

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Department of Physics, Informatics and Mathematics
Master's Degree in Computer Science

A Spherical Representation of HDR Environment Maps with Adaptive Gaussian Primitives

---

Candidate
Alessandro Catenacci
ID number 207458

Supervisor
Prof. Fabio Pellacini

Academic Year 2025/2026

Abstract

HDR environment maps are widely used in computer graphics to represent complex illumination and reproduce realistic lighting and reflections in 3D scenes. Representing these signals compactly is challenging because they combine smooth low-frequency illumination, highly localized light sources, and structured regions with sharp transitions. Conventional image-based representations also rely on planar parameterizations of the sphere, introducing distortions and discontinuities that are unrelated to the underlying spherical signal.

This thesis presents a compact and continuous representation of HDR environment maps defined directly on the sphere. The proposed model extends Anisotropic Spherical Gaussians (ASGs) with two independent shape parameters controlling radial falloff and contour geometry. These additional degrees of freedom allow each primitive to adapt to local signal structure beyond the Gaussian profile and elliptical contours of standard ASGs. A moment-preserving normalization in the tangent plane keeps the interpretation of kernel scale and coefficient magnitude consistent as the kernel shape changes.

Reconstruction is formulated as the optimization of a positive mixture of spherical kernels. The fitting procedure jointly optimizes position, orientation, scale, color, and shape, while adaptive spherical sampling, an HDR-aware objective, and residual-driven densification progressively allocate additional primitives to regions that require greater representational capacity.

A compact codec is also introduced to quantize and store the resulting model parameters. The method is evaluated through reconstruction quality, rate-distortion analysis, shape ablations, visual comparisons, and image-based lighting experiments. Comparisons against standard ASGs and Image-GS, an adaptive image-space representation based on 2D Gaussian primitives, are used to assess the effect of shape adaptation and the benefits of operating directly on the sphere.

The results show that adapting radial profile and contour geometry increases the expressive capacity of spherical Gaussian kernels while preserving a compact, continuous, and parameterization-independent representation. Radial and contour adaptation play complementary roles, allowing the model to better capture localized emitters, elongated structures, and sharp boundaries that are difficult to represent with fixed Gaussian footprints.

Ai miei nonni

Ringraziamenti

Ringrazio il mio relatore, il Prof. Fabio Pellacini, per avermi guidato con attenzione e per la libertà che mi ha lasciato nel portare avanti questo lavoro. Ha seguito le mie scelte senza mai sostituirsi a me, e gliene sono grato.

Mamma e papà hanno fatto in modo che potessi studiare senza sentirne il peso. Mi hanno dato fiducia senza mai trasformarla in aspettative, e la libertà di capire da solo dove volessi andare. Con loro alle spalle, nessuna decisione mi è mai sembrata troppo grande.

Con mio fratello Luca non ho mai avuto bisogno di spiegarmi troppo. Quando contava, sapevo già da che parte l'avrei trovato, e non mi ha mai dato motivo di dubitarne.

Nonna Diana ha avuto in me una fiducia molto più ostinata della mia. I miei dubbi, per lei, erano sempre provvisori: dove io vedevo soprattutto quello che poteva andare storto, lei vedeva senza nessuna fatica quello che poteva andare bene.

Tra i miei amici, due meritano più di una riga, entrambi lontani solo sulla carta. Andrea è stato per anni una costante delle mie giornate, fino a farmi dimenticare quanto poco fosse scontato. Giovanni c'è da prima di quasi tutto il resto, ed è una di quelle persone che il tempo non ha fatto altro che confermare.

Grazie anche a chi non ho nominato, ma sa di aver avuto un posto in tutto questo.

Contents

Abstract . . . . . **iii**

Ringraziamenti . . . . . **vii**

List of Figures . . . . . **xi**

List of Tables . . . . . **xiii**

1 Introduction **1**
1.1 Contributions 4

2 Background & Related Work **5**
2.1 HDR Environment Maps and Image-Based Lighting 5
2.2 Spherical Signals and Their Parameterizations 7
2.3 Representations of Spherical Illumination 11
2.4 Spherical Gaussian Representations 13
2.5 Adaptive Primitive Representations 14
2.6 Neural Representations of Illumination 16
2.7 Positioning of This Work 16

3 Shape-Adaptive Spherical Primitives **19**
3.1 Positive Mixtures on the Sphere 19
3.2 The ASG as a Lifted Planar Footprint 20
3.3 Adapting the Radial Profile and the Contour 22
3.4 Moment-Preserving Normalization 23
3.5 The Primitive on the Sphere 26
3.6 Properties of the Representation 27

4 HDR Environment Map Reconstruction **29**
4.1 Reconstruction as Optimization on the Sphere 29
4.2 Spherical Quadrature 30
4.3 A Perceptual HDR Objective 31
4.4 Optimizing Constrained Primitives 33
4.5 Initialization and Residual-Guided Allocation 35
4.6 Compact Representation 38

5 Experimental Evaluation **41**
5.1 Experimental Setup 41
5.2 Shape Adaptation at a Matched Primitive Count 43
5.3 Radial and Contour Adaptation 45
5.4 How the Fitted Primitives Use Shape 46
5.5 Rate–Distortion 48
5.6 Comparison with an Image-Space Representation 51
5.7 Rendering with Reconstructed Environments 53

6 Discussion **57**
6.1 Shape as Capacity per Primitive 57
6.2 Radiometric and Perceptual Fidelity 59
6.3 What Fitting on the Sphere Provides 60
6.4 Limitations 62
6.5 Future Work 63

7 Conclusion **67**

A Implementation and Experimental Details **69**
A.1 Software and Data 69
A.2 Fitting Configuration 69
A.3 Image-GS 69
A.4 Evaluation Measures 71
A.5 Rendering 71

B Per-Scene Results **73**

References **75**

List of Figures

1.1 From an HDR raster to a spherical signal 2

2.1 Equal raster cells cover unequal solid angles 9
2.2 One environment in three rasters 10

3.1 Lifting a planar footprint to the sphere 21
3.2 Shape adaptation on the sphere 23
3.3 Contours and profiles of the normalized footprint 24

4.1 Adaptive refinement of the fitting quadrature 31
4.2 Positive residual and insertion 36

5.1 The fifteen evaluation environments 42
5.2 Local reconstruction at 10,000 primitives 45
5.3 Fitted exponents of the full model 47
5.4 Shapes of the most energetic primitives 48
5.5 Quality against storage 49
5.6 Reconstruction across budgets 50
5.7 Whole-sphere reconstructions at 10,000 primitives 52
5.8 Rendering with reconstructed environments 53
5.9 Map quality against rendering quality 55

List of Tables

2.1 Representation families and what they adapt 17

3.1 Parameters of a shape-adaptive primitive 28

4.1 Bits per field of a primitive record 39

5.1 ASG and full model at 10,000 primitives 44
5.2 Radial and contour adaptation 46
5.3 Storage and quality across budgets 49
5.4 Comparison with Image-GS 51
5.5 Agreement between map measures and renderings 54

A.1 Fitting parameters 70

B.1 Image-GS and rendering error per scene 73

Chapter 1

Introduction

Much of the realism of a rendered image comes from its lighting, and much of the lighting used in computer graphics is not modeled explicitly but photographed. A high-dynamic-range (HDR) panorama captured at a single point records how much light arrives there from every direction: the sun and the sky, lamps and windows, and the dimmer light reflected by walls, floors, and furniture. Used as an *environment map*, such a panorama lights a virtual object as if it stood in the photographed scene. Since Debevec's work on rendering synthetic objects into real scenes [Deb98; Deb02], image-based lighting has become a standard component of rendering systems, from film production and real-time engines to material design and inverse rendering.

This thesis is concerned with how an environment map should be *represented*. The question is partly one of storage, but it is also one of what kind of object an environment map is. It is a function that assigns radiance to directions, and its natural domain is the sphere of directions rather than the rectangle in which it is usually stored. Every raster used to store it, whether a latitude–longitude panorama, a cube map, or an octahedral map, flattens the sphere with some distortion: it samples some directions more densely than others and introduces seams, folds, or singular points. These are properties of the storage format, not of the illumination, yet they affect every operation performed on the raster. A representation defined on directions avoids them by construction. It can be evaluated in any direction, rasterized into any parameterization at any resolution, and compared with the target through integrals over the sphere rather than sums over pixels. Figure 1.1 illustrates this change of representation.

Such a representation is difficult to design because of the structure of HDR illumination. A typical environment combines content at very different spatial scales and intensities. A clear sky varies slowly over most of the sphere, while the solar disk subtends about half a degree and is several orders of magnitude brighter than the sky around it. An interior is lit by a few windows and luminaires with sharp, often rectilinear boundaries, surrounded by large regions of dim reflected light that still carry visible texture. The brightest directions dominate the energy that reaches a surface and determine the shape of shadows and highlights;

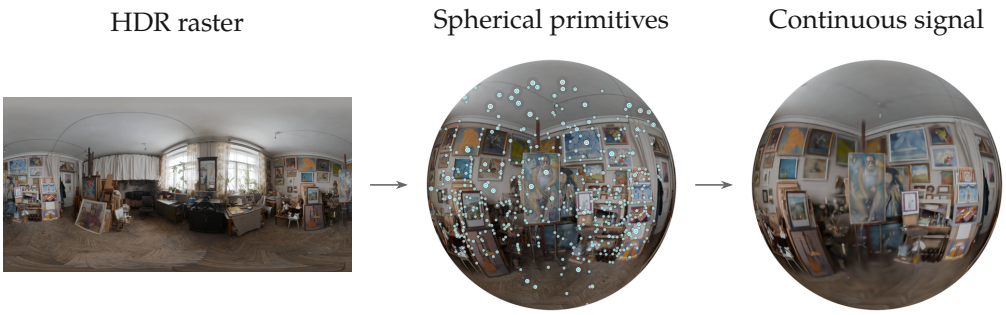


Figure 1.1. ***From an HDR raster to a spherical signal.*** *The input panorama stores samples of the illumination. The fitted representation places localized primitives directly on the sphere; the middle view marks a subset of their centers. Their positive sum can be evaluated in any direction, here on the visible hemisphere of the reconstruction. Artist Workshop environment, display-mapped for viewing.*

the dimmer structure is what a viewer sees in glossy and mirror-like reflections. A compact representation has to serve both.

Uniform representations handle this heterogeneity poorly. A raster spends as many values on a featureless patch of sky as on the edge of the sun, and a global frequency basis such as spherical harmonics represents the smooth part efficiently but needs a very large number of coefficients to reproduce a small, bright source without ringing [RH01; NRH03]. An alternative is a set of *explicit primitives*: localized lobes with their own positions, orientations, and sizes, optimized so that capacity moves to wherever the signal requires it. In the image domain, adaptively allocated two-dimensional Gaussians have proved very effective [Zha+24; Zha+25]. On the sphere, the natural primitives are spherical Gaussians, compact and easily rotated lobes with closed-form products and integrals [Gre+06; Wan+09; Zha+21], and their anisotropic extension, the anisotropic spherical Gaussian (ASG) of Xu et al. [Xu+13], which gives each lobe an orientation and independent widths along two tangent axes.

Anisotropy sets the orientation and aspect ratio of a lobe but not its shape. Every ASG falls off as a Gaussian from its center and has elliptical contours: a rounded peak, smooth tails, and no plateau or well-defined edge. Many of the most important structures in an environment map look different. A diffuser panel or the solar disk is nearly uniform across its interior and ends abruptly, and a fluorescent fixture or a window seen from inside has a boundary closer to a rounded rectangle than to an ellipse. A Gaussian mixture approximates such a structure with several overlapping lobes, some building up the flat interior and others filling the corners, all of them spreading their tails beyond the true boundary. Since every lobe adds nonnegative radiance, the light that spills over the boundary cannot be removed by the others, and primitives are spent on correcting the kernel shape rather than on describing the scene.

The representation developed in this thesis keeps what makes the ASG attractive and relaxes what makes it rigid. An environment is approximated by a positive sum of localized primitives on the sphere, each with a center direction, a tangent orientation, two concentrations that set its extent, and an RGB coefficient, exactly as in an ASG mixture. In addition, each primitive has two *shape exponents*. The first changes the radial falloff, moving it continuously from the Gaussian toward a flat-topped profile with a sharper edge; the second changes the contours, from diamond-like shapes through ellipses to increasingly rectangular ones. One choice of the exponents gives exactly the conventional ASG. The family is normalized so that changing the shape of a primitive changes neither its integrated radiance nor its spatial extent: the coefficient and the concentrations keep the meaning they have for the ASG, and the exponents describe shape alone. Every primitive remains explicit, with a position, an orientation, a size, a color, and a shape that can be inspected, quantized, or evaluated on its own.

The thesis investigates whether this freedom increases what a mixture can represent, both per primitive and per byte of storage. Because the ASG belongs to the family, the question can be answered under controlled conditions. An ASG mixture and a shape-adaptive mixture are fitted to the same environment by the same procedure and compared at equal primitive count and as a function of storage, so that differences in quality come from the shape of the primitives alone. The answer is positive on both counts. At equal primitive count, shape adaptation improves every mean metric and almost every individual environment, most of all those lit by bounded sources. The radial exponent mainly corrects the energy of such sources, and the contour exponent their visible outline. The advantage persists after compact encoding, when the models are compared by storage. Against an adaptive image-space representation fitted to the panorama, the spherical mixture is slightly better as an image and clearly better as lighting, and the comparison shows that the energy-weighted error on the sphere, not the image metrics of the panorama, predicts which reconstruction lights a scene more accurately.

The thesis addresses the reconstruction of a single HDR environment map under a fixed primitive budget. Learning priors over many environments, estimating lighting from photographs of objects, and deriving rendering operators for the new primitives, such as products or convolutions with reflectance lobes, are outside its scope. The reconstructed environment is evaluated directly with radiometric and perceptual measures, and indirectly through renderings of simple scenes lit by the reconstructions.

The thesis is organized as follows. Chapter 2 introduces HDR environment maps, the geometry of the sphere of directions and of the rasters that store it, and reviews the representations on which this work builds. Chapter 3 defines the shape-adaptive primitive and its normalization, and Chapter 4 describes how a mixture of such primitives is fitted to an environment and encoded. Chapter 5

evaluates the representation, Chapter 6 interprets the results and states the limitations of the work, and Chapter 7 concludes. Appendix A collects implementation and experimental details, and Appendix B per-scene results.

1.1 Contributions

- A *shape-adaptive spherical primitive* for positive mixtures of HDR illumination. It extends the anisotropic spherical Gaussian with independent exponents for radial falloff and contour geometry, contains the ASG exactly, and is normalized so that its mass and second moments do not depend on its shape. Reading the ASG as a planar footprint lifted to the sphere motivates the construction and yields exact statements about the mass, mean direction, and principal axes of every primitive (Chapter 3).
- A *sphere-native reconstruction procedure* that fits the representation to an HDR environment through solid-angle quadrature, a perceptual HDR objective, and residual-guided allocation of primitives, and that fits the fixed-shape and shape-adaptive models identically. A compact fixed-rate encoding makes the storage cost of shape explicit (Chapter 4).
- An *experimental study* of what shape adds: reconstruction quality at matched primitive count and after compact encoding, the separate roles of the two exponents, how the fitted primitives use their shape, a comparison with an adaptive image-space representation, the effect of reconstruction differences on rendered appearance, and an analysis of which measures of the map predict that effect (Chapter 5).

Chapter 2

Background & Related Work

This chapter introduces the signal studied in this thesis and the work on which it builds, in two parts. Sections 2.1 and 2.2 contain foundational material: what an environment map measures, how it is used in rendering, and the geometry of the sphere of directions on which it is defined and of the rasters in which it is stored. These results are standard; they are stated and explained rather than proved, and they fix the notation used throughout. Sections 2.3 to 2.7 review related work and define each representation where it first appears. They trace the development of representations for such signals, from fixed bases through spherical Gaussian lobes to the adaptive primitives and neural representations of recent years, and conclude by placing the thesis among them.

2.1 HDR Environment Maps and Image-Based Lighting

2.1.1 Radiance and Solid Angle

Light arriving at a point comes from a set of directions, and the natural measure of the size of such a set is the *solid angle*. It generalizes the planar angle: just as an angle in radians is the length of the arc it cuts from a unit circle, the solid angle of a cone of directions is the area it cuts from the unit sphere, measured in steradians (sr). The whole sphere subtends 4π sr, a hemisphere 2π sr. A small surface patch of area $\mathrm{d}A$ at distance r, whose normal makes an angle β with the line of sight, subtends the solid angle $\mathrm{d}\omega = \cos\beta\,\mathrm{d}A/r^2$: the farther away or the more obliquely a surface is seen, the smaller the solid angle it occupies.

The quantity that describes light traveling along a ray is *radiance*, the power per unit projected area and per unit solid angle, in $\mathrm{W\,m^{-2}\,sr^{-1}}$ [PJH23]. Radiance is constant along a ray in empty space, which makes it the quantity that cameras measure and renderers propagate. Color images store radiance in a small number of spectral bands; throughout this thesis these are the three linear RGB channels.

2.1.2 Distant Illumination

An environment map describes the radiance arriving at a point $\mathbf{p}$ from every direction. In general this incident radiance depends on both the point and the

direction. If the sources of illumination are far away compared with the size of the objects being lit, the dependence on the point can be neglected: every point sees the same surroundings in the same directions. The incident radiance is then a function of direction alone,

$$L : \mathbb{S}^2 \to \mathbb{R}^3_+, \qquad \omega \mapsto L(\omega), \tag{2.1}$$

where ω is the direction *from which* light arrives, a unit vector on the sphere $\mathbb{S}^2$. This *distant-illumination* assumption underlies environment mapping. It ignores parallax, so a panorama captured at one point is exact only at that point, but for objects that are small compared with their surroundings it is an excellent approximation.

Everything a renderer needs from the environment is an integral or a lookup of L. The light reflected toward a direction ω_o by a surface point with normal $\mathbf{n}$ and reflectance f_r is given by the reflection equation,

$$L_o(\omega_o) = \int_{\mathbb{S}^2} f_r(\omega, \omega_o)\, L(\omega)\, V(\omega)\, (\omega \cdot \mathbf{n})_+ \, \mathrm{d}\omega, \tag{2.2}$$

a special case of the rendering equation [Kaj86] in which all incident light comes from the environment. Here V is the visibility of the environment in direction ω, and $(\tau)_+ = \max(\tau, 0)$. A perfectly diffuse surface integrates L against a broad cosine lobe, so its appearance depends mostly on the low-frequency content of the environment. A mirror reads L in a single direction and reveals every detail. Glossy materials lie in between, and cast shadows depend on where the bright directions are and how large they are. An environment representation is therefore judged by how well it serves this whole range of uses, not by a single criterion.

One further quantity will be used to interpret the coefficients of a mixture model: the integral of the radiance over all directions,

$$\int_{\mathbb{S}^2} L(\omega)\, \mathrm{d}\omega, \tag{2.3}$$

known in radiative transfer as the fluence rate. We call it the *integrated radiance*. It summarizes the radiance of an environment, or of a part of it, without the cosine weighting of irradiance, and like L it has one value per color channel.

2.1.3 High-Dynamic-Range Capture and Image-Based Lighting

Environment mapping was introduced as a way of rendering reflections by looking up a precomputed image of the surroundings in the reflected direction [BN76]. Greene generalized the idea to other world projections, most notably the cube map, and discussed how such maps should be filtered [Gre86]. Environment maps became measurements of real illumination with the capture of high-dynamic-range radiance. The radiance of real environments spans an enormous range: in

a daylit interior, the view through a window can be hundreds of times brighter than a shadowed wall, and outdoors the solar disk is several orders of magnitude brighter than the surrounding sky. An ordinary photograph clips this range, and a clipped sun would light a scene with only a small fraction of its true radiance. HDR environment maps are therefore captured by merging several exposures of each view into a single radiance estimate [DM97] and stored in floating-point formats such as RGBE [War91] or OpenEXR. The result is a *scene-referred* image whose values are proportional to physical radiance and have no fixed upper bound. Debevec introduced the use of such maps to light synthetic objects, known as image-based lighting [Deb98; Deb02]; Reinhard et al. give a broader treatment of HDR imaging [Rei+10].

The dynamic range is also what makes HDR environments demanding to represent and to compare. The brightest directions dominate the energy that reaches a surface, while the dim structure is what a viewer sees in reflections, and an error measured in linear radiance is dominated by the former. How reconstruction error is weighted across this range is therefore a design decision in its own right. It is taken up with the fitting objective in Chapter 4 and with the evaluation protocol in Chapter 5.

2.2 Spherical Signals and Their Parameterizations

2.2.1 Directions, Measure, and Rotations

A direction is a unit vector $\omega \in \mathbb{S}^2$. Throughout the thesis, the world frame is right-handed with $+z$ pointing up, and a direction is written in spherical coordinates as

$$\omega(\theta, \phi) = (\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta), \qquad \theta \in [0, \pi],\ \phi \in [0, 2\pi), \tag{2.4}$$

with the polar angle θ measured from $+z$ and the azimuth ϕ measured from $+x$. The solid-angle element in these coordinates is

$$\mathrm{d}\omega = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi. \tag{2.5}$$

The factor $\sin\theta$ matters for much of what follows. A band of fixed angular height $\mathrm{d}\theta$ is a wide ring near the equator but shrinks to a point at the poles, so equal steps in θ and ϕ do not cover equal amounts of the sphere. Integrals on the sphere are taken with respect to $\mathrm{d}\omega$. They express the integrated radiance of an environment, the irradiance on a surface, and the error of a reconstruction, and they make solid angle, not pixel count, the right way to weigh directions.

The sphere has no preferred point and no preferred direction: it looks the same after any rotation. A rotation R acts on a function by $f \mapsto f \circ R^{-1}$, which turns the environment without otherwise changing it. A representation defined in terms of directions and dot products between them inherits this symmetry, since rotating

the environment amounts to rotating the parameters that describe directions. A representation defined on a particular raster does not, because the raster has distinguished points such as its poles, seams, and corners.

2.2.2 Tangent Frames and Local Charts

Much of the geometry of localized functions on the sphere takes place near a single direction $\mathbf{z}$. The plane through $\mathbf{z}$ perpendicular to it, the *tangent plane*, touches the sphere at $\mathbf{z}$ and contains all directions of motion along the sphere at that point. A *local frame* at $\mathbf{z}$ is a right-handed orthonormal basis $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ in which $\mathbf{x}$ and $\mathbf{y}$ span the tangent plane and $\mathbf{z} = \mathbf{x} \times \mathbf{y}$. It provides the axes along which a function centered at $\mathbf{z}$ can be stretched or oriented.

For a given $\mathbf{z}$, the tangent axes are determined up to a rotation about $\mathbf{z}$. They cannot be chosen as a continuous function of $\mathbf{z}$ over the whole sphere: by the hairy-ball theorem, every continuous tangent vector field on $\mathbb{S}^2$ vanishes somewhere. Practical constructions therefore build a frame from $\mathbf{z}$ with a formula that is continuous almost everywhere and discontinuous along a curve or at a point. The construction of Frisvad [Fri12] and its numerically robust revision by Duff et al. [Duf+17] are standard examples in rendering. Once such a reference frame is fixed, any other frame at $\mathbf{z}$ is described by a single angle of rotation about $\mathbf{z}$.

Given a frame, a direction ω near $\mathbf{z}$ can be described by its angular distance $\theta = \arccos(\omega \cdot \mathbf{z})$ from the center and an azimuth ψ about $\mathbf{z}$, measured from $\mathbf{x}$. Planar coordinates require flattening the neighborhood of $\mathbf{z}$ onto the tangent plane, and, as with maps of the Earth, there are several ways to do so. The *gnomonic* projection maps ω to the point where the ray through ω meets the tangent plane; it is the projection used by each face of a cube map. The *azimuthal equidistant* projection preserves angular distance from the center. The *orthographic* projection drops ω perpendicularly onto the tangent plane, giving the coordinates

$$(s, t) = (\omega \cdot \mathbf{x},\ \omega \cdot \mathbf{y}) = (\sin\theta\cos\psi,\ \sin\theta\sin\psi). \tag{2.6}$$

It maps the hemisphere in front of $\mathbf{z}$ one-to-one onto the unit disk, and its area element is related to the solid angle by

$$\mathrm{d}s\,\mathrm{d}t = (\omega \cdot \mathbf{z})\,\mathrm{d}\omega. \tag{2.7}$$

The factor $\omega \cdot \mathbf{z} = \cos\theta$ is the familiar foreshortening of a hemisphere seen from above: near the center, planar area and solid angle agree, while toward the horizon a large solid angle is compressed into a thin ring at the edge of the disk. The three projections place a direction at distances $\sin\theta$, θ, and $\tan\theta$ from the center of the chart, and these differ only at third order in θ, so near the center the projections nearly coincide. The choice between them matters little for concentrated lobes and determines how broad lobes are distorted. Anisotropic spherical Gaussians

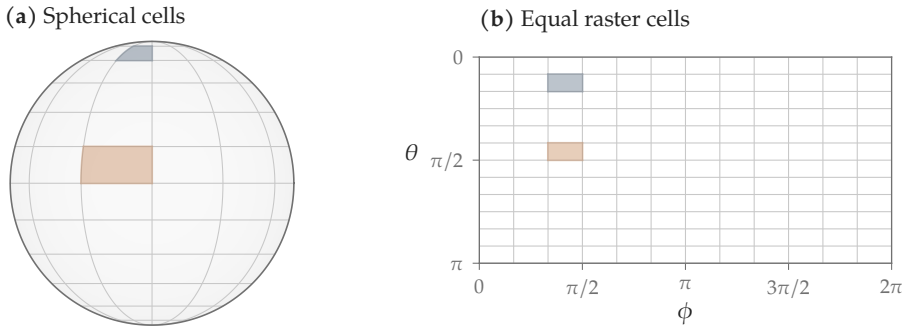


Figure 2.1. ***Equal raster cells cover unequal solid angles.*** *Both highlighted cells span* $\Delta\theta = 15°$ *and* $\Delta\phi = 30°$ *and therefore have the same area in the latitude–longitude raster* (*b*)*. On the sphere* (*a*)*, the blue cell* ($\theta \in [15°, 30°]$) *covers about* 0.39 *times the solid angle of the orange cell* ($\theta \in [75°, 90°]$)*, by Equation* (2.8)*.*

(Section 2.4) are defined in orthographic coordinates, and the same chart is central to Chapter 3, where it is illustrated in Figure 3.1.

2.2.3 Raster Parameterizations

To be stored as an image, a spherical function must be mapped onto a planar domain. Such a map is a *parameterization*, and because the sphere cannot be flattened without distortion, every parameterization trades off uniformity, continuity, and simplicity. Three are common for environment maps; Figure 2.2 shows one environment stored in each.

The *latitude–longitude*, or equirectangular, map uses the spherical angles directly as image coordinates: a texel in row i and column j of a $W \times H$ image covers the angular rectangle $\theta \in [i\pi/H, (i+1)\pi/H]$, $\phi \in [2\pi j/W, 2\pi(j+1)/W]$. Its solid angle follows from Equation (2.5) as

$$\Omega_{ij} = \frac{2\pi}{W}\left(\cos\theta_i - \cos\theta_{i+1}\right), \tag{2.8}$$

which is proportional to $\sin\theta$ for small texels. The map is simple and intuitive, and it is the usual format for capturing and distributing panoramas. It is also the most distorted of the three. Texels at the equator cover about $\pi/2$ times the mean texel solid angle, while texels near the poles cover almost nothing, so an entire row of texels is spent on a tiny cap around each pole. The left and right edges of the image are adjacent on the sphere and form a seam. Figure 2.1 compares equal raster cells with the solid angles they cover.

A *cube map* projects the sphere gnomonically onto the six faces of a cube centered at the viewer [Gre86]. It has no singular points and is supported directly by graphics hardware, but the solid angle of a texel varies across each face, by a factor of $3\sqrt{3} \approx 5.2$ between the center and the corners, and the faces meet along twelve edges. An *octahedral* map projects the sphere onto an octahedron by normalizing

(**a**) Latitude–longitude (**b**) Cube map (**c**) Octahedral map



Figure 2.2. ***One environment in three rasters.*** (***a***) *Latitude–longitude: rows follow the polar angle and columns the azimuth, so the top and bottom rows each collapse to a single direction.* (***b***) *Cube map, unfolded into a cross; lines mark the face edges.* (***c***) *Octahedral map: the upper hemisphere, here the ceiling, fills the central diamond, and the lower hemisphere is folded into the corners.*

directions in the ℓ_1 norm, $\omega \mapsto \omega/(|\omega_x| + |\omega_y| + |\omega_z|)$, and unfolds the octahedron onto a single square: the upper hemisphere fills a central diamond and the lower hemisphere is folded into the four corners [PH03]. Its texel solid angles vary by the same factor of about five, but the whole sphere is stored in one square image without singular points, which also makes the octahedral mapping a popular compact encoding of unit vectors [Cig+14]. Where uniform sampling matters more than convenience, equal-area pixelizations such as HEALPix [Gór+05] divide the sphere into cells of exactly equal solid angle, at the cost of a less regular image layout.

These properties belong to the storage format, not to the illumination. Two rasters of the same environment in different parameterizations describe the same function but sample it differently. Any operation that treats texels as equal, such as a mean squared error over pixels, a pixel-space filter, or a representation fitted in image coordinates, implicitly weights directions by the texel density of the chosen parameterization.

2.2.4 Quadrature and Error on the Sphere

Integrals over the sphere are computed numerically by *quadrature*, which approximates the integral by a weighted sum over sample directions,

$$\int_{\mathbb{S}^2} f(\omega)\,\mathrm{d}\omega \;\approx\; \sum_j \Omega_j\, f(\omega_j), \qquad \sum_j \Omega_j = 4\pi, \tag{2.9}$$

where each weight Ω_j is the solid angle of the cell that the sample ω_j represents. A raster already defines such a rule: the samples are the texel centers and the weights are the texel solid angles, as in Equation (2.8). Replacing the weights by a constant turns the sum into an integral with respect to a distorted measure, the sampling density of the parameterization, which is exactly the implicit reweighting described above.

Quadrature rules that do not depend on a raster can be built directly on the sphere. A widely used construction starts from a regular icosahedron, subdivides each of its twenty triangular faces recursively, and projects the new vertices onto the sphere. The result is a *geodesic grid* whose cells are nearly uniform in size and shape, with no poles or seams. Taking one sample per spherical triangle and weighting it by the triangle's exact solid angle gives a quadrature rule whose weights sum to exactly 4π. Such grids can also be refined locally, by subdividing only the cells in which the integrand varies strongly. Monte Carlo integration, with directions drawn uniformly on the sphere and each sample weighted by 4π divided by the number of samples, is the stochastic counterpart.

The same reasoning applies to the error of a reconstruction. The discrepancy between a reconstruction $\widehat{L}$ and the target L is naturally measured by an integral over directions, such as the squared error $\int_{\mathbb{S}^2} \left\| \widehat{L}(\omega) - L(\omega) \right\|^2 \, \mathrm{d}\omega$, evaluated with a quadrature rule. This measure does not depend on any parameterization, whereas the mean squared error over the texels of a raster is the same integral taken with respect to the raster's sampling density. Both the fitting procedure of Chapter 4 and the radiometric error of Chapter 5 are built on geodesic quadrature of this kind.

2.3 Representations of Spherical Illumination

The raster is itself the simplest representation of an environment. It stores samples at fixed directions with uniform capacity: storage grows with the square of the angular resolution regardless of content, and the samples inherit the uneven density and the seams of the parameterization. Compression reduces the storage cost but keeps the raster as the underlying representation. A more compact representation expresses the function in terms of a set of basis functions or primitives, and different choices trade off compactness, locality, and ease of use in rendering.

The spherical harmonics $Y_{\ell m}$ are the analogue of the Fourier basis on the sphere. Any square-integrable function can be expanded as

$$f(\omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{\ell m} \, Y_{\ell m}(\omega), \tag{2.10}$$

where the band index ℓ plays the role of frequency: a band-ℓ harmonic varies on angular scales of roughly π/ℓ. Truncating the expansion after band n gives the best least-squares approximation with $(n+1)^2$ coefficients, and rotations mix coefficients only within each band. Spherical harmonics made real-time diffuse lighting from environment maps practical: nine coefficients describe the irradiance of a Lambertian surface accurately, because the cosine lobe in Equation (2.2) suppresses high frequencies [RH01]. Precomputed radiance transfer extended the idea to shadows and interreflections under low-frequency lighting by representing

both lighting and transfer in the same basis [SKS02]. Spherical harmonics are poorly suited, however, to reproducing the environment itself. A small bright source contains all frequencies, the number of coefficients grows quadratically with the required angular resolution, and a truncated expansion rings around sharp features and can even become negative, which is physically meaningless for radiance.

Localized bases address this by combining frequency with position. Each wavelet is supported on a region of the sphere whose size depends on its level, so a sharp feature affects only a few coefficients. Spherical wavelets can be built on subdivided polyhedra [SS95], and Haar wavelets on the faces of a cube map have been used to keep only the largest coefficients of each environment, a nonlinear approximation that adapts to the content and supports all-frequency shadows [NRH03]. Spherical piecewise constant basis functions partition the sphere into regions and admit fast rotations and products [Xu+08]. The positions and sizes of these basis functions remain tied to a fixed hierarchy: the representation adapts by selecting coefficients, not by moving its functions.

A different approach represents a function as a weighted sum of a small number of localized kernels whose own parameters are free:

$$f(\omega) \approx \sum_{i=1}^{N} \mathbf{a}_i \, K(\omega; \vartheta_i), \tag{2.11}$$

where ϑ_i collects the center, width, and possibly orientation of kernel i, and $\mathbf{a}_i$ is its weight. Unlike a basis expansion, this approximation is nonlinear in the kernel parameters, so fitting it is an optimization problem rather than a projection, but the kernels can move to wherever they are needed. *Spherical radial basis functions* are kernels of this kind that depend only on the angle between ω and a center $\mathbf{z}$; Tsai and Shih combined them, with movable centers and bandwidths, with a tensor approximation of the transfer [TS06].

These representations were designed to make a particular light-transport computation cheap, such as diffuse shading or visibility-weighted integrals, and their accuracy is measured through that computation rather than through the environment itself. A basis that reproduces diffuse shading perfectly may still render a poor mirror reflection. This thesis instead fits the incident radiance directly, so that the result represents the environment itself and can be queried in any direction or rasterized into any projection.

2.4 Spherical Gaussian Representations

2.4.1 Isotropic Lobes

The most widely used localized kernel on the sphere is the *spherical Gaussian* (SG),

$$G(\omega; \mathbf{z}, \kappa) = \exp\big(\kappa\,(\omega \cdot \mathbf{z} - 1)\big), \tag{2.12}$$

with center $\mathbf{z}$ and sharpness $\kappa > 0$. It equals one at the center and decays with angular distance. Near the center, $\kappa(\cos\theta - 1) \approx -\kappa\theta^2/2$, so the SG behaves like a Gaussian in the angle θ with variance $1/\kappa$, and its level sets are circles around the center. Normalized to unit integral, it is the density of the von Mises–Fisher distribution, the basic model of a concentrated distribution of directions in directional statistics [Fis53; MJ00].

SGs have properties that make them attractive in rendering [Wan+09]. They rotate by rotating their center, their integral over the sphere is available in closed form, $\int_{\mathbb{S}^2} G\,\mathrm{d}\omega = 2\pi(1 - e^{-2\kappa})/\kappa$, and the product of two SGs is again an SG, so that integrals of products such as those in Equation (2.2) can be approximated analytically. Mixtures of Gaussian lobes were used early on to approximate directional reflectance and transfer compactly [Gre+06], and Wang et al. built an all-frequency rendering system for dynamic, spatially varying reflectance on the closed-form rotation, product, and integration rules of SGs [Wan+09]. The same algebraic convenience has made SG mixtures a popular illumination model in inverse rendering. PhySG represents both environment lighting and reflectance with SGs, so that rendering remains analytic and differentiable during the joint estimation of geometry, material, and lighting [Zha+21], and NeRD uses an SG mixture for the illumination of each image in a collection [Bos+21]. In these settings the mixture is a compact lighting model with a small number of lobes, not a faithful reconstruction of a given HDR environment.

How the lobes of a mixture are fitted depends on what the mixture represents. For a probability distribution of directions, expectation maximization estimates centers, concentrations, and weights from samples [Han+07]. For a function, such as a transfer function or an environment, the parameters are found by nonlinear least-squares fitting to sampled values [Gre+06; TS06], or, in inverse rendering, by gradient descent through a differentiable renderer [Zha+21; Bos+21]. For a fixed set of lobe geometries the weights enter linearly and can be estimated by linear, or nonnegative, least squares. The difficulty lies in the geometry, which enters nonlinearly and makes the fit sensitive to initialization. With a few dozen lobes this is manageable. A detailed HDR environment requires thousands of lobes, and deciding where to place them becomes as important as refining them.

Isotropic lobes are well suited to point-like features. Their circular symmetry is a limitation for elongated features such as strip lights, window frames, or a bright horizon, which must be covered by chains of overlapping lobes.

2.4.2 Anisotropic Lobes

The *anisotropic spherical Gaussian* (ASG) of Xu et al. [Xu+13] gives up radial symmetry to represent elongated lobes. It uses a local frame $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ at the center and two positive bandwidths λ and μ, which we call *concentrations*:

$$G_{\mathrm{ASG}}(\omega) = C\ (\omega \cdot \mathbf{z})_+\ \exp\!\big(-\lambda(\omega \cdot \mathbf{x})^2 - \mu(\omega \cdot \mathbf{y})^2\big), \tag{2.13}$$

where $C > 0$ is the amplitude of the lobe. In the orthographic coordinates of Equation (2.6), the exponential is a Gaussian $\exp(-\lambda s^2 - \mu t^2)$ with elliptical level sets aligned with the frame. Larger concentrations give narrower lobes, and unequal concentrations give an elongated lobe whose long axis is set by the orientation of the frame. Xu et al. build the ASG on the Bingham distribution, whose density is the exponential of a quadratic form in the direction and is therefore antipodally symmetric [Bin74]; the clamped cosine removes the antipodal copy and confines the lobe to the hemisphere around its center. The Kent distribution adds an elliptical quadratic term of the same kind to the von Mises–Fisher exponent [Ken82]. Xu et al. also derive approximate closed-form expressions for the integral of an ASG, for products, and for convolution with other lobes, which make it usable for rendering anisotropic reflectance and lighting.

ASG mixtures have also been fitted directly to captured illumination. Clausen et al. [Cla+24] approximate each frame of a sequence of HDR environment maps with a fixed number of ASGs, between 9 and 60 in their experiments, by gradient descent on two solid-angle-weighted ℓ_1 terms: one compares the radiance itself, the other a low-order spherical-harmonic approximation of the diffuse shading, which keeps the total energy correct. A temporal term ties the lobes of consecutive frames together. With about 15 lobes, renderings of spheres across most roughness levels come close to the reference, but the authors report that mirror reflections lose fine detail and that sharp-cornered lights, such as rectangular ones, are not represented accurately.

The ASG is the starting point of this thesis. It lets a lobe adapt its position, orientation, extent, and aspect ratio to a feature, but, like every lobe discussed so far, not its shape: its radial falloff is Gaussian and its contours are ellipses, whatever the feature it describes.

2.5 Adaptive Primitive Representations

2.5.1 Gaussian Primitives for Images and Scenes

Outside the sphere, explicit primitives have become one of the most successful ways of representing visual signals. Most of them rest on the anisotropic Gaussian in the plane or in space, whose parameters have a direct geometric meaning. Normalized to unit integral, its covariance is exactly its matrix of second moments, so the covariance describes the spatial extent of the kernel, and its factorization into

a rotation and two principal variances separates orientation from size. A closed-form normalization, geometrically meaningful parameters, and cheap evaluation make Gaussian kernels the default choice for explicit primitive representations.

Anisotropic Gaussian kernels became a central tool in graphics through point-based rendering, where elliptical weighted average splatting reconstructs surfaces from anisotropic Gaussian footprints [Zwi+01]. 3D Gaussian splatting revived the idea for radiance fields, representing a scene by millions of optimized anisotropic 3D Gaussians that are projected and alpha-composited into images [Ker+23]. Its success depends as much on the allocation of primitives as on their optimization. A gradient step can move a primitive but cannot create one where none exists, so the method periodically clones, splits, and prunes primitives according to their gradients and opacities.

In the image domain, GaussianImage represents an image by a set of anisotropic 2D Gaussians whose color contributions are accumulated without depth sorting, and develops a codec for them with quantization-aware fine-tuning and vector quantization [Zha+24]. Image-GS represents images with colored anisotropic 2D Gaussians and allocates them adaptively [Zha+25]. Gaussians are initialized according to image gradients and added progressively where the reconstruction error is largest, and each pixel is rendered as a weighted average of a few nearby Gaussians, normalized by their total weight. The progressive optimization yields a level-of-detail hierarchy, and the local decoding permits fast random access. Both methods show that explicit Gaussians, allocated according to image content, give compact and quickly decodable image representations.

Applied to an environment map, however, these methods operate on the raster rather than on the sphere. The image plane of a panorama has a seam that the representation does not know about, and the texels of a latitude–longitude map cover solid angles that differ without bound between the equator and the poles. An unweighted image loss therefore spends primitives where the parameterization is dense rather than where the illumination is complex. Both methods were also developed for standard images with bounded values, so HDR radiance must first be mapped into a bounded range. Image-GS is the closest adaptive image-space counterpart of the representation developed here, and Chapter 5 uses it, fitted to the latitude–longitude panorama, as the image-space reference.

2.5.2 Relaxing the Gaussian Kernel

Relaxing the Gaussian shape is an old idea outside rendering. The generalized normal, or exponential-power, distribution replaces the squared exponent of the Gaussian by an arbitrary power, interpolating between peaked, Laplace-like profiles and flat-topped, nearly uniform ones [Nad05]. Superellipses and their three-dimensional generalization, the superquadrics introduced to graphics by Barr, use an ℓ_q-type exponent to move shapes continuously between diamonds, ellipses, and rounded rectangles [Bar81].

The success of Gaussian splatting prompted several attempts to relax its kernel, motivated by the observation that sharp edges and flat regions require many Gaussians. Generalized exponential splatting replaces the Gaussian by a generalized exponential function with a learnable shape parameter [Ham+24]. Deformable radial kernel splatting learns directional radial bases and boundary controls for each primitive [Hua+25]. 3D convex splatting uses smooth convex shapes with learnable smoothness and sharpness [Hel+25], and deformable Beta splatting uses bounded-support Beta kernels whose shape is controlled by a learnable parameter [Liu+25]. These methods show that primitive shape is a useful degree of freedom. Their kernels, however, are designed for the projection and alpha compositing of three-dimensional primitives, not for a positive superposition of radiance on the sphere.

2.6 Neural Representations of Illumination

Implicit neural representations encode a signal in the weights of a network that maps coordinates to values. Neural radiance fields established the approach for scenes [Mil+20], and networks with sinusoidal activations showed that such representations can capture fine detail in images and other signals [Sit+20]. Two lines of work are relevant to environment illumination. RENI learns a rotation-equivariant prior over a corpus of outdoor HDR environments with conditional spherical neural fields; a low-dimensional latent code then describes an environment, which supports generation, completion, and lighting estimation in inverse rendering [GES22]. NEnv fits a single environment map with two networks, a normalizing flow used for importance sampling and an implicit network that reconstructs the radiance [Rod+23].

Neural models and explicit primitive mixtures make different trade-offs. A learned prior can regularize ill-posed estimation problems that a per-map fit cannot, and a network can represent complex texture without allocating primitives explicitly. An explicit mixture keeps the location, extent, orientation, color, and shape of every primitive directly accessible: it can be edited, quantized field by field, evaluated one primitive at a time, and fitted without training data. RENI depends on a training corpus and NEnv on network decoders, so neither can be compared with an explicit mixture at an equal number of primitives; both are discussed here as related work.

2.7 Positioning of This Work

Table 2.1 summarizes the representations of this chapter by what they let adapt to the content of a signal. Read from top to bottom, it shows two lines of development that have remained largely separate. On the sphere, lobe mixtures were designed mainly as tools for light-transport computations, usually with few lobes, and their coefficients and widths keep a direct physical and geometric meaning. In the

Table 2.1. ***Representation families and what they adapt.*** *Families are characterized by their design, not by the quality of particular methods. Within each group, a plus sign marks what a family lets adapt in addition to the row above it.*

Family	References	Element	Adapts to content
Rasters and fixed bases			
Raster maps	[Gre86; PH03]	planar texel	nothing
Spherical harmonics	[RH01; SKS02]	global harmonic	truncation band
Wavelets, SPCBF	[NRH03; Xu+08]	fixed hierarchy	selected coefficients
Lobe mixtures on the sphere			
SRBF, SG mixtures	[TS06; Wan+09]	isotropic lobe	centers, widths, weights
ASG mixtures	[Xu+13; Cla+24]	elliptical lobe	+ orientation, anisotropy
Primitives in the plane and in space			
2D Gaussian images	[Zha+24; Zha+25]	planar Gaussian	+ error-driven allocation
Relaxed splatting	[Ham+24; Liu+25]	non-Gaussian	+ kernel shape
Learned representations			
Neural fields	[GES22; Rod+23]	network	learned weights
This thesis		adaptive lobe	allocation, pose, extent, shape

image plane and in three-dimensional radiance fields, representations allocate many primitives according to the content and, more recently, let the kernel shape itself adapt. Among the representations reviewed here, no spherical lobe mixture lets the shape of its lobes adapt: all keep a Gaussian profile and elliptical contours.

This thesis brings the two lines together on the sphere. Its representation is a positive mixture of primitives defined directly on directions and fitted to a single HDR environment with solid-angle quadrature and residual-guided allocation, much as adaptive image representations are fitted to images. Each primitive extends the ASG with two exponents, one for the radial profile and one for the contour, much as relaxed splatting kernels extend the Gaussian. Unlike those kernels, the family is normalized so that mass and second moments do not depend on shape. The coefficient of a primitive keeps its meaning as integrated radiance, the concentrations keep their meaning as extent, and the exponents describe shape alone. The conventional ASG is the member with a Gaussian profile and elliptical contours, so a fixed-shape and a shape-adaptive mixture can share every other component of the representation and of the fitting procedure. The result is a compact, explicit, and projection-independent representation of the environment, in which every primitive has a position, an orientation, an extent, a color, and a shape of its own.

Chapter 3

Shape-Adaptive Spherical Primitives

This chapter defines the representation used throughout the thesis: a positive mixture of localized kernels on the sphere whose members adapt not only their position, orientation, and extent, but also their shape. Everything discussed here is a property of the function class itself; how its parameters are estimated from an HDR environment is the subject of Chapter 4.

3.1 Positive Mixtures on the Sphere

An HDR environment is a function $L : \mathbb{S}^2 \to \mathbb{R}^3_+$ of direction (Equation (2.1)). We approximate it by a mixture of N localized primitives of the form (2.11),

$$\widehat{L}(\omega) = \sum_{i=1}^{N} \mathbf{a}_i \, K_i(\omega), \qquad \mathbf{a}_i \in \mathbb{R}^3_+, \quad K_i : \mathbb{S}^2 \to \mathbb{R}_+, \tag{3.1}$$

in which each primitive consists of a nonnegative scalar kernel K_i, which determines where and how the primitive contributes, and a nonnegative RGB coefficient $\mathbf{a}_i$, which sets the magnitude of that contribution. An ASG mixture has this form, with each K_i an ASG.

Three choices are built into Equation (3.1). First, contributions are summed directly in linear radiance, without the opacity, compositing, or division by the total kernel weight used by splatting-based scene and image representations [Ker+23; Zha+25]. An environment map describes incident radiance without occlusion, so superposition is the natural way to combine contributions, and it makes $\widehat{L}$ linear in the coefficients. Second, the three color channels share one scalar kernel per primitive, so spatial variation of color is represented by several primitives. Third, the mixture is positive: $\widehat{L}$ is nonnegative in every direction, but primitives cannot cancel one another.

Positivity has a consequence for the shape of the kernels that is central to this thesis. Consider a bounded source with a nearly uniform interior and a sharp edge, such as a diffuser panel, a window seen from inside, or the solar disk. A

mixture of Gaussian lobes builds up the flat interior by overlapping several lobes, and every lobe that contributes near the boundary also spreads a tail beyond it. In a signed expansion this spillover could be canceled by a negative term; in a positive mixture it cannot, and the only remedy is to use narrower lobes near the boundary, which cover the interior less efficiently and must therefore be more numerous. The same holds for contours: an elliptical lobe cannot fill the corner of a rectangular source without either leaving a gap or spilling over the neighboring edges. A sharp transition that cannot be produced by subtraction has to be carried by the shape of the primitives themselves.

This suggests making the shape of the kernel adaptive, and the construction that follows is guided by what should be preserved in doing so. The representation should remain a positive superposition of kernels defined on the sphere, evaluable in any direction and rotated by rotating its frames. It should contain the ASG exactly, so that the fixed-shape model is a special case and the effect of shape can be isolated by comparison. Changing the shape of a primitive should not change its size or its brightness. And each primitive should remain explicit and cheap to evaluate, with a closed-form kernel and a small, fixed number of parameters.

3.2 The ASG as a Lifted Planar Footprint

Each primitive has a local frame $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ at its center $\mathbf{z}$ (Section 2.2.2), two concentrations $\lambda, \mu > 0$, and an RGB coefficient $\mathbf{a}$, exactly as an ASG. To see where shape can be added without disturbing these parameters, it helps to rewrite the ASG of Equation (2.13) so that their roles separate.

Rescaling the orthographic coordinates (s, t) of Equation (2.6) by the square roots of the concentrations gives the *standardized coordinates*

$$u = \sqrt{\lambda}\ \omega \cdot \mathbf{x}, \qquad v = \sqrt{\mu}\ \omega \cdot \mathbf{y}, \tag{3.2}$$

which we collect in $\mathbf{u} = (u, v)$. In these coordinates the exponent of every ASG is $-(u^2 + v^2)$, whatever its concentrations, which therefore act only as an anisotropic rescaling of the chart. With the amplitude $C = \sqrt{\lambda\mu}/\pi$, the ASG becomes

$$K_{\mathrm{G}}(\omega) = \sqrt{\lambda\mu}\ (\omega \cdot \mathbf{z})_+\ \varphi_{\mathrm{G}}(\mathbf{u}), \qquad \varphi_{\mathrm{G}}(\mathbf{u}) = \frac{1}{\pi}\, e^{-(u^2+v^2)}. \tag{3.3}$$

The function φ_{G} is the planar Gaussian density with covariance $\frac{1}{2}I$: it has unit mass and second moment $\frac{1}{2}$ along each axis. We call it the *planar footprint* of the ASG. The factor $\sqrt{\lambda\mu}$ is the Jacobian of the rescaling.

Equation (3.3) is one instance of a general construction. Any integrable planar footprint $\varphi \geq 0$ defines a spherical kernel

$$K_{\varphi}(\omega) = \sqrt{\lambda\mu}\ (\omega \cdot \mathbf{z})_+\ \varphi\big(\sqrt{\lambda}\ \omega \cdot \mathbf{x},\ \sqrt{\mu}\ \omega \cdot \mathbf{y}\big), \tag{3.4}$$

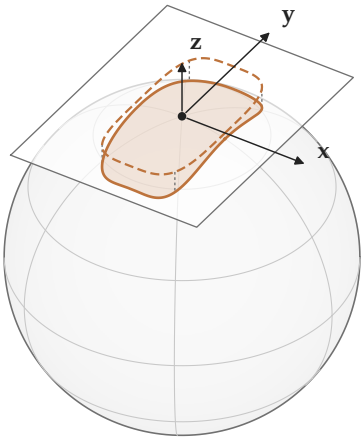


Figure 3.1. ***Lifting a planar footprint to the sphere.*** *The tangent plane at* $\mathbf{z}$ *carries the local axes* $\mathbf{x}$ *and* $\mathbf{y}$*, along which the orthographic coordinates* (s, t) *are measured. The dashed curve is the* $1/e$*-of-peak contour of the footprint* $\varphi_{2,4}$*, rescaled by* $(\lambda, \mu) = (18, 5)$*. Dotted guides parallel to* $\mathbf{z}$ *carry it onto the sphere (solid curve). The solid curve marks corresponding footprint coordinates, not a level set of the kernel, which also includes the clamped cosine.*

which we call its *orthographic lift*. The clamped cosine restricts K_φ to the hemisphere in front of $\mathbf{z}$, which the orthographic chart maps one-to-one onto the unit disk. On that hemisphere, the area element of Equation (2.7) and the rescaling (3.2) give

$$\mathrm{d}\omega = \frac{\mathrm{d}s\,\mathrm{d}t}{\omega \cdot \mathbf{z}} = \frac{\mathrm{d}u\,\mathrm{d}v}{\sqrt{\lambda\mu}\;\omega \cdot \mathbf{z}},$$

so the two factors in front of φ in Equation (3.4) are exactly the ones that cancel this change of variables. The rescaling also maps the unit disk of the chart onto the ellipse

$$\mathcal{E}_{\lambda,\mu} = \left\{\mathbf{u} \in \mathbb{R}^2 :\; u^2/\lambda + v^2/\mu < 1\right\}.$$

Integrals of the kernel against any function g of the tangent coordinates therefore reduce to planar integrals of the footprint:

$$\int_{\mathbb{S}^2} K_\varphi(\omega)\, g(\omega \cdot \mathbf{x}, \omega \cdot \mathbf{y})\, \mathrm{d}\omega = \int_{\mathcal{E}_{\lambda,\mu}} \varphi(\mathbf{u})\, g\left(u/\sqrt{\lambda},\; v/\sqrt{\mu}\right) \mathrm{d}\mathbf{u}. \tag{3.5}$$

Seen through the orthographic chart, a spherical primitive is its planar footprint, stretched by the concentrations and restricted to the image of the front hemisphere (Figure 3.1). This gives the clamped cosine a second role besides removing the antipodal copy of the even exponential: it is the Jacobian that turns a planar density into a spherical one.

Through the lift, a primitive separates into components with distinct roles. The frame fixes where the primitive is and how it is oriented, the concentrations fix its extent along the two tangent axes, the footprint fixes its shape in standardized coordinates, and the coefficient fixes the magnitude of its contribution. In the ASG

the footprint is the Gaussian φ_G, which imposes the two assumptions discussed in Section 3.1: a Gaussian radial falloff and elliptical contours. Shape adaptation, in the sense of this thesis, means replacing φ_G by a richer family of footprints while leaving the rest of Equation (3.4) unchanged. Equation (3.5) also shows what the replacement must respect for the other components to keep their meaning: the spherical mass of the primitive is the planar mass of the footprint inside $\mathcal{E}_{\lambda,\mu}$, and its spread in the chart is given by the planar second moments of the footprint, divided by λ and μ.

3.3 Adapting the Radial Profile and the Contour

The new footprint relaxes the two assumptions of the Gaussian separately, each with one exponent. A generalized radius sets the geometry of the contours, and a generalized profile applied to that radius sets how the footprint falls from its center to its tails.

For the contours, let $q > 0$ and define the q-radius

$$r_q(\mathbf{u}) = \left(|u|^q + |v|^q\right)^{1/q}, \tag{3.6}$$

which is the ℓ_q norm of $\mathbf{u}$ for $q \geq 1$. Its level sets are superellipses, the curves behind the superquadrics of Section 2.5.2. They are circles at $q = 2$; as q decreases toward one they approach a diamond, and as q grows they approach a square whose corners, on the diagonals, become increasingly sharp. The admissible range $q \in [1.25, 16]$ stays within the convex regime and keeps r_q continuously differentiable away from the origin, which the ℓ_1 norm, with its creases along the axes, would not be. Mapped back to the chart, a level set $r_q = \rho$ is a superellipse aligned with the frame, with semi-axes $\rho/\sqrt{\lambda}$ and $\rho/\sqrt{\mu}$. For $q > 2$ and unequal concentrations it is a rounded rectangle of arbitrary aspect ratio and orientation.

For the radial falloff, the Gaussian profile $e^{-\rho^2}$ is generalized by the exponential-power profile $\exp(-(b\rho^2)^p)$, the radial part of the generalized normal distribution [Nad05], in which $b > 0$ sets the scale and $p > 0$ the shape. The profile is Gaussian at $p = 1$. As p increases, it stays nearly flat for $\rho^2 < 1/b$ and drops steeply around $\rho^2 = 1/b$, and as $p \to \infty$ it tends to a flat top with a hard edge. The admissible range $p \in [1, 4]$ runs from the Gaussian toward such flat-topped profiles. It excludes $p < 1$, whose tails are heavier than Gaussian, so no admissible footprint has heavier tails than the ASG.

Applying the profile to the q-radius gives the *shape-adaptive footprint*

$$\varphi_{p,q}(\mathbf{u}) = c_{p,q}\ \exp\left[-\left(b_{p,q}\, r_q(\mathbf{u})^2\right)^p\right], \qquad (p, q) \in \mathcal{P} = [1, 4] \times [1.25, 16], \tag{3.7}$$

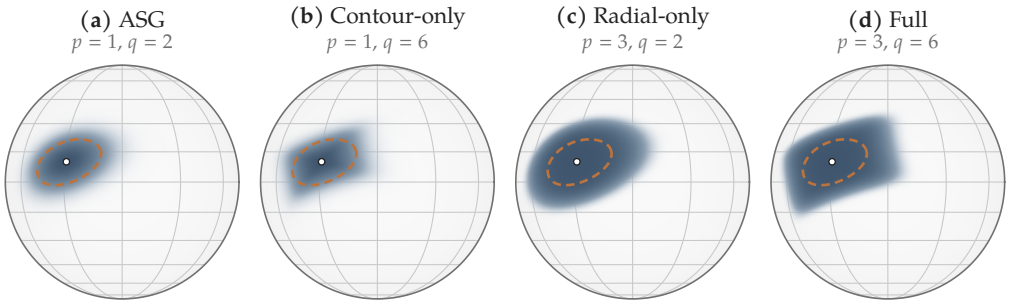


Figure 3.2. ***Shape adaptation on the sphere.*** *The four nested models with the same frame and concentrations* $(\lambda, \mu) = (5, 15)$. (***a***) *The ASG.* (***b***) *A larger contour exponent turns the contours into rounded rectangles.* (***c***) *A larger radial exponent flattens the core into a plateau with a sharper edge.* (***d***) *Both. Each kernel is shaded relative to its own peak. The dashed curve, identical in all panels, is the ellipse of one standard deviation along the frame axes: the shape changes while the second moments stay fixed.*

where the constants $b_{p,q}, c_{p,q} > 0$ depend only on the exponents and are fixed in Section 3.4. Its orthographic lift,

$$K(\omega) = \sqrt{\lambda\mu}\ (\omega \cdot \mathbf{z})_+\ \varphi_{p,q}\big(\sqrt{\lambda}\ \omega \cdot \mathbf{x},\ \sqrt{\mu}\ \omega \cdot \mathbf{y}\big), \tag{3.8}$$

is the *shape-adaptive kernel*, and it is the kernel K_i of the mixture (3.1) used throughout the thesis. Figure 3.2 shows the separate effects of the exponents on the sphere, while Figure 3.3 isolates the contour geometry and radial profile in standardized coordinates.

The two exponents act on different aspects of the footprint. Because $\varphi_{p,q}$ decreases with r_q, its level sets are exactly the superellipses $r_q = \text{const}$ and depend on q alone. The radial exponent changes no contour; it changes the value that the footprint takes on each contour, that is, how the mass is distributed across the nested superellipses. The exponents interact only through the normalization constants.

At $(p, q) = (1, 2)$ the q-radius is the Euclidean radius and the profile is Gaussian. The normalization of the next section gives $b_{1,2} = 1$ and $c_{1,2} = 1/\pi$ there, so that $\varphi_{1,2} = \varphi_{\mathrm{G}}$ and Equation (3.8) reduces to the ASG in the form (3.3). The conventional ASG is a member of the family, not an approximation to one.

3.4 Moment-Preserving Normalization

The constants $b_{p,q}$ and $c_{p,q}$ decide whether the exponents describe shape alone or also change the size and brightness of a primitive. The unnormalized footprint $\exp(-r_q^{2p})$ illustrates the problem. Its $1/e$ contour is the unit ℓ_q ball for every shape, which may look like a natural way of holding its size fixed, but its mass and its root-mean-square radius both change with the exponents. At $(p, q) = (1, 16)$ it has 27% more mass than the Gaussian and a 15% larger root-mean-square radius;

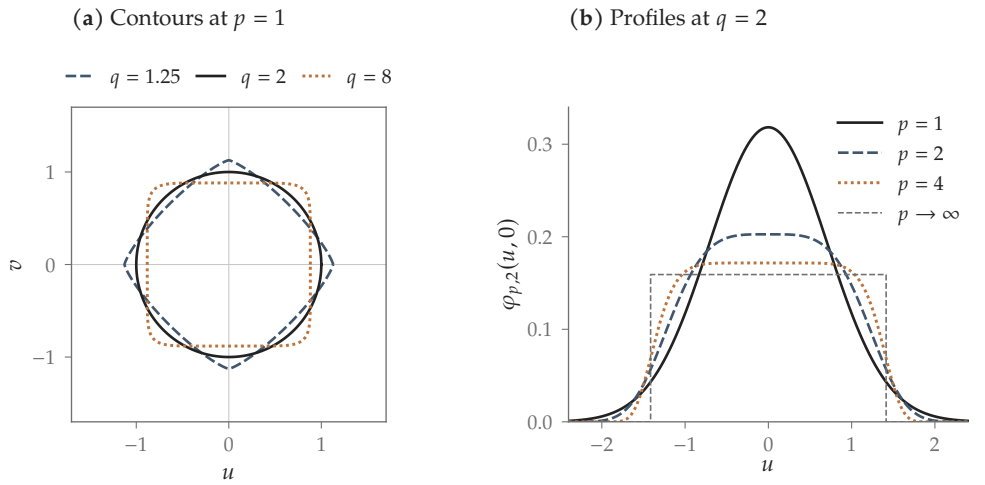


Figure 3.3. ***Contours and profiles of the normalized footprint.*** (***a***) *Contours at* $1/e$ *of the peak of* $\varphi_{1,q}$ *in standardized coordinates, isolating the effect of* q*. The circular contour at* $q = 2$ *belongs to the ASG.* (***b***) *Axial profiles of circular footprints, isolating the effect of* p*. Every footprint has unit mass and the same second moments. A flatter core forces a lower peak and a steeper shoulder; the thin gray outline is the limit* $p \to \infty$*, a uniform disk of radius* $\sqrt{2}$*.*

at $(4, 2)$ it has 9% less mass and a 30% smaller radius. Moving away from the Gaussian would then also rescale the integrated radiance and the extent of the primitive. The coefficient and the concentrations would describe magnitude and extent only in combination with the shape, and the same quantized concentration would mean different widths for different primitives. An optimizer that flattens a lobe by increasing p would also dim and shrink it, and would have to compensate through $\mathbf{a}$, λ, and μ.

The normalization therefore requires every footprint of the family to share the mass, mean, and covariance of the Gaussian footprint φ_{G}:

$$\int_{\mathbb{R}^2} \varphi_{p,q}(\mathbf{u})\,\mathrm{d}\mathbf{u} = 1, \quad \int_{\mathbb{R}^2} \mathbf{u}\,\varphi_{p,q}(\mathbf{u})\,\mathrm{d}\mathbf{u} = \mathbf{0}, \quad \int_{\mathbb{R}^2} \mathbf{u}\,\mathbf{u}^{\mathsf{T}}\varphi_{p,q}(\mathbf{u})\,\mathrm{d}\mathbf{u} = \tfrac{1}{2}\,I. \tag{3.9}$$

The mass condition fixes the meaning of the coefficient: by Equation (3.5), a unit-mass footprint makes $\mathbf{a}_i$ the integrated radiance (Equation (2.3)) of primitive i, whatever its shape, up to the hemisphere truncation discussed in Section 3.5. The second moments fix the meaning of the concentrations. The second moment along an axis is the mean squared distance of the footprint's mass from the center along that axis, a measure of spread that does not refer to any particular contour. Holding it at $\frac{1}{2}$ makes the spread of every primitive in the orthographic chart equal to $\operatorname{diag}\big(1/(2\lambda),\, 1/(2\mu)\big)$, which is what λ and μ describe for the ASG. Fixing the peak value instead would leave the mass free to vary with the shape, and fixing a contour, as the unnormalized footprint does, would leave both mass and spread free.

Equation (3.9) amounts to six scalar conditions, but the footprint has only two free constants. Symmetry resolves the mismatch. The q-radius is unchanged by reflecting either coordinate and by exchanging u and v, so every footprint of the form (3.7) has zero mean, zero mixed moment, and equal axial moments. What remains are two conditions, unit mass and an axial second moment of $\frac{1}{2}$: the constant $b_{p,q}$ rescales the argument and sets the spread, and $c_{p,q}$ scales the value and sets the mass.

Both integrals separate into a factor that depends only on the contour and a factor that depends only on the radial profile. Because r_q is homogeneous of degree one, the region $r_q \le \rho$ is the unit ℓ_q ball $B_q = \{r_q \le 1\}$ scaled by ρ. Its area is therefore $\rho^2 A_q$ and its axial second moment $\rho^4 A_q M_q$, where A_q is the area of B_q and M_q the mean of u^2 over it. Both follow from Beta integrals over the ball:

$$A_q = \frac{4\,\Gamma(1+1/q)^2}{\Gamma(1+2/q)}, \qquad M_q = \frac{\Gamma(3/q)\,\Gamma(1+2/q)}{\Gamma(1/q)\,\Gamma(1+4/q)}, \tag{3.10}$$

with $A_2 = \pi$ and $M_2 = \frac{1}{4}$ for the disk. The footprint is constant on each level set of r_q, so both integrals can be taken shell by shell. Writing $x = \rho^2$, the region of the plane where $x < r_q^2 \le x + \mathrm{d}x$ has area $A_q\,\mathrm{d}x$ and axial second moment $2A_qM_q\,x\,\mathrm{d}x$, and the two-dimensional integrals reduce to one-dimensional ones,

$$\begin{aligned} \int_{\mathbb{R}^2} \varphi_{p,q}\,\mathrm{d}\mathbf{u} &= c_{p,q}A_q \int_0^\infty e^{-(b_{p,q}x)^p}\,\mathrm{d}x, \\ \int_{\mathbb{R}^2} u^2\varphi_{p,q}\,\mathrm{d}\mathbf{u} &= 2c_{p,q}A_qM_q \int_0^\infty x\,e^{-(b_{p,q}x)^p}\,\mathrm{d}x. \end{aligned} \tag{3.11}$$

The substitution $\tau = (b_{p,q}x)^p$ turns both into gamma integrals:

$$\int_{\mathbb{R}^2} \varphi_{p,q}\,\mathrm{d}\mathbf{u} = \frac{c_{p,q}A_q\Gamma(1+1/p)}{b_{p,q}}, \quad \int_{\mathbb{R}^2} u^2\varphi_{p,q}\,\mathrm{d}\mathbf{u} = \frac{c_{p,q}A_qM_q\Gamma(1+2/p)}{b_{p,q}^2}. \tag{3.12}$$

Setting the first to one and the second to $\frac{1}{2}$ gives

$$b_{p,q} = 2M_q\,\frac{\Gamma(1+2/p)}{\Gamma(1+1/p)}, \qquad c_{p,q} = \frac{b_{p,q}}{A_q\,\Gamma(1+1/p)}. \tag{3.13}$$

Both constants depend only on (p, q), so they are computed once per primitive rather than per evaluated direction. At the Gaussian member, $M_2 = \frac{1}{4}$ and $\Gamma(3)/\Gamma(2) = 2$ give $b_{1,2} = 1$, and $A_2 = \pi$ then gives $c_{1,2} = 1/\pi$, recovering φ_{G}.

The constants have a direct interpretation. By Equation (3.12) with $b = c = 1$, the per-axis variance of the unnormalized profile $\exp(-r_q^{2p})$ is exactly $\frac{1}{2}b_{p,q}$. The constant $b_{p,q}$ is therefore the ratio of that variance to the variance $\frac{1}{2}$ of the Gaussian footprint, and multiplying r_q^2 by $b_{p,q}$ brings the profile back to the Gaussian spread.

Once the width is corrected, $c_{p,q}$ is the reciprocal of the remaining mass; since the footprint peaks at the center, it is also the peak value. The flat-top limit makes this concrete: as $p \to \infty$ the footprint tends to a uniform density on an ℓ_q ball whose second moments are still $\frac{1}{2}$, which for $q = 2$ is the uniform disk of radius $\sqrt{2}$ and density $1/(2\pi)$ (Figure 3.3b).

The normalization fixes the mass, mean, and covariance of the footprint and nothing else. In the normalized family, p and q therefore control exactly the structure beyond second order: how flat the core is and how sharp the boundary, and whether the mass concentrates along the axes or toward the corners. At fixed mass and variance a flatter core must be lower, and the peak $c_{p,q}$ drops from $1/\pi \approx 0.32$ for the Gaussian to about 0.17 at $p = 4$, whereas the contour exponent barely changes it.

3.5 The Primitive on the Sphere

The normalization is exact in the plane. What it implies for the spherical primitive follows from the lift identity (3.5).

With $g = 1$, the identity gives the spherical mass of a primitive,

$$m = \int_{\mathbb{S}^2} K(\omega)\,\mathrm{d}\omega = \int_{\mathcal{E}_{\lambda,\mu}} \varphi_{p,q}(\mathbf{u})\,\mathrm{d}\mathbf{u} \;\le\; 1, \tag{3.14}$$

the planar unit mass clipped to the ellipse $\mathcal{E}_{\lambda,\mu}$. The deficit $1 - m$ is the part of the planar footprint that would lie beyond the horizon, and it is the only difference between the planar normalization and the spherical primitive. It shrinks as either concentration grows, because the ellipse grows, and it can be bounded because the ellipse contains the disk of radius $\sqrt{\min(\lambda, \mu)}$. For the Gaussian member the mass outside that disk is known in closed form, which gives $m \ge 1 - e^{-\min(\lambda,\mu)}$. The fitting procedure of Chapter 4 keeps $\lambda, \mu \ge 5$, so the Gaussian deficit is at most $e^{-5} \approx 0.7\%$. For general shapes Equation (3.14) has no closed form, but it is easily evaluated numerically. At the broadest admissible concentrations, $\lambda = \mu = 5$, quadrature in polar coordinates over the admissible exponents gives a smallest spherical mass of 0.9911, at $(p, q) = (1, 16)$, where Gaussian tails combine with nearly square contours that carry mass farthest out along the diagonals. Larger or unequal concentrations only reduce the deficit. Every admissible primitive therefore has a spherical mass between 0.991 and one, and its coefficient $\mathbf{a}_i$ equals its integrated radiance to within one percent. In this sense $\mathbf{a}_i$ is a *nominal* integrated-radiance coefficient.

Normalizing on the sphere instead, by requiring $m = 1$, would mean evaluating the two-dimensional integral (3.14) numerically for every primitive and differentiating through it during fitting. The planar normalization is closed-form, and the truncation it leaves is small and explicit.

The lift also determines the low-order directional moments of a primitive exactly. Taking g to be s, t, or st in Equation (3.5) gives an odd integrand over a centrally symmetric ellipse, and taking $g = s^2$ or $g = t^2$ gives the planar second moments clipped to the ellipse:

$$\begin{gathered}\int_{\mathbb{S}^2} K(\omega)\,\omega\cdot\mathbf{x}\,\mathrm{d}\omega = \int_{\mathbb{S}^2} K(\omega)\,\omega\cdot\mathbf{y}\,\mathrm{d}\omega = \int_{\mathbb{S}^2} K(\omega)\,(\omega\cdot\mathbf{x})(\omega\cdot\mathbf{y})\,\mathrm{d}\omega = 0,\\ \int_{\mathbb{S}^2} K(\omega)\,(\omega\cdot\mathbf{x})^2\,\mathrm{d}\omega \le \frac{1}{2\lambda}, \qquad \int_{\mathbb{S}^2} K(\omega)\,(\omega\cdot\mathbf{y})^2\,\mathrm{d}\omega \le \frac{1}{2\mu},\end{gathered} \tag{3.15}$$

with equality in the second line up to the same truncation. The center $\mathbf{z}$ is therefore the mean direction of the primitive, and the frame axes are the principal axes of its spread in the tangent plane, exactly and for every shape and pair of concentrations. For concentrated primitives, where the orthographic coordinates approximate angular distance, the root-mean-square angular spread along $\mathbf{x}$ is close to $(2\lambda)^{-1/2}$ radians; the concentration bounds of the fitter correspond to spreads from about 18° down to about 0.04°. Because this second-order structure does not depend on the shape, the frame and concentrations of a new primitive can be matched to a measured local covariance of the signal while its shape is left to the optimizer. The allocation procedure of Section 4.5 relies on this when inserting primitives.

For concentrated primitives the chart is nearly isometric where the footprint has appreciable mass, and the primitive looks like its planar footprint mapped onto a small patch of the sphere: an elliptical Gaussian lobe, a flat-topped ellipse, a rounded rectangle, or a flat-topped rounded rectangle. For broad primitives the lift compresses the outer part of the footprint toward the horizon, and the clamped cosine takes the kernel to zero at the edge of the hemisphere (Figure 3.1). Since only dot products between ω and the frame enter the kernel, the representation is also equivariant under rotations of the sphere: rotating the environment amounts to rotating the frames (Section 2.2).

3.6 Properties of the Representation

Each primitive carries the parameters listed in Table 3.1: a frame with three degrees of freedom, two concentrations, an RGB coefficient, and the two exponents introduced in this chapter.

An ASG of the form (2.13) with amplitude C_i and color $\mathbf{c}_i$ is represented exactly by the same frame and concentrations, the exponents $(1, 2)$, and the coefficient $\mathbf{a}_i = \pi C_i \mathbf{c}_i / \sqrt{\lambda_i \mu_i}$, so any ASG mixture is a special case of the adaptive mixture. Restricting the exponents defines a hierarchy of nested models that share every other part of the representation. The *ASG* model fixes $(p, q) = (1, 2)$. The *radial-only* model learns p with $q = 2$, giving elliptical contours with an exponential-power profile. The *contour-only* model learns q with $p = 1$, giving superelliptic contours with a Gaussian profile in the q-radius. The *full* model learns both. Because frames,

Table 3.1. ***Parameters of a shape-adaptive primitive.*** *The admissible ranges of the exponents define the kernel family; bounds on the concentrations are imposed by the fitting procedure of Chapter 4.*

Parameter	Domain	DoF	Role
frame $(\mathbf{x}, \mathbf{y}, \mathbf{z})$	$SO(3)$	3	center $\mathbf{z}$ and orientation of the tangent axes
concentrations λ, μ	$\mathbb{R}_{>0}$	2	extent along $\mathbf{x}$ and $\mathbf{y}$
coefficient $\mathbf{a}$	$\mathbb{R}^3_+$	3	nominal integrated RGB radiance
radial exponent p	$[1, 4]$	1	falloff from core to tail
contour exponent q	$[1.25, 16]$	1	geometry of the level sets
total		10	ASG: 8

concentrations, and coefficients mean the same in all four, differences between them can be attributed to the shape of the footprint alone. This is the basis of the controlled comparisons of Chapter 5, for which Chapter 4 fits the four models with the same procedure.

Different parameter values can describe the same kernel. Reversing both tangent axes, or exchanging them together with their concentrations, leaves Equation (3.8) unchanged, since $\varphi_{p,q}$ depends only on $|u|$ and $|v|$ and r_q is symmetric in its arguments. Every kernel therefore has several equivalent parameterizations, of which the compact encoding of Section 4.6 stores a canonical one.

Evaluating a primitive in one direction takes three dot products, two q-th powers and a root for the q-radius, one p-th power, and one exponential; the gamma functions of Equation (3.13) are evaluated once per primitive. Compared with an ASG, which needs the dot products and one exponential, the adaptive kernel adds a small, constant number of operations per evaluation. The kernel is nonnegative and continuous on the sphere, continuously differentiable on the open front hemisphere, and differentiable with respect to all of its parameters, including the exponents, since the normalization constants are smooth functions of (p, q).

The price of the extension is that the closed-form product and convolution approximations of the ASG [Xu+13] rely on its Gaussian footprint and do not carry over to general (p, q). The representation is designed for reconstructing and evaluating the environment itself; analytic rendering operators for the adaptive family are left open (Section 6.4).

The resulting representation is still a positive superposition of kernels on the sphere, equivariant under rotations and evaluable in any direction, with the ASG as the member $(p, q) = (1, 2)$ and ten parameters per primitive. The normalization keeps mass and second moments fixed, so the coefficients and concentrations retain their meaning and the exponents describe shape alone. How the number, placement, geometry, shape, and color of the primitives are chosen for a particular environment is the subject of Chapter 4.

Chapter 4

HDR Environment Map Reconstruction

Chapter 3 defined a family of functions on the sphere. This chapter describes how a member of that family is chosen for a given environment: how an HDR raster becomes a positive mixture of shape-adaptive primitives, and how the fitted mixture is stored. Two principles guide the design. The fit is faithful to the sphere: directions are weighted by solid angle, and no step depends on the parameterization of the input. And the fit is the same for every model: the ASG and the three shape-adaptive models are fitted by an identical procedure, so that the comparisons of Chapter 5 measure the primitives and nothing else.

Several components of the procedure are standard and are adopted as such: the perceptual quantizer of HDR video as the encoding of radiance [MND13], the Charbonnier penalty [Cha+94], the Adam optimizer [KB15], the reference tangent basis of Duff et al. [Duf+17], the octahedral encoding of directions [Cig+14], and the strategy, familiar from Gaussian splatting and Image-GS [Ker+23; Zha+25], of adding primitives during optimization where the reconstruction falls short. What this thesis adds concerns the spherical domain and the positive mixture. The quadrature refines itself around edges and emitters. Scene-referred radiance is calibrated for the perceptual objective, which is extended beyond its nominal range. The frame of a primitive stays continuous when its center crosses the chart boundary of the reference basis. New primitives are placed where the residual is positive, matched to its second moments, and accepted only if the objective decreases. And the encoding stores each physical parameter in a canonical form at a fixed rate.

4.1 Reconstruction as Optimization on the Sphere

Given a target environment L and a budget of N primitives, reconstruction seeks the parameters Θ of the mixture (3.1), that is, the frame, concentrations, coefficient,

and exponents of every primitive, that minimize a discrepancy between $\widehat{L}$ and L integrated over the sphere:

$$\mathcal{E}(\Theta) = \frac{1}{4\pi} \int_{\mathbb{S}^2} D\big(\widehat{L}(\omega), L(\omega)\big)\, \mathrm{d}\omega. \tag{4.1}$$

The pointwise penalty D is defined in Section 4.3. The objective is defined on directions and makes no reference to the raster in which L is stored. The raster still matters. It supplies the values of L through interpolation at arbitrary directions, and its resolution and parameterization bound the information available to the fit: detail finer than a texel cannot be recovered, and a latitude–longitude panorama carries less information per unit solid angle near its equator than near its poles. What the spherical formulation removes is the other influence of the raster, the implicit weighting of directions by texel density (Section 2.2.3).

The four models of Section 3.6 differ only in which exponents are free. The ASG model fixes $(p_i, q_i) = (1, 2)$ for every primitive, the radial-only model optimizes p_i with $q_i = 2$, the contour-only model optimizes q_i with $p_i = 1$, and the full model optimizes both. Frozen exponents stay exactly at their Gaussian values throughout the fit, so the ASG model is a conventional ASG mixture. Everything else is shared: the quadrature, the objective, the initialization, the optimizer and its schedule, the insertion of new primitives, the budget, and the random seed.

4.2 Spherical Quadrature

The objective is discretized with the geodesic quadrature of Section 2.2.4, with one sample at the center of each spherical triangle and its exact solid angle as weight. The level of the base grid is chosen so that the number of cells is as close as possible to the number of texels of the input raster, so that the quadrature samples the sphere about as densely as the raster does on average.

A uniform grid of this density is adequate for most of the sphere but not for the features that matter most. A solar disk subtends a fraction of a degree, and the edge of a luminaire cuts through cells. The base grid is therefore refined adaptively by one level. Each base cell is split into its four child triangles, and the target is sampled at their centers. The cell is refined when either of two conditions holds. The first detects edges and texture: it compares the variation of the RGB values among the four children, weighted by their solid angles, with the energy of the cell, plus a small floor proportional to the energy of an average cell so that faint noise in dark regions does not trigger refinement. The second detects concentrated peaks: it refines a cell when one child exceeds the mean luminance of the cell by an amount comparable to the energy of an average cell. A refined cell is replaced by its four children. An unrefined cell keeps a single sample at its center, whose value is the solid-angle average of its four children. The result is a partition of the sphere into spherical triangles whose exact solid angles sum

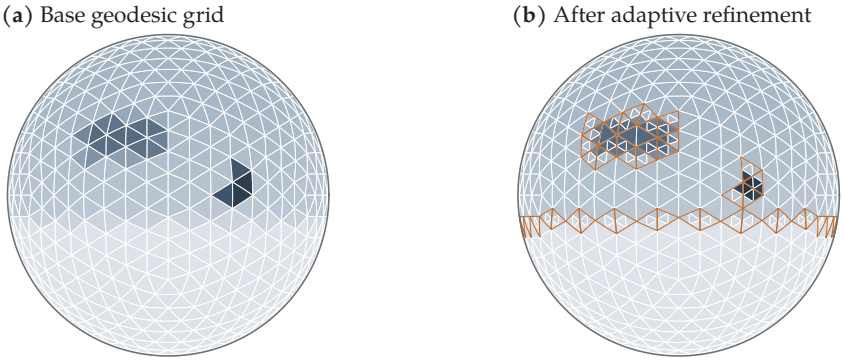


Figure 4.1. ***Adaptive refinement of the fitting quadrature.*** *A synthetic environment with a sky gradient, a darker ground, a rectangular luminaire, and a small bright source, on a geodesic grid much coarser than in an actual fit. Each cell is shaded, on a logarithmic scale, by the average radiance over it.* (***a***) *Base grid.* (***b***) *The cells outlined in orange satisfy the refinement criterion and are replaced by their four children: they follow the horizon, the edges of the luminaire, and the small source, while the uniform interior of the luminaire and the smooth sky remain coarse.*

to 4π, denser around edges and emitters, and in which every sample value is an average over its cell rather than a point sample of the raster. Figure 4.1 shows the refinement on a synthetic environment.

The target values are computed once, by bilinear interpolation in the input raster at the sample directions, and every part of the fit uses the same samples: the objective, the initialization, and the insertion of new primitives. Each optimization step draws a mini-batch uniformly from the M samples of the quadrature, cycling through random permutations so that every sample is visited once per pass. A sample keeps its solid angle as a weight, rescaled to

$$w_j = \frac{M\,\Omega_j}{4\pi}, \tag{4.2}$$

so that the weights average to one and the weighted mean over a batch is an unbiased estimate of the solid-angle mean that the full quadrature computes. Refined regions contribute more samples, each with a smaller weight: the refinement increases the resolution of the objective where it is needed without changing what it measures.

A separate geodesic grid, rotated relative to the fitting grid, is used only to evaluate fitted representations in Chapter 5. The rotation keeps evaluation samples from coinciding with the samples the fit has seen.

4.3 A Perceptual HDR Objective

The pointwise penalty D decides how errors at different radiance levels are weighed against each other. In linear radiance, the error is dominated by the brightest directions: a one percent error on the radiance of the sun outweighs any

error on the walls of a room, so a fit driven by linear error spends its capacity on emitters and leaves the visible structure of the rest of the environment unresolved. A logarithmic error reverses the imbalance. It treats relative errors equally at every level, but it magnifies differences between very small values, where the input is dominated by noise and quantization, and it needs an arbitrary offset to be defined at zero.

Display and imaging practice faces the same problem and addresses it with perceptually motivated encodings, which map absolute luminance to code values in which equal steps are approximately equally visible [AMS08; MA21]. The perceptual quantizer (PQ), standardized as SMPTE ST 2084 and adopted in ITU-R BT.2100 [MND13; SMP14; ITU25], was derived from a model of contrast sensitivity and covers luminance up to 10,000 cd/m^2. With Y the luminance divided by 10,000 cd/m^2, it is

$$T(Y) = \left(\frac{c_1 + c_2\, Y^{m_1}}{1 + c_3\, Y^{m_1}} \right)^{m_2}, \tag{4.3}$$

with fixed constants[1] chosen so that T maps the nominal range onto $[0, 1]$. It spends code values densely in the dark range, where the eye is sensitive to small absolute differences, and progressively more sparsely toward bright highlights. Measured in PQ, a small visible error on a bright source and a small visible error in a shadowed region carry comparable weight.

PQ is defined on absolute luminance, whereas the panoramas used here are scene-referred with an unknown absolute scale, so a calibration convention is needed. The thesis maps a linear value of one to 203 cd/m^2, the HDR reference white of broadcast practice [ITU26], and encodes each RGB channel with this fixed factor for every environment. The convention is arbitrary and affects absolute PQ values, but it is identical for all models. With this calibration, linear values above about 49 exceed the nominal range, and bright emitters routinely do, the solar disk by orders of magnitude. Clipping them would make the objective indifferent to the energy of the brightest sources, since a reconstruction and a target that both exceed the limit would have zero residual. During fitting, Equation (4.3) is therefore evaluated beyond its nominal endpoint. The extended curve remains smooth and increasing and approaches the finite limit $(c_2/c_3)^{m_2} \approx 2$, about twice the code value of the nominal endpoint. Very bright directions therefore keep contributing, with diminishing weight, without dominating the objective. The PQ-domain metrics of Chapter 5 follow the standard and clip at the nominal endpoint instead.

Residuals of encoded values are penalized by the Charbonnier function $\rho_\epsilon(t) = \sqrt{t^2 + \epsilon^2} - \epsilon$ [Cha+94], a smooth approximation of the absolute value that is quadratic only for $|t| \lesssim \epsilon$. With a small ϵ it behaves essentially as an ℓ_1 penalty, so the large residuals that concentrate along the edges of emitters do not dominate

[1] $m_1 = 2610/16384$, $m_2 = 128 \cdot 2523/4096$, $c_1 = 3424/4096$, $c_2 = 32 \cdot 2413/4096$, and $c_3 = 32 \cdot 2392/4096$.

the gradient as they would under a squared error. The penalty is averaged over the three channels, and the discrete objective is

$$\mathcal{E}(\Theta) = \frac{1}{4\pi} \sum_{j=1}^{M} \Omega_j \, \frac{1}{3} \sum_{c=1}^{3} \rho_\epsilon \Big(T\big(\widehat{L}_c(\omega_j)\big) - T\big(L_c(\omega_j)\big) \Big), \tag{4.4}$$

where $\widehat{L}_c$ and L_c denote the channels of the reconstruction and the target, each scaled by the calibration factor before encoding. A mini-batch estimates Equation (4.4) by averaging w_j times the inner sum over its samples.

Equation (4.4) is what the fit minimizes. It is not the main radiometric measure of Chapter 5, which is a relative error in linear radiance, chosen there because it exposes errors in the energy of emitters that a perceptual encoding compresses. The two measures need not rank reconstructions identically, and Chapter 6 returns to the cases in which they do not.

4.4 Optimizing Constrained Primitives

The parameters of a primitive are constrained: its center is a unit vector, its frame is orthonormal, its concentrations and coefficients are positive, and its exponents lie in bounded intervals. The fit optimizes an unconstrained parameterization of each primitive by stochastic gradient descent and projects onto the constraints after every update.

The center is stored as an unnormalized vector, normalized whenever the kernel is evaluated and renormalized after each update. The orientation of the frame is a single angle α measured from a reference tangent basis $(\mathbf{t}_1, \mathbf{t}_2)$ at the center,

$$\mathbf{x} = \cos\alpha \, \mathbf{t}_1 + \sin\alpha \, \mathbf{t}_2, \qquad \mathbf{y} = -\sin\alpha \, \mathbf{t}_1 + \cos\alpha \, \mathbf{t}_2, \tag{4.5}$$

which keeps the frame orthonormal and right-handed by construction. The reference basis is the two-chart construction of Duff et al. [Duf+17], which is smooth on each hemisphere and switches charts at the equator (Section 2.2.2). When a center crosses the equator during optimization, the reference basis jumps by a rotation about the center that depends only on the longitude of the crossing. The fitter adds the same rotation to α, so the frame itself moves continuously and the chart switch has no effect on the represented function. The concentrations are optimized through their logarithms, the coefficients through the softplus function $a = \log(1 + e^{\beta})$ applied per channel, which keeps them positive and is nearly linear for large values, and the exponents through $\log p$ and $\log(q/2)$. Both shape parameters are zero at the Gaussian member, so the ASG lies at the origin of the shape parameters, and equal steps correspond to equal relative changes of p and q.

The parameterization inherits the redundancies of the kernel noted in Section 3.6. The angle α matters only modulo π, since reversing both tangent axes

leaves the kernel unchanged. For an isotropic primitive with circular contours, $\lambda = \mu$ and $q = 2$, the kernel does not depend on α at all and α receives no gradient; with $\lambda = \mu$ and $q \neq 2$, it matters modulo $\pi/2$. None of this hinders gradient descent, which does not need a unique parameterization. The encoding of Section 4.6 removes the redundancy before storage.

After each update, centers are renormalized, exponents are clamped to their admissible ranges $[1, 4]$ and $[1.25, 16]$, and concentrations are clamped to an interval with lower bound five, which keeps every primitive in the regime where its coefficient equals its integrated radiance to within one percent (Section 3.5). The ratio of the two concentrations is limited at the same time, so that no primitive degenerates into an arbitrarily thin streak. In addition, a small quadratic penalty acts on the logarithm of any concentration above a soft cap, which rises to the hard upper bound during the early part of the fit. It discourages primitives from collapsing onto individual bright samples before the arrangement of the mixture has settled. The function minimized by gradient descent is Equation (4.4) plus this penalty.

The optimizer is Adam [KB15], with separate learning rates for centers, angles, concentrations, coefficients, and exponents. The exponents use the smallest rate, so shape evolves more slowly than position and extent. Gradients are accumulated over mini-batches and clipped in global norm, and the learning rates are held constant for the first half of the fit and then decay along a cosine schedule. Because primitives are added during the fit, the optimizer keeps its state per primitive: a new primitive starts with zero moment estimates and its own bias correction, while existing primitives keep theirs. Every model runs for the same number of steps; the values are listed in Section A.2.

Evaluating Equation (3.8) for about 10^8 pairs of directions and primitives per step requires some care in floating point. The normalization constants (3.13) are computed once per primitive and per step, in double precision and in logarithmic form through the log-gamma function, which avoids cancellation near the Gaussian member, where $b_{p,q}$ and $c_{p,q}$ are ratios of nearly equal gamma values. The exponent $(b_{p,q}\, r_q^2)^p$ is evaluated in the log domain after factoring out the larger standardized coordinate $m = \max(|u|, |v|)$,

$$\log\left[(b_{p,q}\, r_q^2)^p\right] = p\left(\log b_{p,q} + 2\log m + \frac{2}{q}\log s\right), \qquad s = \left(\frac{|u|}{m}\right)^q + \left(\frac{|v|}{m}\right)^q, \quad (4.6)$$

so that $s \in [1, 2]$ and no large power is formed even at $q = 16$. The exponent is clamped where the kernel becomes negligible, and safe operands replace the undefined logarithms at the center of a primitive, where $u = v = 0$, so that gradients are finite everywhere.

4.5 Initialization and Residual-Guided Allocation

The fit starts from a fraction of the budget and adds the remaining primitives in stages. Initial primitives are placed according to the content of the target. At every quadrature sample, a structure tensor of the log-compressed luminance is estimated from central differences along the axes of a reference tangent frame, at the scale of one and two quadrature cells. Its trace measures the amount of local detail, its leading eigenvector gives the dominant gradient direction, and the normalized difference of its eigenvalues, the coherence, measures how strongly the variation is oriented. A content score that combines local detail with log-luminance is clipped at a high quantile, so that a few extreme samples cannot absorb the budget, normalized to a density on the sphere, and mixed with a small uniform component that guarantees coverage of featureless regions. Initial centers are drawn from this density without replacement.

Each initial primitive receives a size matched to the density ϱ at its center. With N_0 initial primitives, a primitive at $\mathbf{z}$ represents a share $1/(N_0\varrho(\mathbf{z}))$ of the sphere, which is the area of a disk of radius

$$w = \big(\pi N_0\, \varrho(\mathbf{z})\big)^{-1/2}, \tag{4.7}$$

and its concentrations are set to w^{-2}. The frame is aligned with the direction of least variation, along an edge rather than across it, and the concentrations are made anisotropic in proportion to detail and coherence, up to a modest ratio and at a fixed geometric mean. Every primitive starts at the Gaussian member $(p, q) = (1, 2)$: shape is not estimated from the data but left to the optimizer. Finally, the coefficients distribute the integrated radiance of the target among the primitives, channel by channel, in proportion to the target radiance at their centers, so that the initial mixture has the same integrated radiance as the target.

A fixed initial arrangement cannot anticipate where a detailed environment needs capacity, and gradient descent can move, stretch, and reshape primitives but cannot create them where none are present. The remaining primitives are therefore inserted during the fit where the reconstruction falls short. In a positive mixture, the two signs of the residual play different roles. Where the reconstruction is too dark, adding a primitive with a suitable weight can only help. Where it is too bright, a new primitive can only make things worse: the excess must be removed by dimming, narrowing, or reshaping the primitives that produce it, which is the work of the optimizer. Allocation therefore considers only the positive part of the residual, $L - \widehat{L}$ where it is positive, and only inserts. No primitive is removed, split, or cloned, and the insertion step never edits an existing primitive (Figure 4.2). Insertion stages occur at regular intervals during the first part of the fit, each adding a fixed fraction of the budget until the budget is reached. They stop well before the end, so that the last primitives are optimized jointly with the others for many steps.

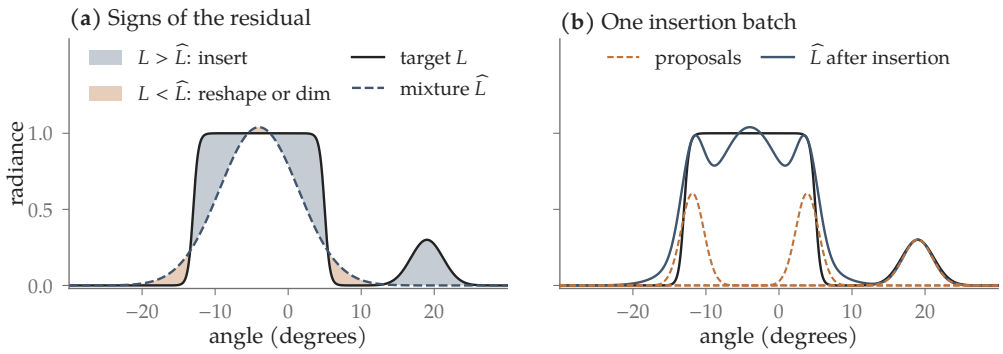


Figure 4.2. ***Positive residual and insertion.*** *A one-dimensional section through a flat-topped source and a dim feature.* (***a***) *A single Gaussian lobe approximates the source. Where the mixture is too dark, a new primitive can only help; where it is too bright, the excess must be removed by reshaping or dimming the existing lobe.* (***b***) *Proposals at the maxima of the positive residual, with widths from its local spread, scaled by one common factor. The subsequent optimization adjusts all primitives jointly.*

Where to insert. Where capacity is missing depends on how the residual is measured. In linear radiance almost all of the missing energy lies around emitters, whereas a perceptual measure spreads it over the visible structure of the environment. The discovery density combines both. Its perceptual part averages the positive differences between target and reconstruction after a logarithmic compression and after a simple tone mapping at a few exposures, clipped at a high quantile and normalized to a probability over the quadrature. Its safeguard part is the positive linear residual luminance of each cell, weighted by its solid angle and divided by the integrated luminance of the target, that is, the fraction of the target energy that the cell is missing. The discovery density is the normalized pointwise maximum of the two. The perceptual part spreads insertions over visible structure, and the safeguard guarantees that a cell missing a substantial share of the energy, such as an unresolved emitter, is discovered. Cells whose linear residual is negligible, in absolute terms or relative to the target, are excluded.

Candidate centers are drawn from the discovery density, several times more than the number to be inserted, and ordered by density. Each candidate receives an allocation width from the perceptual density alone, by Equation (4.7) with the primitive count after insertion, so that the size of a new primitive reflects the scale of the visible structure it is meant to fill rather than the absolute energy of the residual. A greedy spherical non-maximum suppression visits the candidates in order and discards any candidate closer to an already accepted one than a fixed fraction of their allocation radii, until the requested number remains.

Initial geometry of a new primitive. A new primitive starts at the candidate direction, isotropic with the allocation width, and with a provisional coefficient equal to the positive residual at its center times the solid angle of its allocation cap.

Its orientation and anisotropy are then estimated from the residual it is meant to fill. Within the allocation cap, the residual luminance times the solid angle of each cell defines a distribution over the quadrature samples. To this distribution the fitter matches, by maximum likelihood, a density proportional to the clamped cosine $\omega \cdot \mathbf{z}$ times the exponential of a quadratic polynomial in the orthographic coordinates of the candidate's frame: an ASG-shaped density with an additional linear tilt. In this exponential family, maximum likelihood is moment matching, so the fitted density reproduces the first and second moments of the residual within the cap. The linear tilt absorbs any offset of the residual from the candidate direction, so that the quadratic part describes the spread of the residual rather than its displacement, and only the quadratic part is used. Its eigenvectors set the tangent frame and the ratio of its eigenvalues sets the anisotropy, while the geometric mean of the concentrations keeps the value given by the allocation width and the center stays at the candidate direction. When the cap contains too few samples or the fit is degenerate, the primitive remains isotropic.

This step relies on the second-moment property of Section 3.5. Because the frame and the concentrations describe the second moments of every member of the family, a frame and an anisotropy matched to the covariance of the residual remain correct whatever shape the primitive adopts later. New primitives therefore enter at the Gaussian member with matched second-order geometry and acquire their shape during the subsequent optimization, exactly like the initial ones.

Calibration and acceptance. Whether a batch of proposals improves the fit is decided by the objective itself. Let $g_j = \partial \mathcal{E} / \partial \widehat{L}(\omega_j)$ be the derivative of the discrete objective (4.4) with respect to the reconstruction at each sample. For a proposal k with provisional contribution $\mathbf{a}_k K_k(\omega)$, the directional derivative

$$\delta_k = \sum_{j=1}^{M} g_j \cdot \mathbf{a}_k \, K_k(\omega_j) \tag{4.8}$$

is the first-order change of the objective when the proposal is added with a small positive weight; a negative value means that adding it reduces the objective. The provisional coefficients are only estimates, so the batch is calibrated as a whole: all of its contributions are scaled by one nonnegative factor, chosen by a bounded one-dimensional search that evaluates the exact objective at every trial value. The calibrated batch is then appended tentatively, and the complete mixture is rendered at every quadrature sample by the same evaluator used for optimization. The batch is accepted only if both the objective and the objective including the concentration penalty decrease by more than a tolerance derived from floating-point error. If the batch fails, a second attempt keeps only the proposals with $\delta_k \leq 0$. If that also fails, nothing is inserted at that stage.

Accepted primitives are appended with fresh optimizer state, and existing primitives keep their parameters and their state. For the rest of the fit, each new center is confined to a cone around its insertion direction, with half the radius of its allocation cap. The primitive thus keeps filling the residual for which it was inserted instead of drifting into a region that is already covered.

4.6 Compact Representation

A fitted mixture is a set of floating-point arrays. To measure what shape costs in storage, and to compare models by decoded size, the thesis stores a mixture in a fixed-rate binary format that quantizes each physical parameter directly. Every primitive becomes one record of fixed length, and the records are packed after a short header without padding between them.

Before quantization, each primitive is brought to a canonical form using the symmetries of the kernel (Section 3.6). If $\lambda > \mu$, the tangent axes are exchanged together with the concentrations: the new $\mathbf{x}$ is the old $\mathbf{y}$ and the new $\mathbf{y}$ is the old $-\mathbf{x}$, so the frame stays right-handed. After this step $\lambda \leq \mu$, and the orientation is an angle in $[0, \pi)$, since reversing both axes leaves the kernel unchanged. Neither operation changes the represented function.

Each field is then quantized according to its geometry. The center is stored through its two octahedral coordinates [Cig+14], each quantized uniformly. The angle is quantized uniformly over $[0, \pi)$, relative to the reference tangent basis of the *decoded* center: the encoder first transports the original $\mathbf{x}$ axis to the decoded center by the shortest rotation, so that the decoder, which knows only the decoded center, rebuilds a frame consistent with it. Concentrations, coefficients, and exponents act as scales, and all of them are quantized logarithmically, which gives a constant relative step. The concentrations share a range between the smallest and the largest concentration of the mixture, and the coefficients share a range between their smallest and largest positive values, with one code of each coefficient field reserved for an exact zero. The exponents are quantized over their admissible intervals. The ASG model stores no exponents, and the radial-only and contour-only models store one each.

Table 4.1 lists the field widths. Adjacent codes differ by at most 0.3% in concentration, about 0.5% in p, and 1% in q, and the angle step is 0.044°. A header of 28 bytes identifies the format and the model, and stores the number of primitives and the four quantization endpoints of the concentrations and coefficients as 32-bit floating-point values. A mixture of N primitives with records of b bits therefore occupies

$$B(N) = 28 + \left\lceil \frac{b\,N}{8} \right\rceil \text{ bytes.} \tag{4.9}$$

The two exponents add 16 bits to every record, 15.4% of the ASG record, or 20,000 bytes for a mixture of 10,000 primitives. Equal primitive count is therefore not

Table 4.1. ***Bits per field of a primitive record.*** *The exponents are stored only by the models that learn them, so a record has 104 bits for the ASG model, 112 bits for the radial-only and contour-only models, and 120 bits for the full model.*

Field	bits
center, two octahedral coordinates	2×16
tangent angle	12
concentrations λ, μ	2×12
coefficient $\mathbf{a}$, three channels	3×12
radial exponent p	8
contour exponent q	8

equal storage: a full mixture of N primitives occupies as many record bits as an ASG mixture of $120N/104 \approx 1.154N$ primitives. Chapter 5 compares the models at equal primitive count and as a function of storage.

Decoding inverts each transform, rebuilds a right-handed frame from the decoded center and angle, and returns floating-point arrays that are evaluated by the same kernel evaluator as a fitted model. A decoded mixture is an ordinary member of the representation, with every parameter in its admissible range. The format has a fixed rate per primitive and uses neither entropy coding nor quantization-aware fine-tuning, so decoded quality reflects post-training quantization of the parameters alone.

Chapter 5

Experimental Evaluation

The experiments in this chapter test whether shape adaptation increases what a mixture of spherical primitives can represent. Because the ASG model is a member of the shape-adaptive family and every model is fitted by the same procedure, the main comparison is controlled: the ASG and the shape-adaptive models differ only in the exponents they are allowed to learn. The chapter first compares the ASG and the full model at equal primitive count and separates the contributions of the two exponents. It then looks at how the fitted primitives use their shape, follows the comparison through quantization and across budgets, and places the spherical models next to an adaptive image-space representation, both as maps and as lighting.

5.1 Experimental Setup

The evaluation uses fifteen HDR environments from the Poly Haven library [Pol26], shown in Figure 5.1. They include interiors lit by windows and luminaires, studio setups with artificial lights, outdoor scenes by day and by night, and three sky-only environments without ground. Each environment is prepared as a 1024×1024 octahedral EXR, which is the input raster of the spherical fits, and as a 1024×512 latitude–longitude EXR (Section A.1).

Five models are compared. The ASG, radial-only, contour-only, and full models are the nested spherical models of Section 3.6, fitted by the procedure of Chapter 4 with the settings of Section A.2. Image-GS [Zha+25] serves as the adaptive image-space reference. It is fitted in its native formulation, to the latitude–longitude panorama, with a reversible logarithmic HDR encoding, and its output is decoded to linear radiance for evaluation; all references to Image-GS below concern this configuration. Every model is given a budget of 10,000 primitives, except in the rate–distortion study of Section 5.5, and every fit uses a single random seed.



Figure 5.1. ***The fifteen evaluation environments.*** *Reference latitude–longitude panoramas, display-mapped with the exposure used in all figures of this chapter.*

The reconstructions are measured with one radiometric error on the sphere and three image metrics. The radiometric error is the relative L_2 error in linear RGB radiance,

$$e_{L_2} = \left(\frac{\sum_j \Omega_j \left\| \widehat{L}(\omega_j) - L(\omega_j) \right\|^2}{\sum_j \Omega_j \left\| L(\omega_j) \right\|^2} \right)^{1/2}, \tag{5.1}$$

evaluated on the rotated evaluation grid of Section 4.2, a level-nine geodesic grid with 5,242,880 samples that is shared by all models. It weights directions by solid angle and radiance by energy, so it is dominated by the emitters and exposes errors in their energy. The image metrics are computed on common 1024×512 latitude–longitude rasters of the reference and the reconstruction. PQ-PSNR, PQ-SSIM [Wan+04], and PQ-LPIPS [Zha+18] compare the two rasters after the PQ encoding of Section 4.3, with the same calibration and with clipping

at 10,000 cd/m^2. Because the Poly Haven environments are not photometrically calibrated, the absolute PQ scores depend on this convention. The image metrics weight texels uniformly, and therefore give the polar regions of the panorama more weight than the solid angle they cover, and the PQ encoding compresses the dynamic range: they measure visible structure rather than energy. The two kinds of measure are complementary, and both are reported.

A third measure concerns the use of an environment as lighting. Every reconstruction lights the same scene in Blender's Cycles path tracer: a glass, a diffuse, a mirror, and a glossy sphere on a checkered floor, with a fixed camera and fixed materials. The renderings are display-mapped with one exposure per environment, derived from the rendering lit by the reference. The *rendering error* is the root-mean-square difference between the rendering lit by a reconstruction and the rendering lit by the reference, in display values between zero and one. It is small for every model and is therefore reported multiplied by 100 throughout; Section A.5 lists the render settings.

Every environment has equal weight in the means. Unless stated otherwise, the results describe the fitted parameters before quantization. The figures display spherical views of the environments with the AgX view transform and one exposure per scene, derived from the reference. For visual comparison, insets show FLIP error maps [And+20], computed on the displayed images at 67 pixels per degree and drawn on a common color scale within each figure. They locate errors; they are not used as a score.

5.2 Shape Adaptation at a Matched Primitive Count

Table 5.1 compares the ASG model with the full model. Shape adaptation improves every mean metric. The relative L_2 error falls from 0.1849 to 0.1418, a reduction of 23.3%, PQ-PSNR rises from 42.28 to 42.68 dB and PQ-SSIM from 0.9267 to 0.9324, and PQ-LPIPS falls by 10.3%. The improvement is also consistent across environments. The full model has the lower relative L_2 error and the higher PQ-PSNR in fourteen of the fifteen, and the better PQ-SSIM and PQ-LPIPS in thirteen, with the remaining two equal at the precision of Table 5.1. It is worse only on Industrial Sunset in the relative L_2 error and on Kloppenheim in PQ-PSNR, and the ties occur on Kloppenheim and Qwantani Noon: all exceptions are sky-only environments in which both models are already close to the reference.

The gains are largest where the illumination is concentrated in bounded sources. The relative L_2 error falls by 57.0% on Street Lamp, by 32.7% on Studio Small 07 and on Workshop, by 30.6% on Royal Esplanade, and by 27.4% on Qwantani Noon. The largest PQ-PSNR gains, between 0.5 and 0.8 dB, occur on Royal Esplanade, Artist Workshop, Studio Small 07, Workshop, and Kiara Interior, environments with framed openings, fixtures, and other structure with rectilinear boundaries. On the smooth skies of Kloppenheim and Industrial Sunset a Gaussian lobe

Table 5.1. ***ASG and full model at 10,000 primitives.*** *Per-scene reconstruction quality before quantization. The better value of each pair is set in bold, and both values when they are equal at the displayed precision.*

	Rel. L_2 ↓		PQ-PSNR↑		PQ-SSIM↑		PQ-LPIPS↓	
Environment	ASG	Full	ASG	Full	ASG	Full	ASG	Full
Artist Workshop	0.1328	**0.1153**	37.04	**37.72**	0.9271	**0.9355**	0.2398	**0.2024**
Cannon	0.0270	**0.0250**	37.79	**38.15**	0.9268	**0.9312**	0.1450	**0.1332**
Industrial Sunset	**0.0767**	0.0887	59.59	**60.01**	0.9992	**0.9993**	0.0051	**0.0045**
Kiara Interior	0.2114	**0.1666**	36.76	**37.29**	0.9206	**0.9280**	0.2013	**0.1778**
Kloppenheim	0.0405	**0.0404**	**67.50**	67.45	**0.9999**	**0.9999**	**0.0004**	**0.0004**
Neon Photostudio	0.1330	**0.1213**	35.42	**35.72**	0.9567	**0.9611**	0.2198	**0.1959**
Quarry	0.1783	**0.1765**	41.40	**41.58**	0.9546	**0.9565**	0.0966	**0.0932**
Qwantani Noon	0.2255	**0.1636**	65.68	**65.86**	**0.9999**	**0.9999**	**0.0003**	**0.0003**
Royal Esplanade	0.7196	**0.4996**	34.96	**35.77**	0.9181	**0.9300**	0.1658	**0.1396**
Street Lamp	0.3494	**0.1501**	35.02	**35.45**	0.8527	**0.8640**	0.3472	**0.3161**
Studio Small 07	0.0375	**0.0252**	44.88	**45.48**	0.9818	**0.9838**	0.0741	**0.0611**
Urban Street	0.0935	**0.0877**	29.37	**29.61**	0.8083	**0.8166**	0.3503	**0.3266**
Venice Sunset	0.2582	**0.2445**	41.38	**41.81**	0.9628	**0.9653**	0.1240	**0.1158**
Workshop	0.1875	**0.1263**	35.68	**36.24**	0.8986	**0.9097**	0.2357	**0.2022**
Xanderklinge	0.1025	**0.0966**	31.79	**32.04**	0.7931	**0.8049**	0.4332	**0.3981**
Mean	0.1849	**0.1418**	42.28	**42.68**	0.9267	**0.9324**	0.1759	**0.1578**

already fits the dominant content, and the two models are nearly equivalent. On Industrial Sunset, the full model trades energy for visible structure: all three image metrics improve, while the relative L_2 error is 15.7% higher than that of the ASG.

The absolute values of the relative L_2 error differ widely between environments. They are largest on Royal Esplanade and Street Lamp, where the brightest thousandth of the sphere carries more than 60% of the luminance and practically all of the squared radiance that normalizes Equation (5.1). There the error measures, in effect, how well a few small and very intense sources are reproduced, and both models remain far from exact.

Figure 5.2 shows where these differences come from. In Workshop, the full model recovers the frame, the equipment, and the shelf boundaries with markedly higher contrast than the ASG model, and in Urban Street it resolves the grille and the license plate of the car. In Kiara Interior, the horizontal chair rails remain distinct, and in Neon Photostudio the luminous strips keep their separation and extent and follow their corners closely. These are the bounded, flat-topped, and rectilinear structures for which Section 3.1 found a positive Gaussian mixture inefficient, and the FLIP insets agree: in all four regions the full model has the lowest error of the three models. Image-GS, discussed in Section 5.6, loses much

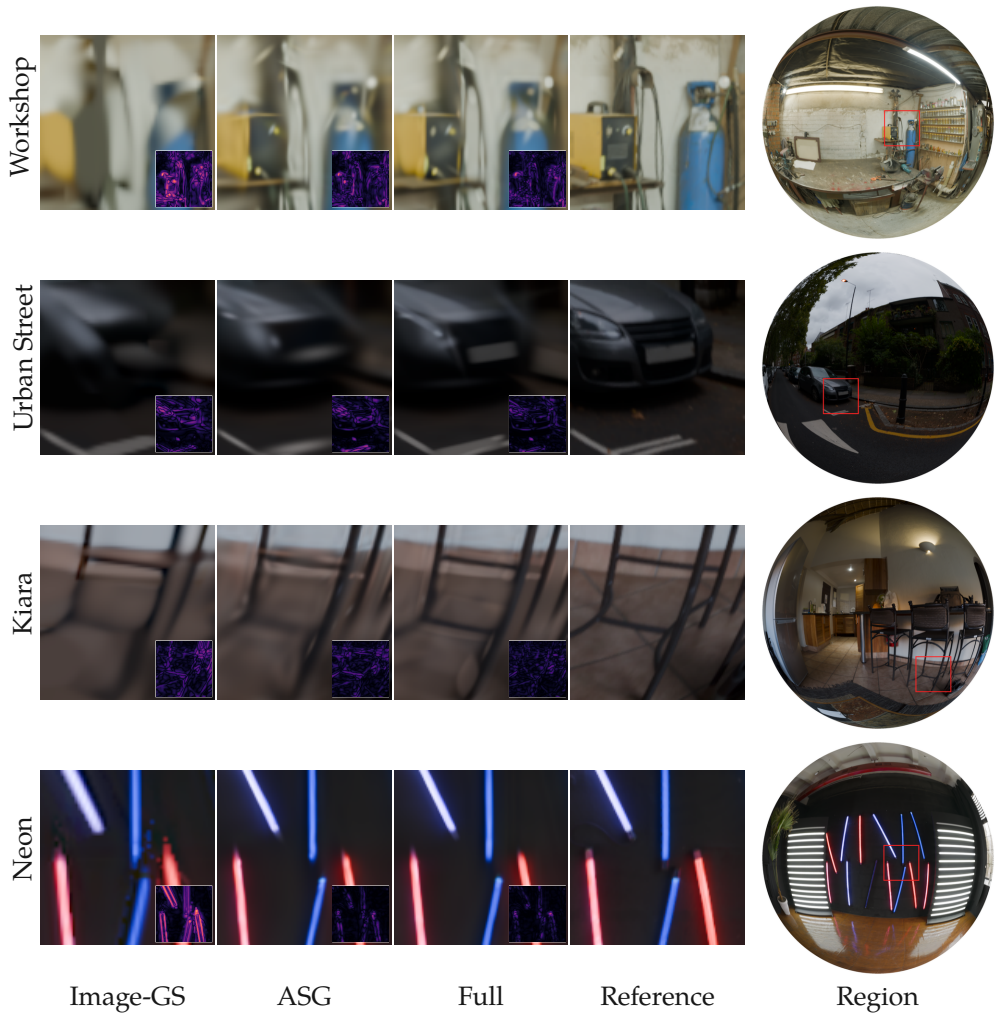


Figure 5.2. ***Local reconstruction at 10,000 primitives.*** *Registered crops of Workshop, Urban Street, Kiara Interior, and Neon Photostudio; the right column marks the region on a spherical view of the reference. Insets show FLIP error maps on a common scale (range 0–1).*

more: most of the equipment in Workshop and the front of the car in Urban Street dissolve, and the neon strips bleed into their surroundings.

5.3 Radial and Contour Adaptation

Table 5.2 separates the two exponents. Learning only the radial exponent p reduces the mean relative L_2 error by 24.6%, from 0.1849 to 0.1394, which accounts for the whole radiometric gain of the full model, and lowers PQ-LPIPS by 8.5%. Learning only the contour exponent q leaves the relative L_2 error essentially unchanged, 1.1% above the ASG, but improves all three image metrics. Adding q to the radial model improves the three image metrics again, and the full model has the best mean in each of them.

Table 5.2. ***Radial and contour adaptation.*** *Means over the fifteen environments at 10,000 primitives, before quantization. The best value of each metric is set in bold.*

Model	Rel. L_2 ↓	PQ-PSNR↑	PQ-SSIM↑	PQ-LPIPS↓
ASG	0.1849	42.28	0.9267	0.1759
Radial-only	**0.1394**	42.58	0.9309	0.1609
Contour-only	0.1870	42.40	0.9278	0.1734
Full	0.1418	**42.68**	**0.9324**	**0.1578**

The two exponents therefore act on different aspects of the reconstruction, as their roles in Chapter 3 suggest. At fixed mass and second moments, p moves mass between the core and the boundary of a primitive: a flat-topped primitive keeps the energy of a bounded emitter inside its outline instead of spreading it into Gaussian tails, and this is what the energy-dominated relative L_2 error rewards. The contour exponent changes the outline itself. It places boundaries and corners where they belong, which matters for visible structure but moves little energy. The full model has a slightly higher relative L_2 error than the radial-only model, 0.1418 against 0.1394, while being better in every image metric. A plausible reading is that, with a fixed number of primitives and an objective defined in the PQ domain (Section 4.3), part of the additional freedom goes to structure that the perceptual encoding rewards and the linear error barely registers.

5.4 How the Fitted Primitives Use Shape

The fitted parameters show how the optimizer uses the two exponents. Figure 5.3a pools the nearly 150,000 primitives of the full model over the fifteen environments. Shape is used broadly but in moderation. Only one primitive in ten stays close to the ASG in both exponents, within 5% of $p = 1$ and of $q = 2$. The radial exponent has a median of 1.18 and stays below 1.43 for three quarters of the primitives: most primitives are somewhat flatter than a Gaussian, and fewer than 1% reach the flat-top bound $p = 4$. The contour exponent spreads on both sides of the circle, with a median of 2.02 and the central half of the primitives between 1.83 and 2.37. Rectangular contours ($q > 2$) are slightly more common than diamond-like ones, and the rectangular bound $q = 16$ is almost never reached. Where the content is smooth, the primitives stay Gaussian: on the sky-only environments Kloppenheim and Qwantani Noon, two thirds and three quarters of the primitives keep q within 5% of two.

Shape use depends on the size and the elongation of a primitive. The smallest primitives, with an angular spread below 0.2°, comparable to a cell of the fitting quadrature, remain almost Gaussian: only 4% of them have $p > 1.5$, plausibly because too few quadrature samples lie under them to distinguish one profile from another. Among primitives with a spread between 0.8° and 3.2°, a third have

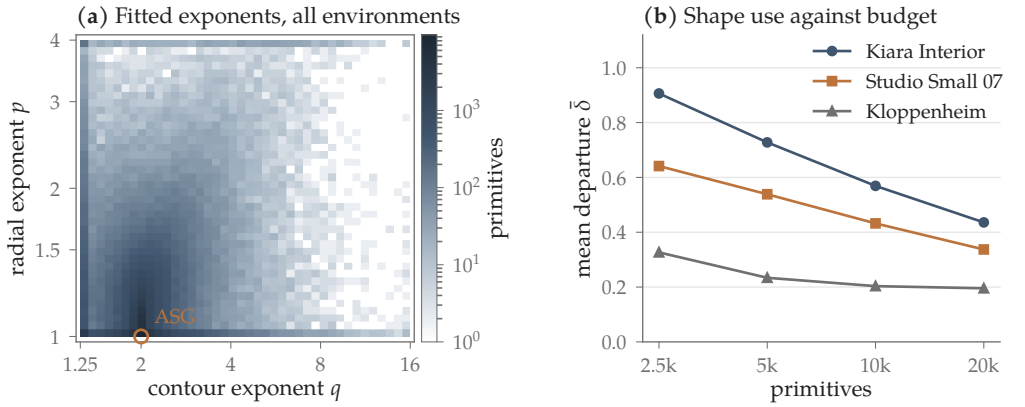


Figure 5.3. ***Fitted exponents of the full model.*** **(*a*)** *Joint distribution of p and q over the primitives of all fifteen environments; the circle marks the ASG.* **(*b*)** *Mean departure from the Gaussian member, $\bar{\delta}$, the average of $\ln p + |\ln(q/2)|$, at four budgets.*

$p > 1.5$ and more than half depart clearly from the circular contour. Elongation favors rectangles: the share of primitives with $q > 2.5$ grows from 16% among nearly isotropic primitives to 25% among those whose widths differ by a factor of three or more, while the share of diamond-like contours stays at about 14%. Elongated, flat-topped, rectangular primitives are what strip lights, window frames, and edges call for.

Figure 5.4 shows where these primitives lie. In Workshop, the fluorescent tubes are covered by elongated primitives with flat profiles, and the bright screen in the middle of the room is represented by a single flat-topped rounded rectangle, while the broad lobes on the ceiling and the floor remain close to Gaussian ellipses. In Neon Photostudio, the mullions of the window are tall, thin, flat-topped rectangles and the sill a long flat one. In both environments the fitted shapes follow the structures they cover, as Section 3.1 anticipated for bounded sources in a positive mixture.

Three further observations connect shape use to the results of the previous sections. First, the environments whose primitives depart most from the Gaussian tend to be those that gain most from shape adaptation. The energy-weighted departure has a Spearman rank correlation of 0.54 with both the reduction of the relative L_2 error and the PQ-PSNR gain over the ASG, and the unweighted departure one of 0.70 with the PQ-PSNR gain, moderate values for fifteen environments. Second, the exponents are used independently. The radial-only and contour-only models distribute their single exponent almost exactly as the full model does, with medians of 1.19 against 1.18 for p and 2.02 for q in both. Neither exponent compensates for the absence of the other, consistent with the additive gains of Section 5.3. Third, shape and primitive count trade off against each other (Figure 5.3b). With fewer primitives, each primitive departs further from the Gaussian: on Kiara Interior the share of primitives with $p > 1.5$ falls from 49% at

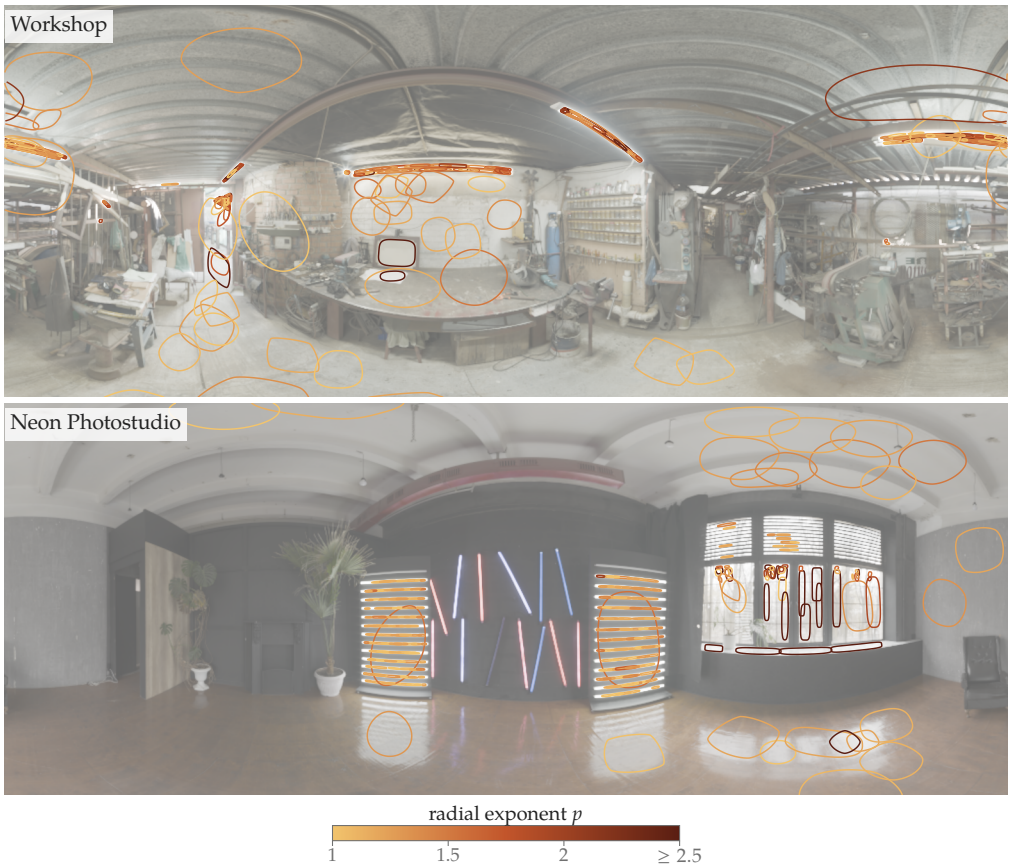


Figure 5.4. ***Shapes of the most energetic primitives.*** *The 300 primitives with the largest coefficients in Workshop and Neon Photostudio, drawn as* $1/e$ *footprint contours on the latitude–longitude reference and colored by their radial exponent. The panoramas are lightened, and their polar caps cropped, for legibility.*

2,500 primitives to 22% at 20,000, and the mean departure halves. The sky-only Kloppenheim stays close to the Gaussian at every budget.

5.5 Rate–Distortion

The comparisons so far count primitives, but a shape-adaptive primitive stores two more fields than an ASG. The rate–distortion study measures quality against storage. Three environments with different content, Kiara Interior, Kloppenheim, and Studio Small 07, are fitted at budgets of 2,500, 5,000, 10,000, and 20,000 primitives. The ASG and full models are encoded with the fixed-rate format of Section 4.6 and evaluated after decoding. Image-GS is placed at its nominal storage of 16 bytes per Gaussian, and all sizes are given in kilobytes of 1,000 bytes. Table 5.3 lists the plotted means together with the relative L_2 error.

Figure 5.5 shows the relative L_2 error and the three image metrics. At every budget the full model is better than the ASG model in each of them, and its

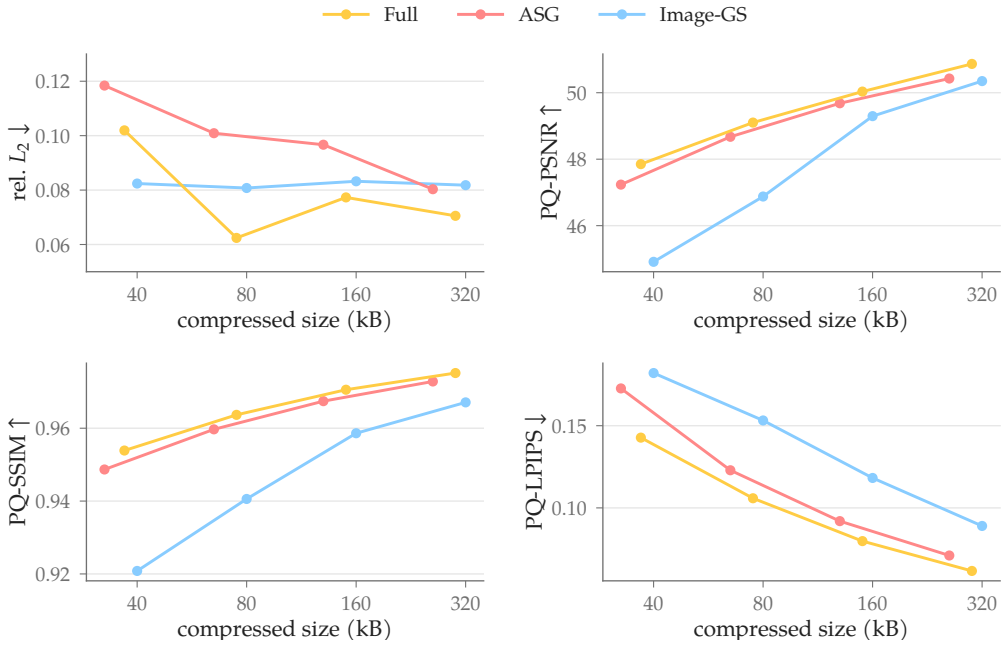


Figure 5.5. ***Quality against storage.*** *Means over Kiara Interior, Kloppenheim, and Studio Small 07 at 2,500, 5,000, 10,000, and 20,000 primitives. The ASG and full models are evaluated after decoding; the size axis is logarithmic.*

Table 5.3. ***Storage and quality across budgets.*** *Means over Kiara Interior, Kloppenheim, and Studio Small 07. The ASG and full models are evaluated after decoding, and their storage is the mean size of the encoded stream; Image-GS is placed at 16 bytes per Gaussian. The best value of each metric is set in bold.*

	Storage (kB)			Rel. L_2 ↓			PQ-PSNR↑		
Primitives	Image-GS	ASG	Full	Image-GS	ASG	Full	Image-GS	ASG	Full
2,500	40.0	32.5	36.9	**0.0824**	0.1184	0.1019	44.91	47.23	**47.85**
5,000	80.0	65.0	75.0	0.0808	0.1009	**0.0624**	46.88	48.67	**49.10**
10,000	160.0	130.0	150.0	0.0832	0.0967	**0.0773**	49.30	49.68	**50.03**
20,000	320.0	260.0	300.0	0.0818	0.0803	**0.0705**	50.35	50.43	**50.87**

curve lies on the better side of the ASG curve over the whole range of sizes, although each of its records is 15% larger. At 10,000 primitives the full model reaches 50.03 dB PQ-PSNR in 150 kB, against 49.68 dB in 130 kB for the ASG; the shape exponents account for the 20 kB difference. Quantization costs little: decoding lowers the mean PQ-PSNR by at most 0.06 dB at any budget, so the advantage of the fitted model survives the conversion to a fixed-rate stream without quantization-aware fine-tuning. Image-GS improves quickly with its budget and comes closest to the spherical models at the largest size, but it stays below the full model in every image metric at every budget. The absolute values

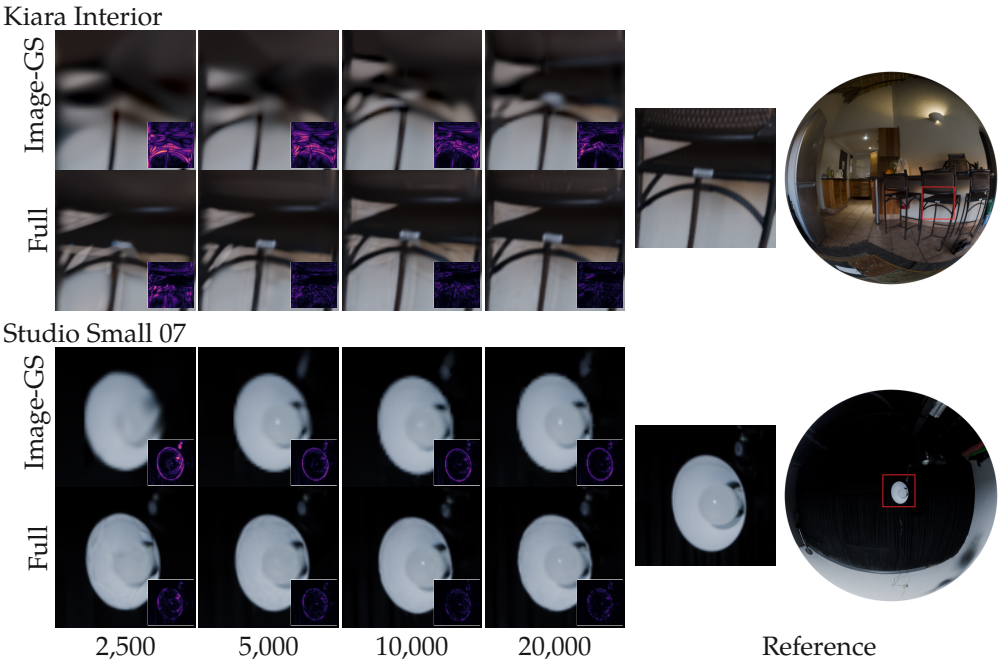


Figure 5.6. ***Reconstruction across budgets.*** *The same region reconstructed by Image-GS and by the full model with 2,500 to 20,000 primitives, next to the reference and its location on the sphere. Insets show FLIP error maps on a common scale (range 0–1).*

are higher than the fifteen-environment means of Table 5.1, mainly because the sky-only Kloppenheim reaches about 67 dB PQ-PSNR; they compare the models within the study, not with Table 5.1.

The relative L_2 error behaves differently from the image metrics. That of the full model is not monotonic in the budget: 0.1019, 0.0624, 0.0773, and 0.0705 from the smallest to the largest budget, with its minimum at 5,000 primitives. It remains below the error of the ASG model at every budget, and the gap is widest at 5,000 primitives: there the full model, in 75 kB, has a lower error than the ASG model reaches with 20,000 primitives in 260 kB. The error of Image-GS barely changes with the budget, between 0.081 and 0.083, and is lower than that of both spherical models only at 2,500 primitives. Its flat curve suggests that the error is set by the brightest sources, which additional Gaussians in the panorama do not place more accurately. The spherical objective is defined in the PQ domain, and additional primitives that improve the encoded reconstruction need not reduce the linear error of the brightest sources either; but the curve of the full model falls overall, and from 5,000 primitives on it stays below that of Image-GS.

Figure 5.6 shows the same trend in two regions. As the budget grows, the chair rails of Kiara Interior and the circular light of Studio Small 07 become sharper, and at every budget the full model resolves them more clearly than Image-GS. Primitive count and primitive shape are complementary sources of capacity: more

primitives resolve more local structure, while the exponents improve what each primitive contributes.

5.6 Comparison with an Image-Space Representation

Image-GS represents the same environments with planar Gaussians fitted to the latitude–longitude panorama. It differs from the spherical models in its domain, its kernel, its logarithmic HDR encoding, and its rendering rule, which normalizes each pixel by the total kernel weight instead of summing positive contributions. The comparison therefore measures two complete representations rather than isolating one design choice.

The full model is ahead of Image-GS on all four mean metrics of the map (Table 5.4). Its relative L_2 error is 29% lower, 0.1418 against 0.2006. In the image metrics the margins are smaller: 0.50 dB in PQ-PSNR, 0.0075 in PQ-SSIM, and 3.2% in PQ-LPIPS, and Image-GS has the higher PQ-PSNR on six of the fifteen environments (Table B.1), among them all three skies, whose smooth gradients its planar Gaussians reproduce well.

Table 5.4. ***Comparison with Image-GS.*** *Means over the fifteen environments at 10,000 primitives, before quantization. The last column is the root-mean-square difference between the material renderings lit by each reconstruction and by the reference, in display values* (×100*; Section 5.1*).

	Environment map				Renderings
Model	Rel. L_2 ↓	PQ-PSNR↑	PQ-SSIM↑	PQ-LPIPS↓	RMSE↓
Image-GS	0.2006	42.18	0.9249	0.1630	1.02
ASG	0.1849	42.28	0.9267	0.1759	0.41
Full	**0.1418**	**42.68**	**0.9324**	**0.1578**	**0.38**

The image metrics, however, describe only part of the difference. They are computed on the panorama, weight its texels uniformly, and after PQ encoding reward local texture more than the correct placement of energy. The relative L_2 error and the renderings measure what the environment does as illumination, and there the gap is wide. Materials lit by Image-GS deviate from the reference renderings 2.7 times more on average than those lit by the full model, and Image-GS produces the largest rendering error of the three models in thirteen of the fifteen environments (Table B.1). Street Lamp shows the contrast most clearly. Image-GS scores better than the full model on all three image metrics, yet its relative L_2 error is nearly four times higher, 0.5680 against 0.1501, and its rendering error more than six times higher. The panorama looks plausible, but the energy of the lamp is misplaced, and the scene it lights looks wrong (Figure 5.8). Even on Cannon, whose fine natural texture Image-GS reproduces best on every map metric, the full model gives the more accurate renderings.



Figure 5.7. ***Whole-sphere reconstructions at 10,000 primitives.*** *One hemisphere of six environments; the marked region is enlarged in the lower right corner.*

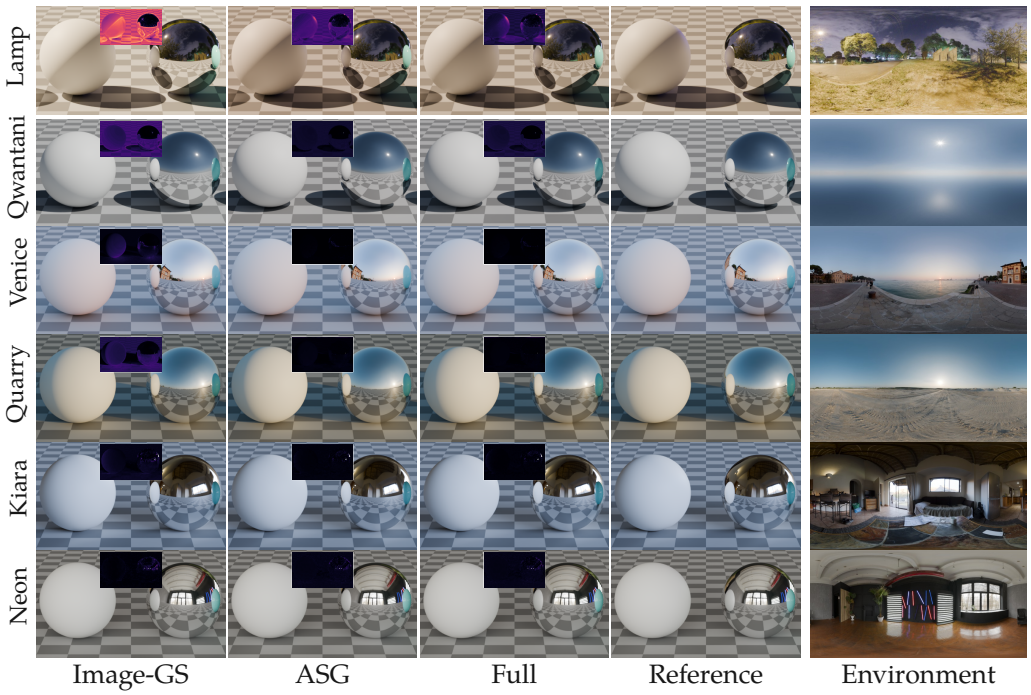


Figure 5.8. ***Rendering with reconstructed environments.*** *The diffuse and the mirror sphere of the material scene, lit by the reconstructions of Street Lamp, Qwantani Noon, Venice Sunset, Quarry, Kiara Interior, and Neon Photostudio at 10,000 primitives, and by the reference, with fixed camera, materials, and exposure. Insets show FLIP error maps on a common scale (range 0–0.5); the right column shows the reference environment.*

Figure 5.7 shows complete hemispheres of six environments. The spherical reconstructions are continuous over the whole sphere and preserve the dominant emitters, the horizon, and the broad gradients of illumination. Fine, irregular texture, as in Cannon and Xanderklinge, is the hardest content for every model, since it must be tiled by many small primitives whatever their shape.

5.7 Rendering with Reconstructed Environments

An environment map is ultimately used as lighting, and its errors are then filtered by the reflection equation (2.2): a diffuse surface integrates broad directional energy, a mirror reveals local structure, and cast shadows respond to the position and extent of the emitters. The rendering error defined in Section 5.1 measures these effects on one scene with four materials.

On average over the fifteen environments, the rendering error ($\times 100$) is 0.38 for the full model, 0.41 for the ASG model, and 1.02 for Image-GS. At the precision of Table B.1, the full model gives the most accurate renderings in nine environments and shares the lowest error with the ASG model in three more; the ASG model is ahead on Qwantani Noon, and Image-GS on Royal Esplanade and Urban Street. Figure 5.8 shows where the differences come from. Under Street Lamp, Image-GS

Table 5.5. ***Agreement between map measures and renderings.*** *Number of environments, out of fifteen, in which a map measure prefers the model of each pair whose rendering error is lower. Models at 10,000 primitives, before quantization; the highest total is set in bold.*

Pair	Rel. L_2	PQ-PSNR	PQ-SSIM	PQ-LPIPS
Image-GS vs. Full	13	7	9	5
Image-GS vs. ASG	11	7	9	4
ASG vs. Full	11	11	12	13
All pairs	**35**	25	30	22

lights the scene too brightly and too coolly: the floor and the diffuse sphere lose the warm tone of the reference, and the shadows become harder. Under Qwantani Noon, Image-GS darkens the floor shadow and changes the shading of the diffuse sphere, and under Quarry it gives the shaded side of the diffuse sphere a cooler tone. In each case the spherical reconstructions stay close to the reference in diffuse shading and in the cast shadows, consistent with their lower relative L_2 error on these environments. All models preserve the dominant structure in the mirror reflections.

The FLIP insets make these differences visible at a glance. On Qwantani Noon, Venice Sunset, and Quarry, Image-GS leaves an error over the floor and the diffuse sphere, while the error maps of both spherical models are nearly black. Street Lamp is the hardest environment. The error map of Image-GS is bright over the entire view, because the whole scene is lit too brightly, and the spherical models also leave a visible error on the floor and the diffuse sphere, smaller for the full model than for the ASG model. Between the two spherical models, the full model has the smaller error in these crops on Street Lamp and Quarry, and the ASG model on Qwantani Noon and Venice Sunset. Kiara Interior and Neon Photostudio are lit by large windows and long tubes, and all three reconstructions light the spheres closely; the insets show mostly small errors in the reflections of the windows. In this view of Neon Photostudio the error of Image-GS is even the smallest, although its rendering error over the four materials is the largest of the three models (Table B.1).

The renderings also allow a more general question: which measure of the map predicts how well a reconstruction lights a scene? For a pair of models and an environment, a map measure *agrees* with the renderings if it prefers the model with the lower rendering error. Table 5.5 counts the agreements over the fifteen environments for every pair of models.

The relative L_2 error is the best predictor. Between Image-GS and the full model, it identifies the model with the more accurate renderings in thirteen of the fifteen environments. The image metrics are close to chance: PQ-PSNR agrees in seven environments, PQ-SSIM in nine, and PQ-LPIPS in five. The comparison between

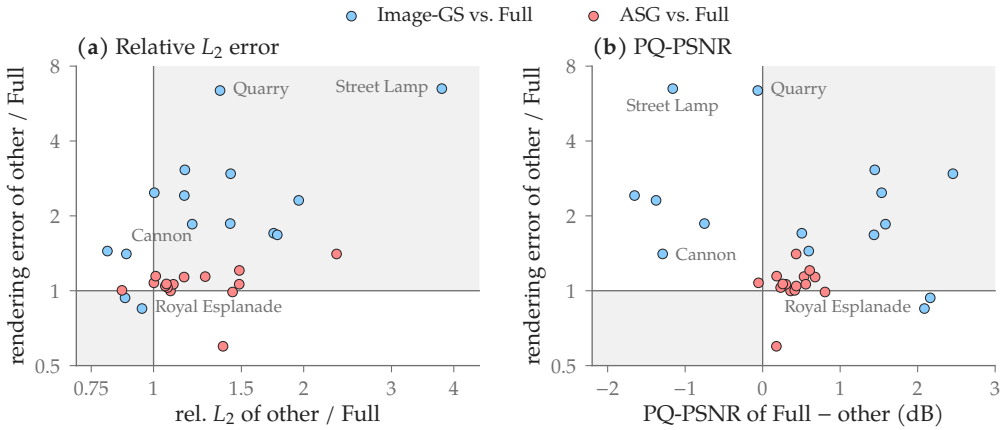


Figure 5.9. ***Map quality against rendering quality.*** *Each point is one environment. The horizontal axis compares Image-GS or the ASG model with the full model on the map, by* (***a***) *the ratio of relative L_2 errors and* (***b***) *the difference in PQ-PSNR; the vertical axis compares their rendering errors. In the shaded quadrants, the map measure and the renderings prefer the same model.*

Image-GS and the ASG model shows the same pattern. Between the two spherical models every measure agrees in eleven to thirteen environments. These models share their domain, their kernel family, and their objective, so their errors differ in degree rather than in kind, and a model that is better on the map tends to be better on every measure at once.

Figure 5.9 shows the individual environments. With the relative L_2 error, almost every point lies in the shaded quadrants, and the two ratios grow together: the largest gap in energy, on Street Lamp, goes with the largest gap in the renderings. The two exceptions are Cannon and Xanderklinge, environments of fine natural texture in which Image-GS has the lower relative L_2 error by 12% and 19% but still lights the scene less accurately. With PQ-PSNR, the points spread over all four quadrants. On six environments, among them Street Lamp, Quarry, and the three skies, Image-GS has the higher PQ-PSNR, yet its rendering error is 1.4 to 6.5 times that of the full model; on Royal Esplanade and Urban Street the full model has a PQ-PSNR more than 2 dB higher, yet Image-GS renders slightly better. The measures that judge the panorama as an image and the measure that judges it as a distribution of energy answer different questions, and only the second one is reliably related to the use of the map as lighting. Section 6.2 returns to this point.

Chapter 6

Discussion

The experiments of Chapter 5 support a consistent picture: shape adaptation makes a spherical mixture more accurate at a matched primitive count and per byte of storage, the two exponents play distinct roles, and the full model is a more accurate lighting representation than an adaptive image-space one. This chapter interprets these findings. It asks what shape buys and what it costs, why the error measures sometimes disagree, and what fitting on the sphere provides, and then traces the limitations of the work to their causes and the directions that follow from them.

6.1 Shape as Capacity per Primitive

A mixture can gain capacity in two ways. It can use more primitives, or it can make each primitive more expressive. The shape-adaptive kernel takes the second route at a small and fixed price: two scalars per primitive, which the compact encoding stores in 16 bits, about 15% of an ASG record. The two kinds of comparison in Chapter 5 answer different questions about this trade. At a matched primitive count the full model has more parameters than the ASG model, so its advantage there shows that shape is useful, but not yet that it is worth its cost. The rate–distortion study answers the second question. The full model is better than the ASG model in the relative L_2 error and in every image metric over the whole range of sizes, so an ASG mixture given the extra bytes that the exponents occupy, about 15% more records, does not reach it. For the environments and budgets of the study, two exponents per primitive are a better use of storage than additional Gaussian primitives.

The most likely reason is the positivity argument of Section 3.1: in a positive Gaussian mixture, the sharp edge of a bounded, flat source is paid for with many narrow lobes along its length. A flat-topped primitive with a rectangular contour describes the same source with its own interior and its own edge. Shape thus adds capacity precisely along the dimension in which a positive Gaussian mixture is least efficient. The per-scene results follow this account closely. The largest gains occur on Street Lamp, Studio Small 07, Workshop, and Royal Esplanade,

environments dominated by luminaires, framed openings, and strip lights, and the fitted primitives in these regions are elongated, flat-topped, and rectangular.

Shape does not help everywhere, and where it does not, the reasons are different. On the smooth skies of Kloppenheim and Industrial Sunset, the Gaussian is already the right shape, and the optimizer leaves most primitives close to it. Since the full model contains the ASG, it has little to gain and little to lose there. The small remaining differences, in either direction, most likely reflect different paths through a nonconvex objective rather than a difference in what the models can represent. Fine natural texture is a different case. Foliage, rock, and gravel, as in Cannon and Xanderklinge, consist of many small, irregular features. The primitives that represent them are small, and the analysis of Section 5.4 shows that the smallest primitives remain almost Gaussian, most likely because too few quadrature samples lie under them to tell one profile from another. Shape is a property of a primitive large enough to have an interior and a boundary. Texture below that scale is a matter of count, whatever the kernel.

The budget study adds a further nuance: shape and primitive count trade off. With fewer primitives, each one covers a larger and more structured feature, and the fitted exponents depart further from the Gaussian. As the budget grows, features are subdivided, primitives shrink, and the mixture assembles flat interiors and sharp edges from more and smaller lobes, so each primitive needs less shape. The full model remains ahead at every budget of the study, but the mechanism suggests that shape is most valuable where compactness matters most, at budgets small enough that each primitive must describe a whole feature.

How the optimizer uses shape is itself informative. Most primitives depart only moderately from the Gaussian, and the bounds of the admissible ranges are rarely reached, so the ranges chosen in Section 3.3 do not constrain the fits in practice. The exponents act as a refinement of the ASG rather than as a different representation, and the fitted primitives remain recognizable as lobes with a center, an orientation, and an extent. Their departures also do not look like noise. Environments whose primitives depart most from the Gaussian tend to be those that gain most from shape adaptation, so the fitted exponents are an interpretable by-product of the fit: a rough map of where the content is not Gaussian. Finally, the radial-only and contour-only models distribute their single exponent almost exactly as the full model does, and the gains of the two exponents add up. Neither exponent stands in for the other. This is the separation that the normalization of Section 3.4 was designed to produce. Since mass and second moments are fixed, the two exponents control only structure beyond second order, and they control different parts of it: the distribution of mass across nested contours, and the geometry of the contours themselves.

6.2 Radiometric and Perceptual Fidelity

Several results of Chapter 5 rank the same reconstructions differently depending on the measure. The radial-only model has the lowest relative L_2 error, while the full model is best in every image metric. On Industrial Sunset the full model improves all three image metrics but has a higher relative L_2 error than the ASG. Across budgets, the relative L_2 error of the full model is not monotonic, while its image metrics improve steadily. And on Street Lamp, Image-GS is better than the full model in all three image metrics, while its relative L_2 error is nearly four times higher. These disagreements are consistent with one explanation: the measures weigh the same errors differently, along two axes.

The first axis is the radiance level. The relative L_2 error is computed in linear radiance and is dominated by the brightest directions, where most of the energy lies. The PQ-domain metrics compress the range with the perceptual quantizer and, following the standard, clip at its nominal endpoint. With the calibration of Section 4.3, every linear value above about 49 is therefore encoded alike, and any error in the energy of a bright emitter beyond that level is invisible to them. Over much of its range the encoding is close to logarithmic, so a given relative error costs roughly as much in a dim corner as on a bright source. The second axis is direction. The relative L_2 error weights directions by solid angle, whereas the image metrics weight the texels of the latitude–longitude panorama uniformly and thus give the polar caps far more weight than the solid angle they cover.

The fitting objective lies between these measures. It is defined in the PQ domain, like the image metrics, but extended beyond the nominal endpoint, so that bright directions keep contributing with diminishing weight, and it weights directions by solid angle, like the relative L_2 error. When a model gains freedom, the fit tends to spend it where this objective rewards it most, on visible structure across the whole range of radiance, and only weakly on the exact energy of the brightest sources. The full model, with one more degree of freedom than the radial-only model, uses part of it for contours that the perceptual encoding rewards and the linear error barely registers. Additional primitives at larger budgets refine the encoded reconstruction and need not lower the linear error of the brightest emitter. These effects are better explained by what is optimized than by a defect of the representation, and they are the price of a deliberate choice: an objective in linear radiance would serve the emitters and leave the rest of the environment unresolved (Section 4.3).

Which measure is right depends on how the environment is used. Seen directly, as a backdrop or in a mirror, an environment is judged by its visible structure, and the image metrics are the relevant ones. Used as illumination, it is judged by its energy and where that energy comes from, and the relative L_2 error and the renderings are. These two views agree in the comparison between the spherical models and Image-GS: the full model has a lower relative L_2 error and lower

rendering errors, and Image-GS has the largest rendering error in thirteen of the fifteen environments. Between the ASG and the full model, the rendering errors are close, 0.41 against 0.38. Diffuse and glossy materials integrate the environment over broad lobes and forgive errors in the precise outline of a source, as long as its energy and extent are right; the gains of shape adaptation are more visible in the environment itself than in the shading it produces.

The agreement analysis of Section 5.7 turns this distinction into numbers. Between Image-GS and the full model, the relative L_2 error predicts which reconstruction lights the scene more accurately in thirteen of fifteen environments, and the image metrics are close to chance. The reason lies in where each measure looks. The image metrics are computed on the panorama, the domain in which Image-GS is fitted and whose texels its objective weights uniformly, and the logarithmic encoding of Image-GS treats relative errors evenly across radiance levels, much as the PQ encoding does. A measure that is close to a model's own objective and domain tends to favor that model. The renderings, in contrast, depend on the integral of the radiance against broad lobes, which the energy-weighted error on the sphere tracks and the encoded panorama hides. Street Lamp shows how far the two can diverge. Image-GS scores better there on all three image metrics, yet the energy of the lamp is misplaced, and the scene it lights looks wrong.

Between the two spherical models, all measures agree with the renderings about equally often. This is the expected case for models that share a domain, a kernel family, and an objective: their errors differ in degree, not in kind. The disagreements between measures therefore matter most when representations of different kinds are compared, which is exactly the comparison an evaluation of a new representation has to make. The lesson extends beyond this thesis. Image metrics on a panorama are necessary to judge an environment representation but not sufficient. An evaluation should include at least one energy-weighted measure on the sphere and the effect of the reconstruction on rendered images, and when the environment is meant as lighting these should carry more weight than the image metrics. This is why this thesis reports all three.

6.3 What Fitting on the Sphere Provides

Several properties of the representation hold by construction rather than by measurement. The reconstruction has no seam and no singular points, since nothing in it refers to a parameterization. It can be evaluated in any direction, so it can be rasterized into any parameterization at any resolution, and it is rotated by rotating the frames of its primitives. Its fit weights directions by solid angle, so capacity follows the content of the illumination rather than the texel density of the input. The experiments relied on these properties without remarking on them. The spherical models were fitted from octahedral rasters and evaluated, without conversion, on a rotated geodesic grid and on latitude–longitude rasters.

A representation fitted in image coordinates is tied to the raster it was fitted in. Its primitives do not wrap across the seam of the panorama, and its loss spends capacity on the polar caps in proportion to their texel count rather than to their content.

The comparison with Image-GS shows what these properties are worth in combination with the rest of each representation, but it does not isolate them. The two representations differ in their domain, in their kernel, in their HDR encoding, in their allocation, and in their rendering rule, and each of these differences plausibly contributes to the gap in linear radiance. The normalized rendering rule of Image-GS is perhaps the most consequential for illumination. It makes each pixel a weighted average of nearby colors rather than a sum of contributions, so a primitive has no integrated radiance of its own, and the energy of a source is not an additive quantity that the fit can allocate. The logarithmic encoding compresses the bright range, and the uniform weighting of texels spends capacity where the panorama is dense. What the comparison does show is that an adaptive image representation, fitted to the panorama with an HDR encoding, does not by itself produce accurate lighting, and that the spherical mixture produces markedly more accurate lighting while also remaining ahead, by smaller margins, on every mean metric computed in the domain of Image-GS. Attributing a share of the difference to the domain alone would require fitting the same kernel with the same objective in the plane and on the sphere, a controlled study that this thesis leaves open.

The scope of the representation is clearer when it is placed next to the families of Chapter 2. For direct lookup, a raster remains the simplest and fastest format: graphics hardware filters and mip-maps it, and one lookup costs the same whatever the content. A mixture costs work in proportion to the number of primitives that reach a direction, and without a culling structure (Section 6.4) it is better used as a source from which rasters of any resolution and parameterization are produced than as a format to be sampled per pixel. When only one light-transport operator is needed, a basis designed for it remains preferable. Nine spherical harmonic coefficients describe diffuse irradiance [RH01], and wavelet bases support all-frequency precomputed transport [NRH03]; a mixture of thousands of primitives would be wasteful in either role. When analytic shading with a few lobes is the goal, as in inverse rendering with spherical Gaussian lighting, the SG and ASG models keep their closed-form products and convolutions, which the adaptive family gives up. And learned priors over many environments address ill-posed problems, such as estimating lighting from photographs of objects or completing partial captures [GES22], that a fit to a single known environment does not attempt. The representation of this thesis serves a different purpose: a compact, explicit, and projection-independent description of one environment, faithful both to its energy and to its visible structure.

6.4 Limitations

The first limitations lie in the kernel family. The footprint (3.7) depends on the standardized coordinates only through $|u|$ and $|v|$, so every primitive is symmetric under the reflection of each tangent axis. A primitive cannot follow a boundary that is sharp on one side and soft on the other, a gradient across its extent, a curved or branching edge, or a junction of several structures. The admissible ranges bound the shapes further: $p \leq 4$ gives a steep but not a hard edge, and $p \geq 1$ excludes tails heavier than Gaussian, so the glow around a bright source, such as the aureole of the sun, must be built from several primitives. The three color channels share one kernel per primitive, so a source whose color changes across its extent needs several primitives as well. And the mixture is positive, so overlapping tails can never cancel. These choices keep each primitive simple, explicit, and physically meaningful, but they leave the representation of complex natural texture to the number of primitives, as Cannon and Xanderklinge show.

The normalization is exact in the plane, not on the sphere. The spherical mass of a primitive is its planar mass truncated at the horizon (3.14), so its coefficient is a nominal integrated radiance, accurate to within one percent only because the fitter bounds the concentrations from below. The same bound limits the angular spread of a primitive to about 18°, and very broad, smooth content must be tiled by several broad primitives. More consequential is what the family gives up in rendering (Section 3.6). The closed-form product and convolution approximations of the ASG rely on its Gaussian footprint and have no counterpart for general exponents, and the thesis provides no importance sampling procedure for the adaptive primitives. Shading with the representation therefore requires numerical integration against its primitives, or conversion to a prefiltered raster, rather than the analytic evaluation that SG and ASG lighting allows.

The fitting procedure has limitations of its own. The objective is nonconvex and the fit is local, starting from a content-driven initialization. Allocation only inserts: no primitive is removed, split, or merged, a primitive that becomes redundant as its neighbors adapt keeps its share of the budget, and each inserted primitive is confined to a cone around its insertion direction. Batches of insertions are accepted greedily, by their immediate effect on the objective. Shape, moreover, is never estimated from the data. Every primitive starts at the Gaussian member, and the exponents are learned at the smallest rate of the optimizer within a fixed number of steps, so the fitted distribution of shapes reflects the optimization as well as the content. The moderate departures from the Gaussian observed in Section 5.4 are what this procedure found, not necessarily what the content would support. The fit is also dense: the fitter evaluates every primitive at every sample of a batch, with no structure that restricts a sample to the primitives that reach it. Fused GPU kernels keep this affordable (Section A.1), and a fit of the full model with 10,000 primitives on the RTX 4090 takes about two minutes, including evaluation and

export. Image-GS fits in under a minute on the same GPU with dedicated CUDA rasterization kernels, so its run time is not directly comparable. Finally, the fit minimizes a perceptual objective and is not designed to minimize the linear error that measures energy, with the consequences discussed in Section 6.2.

The experiments isolate the shape of the primitives and little else. All spherical models share the moment-preserving normalization, the adaptive quadrature, the perceptual objective, and the residual-guided allocation, and none of these was ablated separately: no reported fit uses an unnormalized footprint, a uniform quadrature, a linear objective, or a different allocation rule. Their benefits are argued from their construction in Chapters 3 and 4, not measured.

The encoding is deliberately simple. It stores every field at a fixed rate, with neither entropy coding nor quantization-aware refinement, so the decoded sizes are upper bounds on what the fitted mixtures require. Compactness, moreover, is shown relative to ASG mixtures and to Image-GS, at matched primitive count and storage. It was not compared with production compression of the panorama itself in HDR image or texture formats, so whether a mixture is smaller than a well-compressed raster of comparable quality remains open. The mixture has no culling hierarchy, so evaluating it in one direction touches every primitive unless the evaluator restricts the search, and the cost of evaluation was not benchmarked: the small, constant overhead of a shape-adaptive primitive over an ASG (Section 3.6) rests on counting operations.

The protocol, finally, limits how far the results generalize. Every fit uses a single random seed, so small per-scene differences, such as those on the sky-only environments, cannot be separated from the variation between optimization runs; the consistency of the gains across environments and their size on structured scenes do not depend on such differences. The fifteen environments come from one library and are fitted from inputs of one resolution, and the rate–distortion study covers three of them. The renderings use one scene with four materials and a fixed camera, a small sample of the uses an environment map has, and no part of the evaluation measures the representation inside a larger light-transport or inverse-rendering task.

6.5 Future Work

The most direct extension concerns the footprint itself. Asymmetric footprints, with different exponents or concentrations on the two sides of each tangent axis, would let one primitive represent a boundary that is sharp on one side only, or a source brighter at one end. The insertion step of Section 4.5 already fits a density with a linear tilt in the tangent coordinates and discards it; an asymmetric footprint could use that tilt. Piecewise or mixture-valued footprints, with internal structure inside one frame, could reduce the number of primitives that fine natural texture requires. Each of these departs from central symmetry, so the mean direction of a

primitive would no longer coincide with its center, and the normalization would have to decide which moments remain fixed. An exact spherical normalization, through a precomputed table of the spherical mass (3.14) over the concentrations and exponents, would make every coefficient an exact integrated radiance and would allow broader primitives than the current bound on the concentrations.

The second direction is rendering with the representation. Products and convolutions with reflectance lobes could be approached through tables precomputed in standardized coordinates over the two exponents, or by approximating each adaptive primitive with a few ASGs when an analytic operator is required. Importance sampling appears more accessible, because the lift reduces it to sampling the planar footprint. By the lift identity (3.5), a point $\mathbf{u}$ drawn from $\varphi_{p,q}$ and mapped to the direction with orthographic coordinates $(u/\sqrt{\lambda},\ v/\sqrt{\mu})$ in the frame of the primitive is distributed on the sphere with density K, provided it falls inside the ellipse $\mathcal{E}_{\lambda,\mu}$; the rare points outside it, a fraction $1 - m$ below one percent, can simply contribute nothing. The footprint depends on $\mathbf{u}$ only through r_q, so $\mathbf{u}$ factors into an independent radius and a point on the unit superellipse. By the shell decomposition (3.11), the squared q-radius $x = r_q(\mathbf{u})^2$ of such a point has density proportional to $e^{-(b_{p,q}x)^p}$, and the substitution used there shows that

$$\tau = \left(b_{p,q}\, r_q(\mathbf{u})^2\right)^p \sim \text{Gamma}(1/p), \tag{6.1}$$

with shape $1/p$ and unit scale. A gamma variate τ thus gives the radius $r_q = (\tau^{1/p}/b_{p,q})^{1/2}$. A point on the unit superellipse with the matching distribution is obtained by dividing a pair of independent variates with density proportional to $e^{-|x|^q}$ by their q-radius [SZ90]; each variate is a random sign times the q-th root of a gamma variate with shape $1/q$. Choosing a primitive with probability proportional to the luminance of its coefficient then draws directions with a density proportional to the luminance of $\widehat{L}$, the distribution that environment-map sampling approximates with tabulated or structured densities [Aga+03; PJH23]. It is the unit mass of the footprint that makes this choice exact, since it makes $\mathbf{a}_i$ the integrated radiance of primitive i. The construction has not been implemented or evaluated here, but it suggests that the explicit, normalized form of the primitives can serve rendering directly rather than only through a raster.

The fitting procedure would benefit from allocation that can also remove, split, and merge primitives, so that capacity freed by shape adaptation is returned to the budget, and from initializing shape from the data, for instance from the flatness of the local signal, rather than always from the Gaussian. Restricting each sample to the primitives that reach it would shorten the fit further. A fit with a mixed objective, or with a linear-radiance term restricted to the brightest directions, would show how the trade between visible structure and energy described in Section 6.2 can be controlled. On the storage side, the fitted exponents are concentrated near the Gaussian, which is exactly the situation in which entropy

coding pays off, and quantization-aware refinement and a spatial hierarchy for culling and level of detail would make the representation competitive as a format to be sampled directly. Its evaluation cost should then be measured on graphics hardware.

The evaluation, finally, should move into transport space: renderings with a wider range of materials and scenes, prefiltered glossy lighting, irradiance, and the use of the representation as the lighting model of an inverse-rendering system, in which flat-topped, rectangular primitives would give windows and area lights a natural parameterization. Multiple seeds and a larger and more varied set of environments would sharpen the per-scene conclusions, and the controlled comparison between fitting the same kernels in the plane and on the sphere, proposed in Section 6.3, would measure what the spherical domain contributes on its own.

The agreement analysis of Section 5.7 also points to a better measure for maps that are meant as lighting. Path-traced renderings are expensive, depend on one scene, and add Monte Carlo noise, while the relative L_2 error weights every direction by its energy alone and ignores how materials blur it. Between the two lies a measure computed on the sphere after filtering: the error of the map convolved with a small set of reflectance lobes, from a cosine lobe for diffuse surfaces to narrow lobes for glossy and mirror-like ones, each compared by quadrature over directions. Such a measure would be cheap, deterministic, and independent of any scene, and it would weight each band of angular frequency by how much it matters for a given class of materials. Its agreement with rendered images could be tested exactly as in Table 5.5, over more environments, materials, and pairs of representations, and, if it holds, it could replace the image metrics of the panorama as the primary measure for environment representations used as lighting.

Chapter 7

Conclusion

This thesis set out to represent an HDR environment map as what it is: a function of direction, defined on the sphere rather than on the rectangle in which it is usually stored. Such a function is hard to represent compactly. A small number of bright, bounded sources carries most of its energy and decides how a scene is lit, while large regions of dim light carry the structure that a viewer sees in reflections. Mixtures of explicit primitives suit this heterogeneity, because they place capacity where the signal needs it, and on the sphere the anisotropic spherical Gaussian is their natural element. Every ASG, however, has the same shape: a Gaussian falloff and elliptical contours. In a positive mixture, where the tails of one lobe can never be canceled by another, this makes the flat interiors, sharp edges, and rectilinear outlines of windows, luminaires, and strip lights expensive to represent.

The representation developed in Chapter 3 gives each primitive a shape of its own. It starts from a reading of the ASG as a planar Gaussian footprint lifted to the sphere through the orthographic chart, in which the clamped cosine is the Jacobian of the lift. This view separates the roles of the parameters: the frame places and orients a primitive, the concentrations set its extent, the footprint sets its shape, and the coefficient sets its magnitude. Shape adaptation then means replacing the Gaussian footprint by a richer family, and the family introduced here adds two exponents, one that moves the radial profile from the Gaussian toward a flat top and one that moves the contours from diamonds through ellipses to rounded rectangles. A moment-preserving normalization fixes the mass and the second moments of every footprint, so that the coefficient remains the integrated radiance of a primitive, the concentrations remain its extent, and the exponents describe shape alone. The conventional ASG is the member with a Gaussian profile and elliptical contours. The lift also gives exact statements about every member: its center is its mean direction, its frame axes are the principal axes of its spread, and its spherical mass differs from the planar one only by a small and explicit truncation.

Chapter 4 fitted this representation directly on the sphere. An adaptive geodesic quadrature weights every direction by its solid angle and refines itself around edges and emitters, a perceptual objective in the PQ domain balances errors across the dynamic range, and new primitives are inserted where the positive

residual shows that the reconstruction falls short. Because the second moments of a primitive do not depend on its shape, an inserted primitive can be matched to the covariance of the residual it is meant to fill and left to acquire its shape during optimization. The same procedure fits the ASG model and every shape-adaptive model, and a compact fixed-rate encoding makes the storage cost of shape explicit, so that the comparisons measure the primitives and nothing else.

The experiments of Chapter 5 show that shape adaptation increases what a spherical mixture can represent. At a matched primitive count, the full model improves every mean metric over the ASG, consistently across the environments, and most where the illumination is concentrated in bounded sources. The two exponents play distinct and complementary roles: the radial exponent carries the radiometric gain, by keeping the energy of a source inside its outline, and the contour exponent adds visible structure, by placing boundaries and corners where they belong. The fitted primitives use their shape broadly but in moderation, according to the content they cover, and more when primitives are scarce. The advantage survives quantization and holds per byte of storage, so over the budgets studied, two exponents per primitive are a better use of storage than additional Gaussian primitives. Against an adaptive image-space representation, the full model is ahead on every mean metric computed on the panorama, by small margins, and clearly ahead in linear radiance and in the renderings it lights.

As Chapter 6 argued, these results admit a simple explanation. Shape adds capacity along the very dimension in which a positive Gaussian mixture is least efficient, the boundary of a bounded source, and it adds nothing where the Gaussian is already the right shape or where the content is finer than a primitive. The evaluation also carries a lesson of its own. Measures in linear radiance and in perceptual encodings weigh the same errors differently, and an environment map that looks plausible as a panorama can still misplace the energy that lights a scene. A representation of illumination should therefore be judged on the sphere, by its energy, and through the images it renders, as well as by its appearance.

The work leaves open what its limitations point to. The footprint is centrally symmetric, and asymmetric or structured footprints could take on boundaries and textures that still require many primitives. The adaptive family gives up the closed-form rendering operators of the ASG, and deriving products, convolutions, and sampling procedures for it would let the representation serve rendering directly; the lift, which reduces each primitive to a planar footprint, suggests where to begin. Entropy coding, a hierarchy for culling, and an evaluation in transport space would turn the representation into a practical lighting format. More broadly, the thesis shows that kernel shape is a useful degree of freedom on the sphere, as it has proved to be in the image plane and in space, and that it can be added without giving up what makes spherical lobes attractive. An environment map then becomes a set of explicit primitives, each with a position, an orientation, an extent, a color, and a shape of its own, from which the illumination can be evaluated in any direction.

Appendix A

Implementation and Experimental Details

This appendix collects the settings behind the experiments of Chapter 5: the software and data, the fitting configuration, the Image-GS reference, the evaluation measures, and the renderings.

A.1 Software and Data

The fitting procedure of Chapter 4 is implemented in PyTorch, and `torch.compile` fuses the evaluation and its gradients into a few GPU kernels without changing the computation. Every fit, of the spherical models and of Image-GS, runs with CUDA on a single NVIDIA GeForce RTX 4090. PQ-LPIPS uses the LPIPS package, version 0.1.4. The error maps in the figures are computed with the FLIP evaluator, version 1.7 [And+20], and display mapping uses Blender's AgX configuration through OpenColorIO.

The environments are the Poly Haven latitude–longitude EXRs [Pol26], at 4k resolution where available. Non-finite values are set to zero and negative values are clamped to zero. The 1024×1024 octahedral and 1024×512 latitude–longitude rasters are then resampled from this source: each output texel averages 4×4 bilinear lookups at its subtexel centers, weighted by their solid angles, so that its value is an average of the radiance over the texel.

A.2 Fitting Configuration

Table A.1 lists the settings of every spherical fit reported in Chapter 5. All four spherical models share them; the models differ only in which exponents are optimized (Section 4.1), and every fit uses the same random seed.

A.3 Image-GS

Image-GS [Zha+25] is run with its official implementation, at commit 0308836, in its default configuration and optimization schedule, with gamma correction

Table A.1. ***Fitting parameters.*** *Settings shared by the ASG, radial-only, contour-only, and full models. N is the requested number of primitives; energies are radiance times solid angle.*

Setting	Value
Budget and schedule	
Requested primitives N	10,000; 2,500 to 20,000 in Section 5.5
Initial primitives	$\lceil 0.35N \rceil$
Optimization steps	10,000
Mini-batch	8,192 samples, two batches per step
Insertion stages	every 1,000 steps from step 1,000 to 65% of the fit, each adding up to $0.15N$
Optimizer	
Adam learning rates	centers and angle 8×10^{-3}; log-concentrations 4×10^{-3}; coefficients 2×10^{-2}; $\log p$, $\log(q/2)$ 10^{-3}
Schedule	constant for 50% of the steps, then cosine decay to 10%
Gradient clipping	global norm 1
Constraints and penalty	
Concentrations	$\lambda, \mu \in [5, 10^6]$, ratio at most 64
Exponents	$p \in [1, 4]$, $q \in [1.25, 16]$
Soft concentration cap	from 5×10^4 to 10^6, log-linearly over the first 1,000 steps; weight 10^{-5} on the squared log excess
Objective and quadrature	
Calibration	linear value 1 at 203 cd/m^2; PQ extended above 10,000 cd/m^2
Charbonnier ϵ	10^{-3} in PQ units
Base grid	level 8, 1,310,720 cells, for the 1024×1024 octahedral targets
Refinement tests	RGB variation above 10% of the cell energy plus 1% of the mean cell energy; luminance peak above the mean cell luminance energy
Initialization and insertion	
Initial density	10% uniform, 90% content score clipped at its 99.5% quantile
Initial anisotropy	width ratio at most 4
Perceptual residual	0.65 logarithmic, 0.35 tone-mapped at exposures -2, 0, $+2$ EV
Candidates	four times the insertion count; suppression radius 0.75 allocation radii
Center confinement	cone of half the allocation radius around the insertion direction

disabled and without downsampling of the target. Its input is the 1024×512 latitude–longitude raster, encoded per RGB channel as

$$y = \frac{\log\bigl(1 + L/s\bigr)}{\log\bigl(1 + M/s\bigr)} \in [0, 1],$$

where s is the solid-angle-weighted 95th percentile of the luminance of the octahedral reference and M the largest value of the two reference rasters. The rendered output is decoded by the inverse transform, with negative values set to zero and no upper limit. On the sphere, Image-GS is evaluated by bilinear interpolation of its decoded 1024×512 panorama.

A.4 Evaluation Measures

The relative L_2 error (5.1) is evaluated on a level-nine geodesic grid with 5,242,880 samples, rotated by $\pi/7$ about the axis $(1, 2, 3)$ so that its samples do not coincide with those of the fitting quadrature. The image metrics compare 1024×512 latitude–longitude rasters of the reference and of the reconstruction and weight every texel and color channel equally. For the PQ-domain metrics, each RGB channel is encoded with the perceptual quantizer, a linear value of one mapped to $203\,\mathrm{cd/m^2}$, and clipped at $10{,}000\,\mathrm{cd/m^2}$, so that the encoded values lie in $[0, 1]$; PQ-PSNR uses a peak of one. PQ-SSIM averages the three channels and uses an 11×11 Gaussian window with $\sigma = 1.5$, the constants $K_1 = 0.01$ and $K_2 = 0.03$, and only windows that lie inside the image. PQ-LPIPS uses LPIPS version 0.1 with the AlexNet backbone on the encoded images, mapped from $[0, 1]$ to $[-1, 1]$.

The FLIP error maps shown as insets are not used as a measure. They compare the displayed sRGB images of a reconstruction and of the reference, after the same display mapping, with the low-dynamic-range FLIP at 67 pixels per degree, and are colored with the magma map. The maps of the spherical views span the full FLIP range from zero to one. Those of the renderings (Figure 5.8) are computed on the complete displayed renderings, cropped like them, and colored on the range from zero to 0.5, so that the small errors of the spherical models remain visible.

A.5 Rendering

Every environment, reference or reconstructed, is supplied to Blender's Cycles path tracer as a 2048×1024 latitude–longitude map with linear interpolation. The scene contains four spheres of unit radius in a row on a checkered ground plane: glass, a white diffuse material with roughness 0.55, a chrome mirror with roughness 0.03, and a glossy cyan material with roughness 0.12. A perspective camera with a 55 mm lens frames them at a resolution of 1024×256 pixels. Cycles uses 512 samples per pixel with adaptive sampling at a noise threshold of 0.003, OpenImageDenoise, a fixed seed, up to 16 bounces, of which 6 diffuse and 12 glossy, and no sample clamping. The renderings are mapped to display values with AgX and an exposure set by the 95th percentile of the luminance of the rendering lit by the reference, the same for every reconstruction of an environment. The rendering error of Section 5.1 is the root-mean-square difference of these display values over pixels and channels, computed on the complete rendering. Figure 5.8 shows a 448×256 crop of the same renderings, centered on the diffuse and the mirror sphere, identical for every model and environment.

Appendix B

Per-Scene Results

This appendix lists the per-scene values behind statements of Chapter 5 that are not visible in its tables. The ASG and full models are compared per scene in Table 5.1; Table B.1 adds the map metrics of Image-GS and the rendering error of the three models.

Table B.1. ***Image-GS and rendering error per scene.*** *Map metrics of Image-GS at 10,000 primitives, set in bold where they are better than those of the full model in Table 5.1 at the displayed precision, and the rendering error of the three models* (×100)*, with the lowest in bold.*

	Image-GS				Rendering RMSE↓		
Environment	Rel. L_2 ↓	PQ-PSNR↑	PQ-SSIM↑	PQ-LPIPS↓	Image-GS	ASG	Full
Artist Workshop	0.1379	36.14	0.9182	**0.1887**	0.451	0.277	**0.244**
Cannon	**0.0220**	**39.44**	**0.9511**	**0.0982**	0.169	**0.120**	**0.120**
Industrial Sunset	0.1264	**60.76**	0.9993	**0.0030**	0.059	**0.032**	**0.032**
Kiara Interior	0.1671	35.76	0.9039	0.2265	0.487	0.224	**0.196**
Kloppenheim	0.0466	**69.10**	0.9999	**0.0002**	0.023	**0.010**	**0.010**
Neon Photostudio	0.2113	35.22	0.9539	**0.1691**	0.827	0.516	**0.486**
Quarry	0.2400	**41.65**	0.9539	0.1099	0.737	0.132	**0.116**
Qwantani Noon	0.3198	**67.23**	0.9999	**0.0001**	1.502	**0.389**	0.651
Royal Esplanade	**0.4736**	33.68	0.9136	0.1404	**0.908**	1.058	1.071
Street Lamp	0.5680	**36.61**	**0.8913**	**0.2744**	7.804	1.689	**1.202**
Studio Small 07	0.0360	43.03	0.9720	0.1279	0.282	0.115	**0.096**
Urban Street	**0.0769**	27.44	0.7891	0.3542	**0.589**	0.648	0.629
Venice Sunset	0.2825	40.36	0.9613	0.1236	0.356	0.122	**0.116**
Workshop	0.2236	34.80	0.8829	0.2352	0.550	0.348	**0.328**
Xanderklinge	**0.0781**	31.45	0.7834	**0.3935**	0.581	0.429	**0.402**
Mean	0.2006	42.18	0.9249	0.1630	1.022	0.407	**0.380**

References

[Aga+03] S. Agarwal, R. Ramamoorthi, S. Belongie, and H. W. Jensen. "Structured Importance Sampling of Environment Maps." In: *ACM Transactions on Graphics* 22.3 (2003), pp. 605–612. DOI: 10.1145/882262.882314.

[And+20] P. Andersson, J. Nilsson, T. Akenine-Möller, M. Oskarsson, K. Åström, and M. D. Fairchild. "FLIP: A Difference Evaluator for Alternating Images." In: *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 3.2 (2020), 15:1–15:23. DOI: 10.1145/3406183.

[AMS08] T. O. Aydın, R. K. Mantiuk, and H.-P. Seidel. "Extending Quality Metrics to Full Luminance Range Images." In: *Human Vision and Electronic Imaging XIII*. Vol. 6806. Proceedings of SPIE. 2008, 68060B. DOI: 10.1117/12.765095.

[Bar81] A. H. Barr. "Superquadrics and Angle-Preserving Transformations." In: *IEEE Computer Graphics and Applications* 1.1 (1981), pp. 11–23. DOI: 10.1109/MCG.1981.1673799.

[Bin74] C. Bingham. "An Antipodally Symmetric Distribution on the Sphere." In: *The Annals of Statistics* 2.6 (1974), pp. 1201–1225. DOI: 10.1214/aos/1176342874.

[BN76] J. F. Blinn and M. E. Newell. "Texture and Reflection in Computer Generated Images." In: *Communications of the ACM* 19.10 (1976), pp. 542–547. DOI: 10.1145/360349.360353.

[Bos+21] M. Boss, R. Braun, V. Jampani, J. T. Barron, C. Liu, and H. P. A. Lensch. "NeRD: Neural Reflectance Decomposition from Image Collections." In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021, pp. 12664–12674. DOI: 10.1109/ICCV48922.2021.01245.

[Cha+94] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. "Two Deterministic Half-Quadratic Regularization Algorithms for Computed Imaging." In: *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. Vol. 2. 1994, pp. 168–172. DOI: 10.1109/ICIP.1994.413553.

[Cig+14] Z. H. Cigolle, S. Donow, D. Evangelakos, M. Mara, M. McGuire, and Q. Meyer. "A Survey of Efficient Representations for Independent Unit Vectors." In: *Journal of Computer Graphics Techniques* 3.2 (2014), pp. 1–30.

[Cla+24] P. Clausen, L. Ma, M. He, A. L. Taşel, O. Pilarski, and P. Debevec. "Fitting Spherical Gaussians to Dynamic HDRI Sequences." In: *SIGGRAPH Asia 2024 Posters*. 2024, 96:1–96:3. DOI: 10.1145/3681756.3698208.

[Deb98] P. Debevec. "Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-Based Graphics with Global Illumination and High Dynamic Range Photography." In: *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques* (*SIGGRAPH*). 1998, pp. 189–198. DOI: 10.1145/280814.280864.

[Deb02] P. Debevec. "Image-Based Lighting." In: *IEEE Computer Graphics and Applications* 22.2 (2002), pp. 26–34. DOI: 10.1109/38.988744.

[DM97] P. Debevec and J. Malik. "Recovering High Dynamic Range Radiance Maps from Photographs." In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques* (*SIGGRAPH*). 1997, pp. 369–378. DOI: 10.1145/258734.258884.

[Duf+17] T. Duff, J. Burgess, P. Christensen, C. Hery, A. Kensler, M. Liani, and R. Villemin. "Building an Orthonormal Basis, Revisited." In: *Journal of Computer Graphics Techniques* 6.1 (2017), pp. 1–8.

[Fis53] R. A. Fisher. "Dispersion on a Sphere." In: *Proceedings of the Royal Society of London. Series A* 217.1130 (1953), pp. 295–305. DOI: 10.1098/rspa.1953.0064.

[Fri12] J. R. Frisvad. "Building an Orthonormal Basis from a 3D Unit Vector Without Normalization." In: *Journal of Graphics Tools* 16.3 (2012), pp. 151–159. DOI: 10.1080/2165347X.2012.689606.

[GES22] J. A. D. Gardner, B. Egger, and W. A. P. Smith. "Rotation-Equivariant Conditional Spherical Neural Fields for Learning a Natural Illumination Prior." In: *Advances in Neural Information Processing Systems* (*NeurIPS*). Vol. 35. 2022, pp. 26309–26323. DOI: 10.52202/068431-1908.

[Gór+05] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere." In: *The Astrophysical Journal* 622.2 (2005), pp. 759–771. DOI: 10.1086/427976.

[Gre+06] P. Green, J. Kautz, W. Matusik, and F. Durand. "View-Dependent Precomputed Light Transport Using Nonlinear Gaussian Function Approximations." In: *Proceedings of the Symposium on Interactive 3D Graphics and Games* (*I3D*). 2006, pp. 7–14. DOI: 10.1145/1111411.1111413.

[Gre86] N. Greene. "Environment Mapping and Other Applications of World Projections." In: *IEEE Computer Graphics and Applications* 6.11 (1986), pp. 21–29. DOI: 10.1109/MCG.1986.276658.

[Ham+24] A. Hamdi, L. Melas-Kyriazi, J. Mai, G. Qian, R. Liu, C. Vondrick, B. Ghanem, and A. Vedaldi. "GES: Generalized Exponential Splatting for Efficient Radiance Field Rendering." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). 2024, pp. 19812–19822. DOI: 10.1109/CVPR52733.2024.01873.

[Han+07] C. Han, B. Sun, R. Ramamoorthi, and E. Grinspun. "Frequency Domain Normal Map Filtering." In: *ACM Transactions on Graphics* 26.3, 28 (2007). DOI: 10.1145/1239451.1239479.

[Hel+25] J. Held, R. Vandeghen, A. Hamdi, A. Deliege, A. Cioppa, S. Giancola, A. Vedaldi, B. Ghanem, and M. Van Droogenbroeck. "3D Convex Splatting: Radiance Field Rendering with 3D Smooth Convexes." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2025, pp. 21360–21369. DOI: 10.1109/CVPR52734.2025.01990.

[Hua+25] Y.-H. Huang, M.-X. Lin, Y.-T. Sun, Z. Yang, X. Lyu, Y.-P. Cao, and X. Qi. "Deformable Radial Kernel Splatting." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2025, pp. 21513–21523. DOI: 10.1109/CVPR52734.2025.02004.

[ITU25] ITU-R. *Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange*. Recommendation BT.2100-3. International Telecommunication Union, 2025.

[ITU26] ITU-R. *Guidelines for Operational Practices in High Dynamic Range Television Production*. Report BT.2408-9. International Telecommunication Union, 2026.

[Kaj86] J. T. Kajiya. "The Rendering Equation." In: *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 1986, pp. 143–150. DOI: 10.1145/15886.15902.

[Ken82] J. T. Kent. "The Fisher–Bingham Distribution on the Sphere." In: *Journal of the Royal Statistical Society. Series B* 44.1 (1982), pp. 71–80. DOI: 10.1111/j.2517-6161.1982.tb01189.x.

[Ker+23] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. "3D Gaussian Splatting for Real-Time Radiance Field Rendering." In: *ACM Transactions on Graphics* 42.4 (2023), 139:1–139:14. DOI: 10.1145/3592433.

[KB15] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization." In: *Proceedings of the International Conference on Learning Representations (ICLR)*. 2015.

[Liu+25] R. Liu, D. Sun, M. Chen, Y. Wang, and A. Feng. "Deformable Beta Splatting." In: *ACM SIGGRAPH 2025 Conference Papers*. 2025, 101:1–101:11. DOI: 10.1145/3721238.3730716.

[MA21] R. K. Mantiuk and M. Azimi. "PU21: A Novel Perceptually Uniform Encoding for Adapting Existing Quality Metrics for HDR." In: *Proceedings of the Picture Coding Symposium (PCS)*. 2021, pp. 1–5. DOI: 10.1109/PCS50896.2021.9477471.

[MJ00] K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley, 2000.

[Mil+20] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis." In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2020, pp. 405–421. DOI: 10.1007/978-3-030-58452-8_24.

[MND13] S. Miller, M. Nezamabadi, and S. Daly. "Perceptual Signal Coding for More Efficient Usage of Bit Codes." In: *SMPTE Motion Imaging Journal* 122.4 (2013), pp. 52–59. DOI: 10.5594/j18290.

[Nad05] S. Nadarajah. "A Generalized Normal Distribution." In: *Journal of Applied Statistics* 32.7 (2005), pp. 685–694. DOI: 10.1080/02664760500079464.

[NRH03] R. Ng, R. Ramamoorthi, and P. Hanrahan. "All-Frequency Shadows Using Non-Linear Wavelet Lighting Approximation." In: *ACM Transactions on Graphics* 22.3 (2003), pp. 376–381. DOI: 10.1145/882262.882280.

[PJH23] M. Pharr, W. Jakob, and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. 4th ed. MIT Press, 2023.

[Pol26] Poly Haven. *Poly Haven HDRI Library*. CC0 assets; accessed September 2026. URL: https://polyhaven.com/hdris.

[PH03] E. Praun and H. Hoppe. "Spherical Parametrization and Remeshing." In: *ACM Transactions on Graphics* 22.3 (2003), pp. 340–349. DOI: 10.1145/882262.882274.

[RH01] R. Ramamoorthi and P. Hanrahan. "An Efficient Representation for Irradiance Environment Maps." In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (*SIGGRAPH*). 2001, pp. 497–500. DOI: 10.1145/383259.383317.

[Rei+10] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. 2nd ed. Morgan Kaufmann, 2010.

[Rod+23] C. Rodríguez-Pardo, J. Fabre, E. Garcés, and J. López-Moreno. "NEnv: Neural Environment Maps for Global Illumination." In: *Computer Graphics Forum* 42.4, e14883 (2023). DOI: 10.1111/cgf.14883.

[SZ90] G. Schechtman and J. Zinn. "On the Volume of the Intersection of Two L_p^n Balls." In: *Proceedings of the American Mathematical Society* 110.1 (1990), pp. 217–224. DOI: 10.2307/2048262.

[SS95] P. Schröder and W. Sweldens. "Spherical Wavelets: Efficiently Representing Functions on the Sphere." In: *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (*SIGGRAPH*). 1995, pp. 161–172. DOI: 10.1145/218380.218439.

[Sit+20] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein. "Implicit Neural Representations with Periodic Activation Functions." In: *Advances in Neural Information Processing Systems* (*NeurIPS*). Vol. 33. 2020, pp. 7462–7473.

[SKS02] P.-P. Sloan, J. Kautz, and J. Snyder. "Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments." In: *ACM Transactions on Graphics* 21.3 (2002), pp. 527–536. DOI: 10.1145/566570.566612.

[SMP14] SMPTE. *High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays*. Standard ST 2084:2014. Society of Motion Picture and Television Engineers, 2014.

[TS06] Y.-T. Tsai and Z.-C. Shih. "All-Frequency Precomputed Radiance Transfer Using Spherical Radial Basis Functions and Clustered Tensor Approxi-

mation." In: *ACM Transactions on Graphics* 25.3 (2006), pp. 967–976. DOI: 10.1145/1141911.1141981.

[Wan+09] J. Wang, P. Ren, M. Gong, J. Snyder, and B. Guo. "All-Frequency Rendering of Dynamic, Spatially-Varying Reflectance." In: *ACM Transactions on Graphics* 28.5 (2009), 133:1–133:10. DOI: 10.1145/1618452.1618479.

[Wan+04] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. "Image Quality Assessment: From Error Visibility to Structural Similarity." In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612. DOI: 10.1109/TIP.2003.819861.

[War91] G. Ward. "Real Pixels." In: *Graphics Gems II*. Ed. by J. Arvo. Academic Press, 1991, pp. 80–83. DOI: 10.1016/B978-0-08-050754-5.50025-6.

[Xu+08] K. Xu, Y.-T. Jia, H. Fu, S.-M. Hu, and C.-L. Tai. "Spherical Piecewise Constant Basis Functions for All-Frequency Precomputed Radiance Transfer." In: *IEEE Transactions on Visualization and Computer Graphics* 14.2 (2008), pp. 454–467. DOI: 10.1109/TVCG.2007.70442.

[Xu+13] K. Xu, W.-L. Sun, Z. Dong, D.-Y. Zhao, R.-D. Wu, and S.-M. Hu. "Anisotropic Spherical Gaussians." In: *ACM Transactions on Graphics* 32.6 (2013), 209:1–209:11. DOI: 10.1145/2508363.2508386.

[Zha+21] K. Zhang, F. Luan, Q. Wang, K. Bala, and N. Snavely. "PhySG: Inverse Rendering with Spherical Gaussians for Physics-Based Material Editing and Relighting." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 5449–5458. DOI: 10.1109/CVPR46437.2021.00541.

[Zha+18] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 586–595. DOI: 10.1109/CVPR.2018.00068.

[Zha+24] X. Zhang, X. Ge, T. Xu, D. He, Y. Wang, H. Qin, G. Lu, J. Geng, and J. Zhang. "GaussianImage: 1000 FPS Image Representation and Compression by 2D Gaussian Splatting." In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2024, pp. 327–345. DOI: 10.1007/978-3-031-72673-6_18.

[Zha+25] Y. Zhang, B. Li, A. Kuznetsov, A. Jindal, S. Diolatzis, K. Chen, A. Sochenov, A. Kaplanyan, and Q. Sun. "Image-GS: Content-Adaptive Image Representation via 2D Gaussians." In: *ACM SIGGRAPH 2025 Conference Papers*. 2025, 102:1–102:11. DOI: 10.1145/3721238.3730596.

[Zwi+01] M. Zwicker, H. Pfister, J. van Baar, and M. Gross. "Surface Splatting." In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 2001, pp. 371–378. DOI: 10.1145/383259.383300.