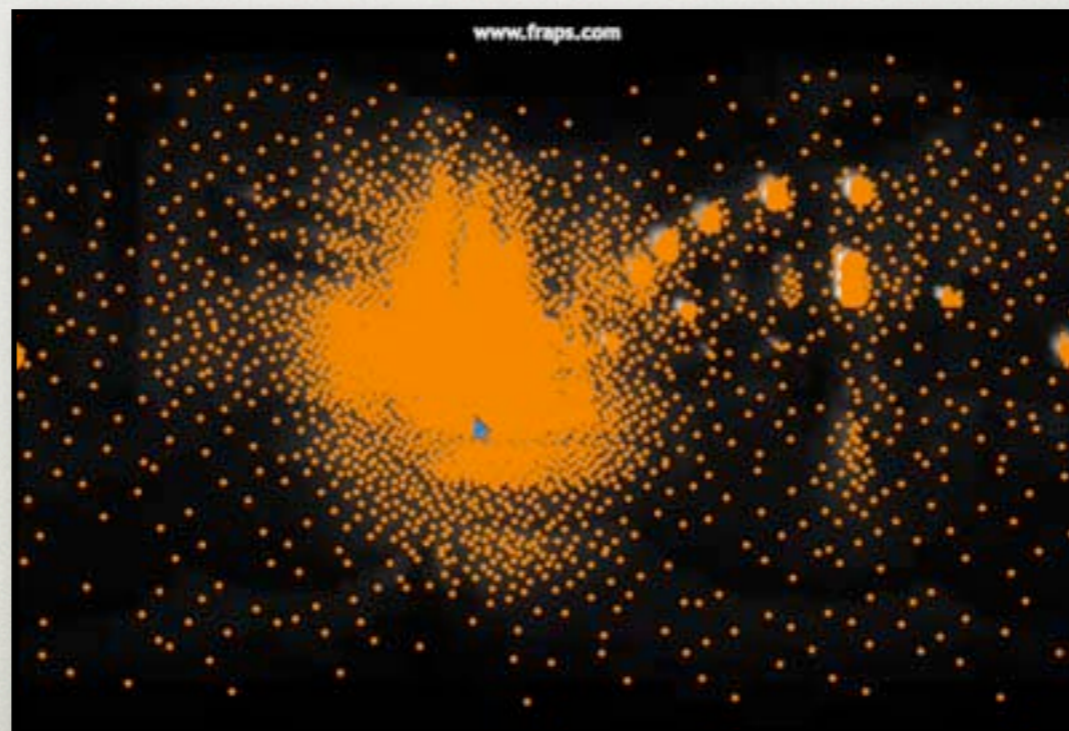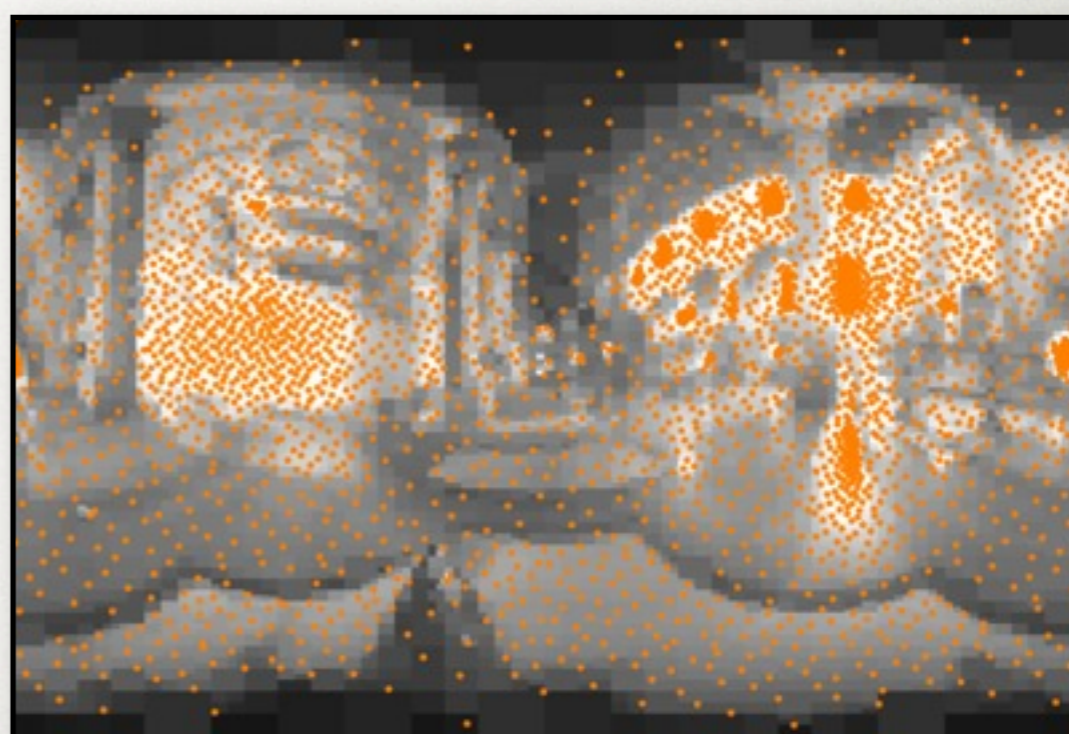* Cons: no rotation

Wavelet Importance Sampling

BRDF

×

Lighting

=

-no rotation

BRDF x Lighting

Clarberg et al. 05

* With WIS we can sample the BRDF or the lighting if we store these functions as wavelets
* However, one of the big benefits of WIS is that we can compute the product of these two functions on-the-fly, obtain the result as a wavelet, and perform HSW on the product directly
* Pros: fast, samples product, high quality
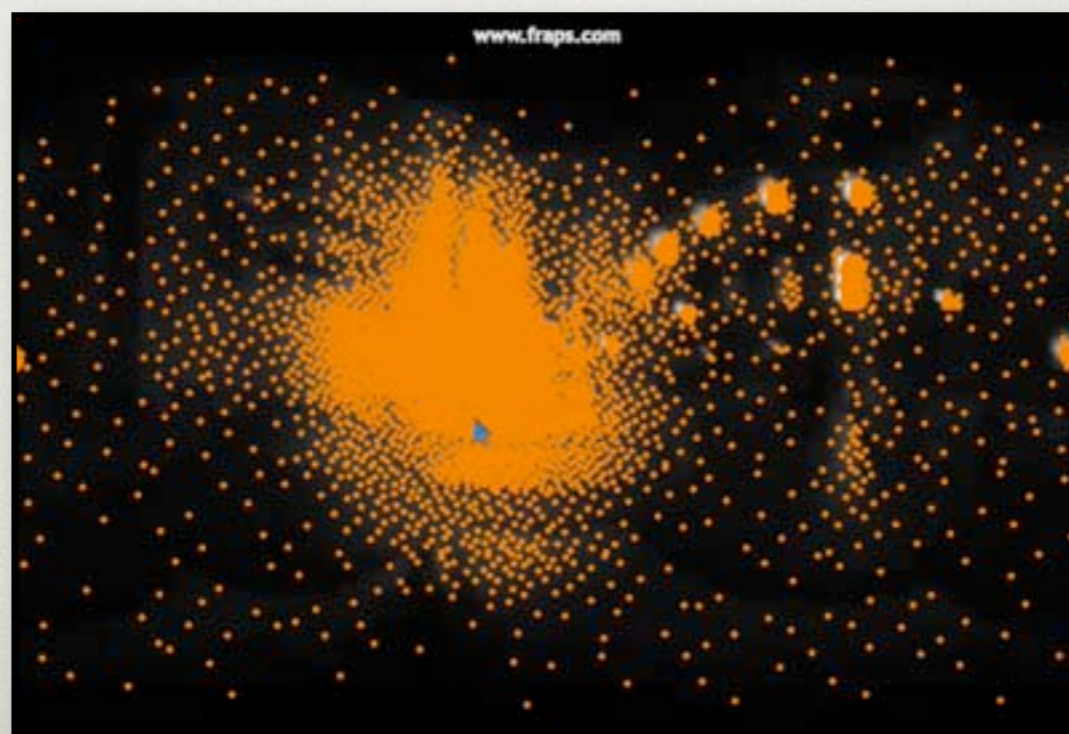* Cons: no rotation

Wavelet Importance Sampling

BRDF

x

Lighting

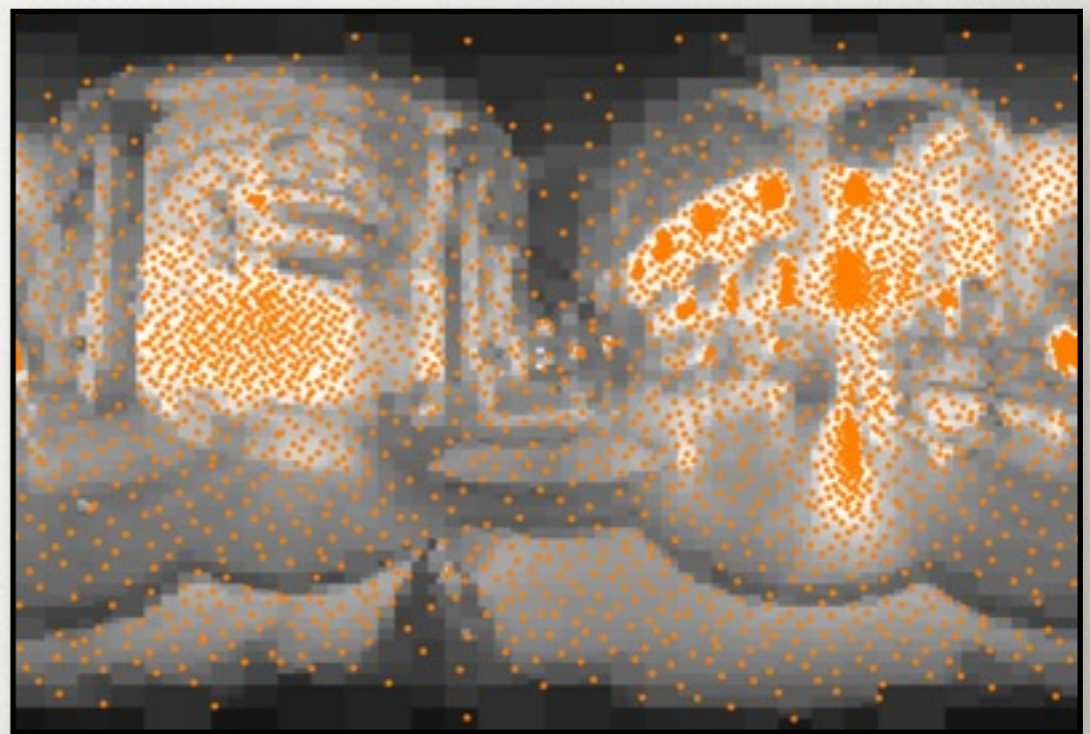-no rotation

-memory usage

BRDF x Lighting

Clarberg et al. 05

* With WIS we can sample the BRDF or the lighting if we store these functions as wavelets
* However, one of the big benefits of WIS is that we can compute the product of these two functions on-the-fly, obtain the result as a wavelet, and perform HSW on the product directly
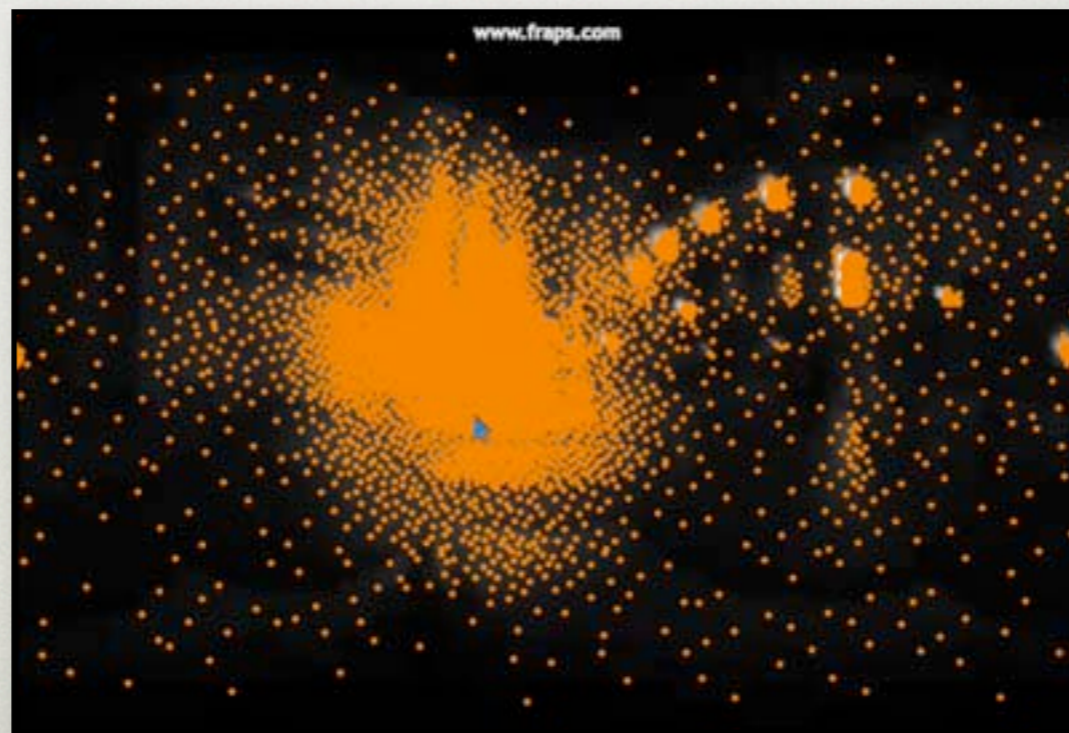* Pros: fast, samples product, high quality
* Cons: no rotation

Motivation

- HSW works well, but Haar wavelets impose limitations (no rotation)

* The motivation for our work is that although hierarchical sample warping works well, the overall technique has some limitations due to the Haar basis not supporting rotation.
* Spherical harmonics support many of the same operations as wavelets and they additionally support rotation.
* however, to date, there have been no efficient ways to importance sample SH.

Motivation

- HSW works well, but Haar wavelets impose limitations (no rotation)

- Spherical harmonics:

  - ✓ Encode low-frequency functions
  - ✓ Product and product integrals
  - ✓ Convolution
  - ✓ Rotation
  - – Importance sampling

* The motivation for our work is that although hierarchical sample warping works well, the overall technique has some limitations due to the Haar basis not supporting rotation.
* Spherical harmonics support many of the same operations as wavelets and they additionally support rotation.
* however, to date, there have been no efficient ways to importance sample SH.

Approach

- Develop a hierarchical sample warping technique for spherical harmonics

* the goal of this paper is to fill this gap in the literature.
* since we saw that the hierarchical sample warping technique works well, our goal is to maintain all of its beneficial properties, but replace wavelets with SH

Spherical Harmonics

$$f(\vec{\omega}) \approx \tilde{f}(\vec{\omega}) = \sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m \, y_l^m(\vec{\omega})$$

Thursday, 6 September 12

* So before diving into the details of our algorithm, lets quickly review spherical harmonics
* Spherical harmonics are a frequency-space basis for representing functions defined over the sphere.
* A low frequency spherical function f can be approximated efficiently using only a few function coefficients f_l^m
* A reconstruction of the original function is a weighted sum of these coefficients with the spherical harmonic basis functions.

Spherical Harmonics

original function

$$\boxed{f(\vec{\omega})} \approx \tilde{f}(\vec{\omega}) = \sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m \, y_l^m(\vec{\omega})$$

* So before diving into the details of our algorithm, lets quickly review spherical harmonics
* Spherical harmonics are a frequency-space basis for representing functions defined over the sphere.
* A low frequency spherical function f can be approximated efficiently using only a few function coefficients f_l^m
* A reconstruction of the original function is a weighted sum of these coefficients with the spherical harmonic basis functions.

Spherical Harmonics

original function

$$\boxed{f(\vec{\omega})} \approx \tilde{f}(\vec{\omega}) = \sum_{l=0}^{n-1} \sum_{m=-l}^{l} \boxed{f_l^m}\, y_l^m(\vec{\omega})$$

function coefficients

Thursday, 6 September 12

* So before diving into the details of our algorithm, lets quickly review spherical harmonics
* Spherical harmonics are a frequency-space basis for representing functions defined over the sphere.
* A low frequency spherical function f can be approximated efficiently using only a few function coefficients f_l^m
* A reconstruction of the original function is a weighted sum of these coefficients with the spherical harmonic basis functions.
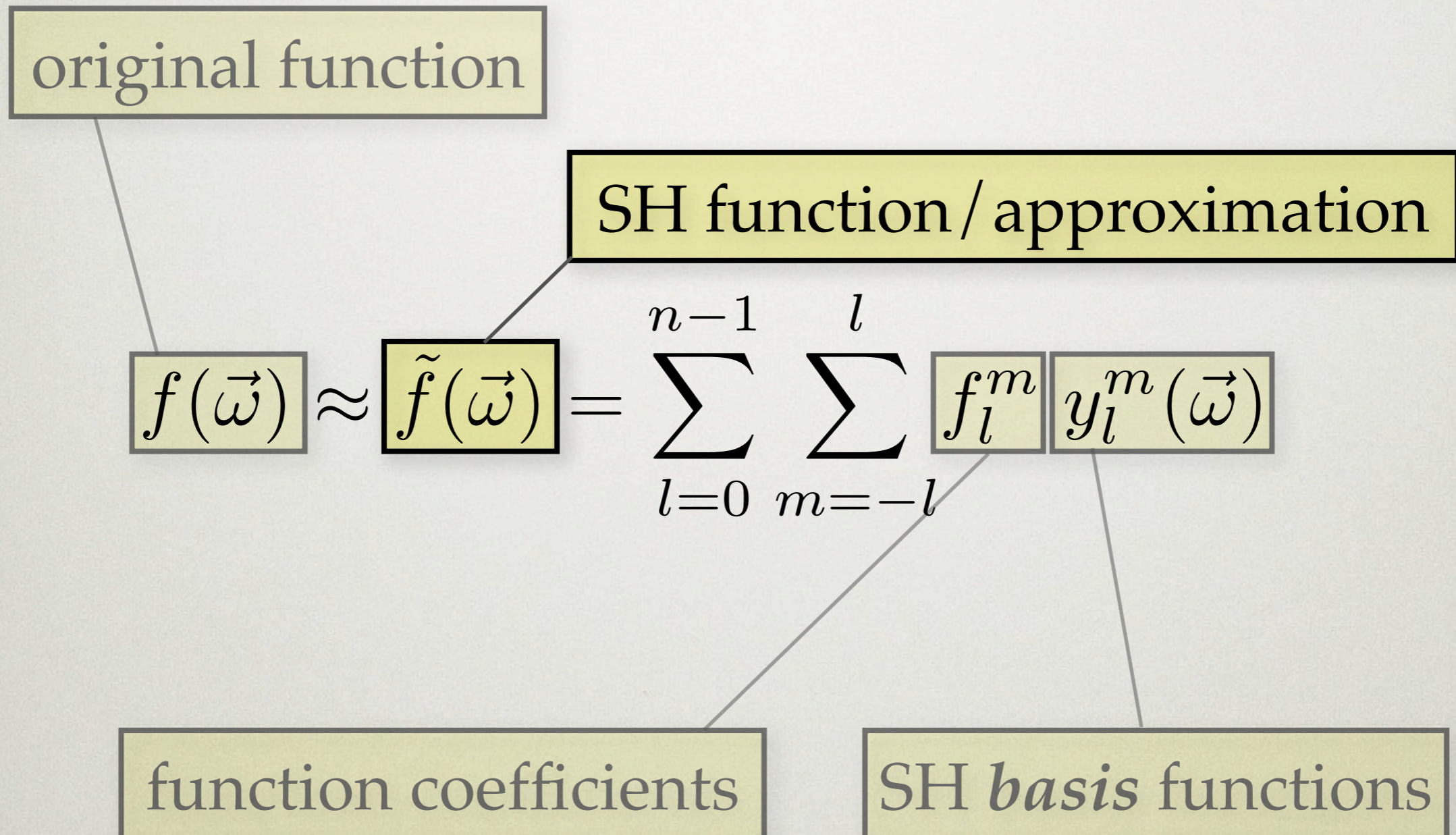
Spherical Harmonics

original function

$$f(\vec{\omega}) \approx \tilde{f}(\vec{\omega}) = \sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m \, y_l^m(\vec{\omega})$$

function coefficients

SH *basis* functions

Thursday, 6 September 12
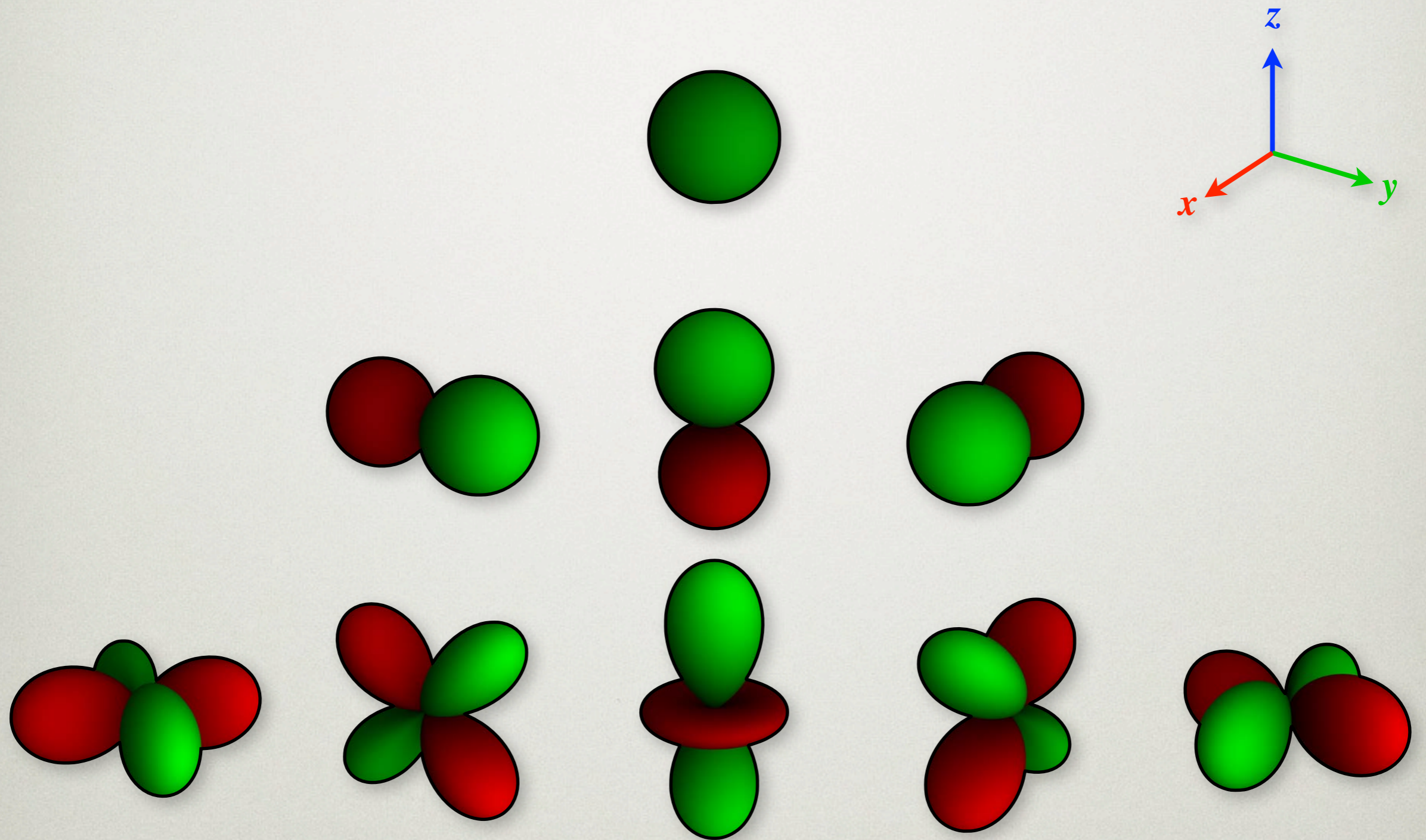
* So before diving into the details of our algorithm, lets quickly review spherical harmonics
* Spherical harmonics are a frequency-space basis for representing functions defined over the sphere.
* A low frequency spherical function f can be approximated efficiently using only a few function coefficients f_l^m
* A reconstruction of the original function is a weighted sum of these coefficients with the spherical harmonic basis functions.

Spherical Harmonics

original function

SH function/approximation

$$f(\vec{\omega}) \approx \tilde{f}(\vec{\omega}) = \sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m \, y_l^m(\vec{\omega})$$

function coefficients

SH *basis* functions

* So before diving into the details of our algorithm, lets quickly review spherical harmonics
* Spherical harmonics are a frequency-space basis for representing functions defined over the sphere.
* A low frequency spherical function f can be approximated efficiently using only a few function coefficients f_l^m
* A reconstruction of the original function is a weighted sum of these coefficients with the spherical harmonic basis functions.
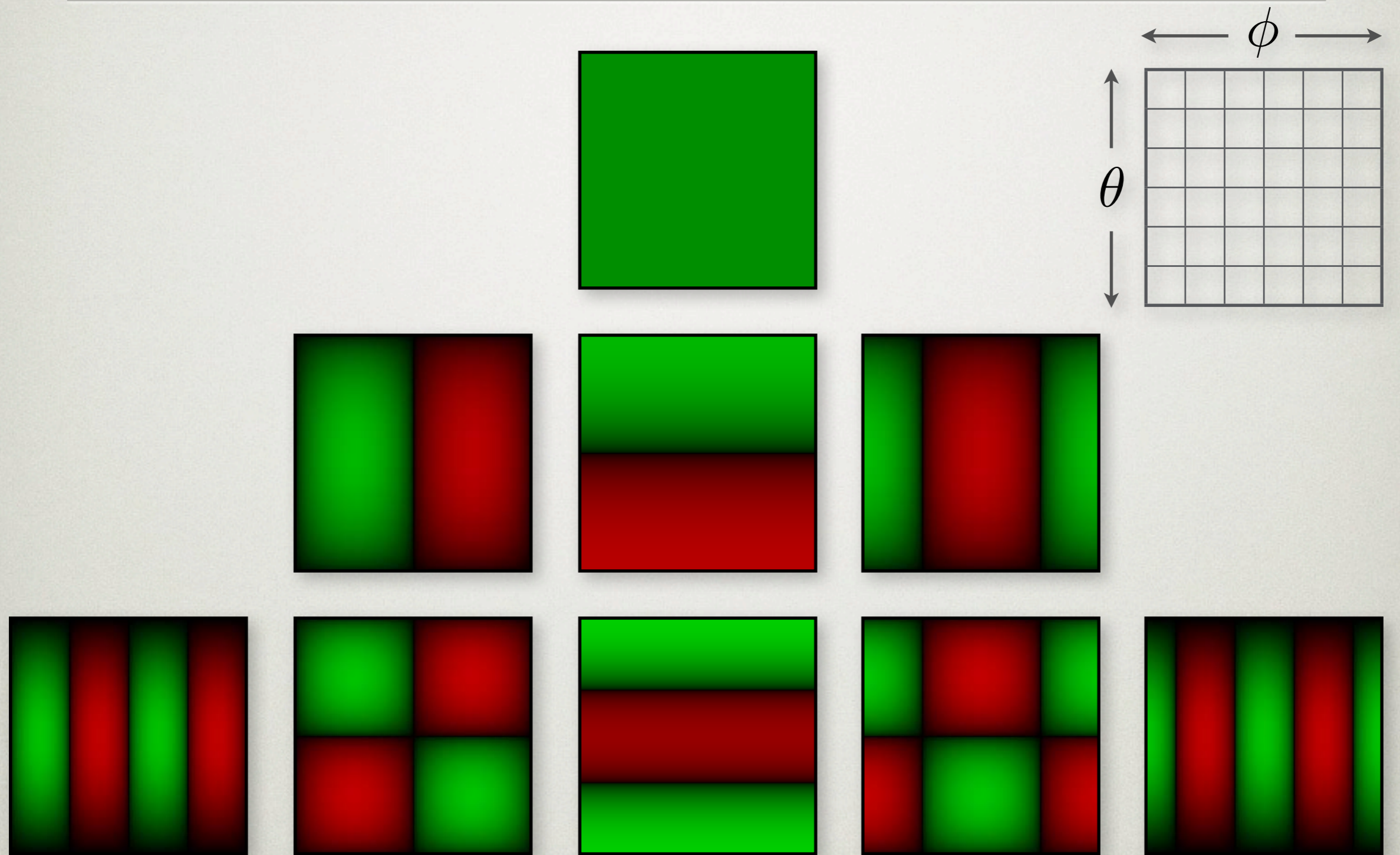
Thursday, 6 September 12

* The SH basis functions are just functions defined over the sphere.
* They can be visualized by plotting the magnitude as a radius. Green regions indicate positive values and red indicates negative.
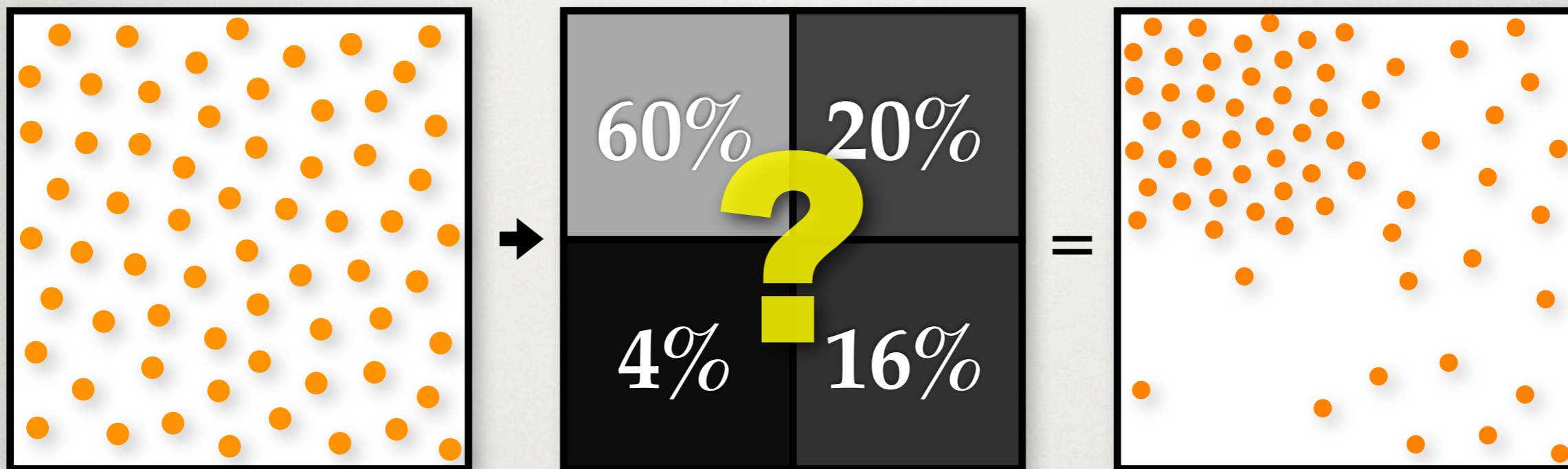* There is an infinite set of these basis functions, here we shown just the first 9

* In spherical coordinates the basis functions can be visualized as 2D images representing the value of the function for each spherical direction.
* Green regions indicate positive values and red indicates negative.

HSW WITH SH

- To use HSW, need quadrant probabilities
  - Haar or mipmap provides these directly
  - how to obtain these from SH function?

* To use HSW, the only information that is required is the quadrant probabilities.
* Haar wavelets and mipmaps provide this information directly
* The question is how to obtain these values from a SH function

Quadrant Probabilities

- Quadrant probability is just an integral of the SH function/approximation:

* Given a mapping from the spherical domain onto the unit square, the quadrant probabilities are just an integral of the SH approximation over the region under the quadrant.

Quadrant Probabilities

- Quadrant probability is just an integral of the SH function/approximation:

$$\int \tilde{f}(\vec{\omega}) \, d\vec{\omega}$$

* Given a mapping from the spherical domain onto the unit square, the quadrant probabilities are just an integral of the SH approximation over the region under the quadrant.

Quadrant Probabilities

- Quadrant probability is just an integral of the SH function/approximation:

$$\int \tilde{f}(\vec{\omega})\, d\vec{\omega}$$

- compute basis integrals once and reuse?

* We can make an important observation by inserting the definition of the SH approximation f tilde into the equation.
* Since f tilde is a weighted sum of basis functions, we see that the integral can be brought inside the summation.
* Effectively what this is saying is that we can compute the integral of f tilde by performing a weighted sum with integrated basis functions. (using the same coefficients)

Quadrant Probabilities

- Quadrant probability is just an integral of the SH function/approximation:

$$\int \tilde{f}(\vec{\omega})\, d\vec{\omega}$$

- compute basis integrals once and reuse?

* We can make an important observation by inserting the definition of the SH approximation f tilde into the equation.
* Since f tilde is a weighted sum of basis functions, we see that the integral can be brought inside the summation.
* Effectively what this is saying is that we can compute the integral of f tilde by performing a weighted sum with integrated basis functions. (using the same coefficients)

Quadrant Probabilities

- Quadrant probability is just an integral of the SH function/approximation:

$$\int \tilde{f}(\vec{\omega})\, d\vec{\omega} = \int \left(\sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m\, y_l^m(\vec{\omega}) \right) d\vec{\omega}$$

- compute basis integrals once and reuse?

* We can make an important observation by inserting the definition of the SH approximation f tilde into the equation.
* Since f tilde is a weighted sum of basis functions, we see that the integral can be brought inside the summation.
* Effectively what this is saying is that we can compute the integral of f tilde by performing a weighted sum with integrated basis functions. (using the same coefficients)

QUADRANT PROBABILITIES

- Quadrant probability is just an integral of the SH function/approximation:

$$\int \tilde{f}(\vec{\omega})\, d\vec{\omega} = \int \left(\sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m \, y_l^m(\vec{\omega}) \right) d\vec{\omega}$$

$$= \sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m \int y_l^m(\vec{\omega})\, d\vec{\omega}$$

- compute basis integrals once and reuse?

* We can make an important observation by inserting the definition of the SH approximation f tilde into the equation.
* Since f tilde is a weighted sum of basis functions, we see that the integral can be brought inside the summation.
* Effectively what this is saying is that we can compute the integral of f tilde by performing a weighted sum with integrated basis functions. (using the same coefficients)

Quadrant Probabilities

- Quadrant probability is just an integral of the SH function/approximation:

$$\int \tilde{f}(\vec{\omega}) \, d\vec{\omega} = \int \left(\sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m \, y_l^m(\vec{\omega}) \right) d\vec{\omega}$$

$$= \sum_{l=0}^{n-1} \sum_{m=-l}^{l} f_l^m \boxed{\int y_l^m(\vec{\omega}) \, d\vec{\omega}}$$

- compute basis integrals once and reuse?

* We can make an important observation by inserting the definition of the SH approximation f tilde into the equation.
* Since f tilde is a weighted sum of basis functions, we see that the integral can be brought inside the summation.
* Effectively what this is saying is that we can compute the integral of f tilde by performing a weighted sum with integrated basis functions. (using the same coefficients)

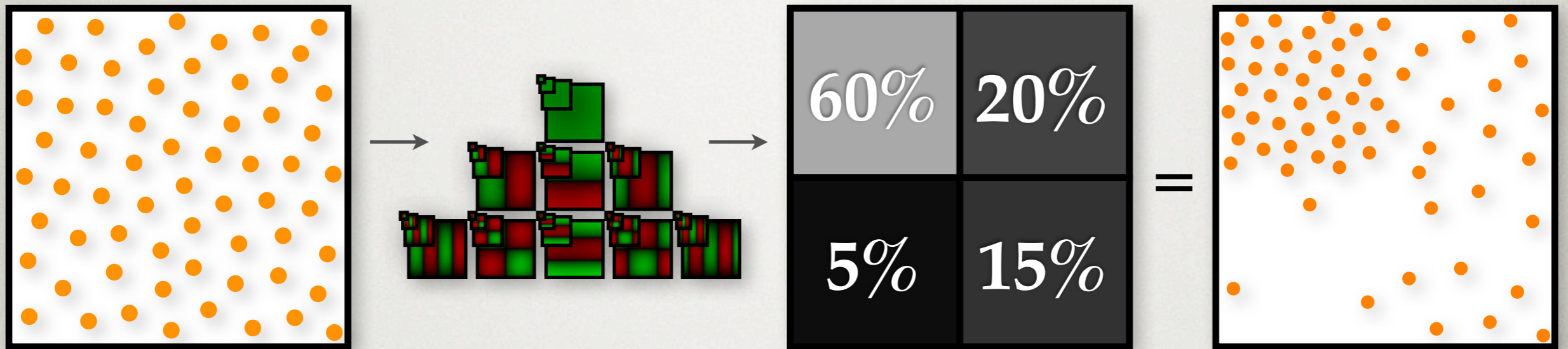# Naive Approach (Basis MipMaps)
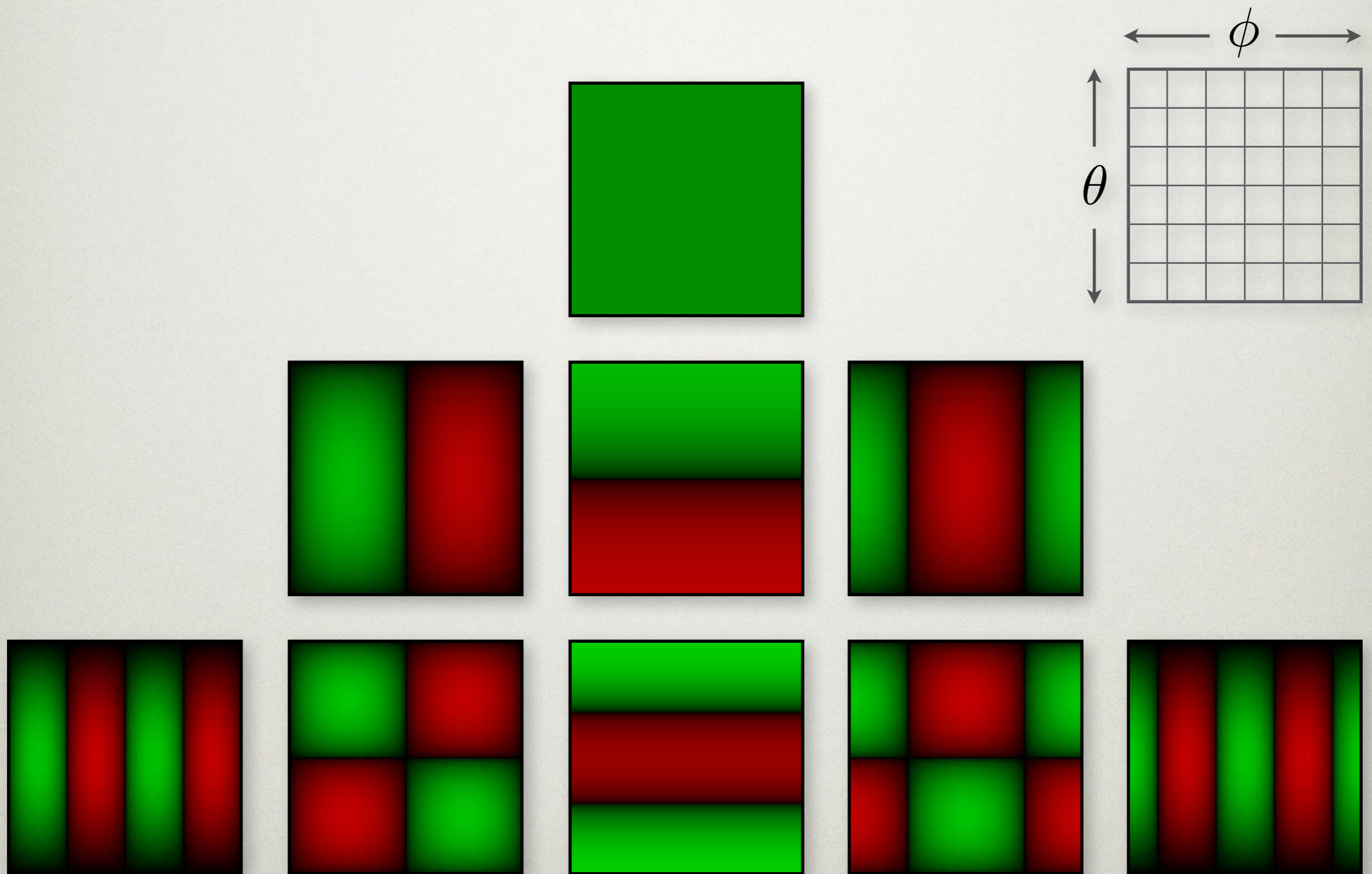
* A naive way to exploit this observation could be to somehow precompute the integrals of the basis functions and perform a weighted sum of the integrals to get the integral of f tilde in any region
* This could be done by precomputing/tabulating the basis functions at a high resolution. (click)
* and then creating lower resolution versions by constructing a mipmap for each basis
* Now, any spherical function can be approximated as a weighted sum of these basis mipmaps,
* So we can apply hierarchical sample warping to any spherical harmonic function

Naive Approach (Basis MipMaps)
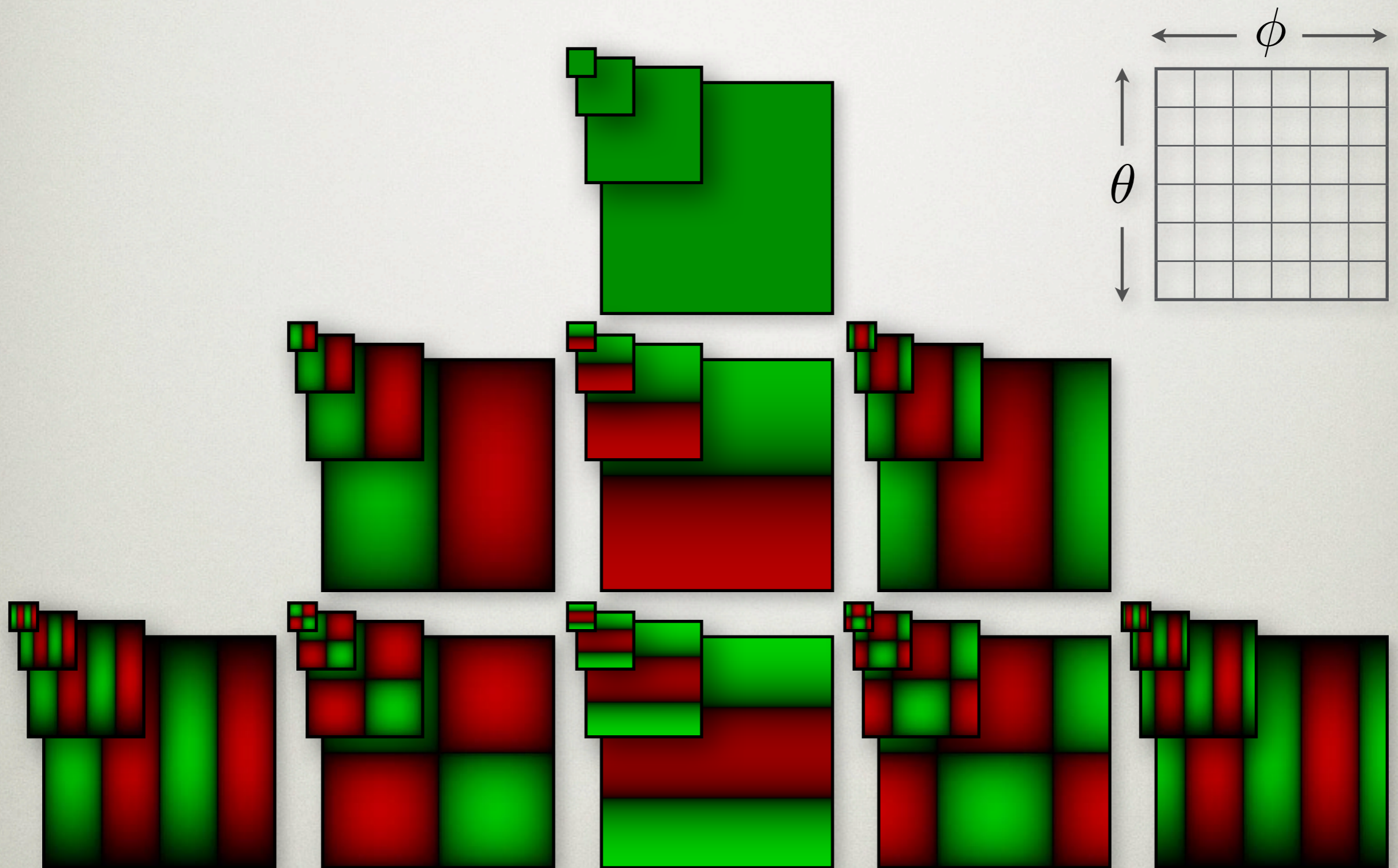
* A naive way to exploit this observation could be to somehow precompute the integrals of the basis functions and perform a weighted sum of the integrals to get the integral of f tilde in any region
* This could be done by precomputing/tabulating the basis functions at a high resolution. (click)
* and then creating lower resolution versions by constructing a mipmap for each basis
* Now, any spherical function can be approximated as a weighted sum of these basis mipmaps,
* So we can apply hierarchical sample warping to any spherical harmonic function

Naive
SH Importance Sampling

* with this naive approach, the HSW algorithm works by using the basis mipmaps to derive the quadrant probabilities.
* however, we can improve upon this simple idea to get a much more efficient algorithm

IMPROVEMENTS

- Separability

- Analytic integrals

* we can make two significant improvements, the first is by exploiting separability

Separability

- In spherical coordinates, SH bases are separable:

$$y_l^m(\theta, \phi) = K_l^m P_l^{|m|}(\cos\theta)\, \Phi^m(\phi)$$

* The naive approach will work for any mapping from the sphere to the unit square.
* However, it turns out that if we choose either spherical coordinates or cylindrical coordinates, we can gain a significant benefit.
* This is because with either of these mappings, the basis functions are separable.
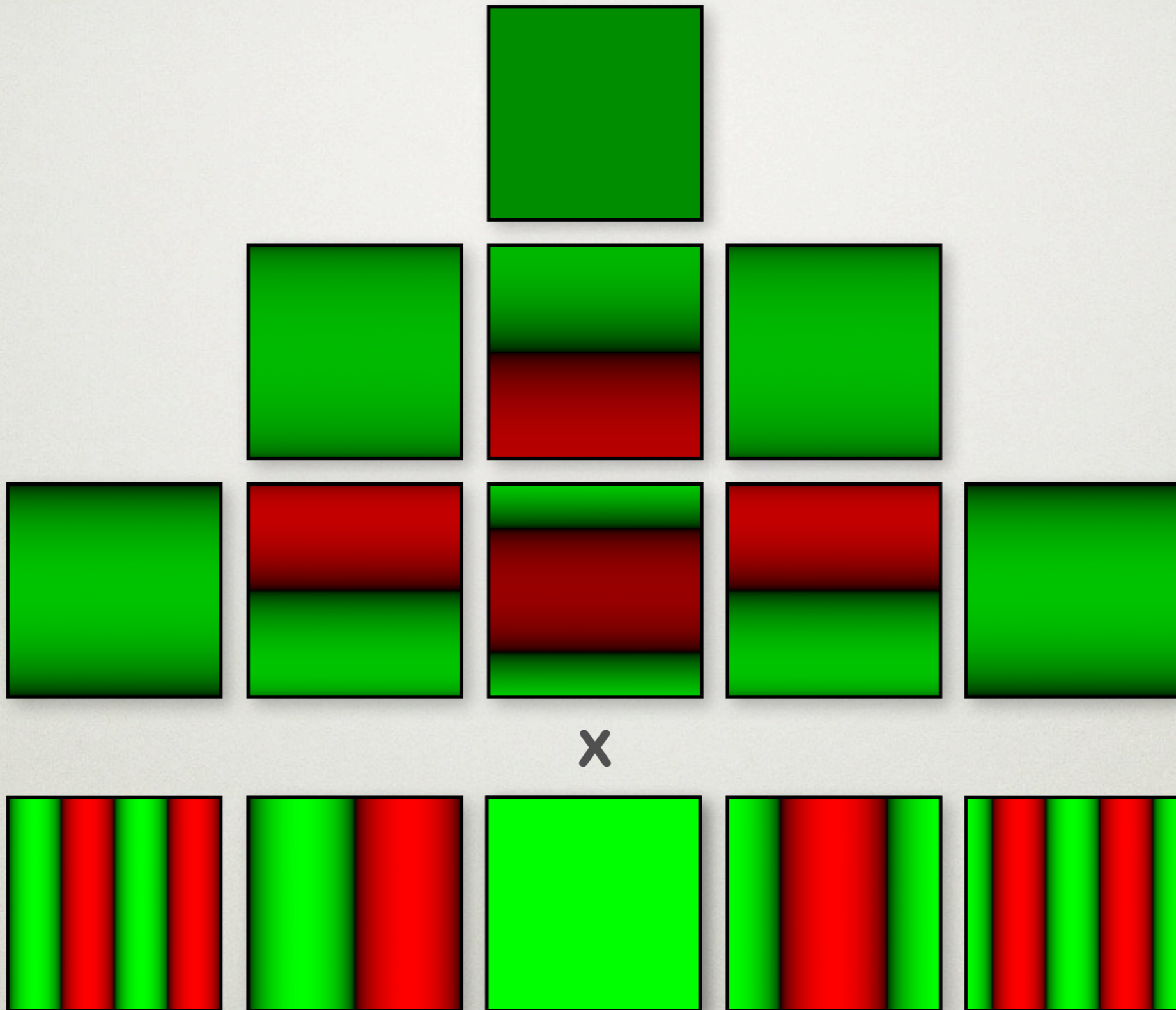
$$K_l^m P_l^{|m|}(\cos\theta)\, \Phi^m(\phi)$$

* Separability allows us to decompose the basis functions into the product of a 1D theta dependent component and a 1D phi dependent component
* in this illustration I am still visualizing the components as 2D images for clarity, but these are really just 1D functions and only have variation in either theta or phi, but not both
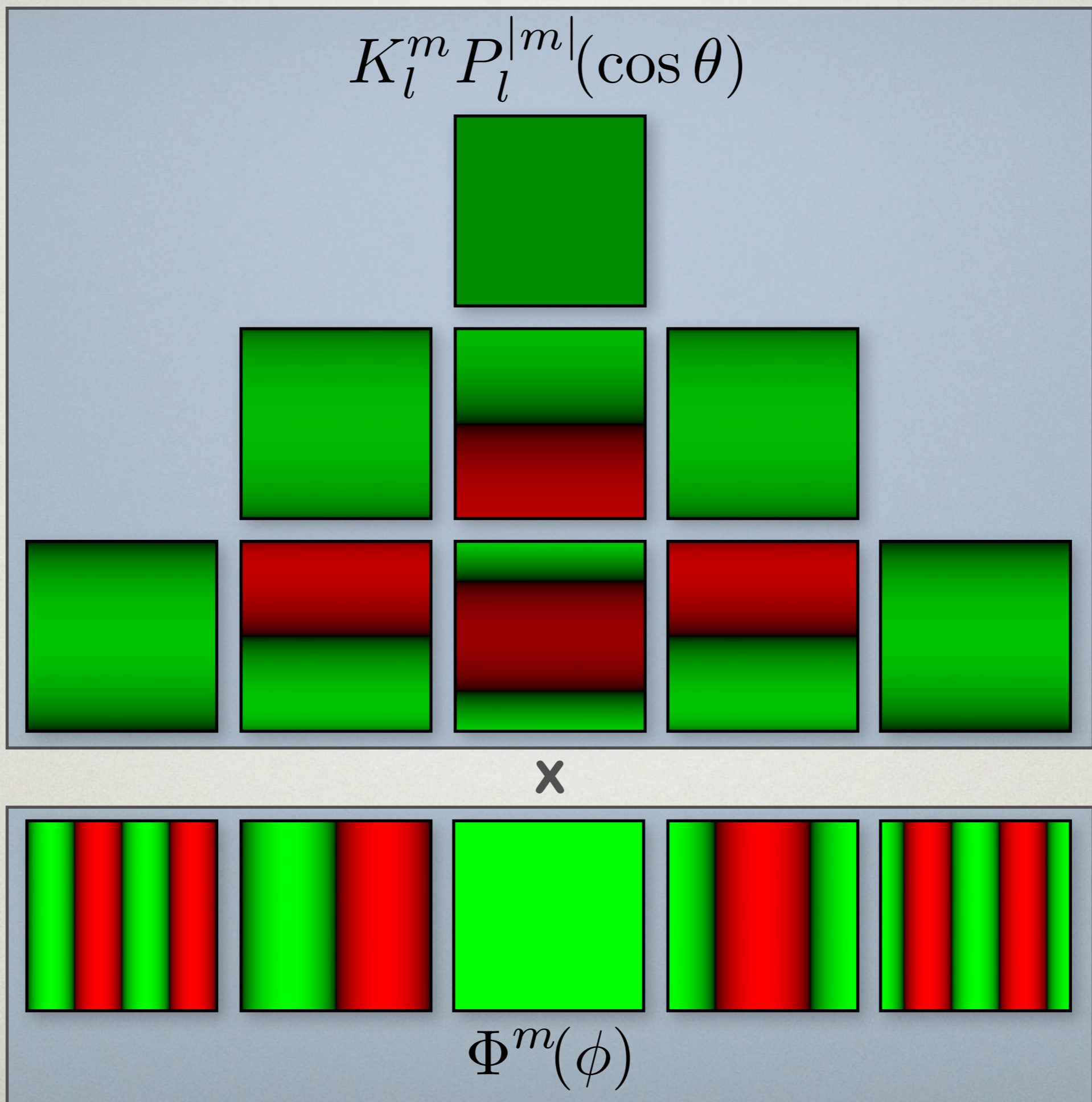
* Separability allows us to decompose the basis functions into the product of a 1D theta dependent component and a 1D phi dependent component
* in this illustration I am still visualizing the components as 2D images for clarity, but these are really just 1D functions and only have variation in either theta or phi, but not both

$$K_l^m P_l^{|m|}(\cos\theta)$$

* The product of these 1D functions results in the 2D basis functions

$$y_l^m(\theta, \phi) = K_l^m P_l^{|m|}(\cos\theta)\, \Phi^m(\phi)$$
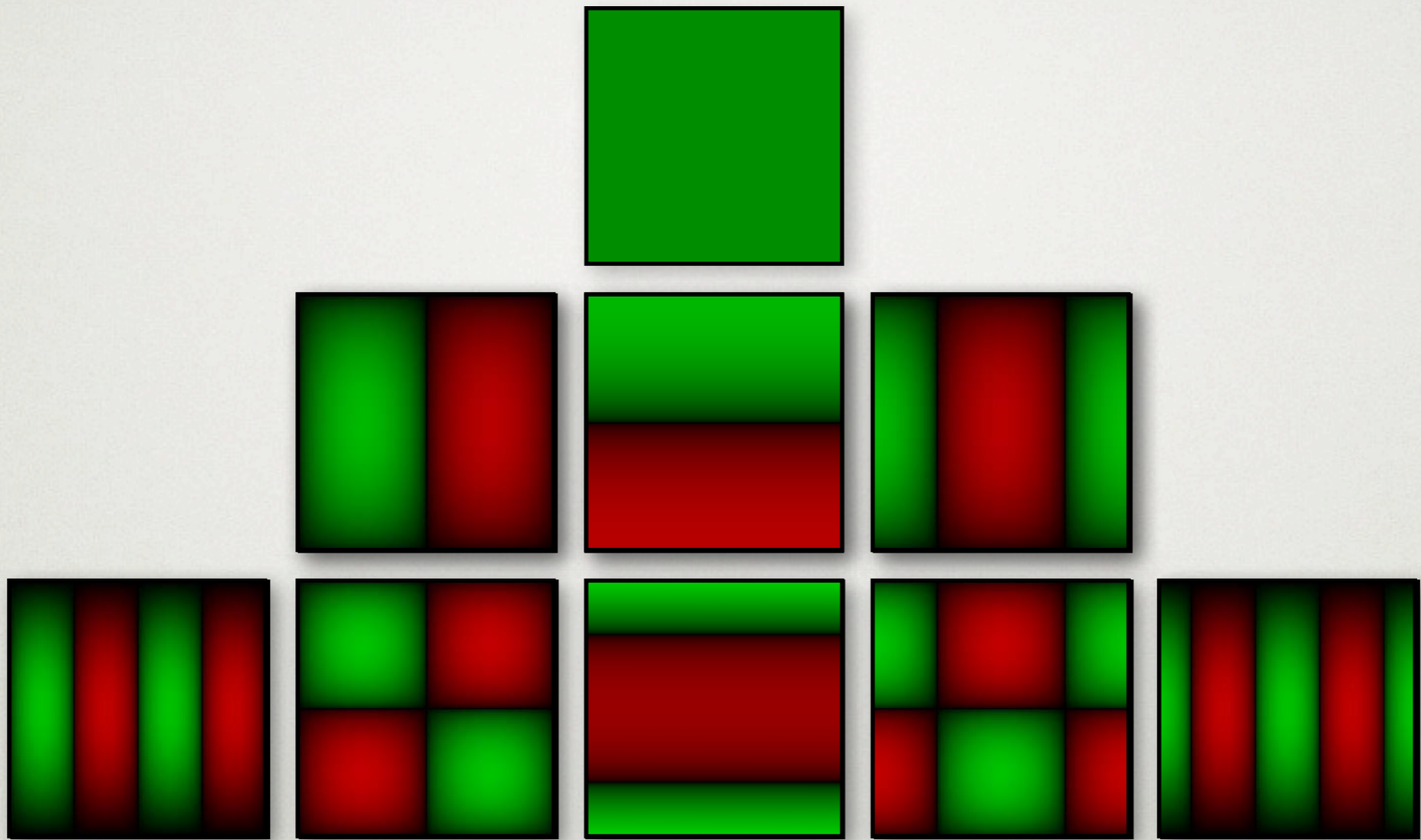
* The product of these 1D functions results in the 2D basis functions

Separable MipMaps

$$\int_{\phi_-}^{\phi_+} \int_{\theta_-}^{\theta_+} y_l^m(\theta, \phi) \sin\theta \; d\theta \; d\phi$$

where

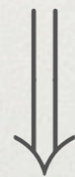$$y_l^m(\theta, \phi) = K_l^m P_l^{|m|}(\cos\theta) \; \Phi^m(\phi)$$

* Mathematically, the reason separability is useful is that it allows us to replace a 2D integral with 2 1D integrals.
* Each mipmap pixel is a 2D integral over a region of the image.
* Separability allows us to compute this 2D mipmap from two 1D mipmaps, which is a significant memory reduction

Separable MipMaps

$$\int_{\phi_-}^{\phi_+} \int_{\theta_-}^{\theta_+} y_l^m(\theta, \phi) \sin \theta \; d\theta \; d\phi$$

where

$$y_l^m(\theta, \phi) = K_l^m P_l^{|m|}(\cos \theta) \; \Phi^m(\phi)$$
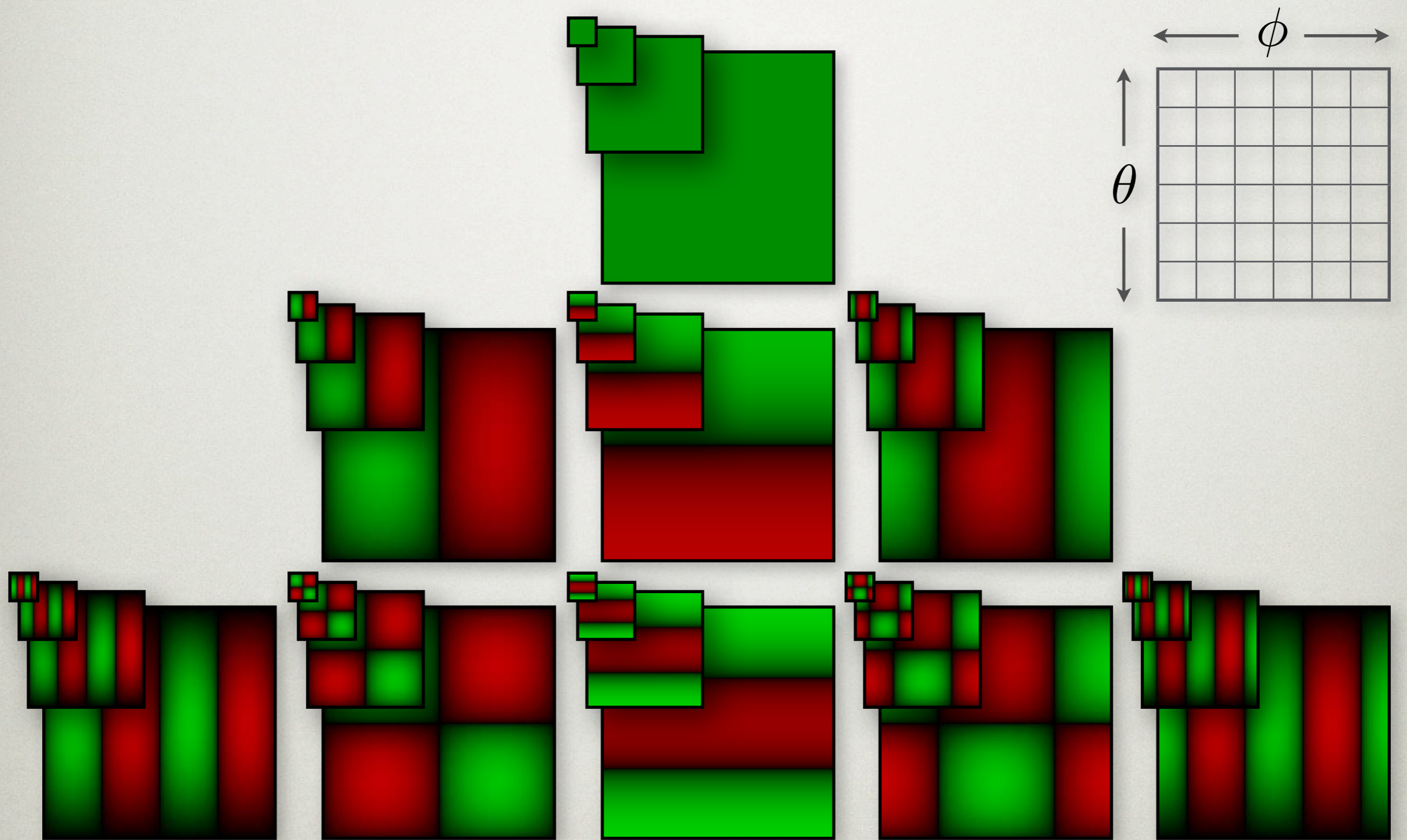
$$\Downarrow$$

$$K_l^m \left(\int_{\theta_-}^{\theta_+} P_l^{|m|}(\cos \theta) \sin \theta \; d\theta \right) \left(\int_{\phi_-}^{\phi_+} \Phi^m(\phi) \; d\phi \right)$$

* Mathematically, the reason separability is useful is that it allows us to replace a 2D integral with 2 1D integrals.
* Each mipmap pixel is a 2D integral over a region of the image.
* Separability allows us to compute this 2D mipmap from two 1D mipmaps, which is a significant memory reduction

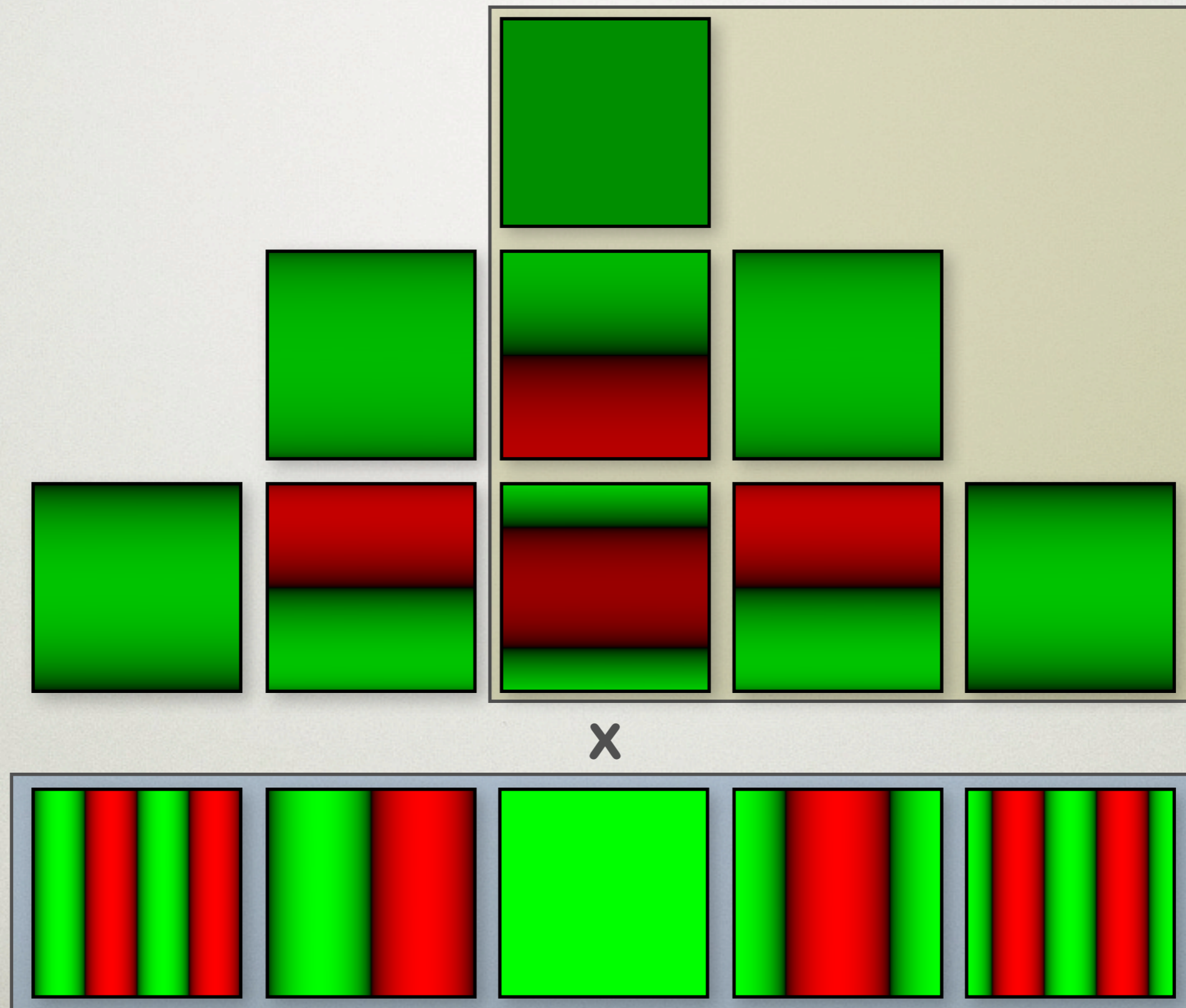# 2D SH Basis MipMaps

* so instead of storing a full set of 2D mipmaps
* we can instead compute a very small collection of 1D mipmaps.

2D = 1D x 1D SH MipMaps

* so instead of storing a full set of 2D mipmaps
* we can instead compute a very small collection of 1D mipmaps.

Analytically Integrable

$$K_l^m \left(\int_{\theta_-}^{\theta_+} P_l^{|m|}(\cos\theta) \sin\theta \, d\theta \right) \left(\int_{\phi_-}^{\phi_+} \Phi^m(\phi) \, d\phi \right)$$

- In cylindrical coordinates, SH basis functions are *analytically* integrable.

  - We can compute the integrals on-the-fly during warping analytically

  - Efficiently evaluated using recurrence relations

* Another observations, we can make is that the 1D mipmap integrals can actually be computed analytically by performing a change of variable to cylindrical coordinates.
* This means that we can compute the mipmaps analytically on–the–fly using recurrence relations presented in the paper.

Analytic Conversion & Product Sampling

- Computing quadrant integrals produces an analytic conversion between SH and spherical mipmaps

  - Same thing can be done for Haar

* one thing I'd like to point out is that computing these quadrant integrals effectively creates an analytic conversion between the SH basis and spherical mipmap basis.
* we could do the same thing with spherical haar wavelets.
* This allows us to multiply the SH function with a Haar wavelet and perform product sampling
* One minor implementation detail is that when doing a wavelet or mipmap product, we need to traverse the quadrant integrals trees twice, so for efficiency, it can still be worthwhile to cache the analytic integrals into 1D mipmaps

Analytic Conversion & Product Sampling

- Computing quadrant integrals produces an analytic conversion between SH and spherical mipmaps

  - Same thing can be done for Haar

- This allows us to multiply the SH function with a Haar wavelet and perform *product* sampling.

* one thing I'd like to point out is that computing these quadrant integrals effectively creates an analytic conversion between the SH basis and spherical mipmap basis.
* we could do the same thing with spherical haar wavelets.
* This allows us to multiply the SH function with a Haar wavelet and perform product sampling
* One minor implementation detail is that when doing a wavelet or mipmap product, we need to traverse the quadrant integrals trees twice, so for efficiency, it can still be worthwhile to cache the analytic integrals into 1D mipmaps

Analytic Conversion & Product Sampling

- Computing quadrant integrals produces an analytic conversion between SH and spherical mipmaps

  - Same thing can be done for Haar

- This allows us to multiply the SH function with a Haar wavelet and perform *product* sampling.

- product traversal -> cache analytic integrals

* one thing I'd like to point out is that computing these quadrant integrals effectively creates an analytic conversion between the SH basis and spherical mipmap basis.
* we could do the same thing with spherical haar wavelets.
* This allows us to multiply the SH function with a Haar wavelet and perform product sampling
* One minor implementation detail is that when doing a wavelet or mipmap product, we need to traverse the quadrant integrals trees twice, so for efficiency, it can still be worthwhile to cache the analytic integrals into 1D mipmaps

Outline

- What is it?

- Why is it desirable?

- How do we do it?

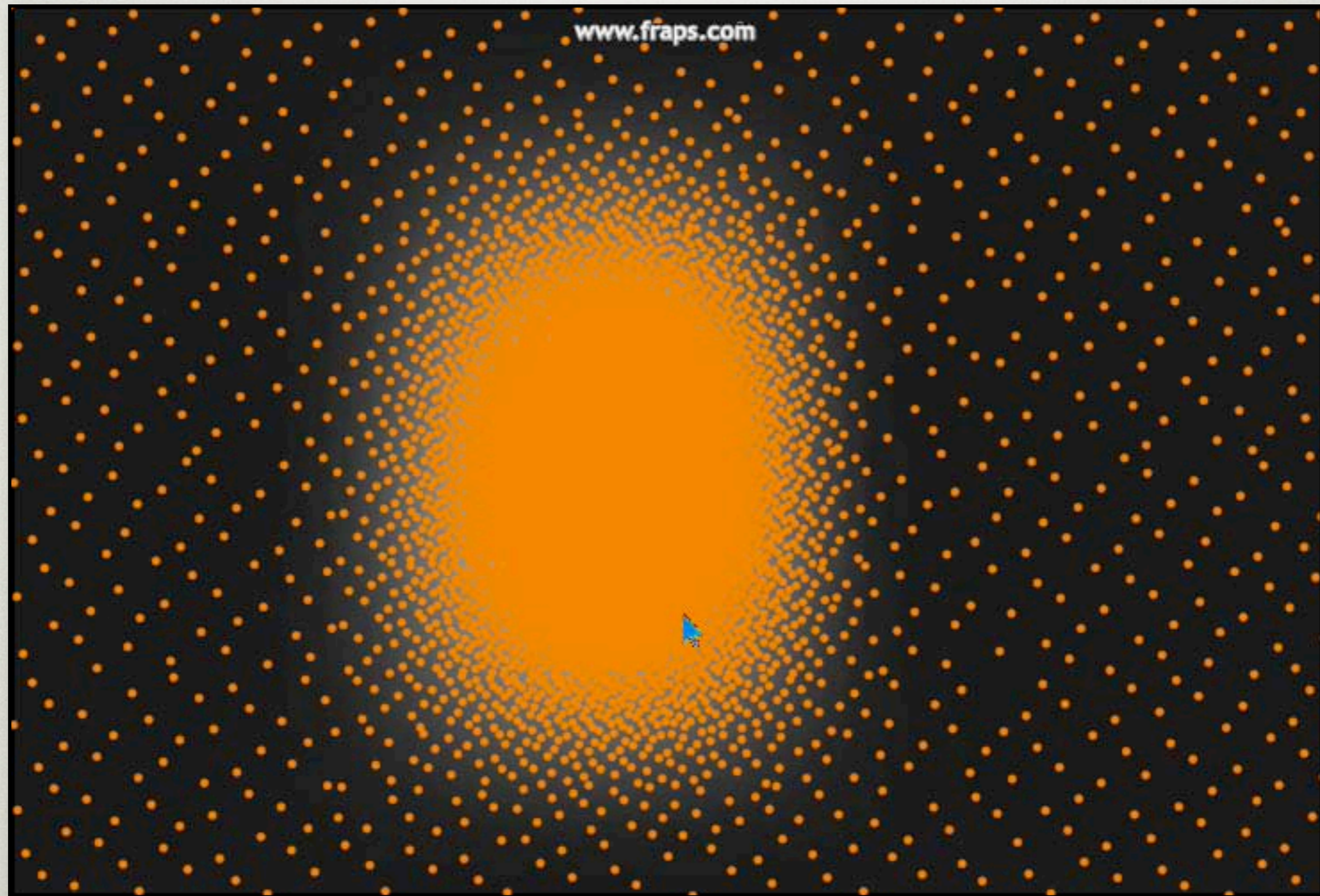- How well does it work?

SH Sampling (random points)
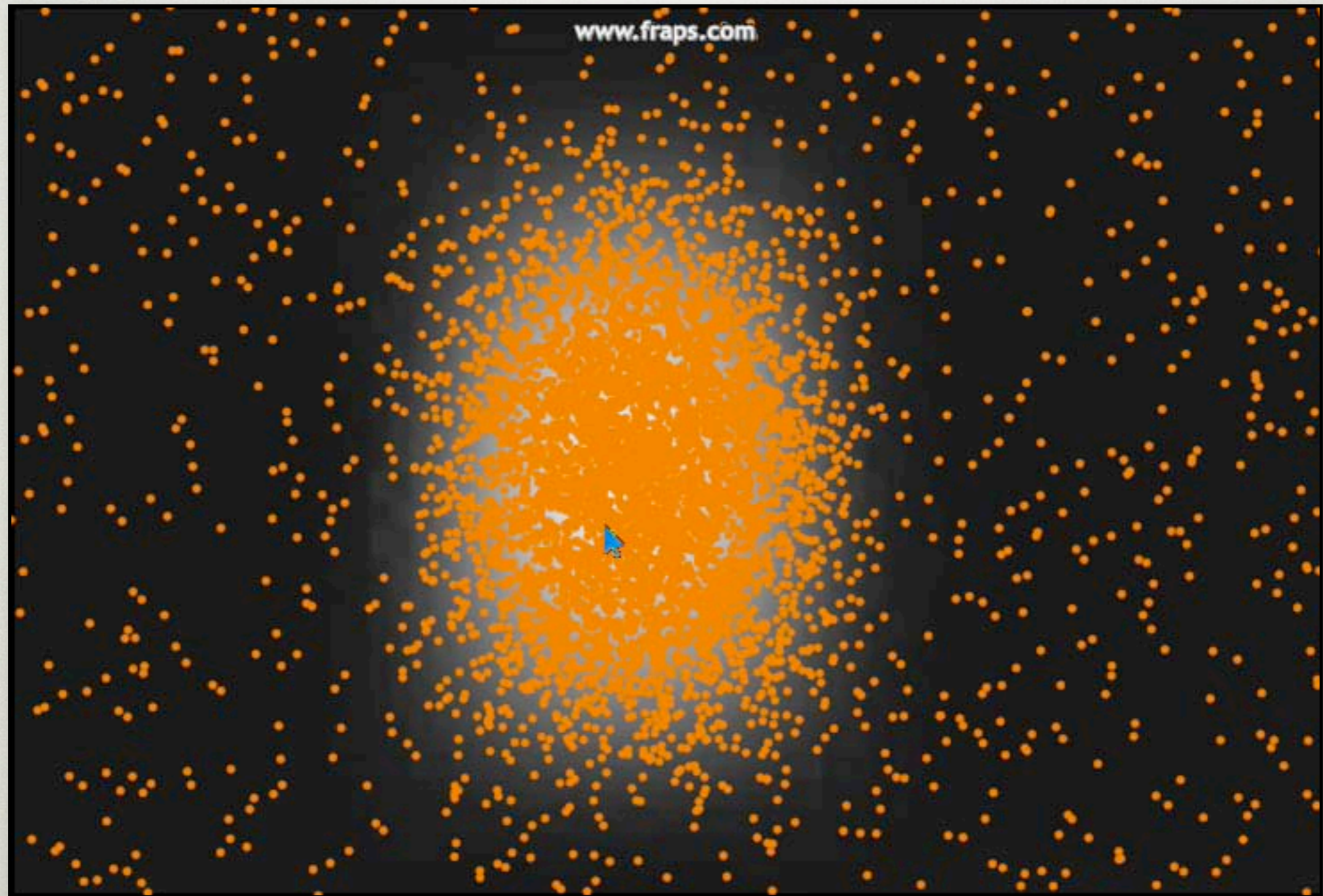
* this is a video capture of our method sampling a spherical harmonic BRDF in real time.
* in this example we used uniform random points as input to the warping
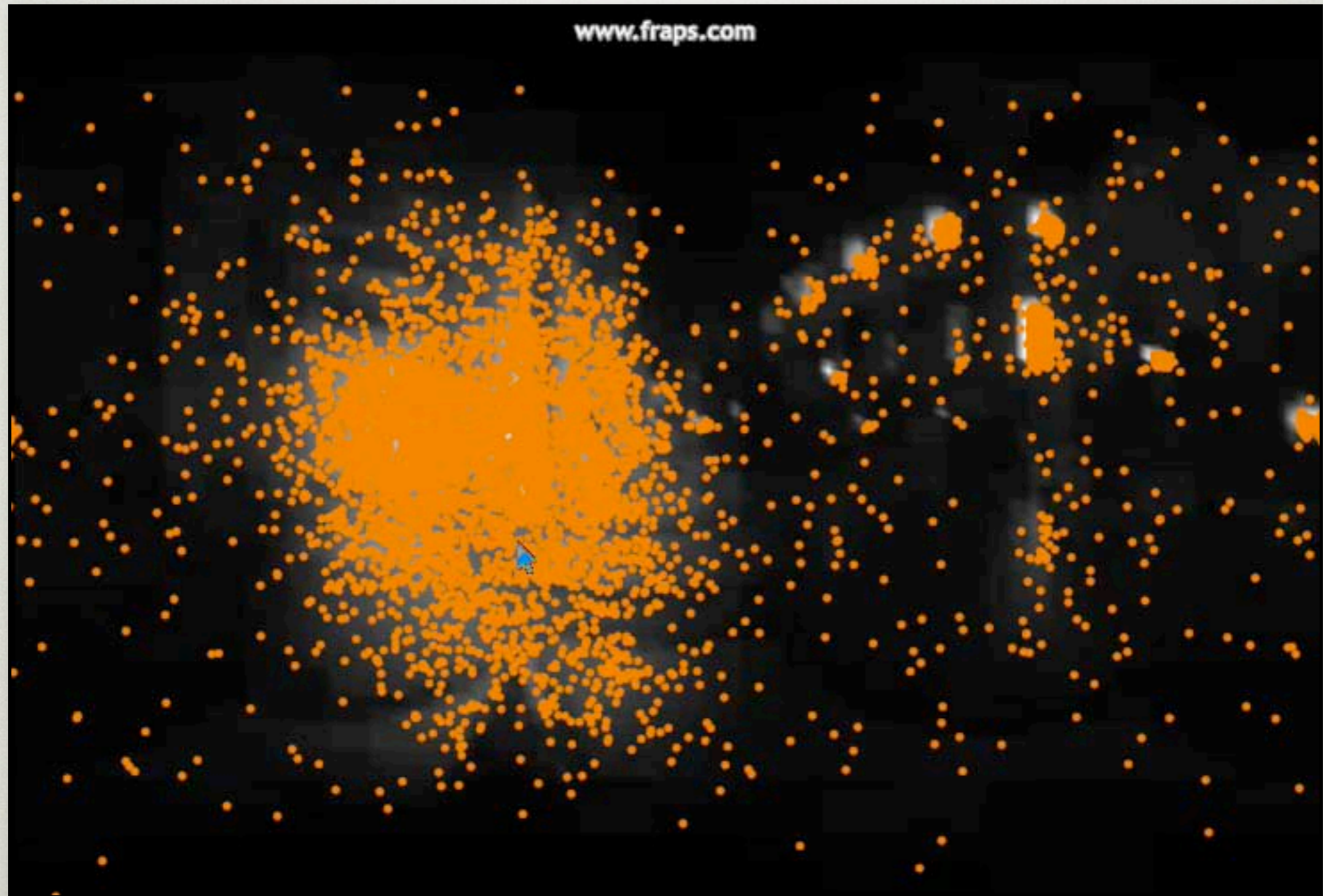
Thursday, 6 September 12

* If we replace those points with quasi-random points, the resulting distribution is of much higher quality at no extra cost, just like with wavelet importance sampling
* In addition to sampling a single SH function, our analytic conversion approach allows us to actually take the product of an SH function and a Haar function and sample this product

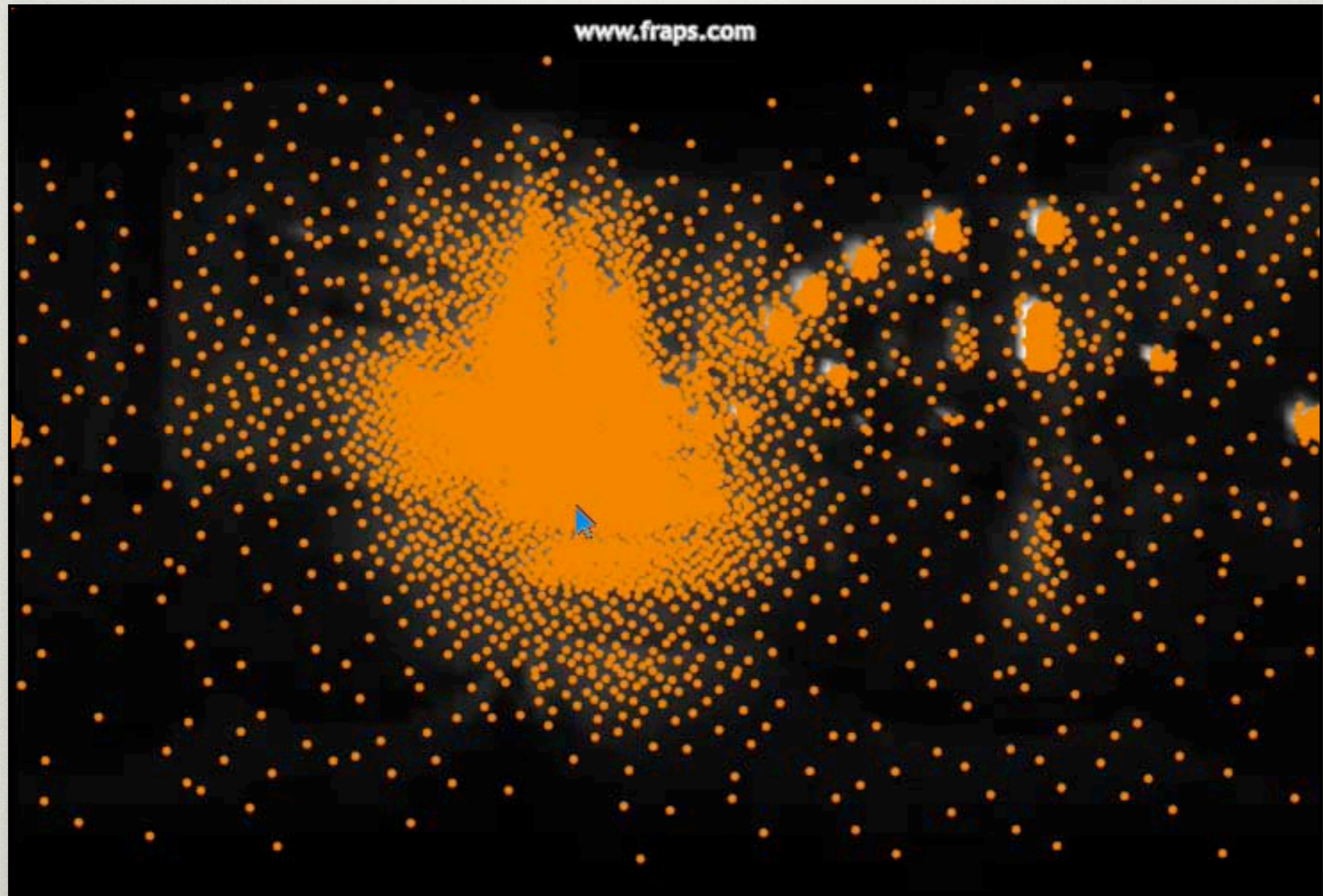# SH x Haar Product
(random points)

Thursday, 6 September 12

* In addition to sampling a single SH function, our analytic conversion approach allows us to actually take the product of an SH function and a Haar function and sample this product
* Here we see a SH BRDF and a Haar environment map product sampled in realtime
* This allows us to combine the benefitial properties of these two distinct bases, retaining high frequency detail in the environment map, and rotation of the BRDF

SH x Haar Product (random points)

* In addition to sampling a single SH function, our analytic conversion approach allows us to actually take the product of an SH function and a Haar function and sample this product
* Here we see a SH BRDF and a Haar environment map product sampled in realtime
* This allows us to combine the benefitial properties of these two distinct bases, retaining high frequency detail in the environment map, and rotation of the BRDF
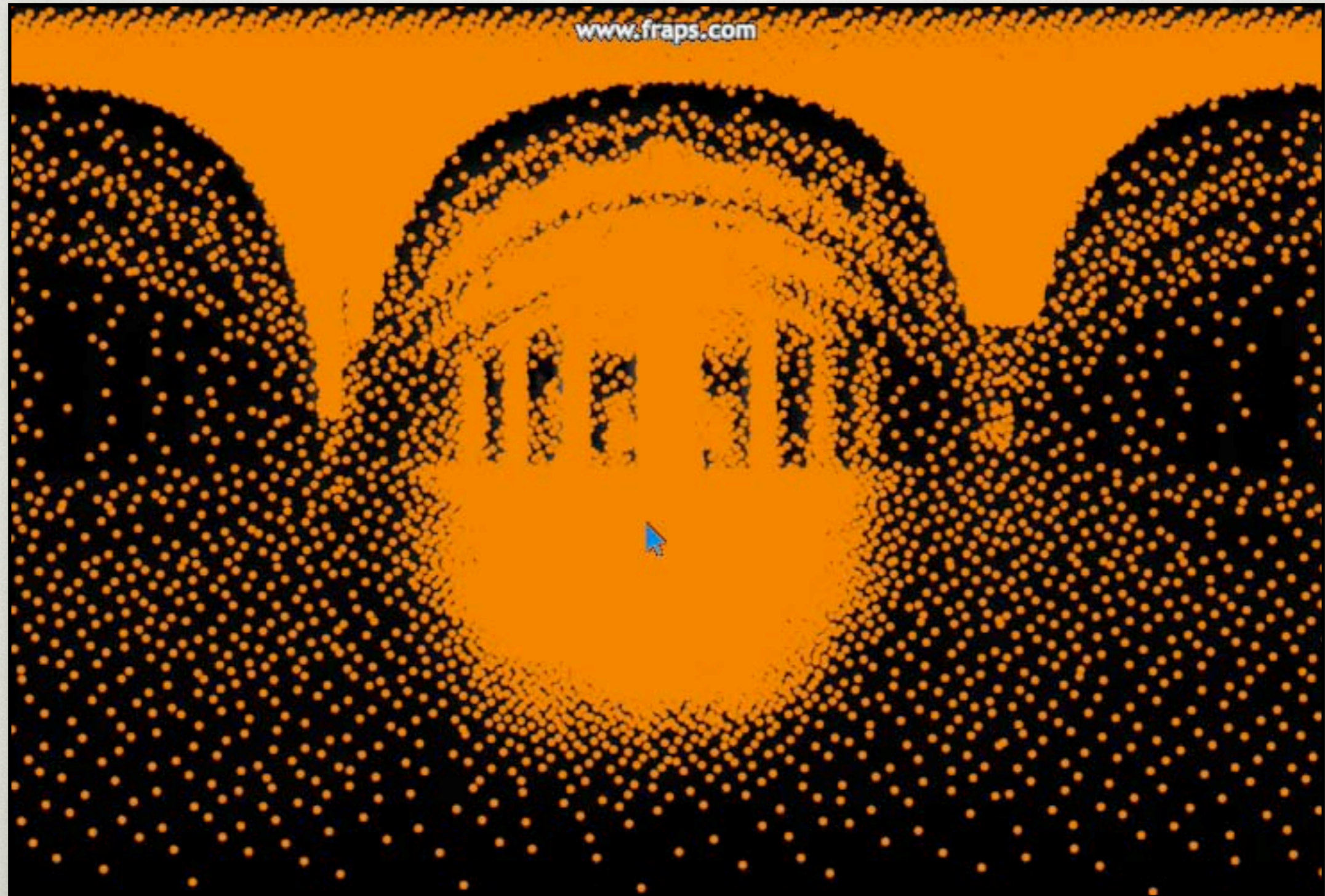
SH x Haar Product (quasi-random points)

Thursday, 6 September 12

* again, by replacing the input points with quasi-random sequences we can obtain a much higher quality distribuction.
* Note that with WIS, we would be required to replicate the BRDF or environment for a number of surface orientations, resulting in much higher memory usage. Whereas here we can rotate the BRDF on-the-fly

SH x Haar Product
(quasi-random points)

Thursday, 6 September 12

* again, by replacing the input points with quasi-random sequences we can obtain a much higher quality distribuction.
* Note that with WIS, we would be required to replicate the BRDF or environment for a number of surface orientations, resulting in much higher memory usage. Whereas here we can rotate the BRDF on-the-fly

SH x Haar Product
(quasi-random points)

* and here is another example with a different environment map.

SH x Haar Product
(quasi-random points)

* and here is another example with a different environment map.

Performance

- Runtime is linear in number of SH coefficients.

* As you saw from the previous videos, our method is fast enough to run in real-time for millions of points, which makes it possible for per-pixel use in a ray tracer.
* The performance of our algorithm is linearly dependent on the number SH coefficients in the BRDF. This can initially seem like a fairly large drawback, however:

Performance

* If we plot the performance as a function of the number of coefficients, we see that even with 400 coefficients we are able to obtain over 5M samples per second.
* So, even though the algorithm is linear in the number of coefficients, the slope of this line is extremely small.

Negative Values

- SH basis functions have negative values

- Does this pose a problem?

* One area which deserves discussion is negative values.
* A question that might come to mind regarding this approach is that SH basis functions have negative values, and whether this poses a problem.
* The answer to this is that it is absolutely OK for the basis functions to have negative values, as long as the final SH approximation is non-negative.
* What this means is that the warping is un-biased for any valid PDF

Negative Values

- SH basis functions have negative values

- Does this pose a problem?

  - This is OK, as long as the final SH approximation is non-negative.

* One area which deserves discussion is negative values.
* A question that might come to mind regarding this approach is that SH basis functions have negative values, and whether this poses a problem.
* The answer to this is that it is absolutely OK for the basis functions to have negative values, as long as the final SH approximation is non-negative.
* What this means is that the warping is un-biased for any valid PDF

Negative Values

- SH basis functions have negative values

- Does this pose a problem?

  - This is OK, as long as the final SH approximation is non-negative.

  - Warping is un-biased for *valid* (non-negative) PDFs

* One area which deserves discussion is negative values.
* A question that might come to mind regarding this approach is that SH basis functions have negative values, and whether this poses a problem.
* The answer to this is that it is absolutely OK for the basis functions to have negative values, as long as the final SH approximation is non-negative.
* What this means is that the warping is un-biased for any valid PDF

Ringing/Artifacts

- SH approximations suffer from ringing, does this present a problem?

* In practice, however, if we start off with a valid PDF, and then perform compression by projecting onto SH, this might cause ringing and negative values.
* This can indeed be a problem with the SH projection, and there are some well-known ways to avoid this which we discuss in the paper.
* When we were faced with this problem initially, we assumed that this was a drawback of our technique compared to WIS (since Haar wavelets do not suffer from ringing), but it actually turns out that this problem also happens with wavelets and sometimes to a larger extent

Ringing/Artifacts

- SH approximations suffer from ringing, does this present a problem?

  - Yes, it can, if ringing introduces negative values

* In practice, however, if we start off with a valid PDF, and then perform compression by projecting onto SH, this might cause ringing and negative values.
* This can indeed be a problem with the SH projection, and there are some well-known ways to avoid this which we discuss in the paper.
* When we were faced with this problem initially, we assumed that this was a drawback of our technique compared to WIS (since Haar wavelets do not suffer from ringing), but it actually turns out that this problem also happens with wavelets and sometimes to a larger extent

Ringing/Artifacts

- SH approximations suffer from ringing, does this present a problem?

  - Yes, it can, if ringing introduces negative values

- Haar wavelets do not have this problem...

* In practice, however, if we start off with a valid PDF, and then perform compression by projecting onto SH, this might cause ringing and negative values.
* This can indeed be a problem with the SH projection, and there are some well-known ways to avoid this which we discuss in the paper.
* When we were faced with this problem initially, we assumed that this was a drawback of our technique compared to WIS (since Haar wavelets do not suffer from ringing), but it actually turns out that this problem also happens with wavelets and sometimes to a larger extent

Ringing/Artifacts

- SH approximations suffer from ringing, does this present a problem?

  - Yes, it can, if ringing introduces negative values

- Haar wavelets do not have this problem...

  - or do they?

* In practice, however, if we start off with a valid PDF, and then perform compression by projecting onto SH, this might cause ringing and negative values.
* This can indeed be a problem with the SH projection, and there are some well-known ways to avoid this which we discuss in the paper.
* When we were faced with this problem initially, we assumed that this was a drawback of our technique compared to WIS (since Haar wavelets do not suffer from ringing), but it actually turns out that this problem also happens with wavelets and sometimes to a larger extent

Ringing/Artifacts
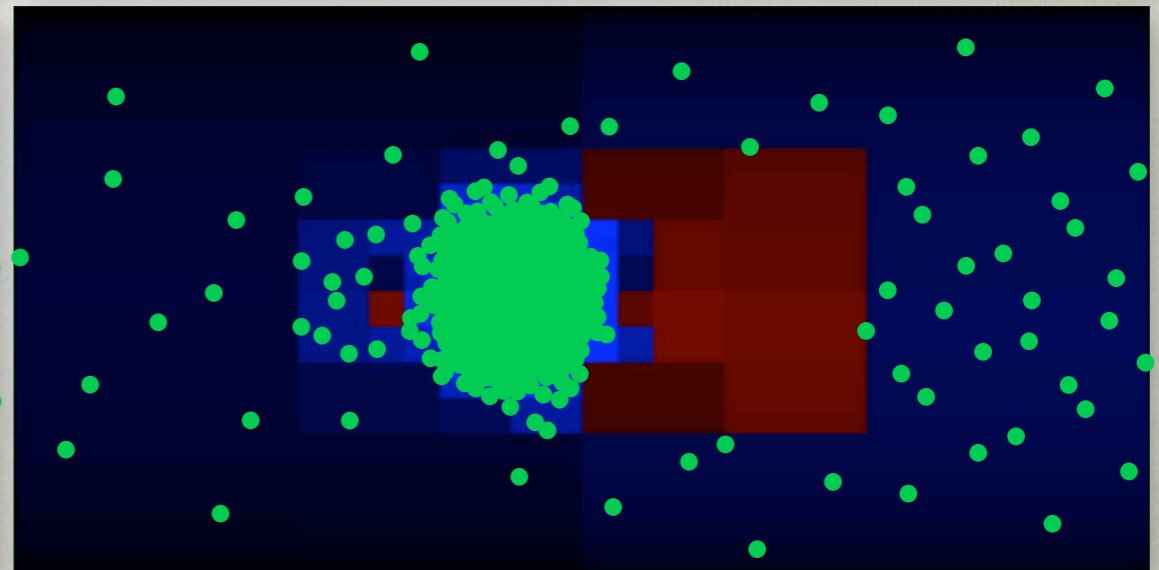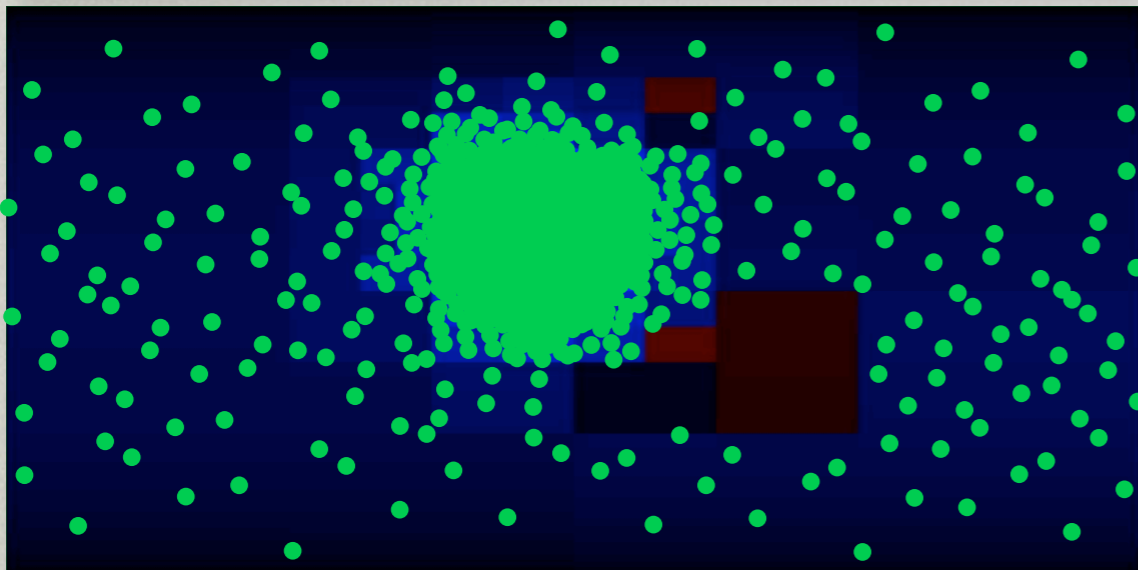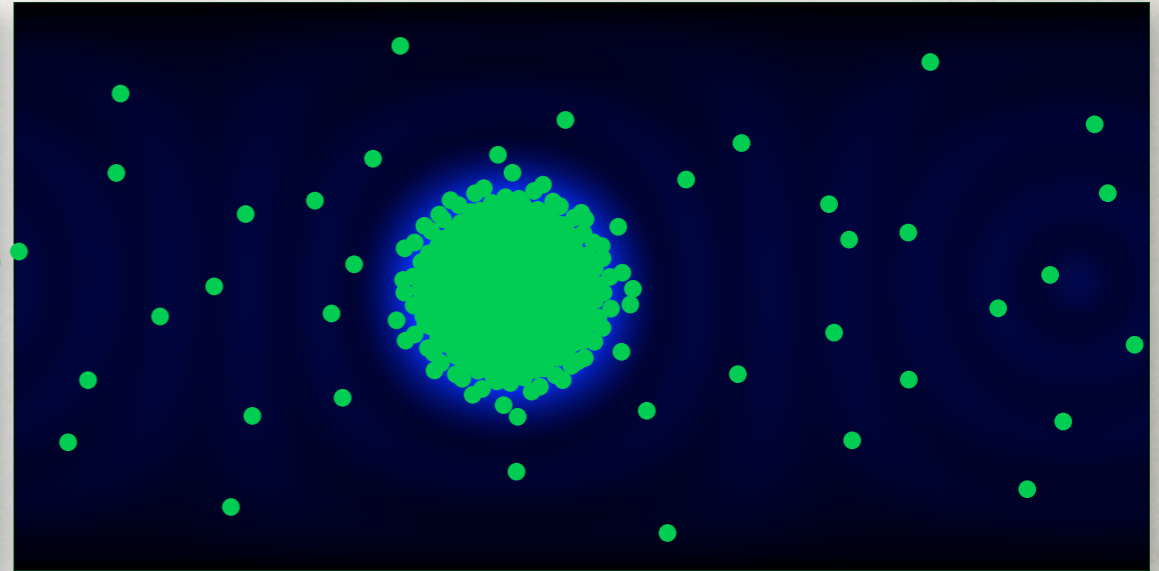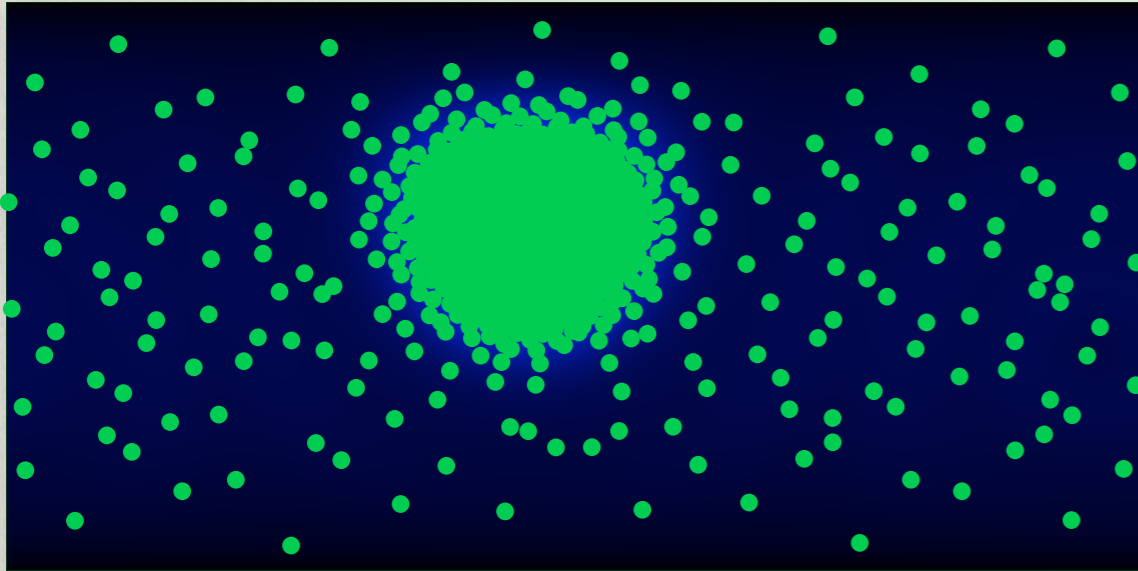
Phong exponent 10 (64 coefficients)   Phong exponent 20 (169 coefficients)

* Even though Haar wavelets do not introduce ringing, they do add error through compression, and anytime the error makes the function negative, the algorithm can completely fail.

* Here we have a comparison of SH sampling on the top, and WIS on the bottom for a Phong BRDF.

* If we use the same amount of memory for the two basis, then it turns out that negative values are actually more of a problem for Haar wavelets than they are for SH.

* whenever the reconstruction results in a non-positive value, no points will be distributed within that region, this will cause errors and bias in the renderings.
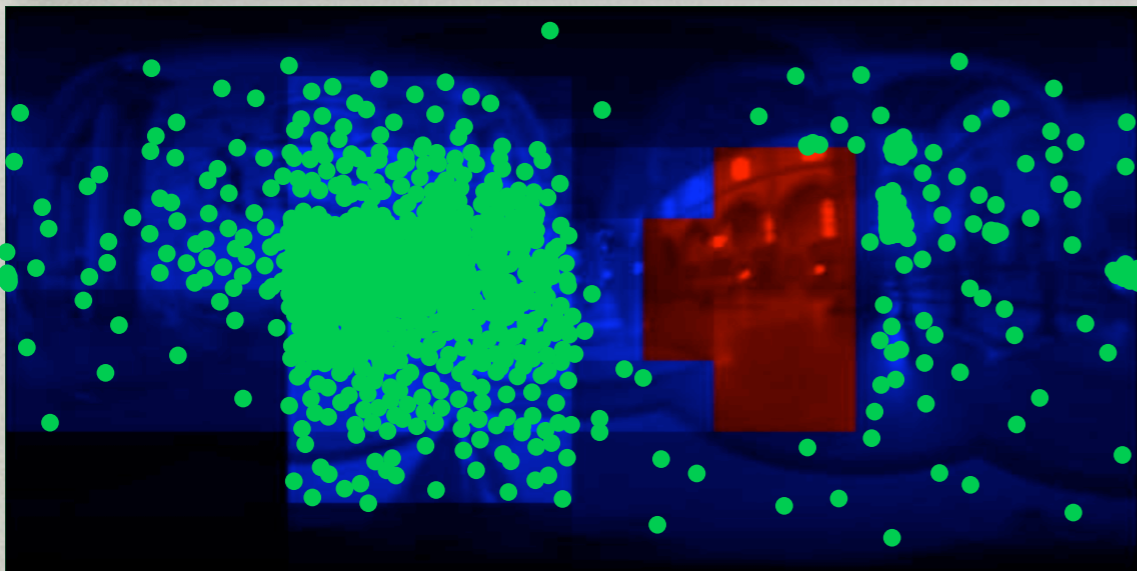
Ringing/Artifacts

Galileo's Tomb

Eucalyptus Grove

Spherical Harmonics
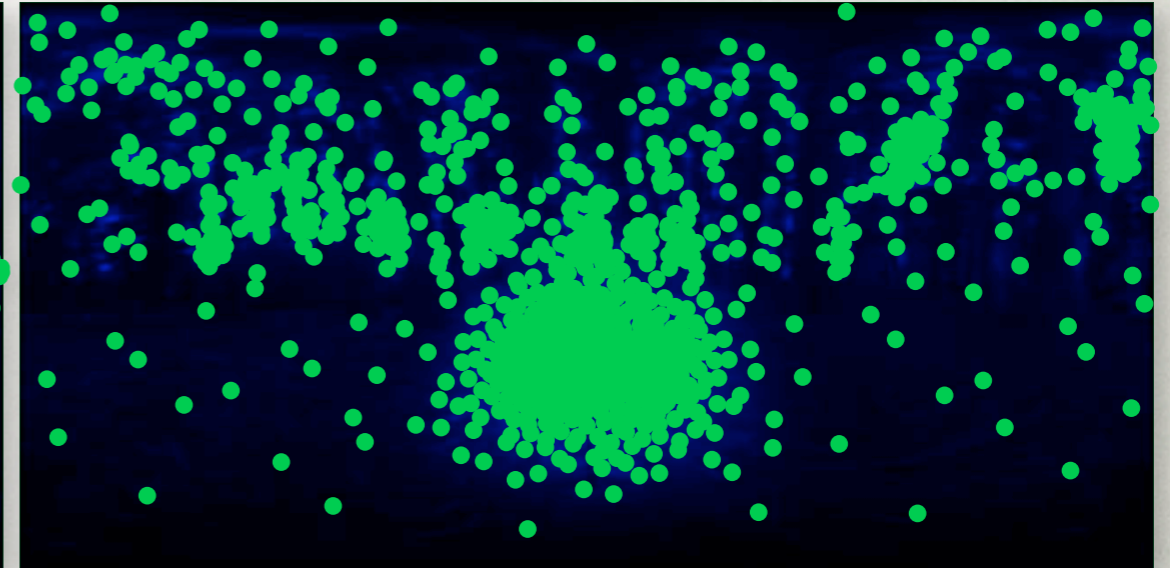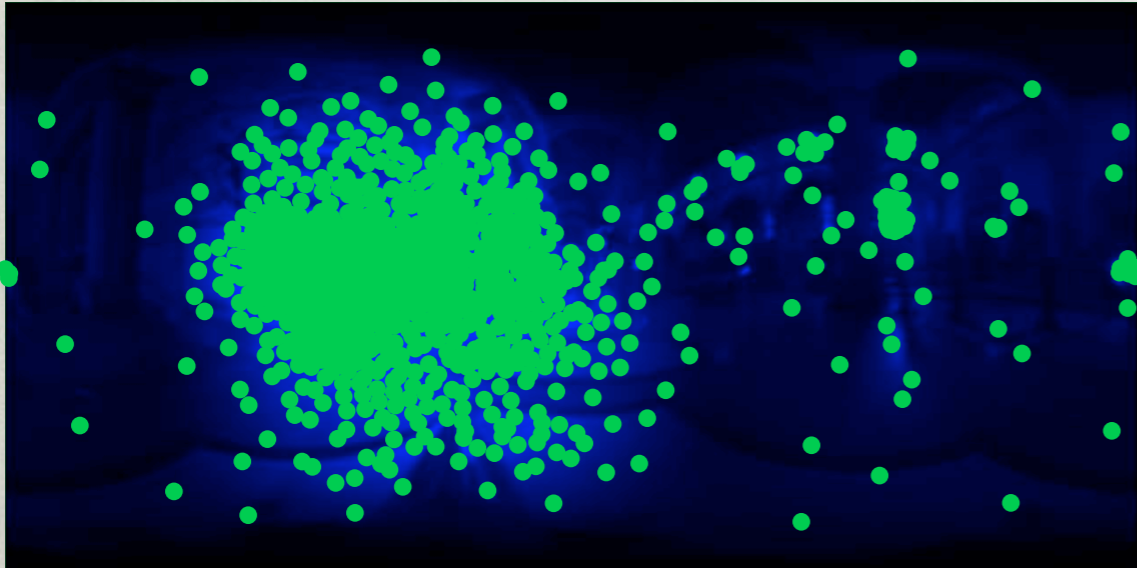
Haar Wavelet
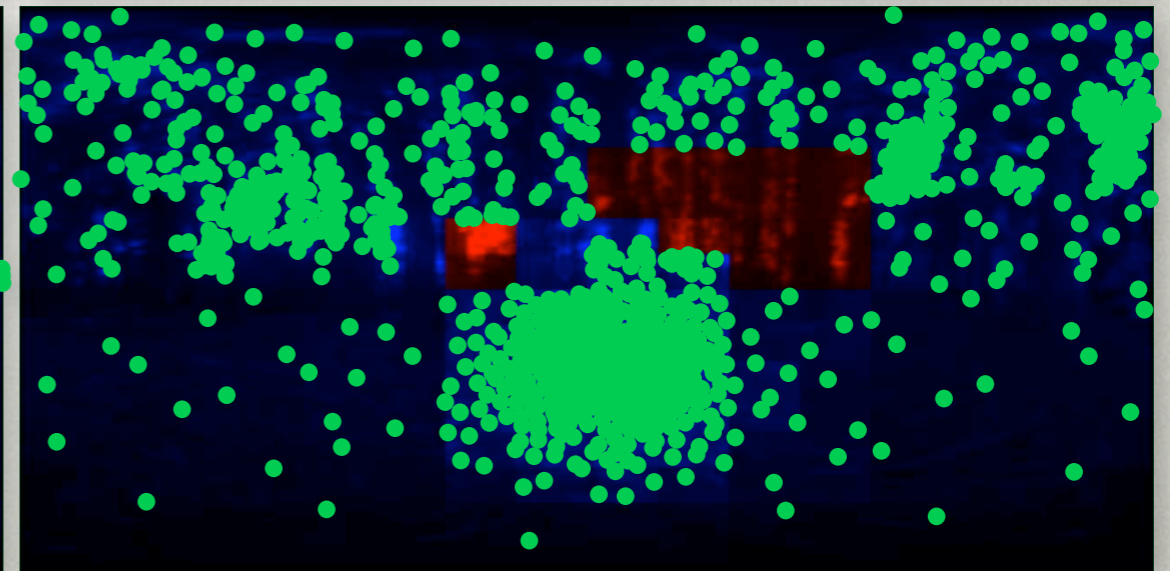
Phong exponent 3 (25 coefficients)

Phong exponent 10 (64 coefficients)

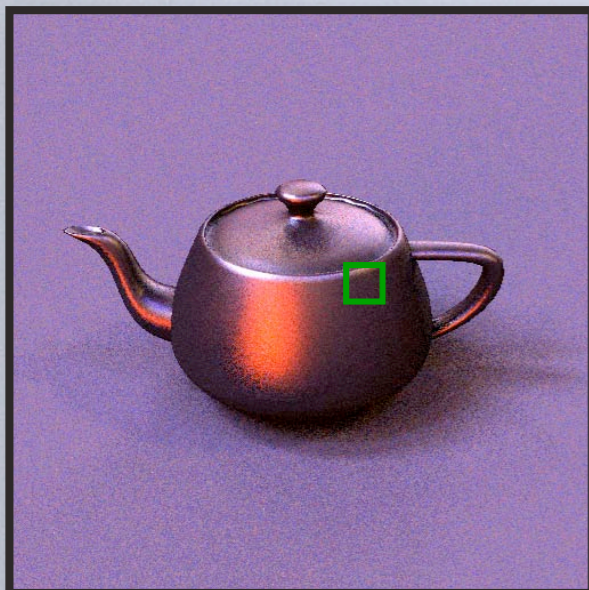* This becomes even more apparent with product sampling.
* as you can see, with Haar wavelets and this extreme compression negative reconstructions result in no samples being placed in regions which should have very high contributions, such as the windows visible in Galileo's tomb.
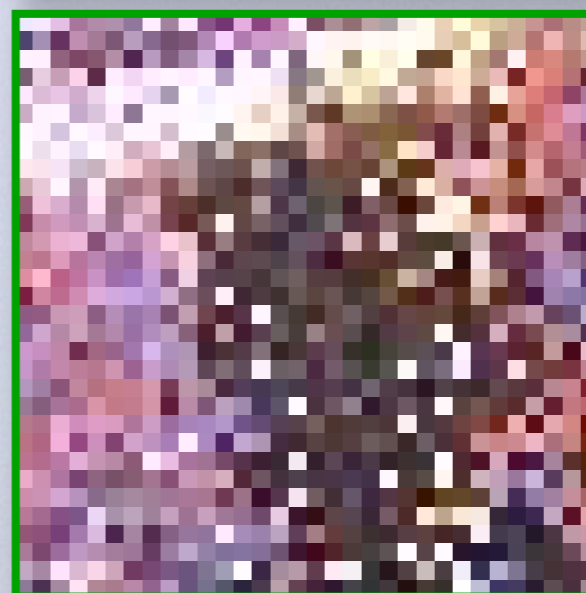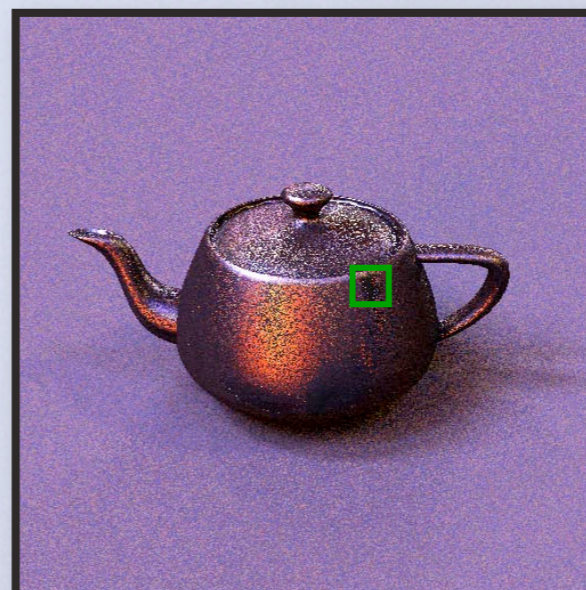
SH x Haar Product (Our Method)

8 samples/pixel   128 samples/pixel
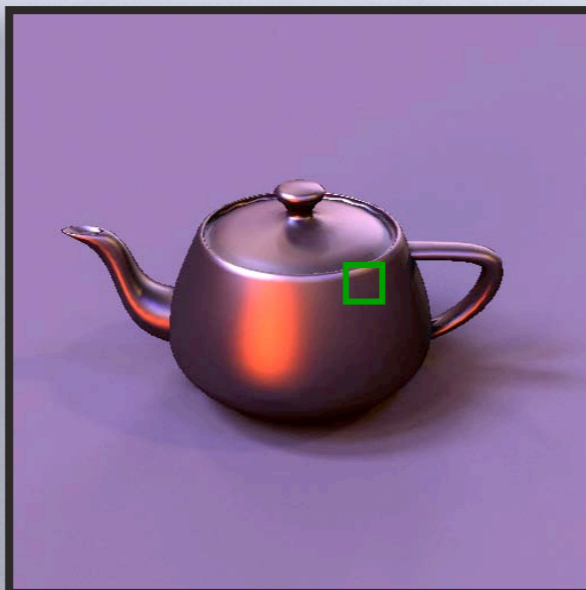
Haar x Haar Product (Clarberg et al. '05)

8 samples/pixel   128 samples/pixel
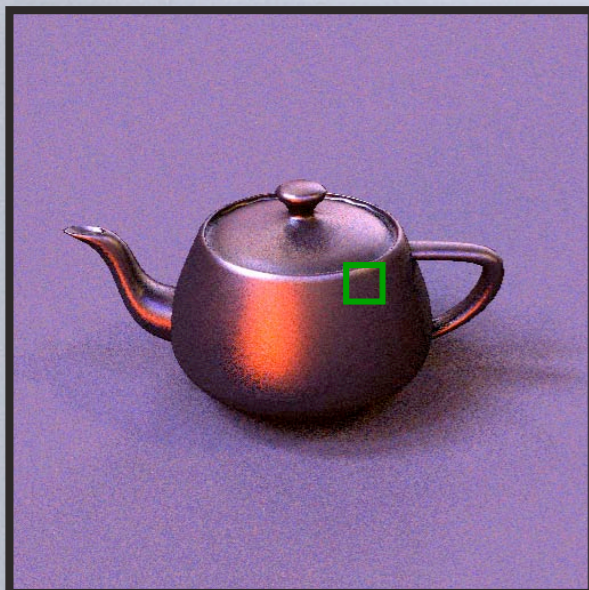
Approx. equal memory usage

* These missing negative regions directly influence the final render.
* Here we show a comparison between our method and a purefly wavelet based approach, where rotation is not supported.
* Since the BRDF must be replicated, extreme compression is necessary to obtain roughly equal memory usage, which results in artifacts due to negative reconstructions in the PDF.

SH x Haar Product
(Our Method)

Haar x Haar Product
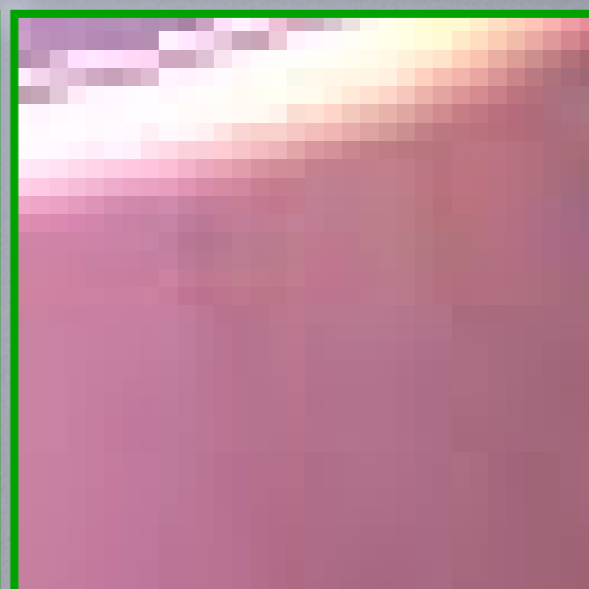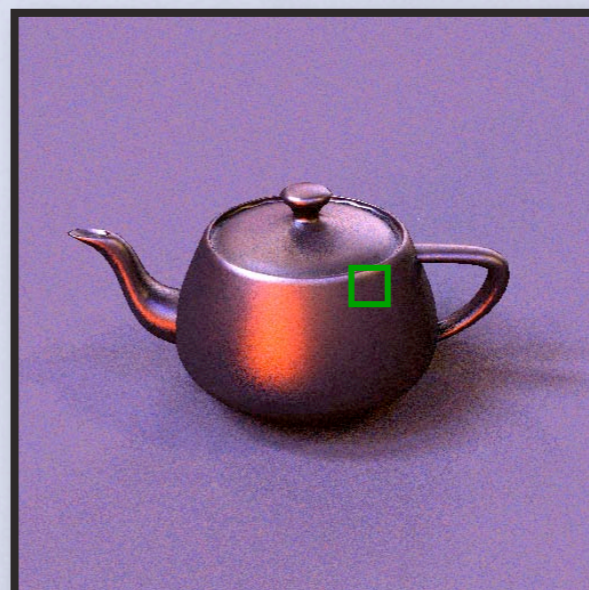**(Clarberg et al. '05)**

8 samples/pixel 128 samples/pixel 8 samples/pixel 128 samples/pixel

Approx. equal quality
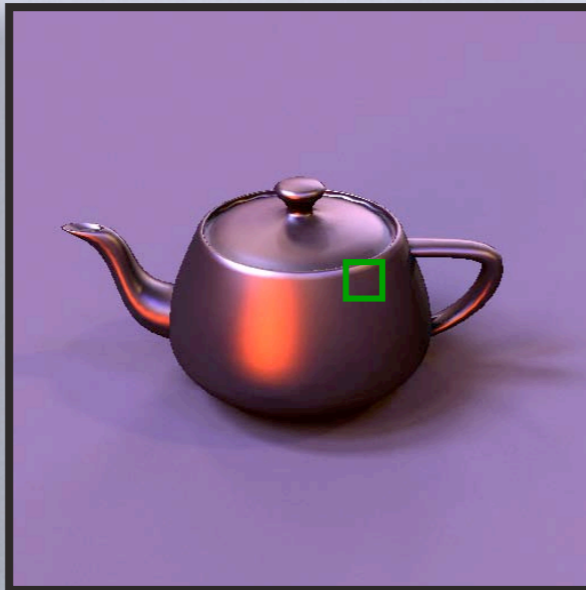
 * We could instead use less compression for the wavelets.
* This would indeed result in images of comparable quality, but would require several orders of magnitude more memory for the wavelets.
* Since our method can rotate one of the two functions, this replication is not necessary
* and we can render images of this quality using only about 300 bytes of storage, vs. 256 MB of storage for the Haar wavelets

SH x Haar Product
(Our Method)

Haar x Haar Product
(Clarberg et al. '05)

8 samples/pixel   128 samples/pixel   8 samples/pixel   128 samples/pixel
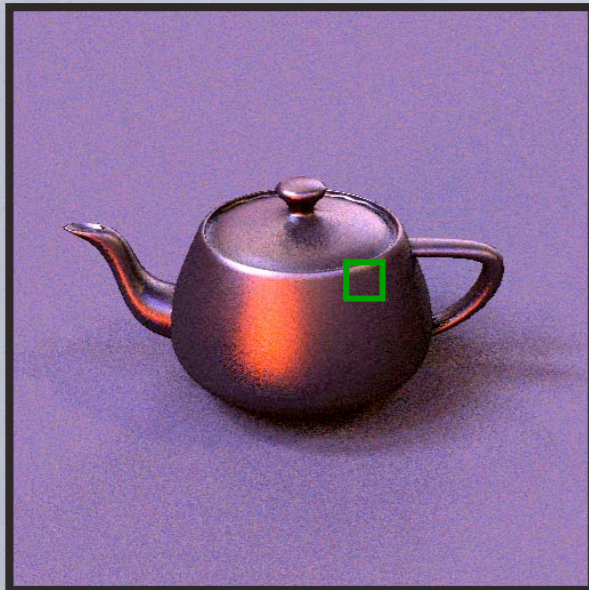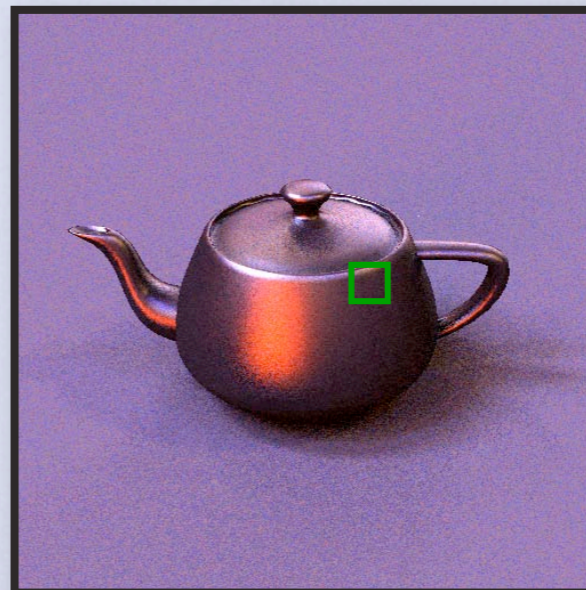
0.3 KB   **256 MB**

Approx. equal quality

 * We could instead use less compression for the wavelets.
* This would indeed result in images of comparable quality, but would require several orders of magnitude more memory for the wavelets.
* Since our method can rotate one of the two functions, this replication is not necessary
* and we can render images of this quality using only about 300 bytes of storage, vs. 256 MB of storage for the Haar wavelets

Summary

- Importance sampling spherical harmonic functions

    - Analytic conversion between SH and spherical mipmaps/Haar wavelets

    - Produces high-quality distributions

- Sampling SH x Haar product.

* So, in summary, our paper presented an efficient importance sampling scheme for spherical harmonics
* it utilized an analytic conversion between SH and mipmaps/haar wavelets
* Due to this conversion we are also able to samples products of SH and haar wavelets

Future Work

- Ensure non-negative approximations

- Explore other bases

- Explore/exploit analytic conversion between SH and Haar further.

* There is still alot of exciting possibilities for future work
* For one, it would be nice to use an automatic method to ensure non–negative approximations. Currently we do this manually, but there are known methods to ensure this automatically. From our results, we see that this is also relevant for regular WIS
* Another are which we think is very exciting is to explore other bases and the conversion between SH and Haar further.
* As it stands now, we have this table of capabilities of Haar and SH, and we have now filled in the importance sampling checkmark for SH.
* It would be extremely useful to try to fill in more of these checkmarks, or come up with other bases that satisfy more of them.

Future Work

- Ensure non-negative approximations

- Explore other bases

- Explore/exploit analytic conversion between SH and Haar further.

	Convolution	Product	Product integral	Rotation	Alias-free	High Frequency	Low Frequency	Sampling
SH	✓	✓	✓	✓	✓	✗	✓	✓
Haar	✗	✓	✓	✗	✗	✓	✓	✓

* There is still alot of exciting possibilities for future work
* For one, it would be nice to use an automatic method to ensure non-negative approximations. Currently we do this manually, but there are known methods to ensure this automatically. From our results, we see that this is also relevant for regular WIS
* Another are which we think is very exciting is to explore other bases and the conversion between SH and Haar further.
* As it stands now, we have this table of capabilities of Haar and SH, and we have now filled in the importance sampling checkmark for SH.
* It would be extremely useful to try to fill in more of these checkmarks, or come up with other bases that satisfy more of them.

Thank You