Hemispherical Lighting Insights

Technical Memo ATVI-TR-24-01

THOMAS ROUGHTON, PETER-PIKE SLOAN, ARI SILVENNOINEN, and PETER SHIRLEY



Fig. 1. In-game render using an IrradZ HHD lightmap, with indirect diffuse lighting overlaid on the left.



Fig. 2. In-game render without (left) and with (right) a per-pixel AO cone multiply for indirect lighting from a volumetric SH grid, where the cone multiply is done by solving to the HLSH model and then evaluating. Indirect diffuse lighting (without albedo) is shown in a cut-out. Evaluating SH directly leads to leakage of light from behind the surface hemisphere on normal-mapped normals; the cone multiply eliminates this and grounds the directional AO in a believable manner.

Precomputed lighting is crucial in many interactive applications, which commonly encode global illumination in compact lightmaps or volumetric data structures. For precomputed lighting to interact with high-frequency geometric detail encoded in normal maps, each surface must be able to evaluate irradiance in any direction in its hemisphere, which requires that the precomputed lighting be represented using directional lighting models. We present two new highly compact hemispherical lighting models, each of which can be efficiently solved to and blended in an advancement over the commonly-used AHD model. We show how these models can be encoded to maintain exact reconstruction of irradiance in the vertex normal/local Z ("IrradZ"), providing a framework for solving to IrradZ models from spherical harmonics (SH) using hemispherical least-squares and correcting prior work where spherical methods have led to unnecessarily high error. Finally, we introduce an efficient method for hemisphere multiplies of quadratic SH and use it to inexpensively apply hemispherical or cone occlusion to volumetric lighting, mitigating light leakage for normal-mapped surfaces and significantly improving the appearance of runtime ambient occlusion.

CCS Concepts: • **Computing methodologies** → *Graphics systems and interfaces.*

Additional Key Words and Phrases: lightmaps, global illumination, spherical harmonics, hemispherical lighting

1 INTRODUCTION

Global illumination is a crucial yet computationally expensive component of lighting. For interactive applications such as games, which operate under restrictive performance budgets on often highly constrained hardware, precomputed lighting is the most efficient way of simulating global illumination effects at runtime, with time-proven techniques such as lightmaps [Carmack et al. 1996] still in use today.

The vast majority of lighting in interactive applications is for surfaces, which at each point have a hemisphere of incident radiance around the surface normal. Commonly, these surfaces are augmented by normal maps [Blinn 1978], where high frequency geometric detail is approximated by a texture. For accurate lighting reconstruction from these varying normals, the lighting model must therefore encode a hemisphere of irradiance. The irradiance for static geometry is commonly represented by *directional lightmap* formats [Chen 2008; McTaggart 2004; Sloan and Silvennoinen 2018], which use a small set of parameters to efficiently represent low-frequency lighting data.

The IrradZ lightmap parameterization [Sloan and Silvennoinen 2018], which encodes vertex-normal-exact hemispherical irradiance in a minimal set of parameters, is highly appealing as a compact directional lightmap format and has been shipped in multiple games. Prior work used Ambient and Highlight Direction (AHD) [Carmack et al. 1999; Iwanicki 2013] as the corresponding lighting model; as our first contribution, we show how the same parameterization can be used to represent multiple other models, such as the novel Hemisphere and Highlight Direction (HHD) model and the linear spherical harmonic-derived HLSH model, and compare the effect of the choice of model on reconstruction accuracy, interpolation properties, and physical reasoning.

In prior work, encoding the AHD lighting model has often been incorrectly done with consideration to spherical error [Iwanicki 2013; Sloan and Silvennoinen 2018], which is sub-optimal for hemispherical reconstruction. We present a framework for hemispherical encoding from spherical harmonics (SH), which exploits the fact that error in the lower hemisphere can be disregarded to provide more accurate reconstruction over the upper hemisphere. This framework requires only a minor increase in computation and no extra storage over spherical encoding, and is fully compatible with maintaining the IrradZ constraint for exact irradiance reconstruction in the hemisphere normal.

Hemispherical reasoning and encoding is also applicable at runtime, such as for lighting dynamic objects. Surface lighting data is often stored volumetrically in (usually quadratic order) spherical harmonics [Ramamoorthi and Hanrahan 2001], which are sampled at the shading location and then evaluated in the direction of the shading normal. When normal mapping is used without hemispherical occlusion this results in light leakage; incident radiance on the hemisphere below the geometric normal is entirely occluded by the surface and so should have no contribution to normal mapped normals, but will leak through if radiance below the hemisphere is not zeroed out. Using our hemispherical encoding frameworks for HHD and HLSH and a novel method for oriented quadratic-to-linear spherical harmonic hemisphere multiplies, we present methods to efficiently encode at runtime the hemispherical irradiance for evaluation with normal-mapped geometry. While approximate, these methods are substantially more accurate than not accounting for hemispherical occlusion and can also be applied per-vertex as a level of detail scheme. We further extend these methods to approximate occlusion by an arbitrary visibility cone such as is commonly computed by ambient occlusion algorithms [Jimenez et al. 2016]; rather than requiring the use of a separate ambient occlusion term and bent normals, this directly multiplies the incident lighting in an efficient manner, producing high quality results.

2 BACKGROUND AND PRIOR WORK

2.1 Hemispherical Lighting Models

There have been many [Chen 2008; Gautron et al. 2004; Habel et al. 2008; Hooker 2016; Iwanicki 2013; Iwanicki and Sloan 2017; Martin 2011; McTaggart 2004; Neubelt and Pettineo 2015; Silvennoinen and Sloan 2021] models used to encode hemispherical lighting data in lightmaps for real-time applications, with different accuracy, computation, and storage tradeoffs. At the higher-quality end of the spectrum representations such as multiple spherical Gaussian lobes [Neubelt and Pettineo 2015] have been used; these encode radiance over the hemisphere in a way that enables both specular and diffuse reconstruction; but are expensive to evaluate and require a high dynamic range color for each lobe (of which there are commonly between nine and twelve). On the lower quality end exist more inexpensive models such as Ambient and Highlight Direction (AHD) [Carmack et al. 1999; Iwanicki 2013], low-order spherical harmonics (SH), and SH's hemispherical variants HSH [Gautron et al. 2004] and the H-Basis [Habel and Wimmer 2010; Ishmukhametov 2011]. These models trade accuracy and radiance reconstruction for evaluation efficiency and compact encoding in a way that is highly attractive for constrained hardware, and are our particular focus in this paper.

2.1.1 Ambient and Highlight Direction (AHD). Ambient and Highlight Direction [Carmack et al. 1999; Iwanicki 2013] is a simple, commonly used directional lighting model. It represents irradiance through the combination of an ambient light with intensity C_a plus a directional light with color C_d and tangent-space direction $\mathbf{d}$. Reconstruction with AHD is

$$I(\mathbf{n}) = C_a + \max(0, \mathbf{n} \cdot \mathbf{d}) C_d, \tag{1}$$

where $I(\mathbf{n})$ is the irradiance and $\mathbf{n}$ is the normal being evaluated.

AHD is of particular interest due to being compatible with the IrradZ encoding [Sloan and Silvennoinen 2018], which uses only a single HDR color and three low-precision luminance values for each lightmap texel, exactly preserves irradiance in the vertex normal ("IrradZ"), and interpolates in a well-behaved manner.

2.1.2 Spherical Harmonics. Spherical harmonics [Ramamoorthi and Hanrahan 2001] are a set of spherical orthonormal basis functions that can be used to encode spherical signals, and are commonly used in rendering to represent radiance or irradiance signals. Increasingly high order spherical harmonics represent increasingly high frequency information; typically, quadratic spherical harmonics (with nine basis functions) are used to represent spherical irradiance, with linear spherical harmonics (four basis functions) sometimes used as an inexpensive alternative. The linear spherical harmonics consist of a constant term and a scaled function corresponding to each of the cardinal axes $-y, z, -x$:

$$\mathrm{Y}(x, y, z) = \left[\frac{1}{2\sqrt{\pi}} \quad -\sqrt{\frac{3}{4\pi}} y \quad \sqrt{\frac{3}{4\pi}} z \quad -\sqrt{\frac{3}{4\pi}} x \right]. \tag{2}$$

The quadratic spherical harmonics add an additional five basis functions:

$$\mathrm{Y}_2(x, y, z) = \left[\sqrt{\frac{15}{4\pi}} xy \quad -\sqrt{\frac{15}{4\pi}} yz \quad \sqrt{\frac{5}{16\pi}} (3z^2 - 1) \quad -\sqrt{\frac{15}{4\pi}} xz \quad \sqrt{\frac{15}{16\pi}} (x^2 - y^2) \right]. \tag{3}$$

Quadratic spherical harmonics can represent irradiance functions with an average error of less than 3% [Ramamoorthi and Hanrahan 2001], and so are often used as an intermediate representation in lightmap baking or in volumetric structures for runtime lighting. Spherical harmonics also have the property that convolutions with rotationally symmetric functions (which project into the *zonal harmonics*) can be done through an element-wise product; for example, to convolve with

a normalized cosine lobe (thereby converting SH coefficients representing radiance to represent irradiance) one simply multiplies the source coefficients by the per-band scale factors $[1, \frac{2}{3}, \frac{1}{4}, \ldots]$.

2.1.3 Hemispherical Spherical Harmonics. Spherical harmonics are not orthonormal over the hemisphere. Shifted basis functions such as the HSH [Gautron et al. 2004] and H-Basis [Habel and Wimmer 2010] address this issue by modifying the spherical harmonic basis functions in varying ways to produce bases that *are* orthonormal over the hemisphere; this enables simpler least-squares projection of hemispherical signals such as hemispherical radiance. Considering only the first four basis functions and disregarding coefficient scales, the H-Basis differs from linear SH only in replacing the third basis function z with $2z - 1$.

Orthonormality is not necessarily helpful when encoding *irradiance* signals over the hemisphere, since the radiance to irradiance convolution must happen in the spherical domain; therefore, since the first four basis functions of HSH and the H-Basis span the same space as linear SH, they represent the same data and, when excluding later basis functions, can be considered equivalent for the purposes of directional lighting reconstruction.

2.1.4 Modified H-Basis. The modified H-Basis [Ishmukhametov 2011] drops the orthonormality constraint in favor of encoding the irradiance in the geometric normal exactly in its constant basis function, with the third basis function becoming (again disregarding scale factors) $1 - z$. While this is again simply a variant of linear SH (i.e. to convert coefficients for $[1, z]$ to coefficients for $[1, 1 - z]$ you simply negate the second coefficient and add the first to it), it is interesting as an early example of encoding the scalar irradiance value directly into a lightmap and having the other basis functions augment it.

2.2 The IrradZ Parameterization

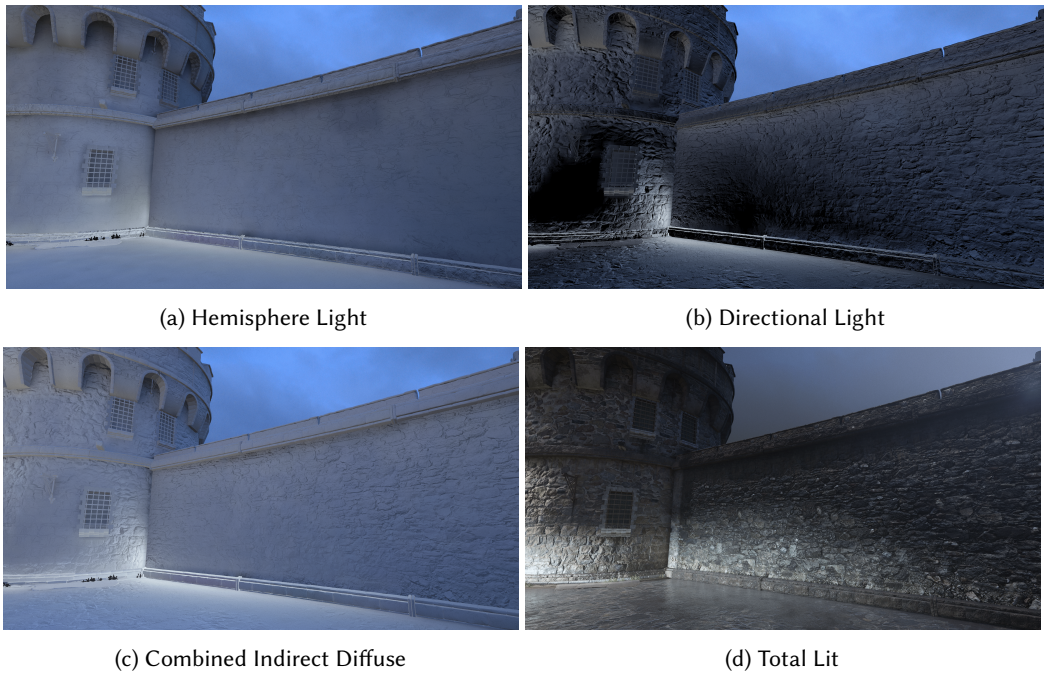
The most appearance-pertinent important property of an indirect irradiance lightmap is its reconstruction of irradiance in local Z ("IrradZ"), since the typical distribution of normals in a normal map is biased towards the hemisphere center. While any lighting model can be encoded such that the irradiance is exact at texel centers, reconstruction of interpolated coefficients (such as from linear texture filtering) is not always well behaved. For the AHD model (Section 2.1.1), for example, direct, independent interpolation of the $\mathbf{d}$ and C_d coefficients can result in artifacts if the precomputed lighting changes quickly; the parameters are coupled in the evaluation and so fully correct interpolation requires interpolating the augmented value $C_d\mathbf{d}$, which, if stored in textures, incurs a significantly higher memory and bandwidth cost. The IrradZ parameterization [Sloan and Silvennoinen 2018], which consists of the scalar irradiance I_z, the same highlight direction $\mathbf{d}$, and a $w_a \in [0, 1]$ "ambient ratio", improves on this issue by directly storing and interpolating the irradiance in local Z, ensuring reconstruction is exact for hemisphere-aligned normals and greatly mitigating interpolation artifacts while simultaneously reducing the encoded memory footprint.

Conversion from the IrradZ encoding to AHD's C_a, C_d, and $\mathbf{d}$ coefficients is given by

$$C_a = I_z w_a, \tag{4}$$

$$C_d = \frac{I_z(1 - w_a)}{\mathbf{d}_z}, \tag{5}$$

where $\mathbf{d}_z$ is the z (normal) component of the tangent-space highlight direction $\mathbf{d}$. This requires a single high-precision color I_z (compared to two for C_a and C_d), the same low-precision tangent space highlight direction $\mathbf{d}$ (in practice represented using hemispherical octahedral encoding [Meyer et al. 2010]), and a low-precision ambient ratio w_a. As a further optimization, $\mathbf{d}$ and w_a can be shared between all color channels, saving six coefficients while still preserving color accuracy for Z-aligned normals. When using block compression, luminance IrradZ requires 2.5 bytes per texel.



(a) Hemisphere Light (b) Directional Light

(c) Combined Indirect Diffuse (d) Total Lit

Fig. 3. The components of the HHD lighting model. The hemisphere light and directional light are summed together to produce the indirect diffuse.

3 EXTENDING IRRADZ: ALTERNATIVE LIGHTING MODELS

The original IrradZ paper [Sloan and Silvennoinen 2018] used only the AHD lighting model for reconstruction. While AHD is simple and efficient, it does not represent a physically plausible lighting setup over the hemisphere, since the irradiance for a constant intensity light over a hemisphere is given by a hemisphere light, not a constant ambient term. Additionally, AHD is a nonlinear model that requires a nonlinear solver for encoding, limiting its flexibility for runtime encoding and blending. We present two new models that make different tradeoffs; the first, HHD (Section 3.1), simply replaces the ambient term in AHD with a physically plausible hemisphere light, while the second, HLSH (Section 3.2), is a linear model.

3.1 Hemisphere and Highlight Direction (HHD)

HHD (Figure 3) is a simple tweak on AHD that replaces the ambient term with a hemisphere in the vertex normal $\mathbf{n}_v$. This accounts for occlusion by the surface plane, providing directional variation even when the lighting is relatively flat and enabling exact reconstruction of a wider range of analytic lighting setups.[1] The irradiance from a unit-intensity hemisphere light with hemisphere normal $\mathbf{n}_v$ is $\frac{1}{2}(\mathbf{n} \cdot \mathbf{n}_v + 1)$; reconstruction with HHD is therefore

$$I(\mathbf{n}) = \frac{(\mathbf{n}_v \cdot \mathbf{n}) + 1}{2} C_a + \max(0, \mathbf{n} \cdot \mathbf{d}) C_d. \tag{6}$$

In tangent space, $\mathbf{n}_v \cdot \mathbf{n}$ is $\mathbf{n}_z$, and dividing by two is a free modifier on most GPUs, so the cost of evaluating a hemisphere is minimal.

[1] HHD can exactly represent any ambient-plus-directional lighting setup on a surface plane; AHD can represent hemispherical lighting only when either C_a is zero or $\mathbf{d}$ is aligned with the vertex normal.

3.1.1 Hemisphere Only. For low-end platforms, using a scalar lightmap that is interpreted as a hemisphere allows for some normal variation while reducing storage and shader costs.

3.1.2 Direct Solve from Linear Spherical Harmonics. One interesting property of HHD is that, since it represents a physically plausible hemispherical lighting model and has four degrees of freedom (C_a, $C_d\mathbf{d}_x$, $C_d\mathbf{d}_y$, and $C_d\mathbf{d}_z$), HHD coefficients can be directly inferred from linear SH (i.e. by reasoning what HHD setup would produce the linear SH coefficients). This is particularly useful for runtime/GPU solves (Section 3.4) and blending (Figure 4). The tangent-space linear SH coefficients $L_{sh} = [l_0, l_1^{-1}, l_1^0, l_1^1]$ for HHD are

$$L_{sh} = C_a \begin{bmatrix} \sqrt{\pi} \\ 0 \\ \frac{\sqrt{3\pi}}{2} \\ 0 \end{bmatrix} + \pi C_d \begin{bmatrix} \frac{1}{2\sqrt{\pi}} \\ -\sqrt{\frac{3}{4\pi}}\mathbf{d}_y \\ \sqrt{\frac{3}{4\pi}}\mathbf{d}_z \\ -\sqrt{\frac{3}{4\pi}}\mathbf{d}_x \end{bmatrix} = \frac{\sqrt{\pi}}{2} \begin{bmatrix} 2C_a + C_d \\ -\sqrt{3}C_d\mathbf{d}_y \\ \sqrt{3}(C_a + C_d\mathbf{d}_z) \\ -\sqrt{3}C_d\mathbf{d}_x \end{bmatrix}. \tag{7}$$

We simplify the inverse mapping by introducing a scaled basis $\mathbf{b}$, where

$$[\mathbf{b}_c, \mathbf{b}_x, \mathbf{b}_y, \mathbf{b}_z] = \frac{2}{\sqrt{\pi}}[l_0, -\frac{1}{\sqrt{3}}l_1^1, -\frac{1}{\sqrt{3}}l_1^{-1}, \frac{1}{\sqrt{3}}l_1^0]. \tag{8}$$

The HHD coefficients have closed forms in terms of $\mathbf{b}$:

$$C_a = \frac{(2\mathbf{b}_c - I_z) - \sqrt{(2\mathbf{b}_c - I_z)^2 - 3(\mathbf{b}_c^2 - \|\mathbf{b}_{xyz}\|^2)}}{3}, \tag{9}$$

$$C_d = \|[\mathbf{b}_x, \mathbf{b}_y, \mathbf{b}_z] - C_a\mathbf{n}\|, \tag{10}$$

$$\mathbf{d} = \frac{[\mathbf{b}_x, \mathbf{b}_y, \mathbf{b}_z] - C_a\mathbf{n}}{C_d}. \tag{11}$$

A full derivation is provided in Appendix C. Note that this technique requires the source linear SH to represent hemispherical data (i.e. that, in tangent space, the $I_z = C_a + C_d\mathbf{d}_z$ coefficient is given by $\frac{2}{\sqrt{3\pi}}l_1^0$); if this is not the case, and data from higher-order bands is available, a hemisphere multiply (such as that described in Section 5.1) should be employed before the solve.

For lighting environments with only moderate energy in higher SH bands, using only linear SH to encode HHD can produce results very similar to a full nonlinear solve (Section 4), with an observed average increase in RMSE of less than 10%.

3.2 IrradZ Hemispherical Linear SH (HLSH)

Those same C_a, C_d, and $\mathbf{d}$ parameters can be interpreted in a manner similar to linear spherical harmonics, with reconstruction given by

$$I(\mathbf{n}) = C_a + (\mathbf{n} \cdot \mathbf{d})C_d \tag{12}$$

$$= I_z\mathbf{n}_z + C_a(1 - \mathbf{n}_z) + c_x\mathbf{n}_x + c_y\mathbf{n}_y, \tag{13}$$

where $C_a = I_z w_a$, $c_x = \mathbf{d}_x C_d$, and $c_y = \mathbf{d}_y C_d$. I_z and $\mathbf{d}$ are the IrradZ encoding parameters; this also uses the identity $C_d = \frac{I_z(1-w_a)}{\mathbf{d}_z}$ from Equation 5. HLSH therefore forms a linear basis

$$\text{HLSH}(x, y, z) = (z, 1 - z, x, y) \tag{14}$$

with coefficients given by $h = [I_z, C_a, c_x, c_y]$.

Unlike AHD and HHD, HLSH can be solved to analytically from any linear basis (Section 3.2.2). Due to its linearity, it can also be trivially blended by averaging the coefficients of h; equivalently,

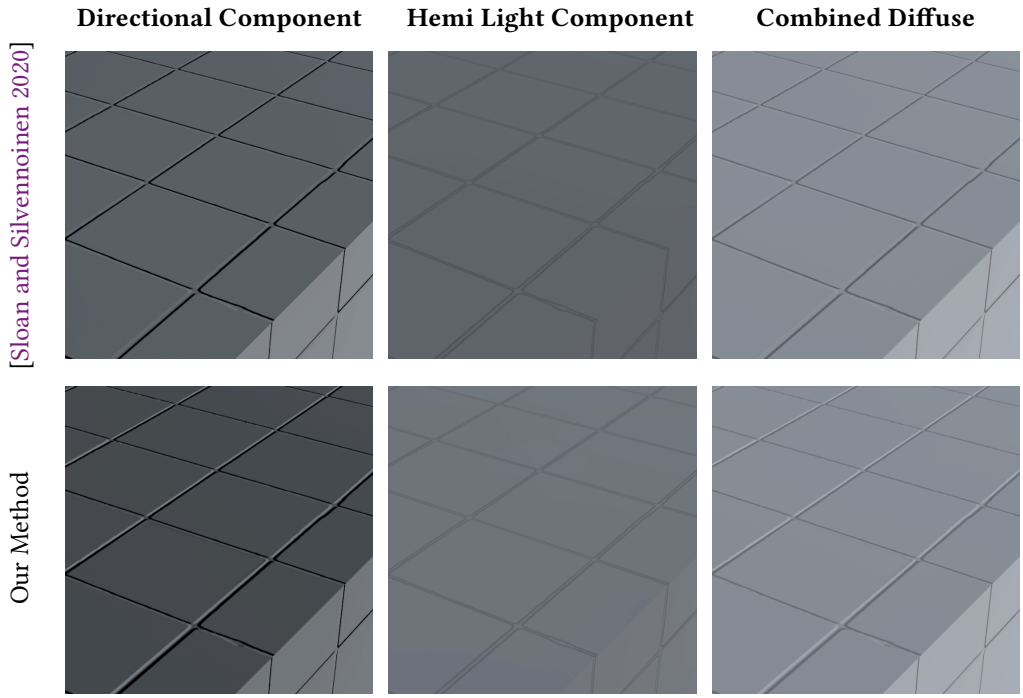


Fig. 4. The result of blending two HHD lightmaps. Previous work [Sloan and Silvennoinen 2020] (top) naively blended IrradZ lightmap values by performing luminance-weighted sums of $\mathbf{d}$ and w_a, which leads to a lack of contrast when blending diverging highlight directions. HHD, in combination with the direct solve from linear SH presented in Section 3.1.2, offers a much higher quality option: project to linear SH, blend there, and then solve back to HHD. Doing so correctly moves energy from the directional light (left) into the hemisphere light (center) and keeps the highlight direction closer to the horizon, preserving highlight contrast while still maintaining the same combined diffuse irradiance (right) for Z-aligned normals.

in AHD parameter form C_a is averaged while $\mathbf{d}$ and C_d are given by the normalized direction and length, respectively, of $\sum_i C_{di}\mathbf{d}_i$.

HLSH is closely related to linear spherical harmonics (differing only by coefficient scale factors), with the projection given by

$$\begin{bmatrix} l_0 \\ l_1^{-1} \\ l_1^0 \\ l_1^1 \end{bmatrix} = \begin{bmatrix} 2\sqrt{\pi}C_a \\ -2\sqrt{\frac{\pi}{3}}c_y \\ 2\sqrt{\frac{\pi}{3}}(I_z - C_a) \\ -2\sqrt{\frac{\pi}{3}}c_x \end{bmatrix}. \tag{15}$$

HLSH is also related to the orthonormal HSH [Gautron et al. 2004] and H-Basis [Habel and Wimmer 2010] families of basis functions and the modified H-Basis [Ishmukhametov 2011]; compared to the latter, HLSH differs only by replacing the constant basis function with z. When considering only the variants with four parameters, these models can be exactly converted between and provide identical reconstruction accuracy; this also means that any of those models can be converted into HLSH and then encoded using the IrradZ parameterization.

3.2.1 IrradZ Parameters. The IrradZ encoding parameters are given from HLSH coefficients by

$$w_a = \frac{C_a}{I_z}, \tag{16}$$

$$\mathbf{d}_z = \left(\frac{c_x^2 + c_y^2}{(I_z - C_a)^2} + 1 \right)^{-\frac{1}{2}}, \tag{17}$$

$$\begin{bmatrix} \mathbf{d}_x \\ \mathbf{d}_y \end{bmatrix} = \frac{\mathbf{d}_z}{I_z - C_a} \begin{bmatrix} c_x \\ c_y \end{bmatrix}. \tag{18}$$

3.2.2 IrradZ-Constrained Direct Solve. HLSH is a linear basis with four coefficients. To preserve the exact irradiance in the hemisphere normal (the IrradZ constraint) in a least-squares solve, we treat the coefficient I_z, corresponding to the z basis function, as fixed. To solve for the remaining coefficients $b = [C_a, c_x, c_y]$, we want to minimize the reconstruction error over the hemisphere $\mathbf{\Omega}^+$

$$E = \int_{\Omega^+} (B(\omega) \cdot b + I_z \omega_z - \mathrm{Y}(\omega) \cdot i)^2 \, \mathrm{d}\omega, \tag{19}$$

where i is the target irradiance, $B(\omega) = (1 - \omega_z, \omega_x, \omega_y)$, the remaining HLSH basis functions, and Y are the SH basis functions such that $\mathrm{Y}(\omega) \cdot i$ gives the target irradiance in the direction ω.

Differentiating and setting the derivatives to zero results in

$$0 = \mathbf{G}_B b + I_z \left(\int_{\Omega} B(\omega) \omega_z \mathrm{d}\omega \right) - \mathbf{G}_{BY} i, \tag{20}$$

where $\mathbf{G}_B = \int_{\Omega^+} B(\omega) B(\omega)^T \mathrm{d}\omega = \frac{2\pi}{3}$ and $\mathbf{G}_{BY} = \int_{\Omega^+} B(\omega) \mathrm{Y}(\omega)^T \mathrm{d}\omega$. If the source basis is quadratic spherical harmonics, $\mathbf{G}_{BY}$ is given by

$$\mathbf{G}_{BY} = \begin{bmatrix} \frac{\sqrt{\pi}}{2} & 0 & \sqrt{\frac{\pi}{12}} & 0 & 0 & 0 & -\frac{\sqrt{5\pi}}{8} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{\frac{\pi}{3}} & 0 & 0 & 0 & \frac{\sqrt{15\pi}}{8} & 0 \\ 0 & \sqrt{\frac{\pi}{3}} & 0 & 0 & 0 & \frac{\sqrt{15\pi}}{8} & 0 & 0 & 0 \end{bmatrix}. \tag{21}$$

Solving for b yields

$$\begin{bmatrix} C_a \\ c_x \\ c_y \end{bmatrix} = \frac{3}{2\pi} \left(\mathbf{G}_{BY} i - I_z \begin{bmatrix} \frac{\pi}{3} \\ 0 \\ 0 \end{bmatrix} \right). \tag{22}$$

3.3 Comparing AHD, HHD, and Hemispherical Linear SH

AHD, HHD, and HLSH are all compact and efficient representations with very similar evaluation, and, when properly solved for, all can produce fairly accurate results, with each having lighting environments it's best suited to (Figure 5); HLSH generally provides the most accurate reconstruction for smooth lighting, while AHD and HHD are more accurate for highly directional environments or those with high amounts of baked light. There are additional considerations: for example, unlike AHD, HHD exactly represents a physical, analytic hemispherical lighting setup (that of a constant ambient light occluded by the surface hemisphere plus a directional light), which means it can round-trip exactly with hemispherical lighting in linear spherical harmonics in a hemispherical-error-minimizing manner (Section 3.1.2). This carries advantages for both runtime solves or blending (Figure 4) and in evaluation, where different material BRDFs can be computed analytically or tabulated for each of the hemisphere and directional light. Similarly, HLSH, being a fully linear basis, can be solved to analytically from spherical harmonics (Section 3.2.2) without requiring a search for the direction vector (Section 4) or separate hemisphere multiply (Section 5),

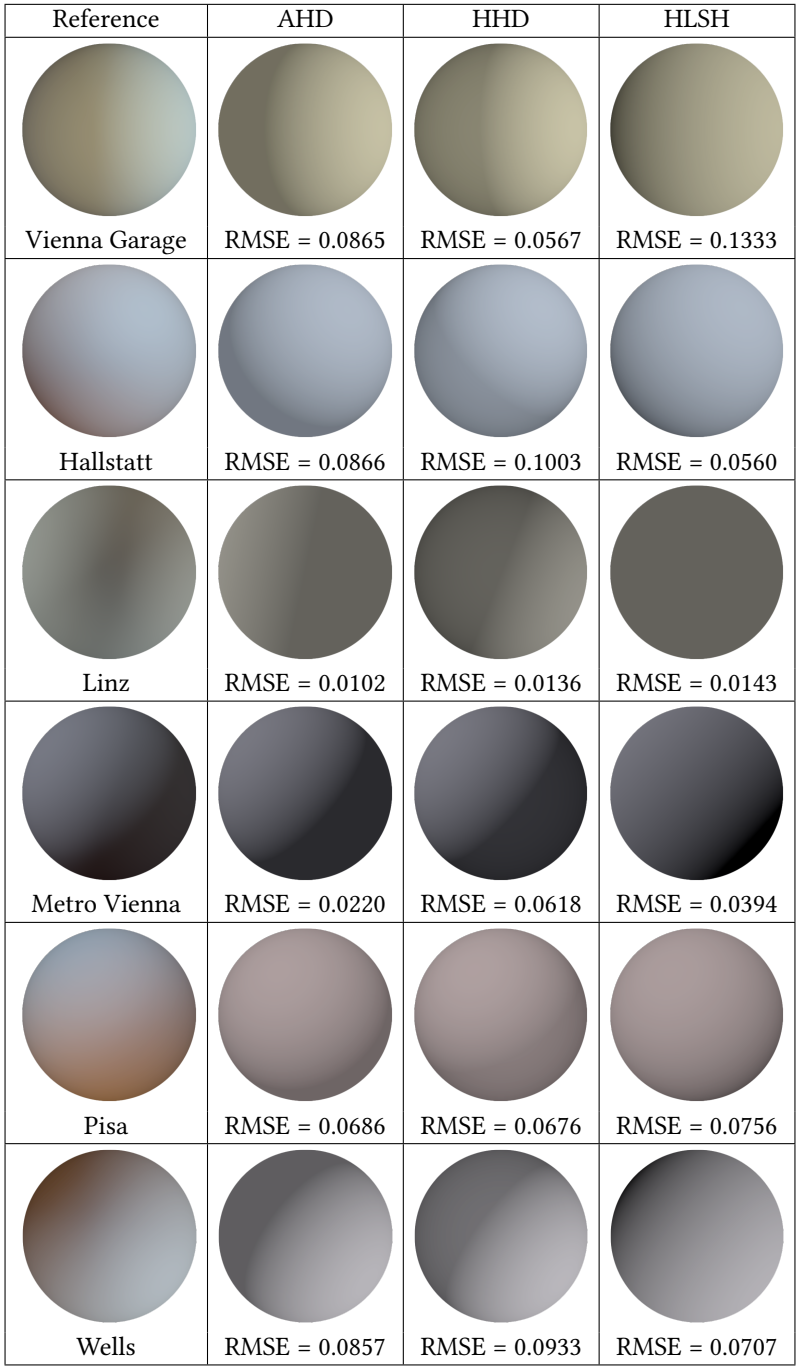


Fig. 5. Irradiance reconstruction on a range of HDR lighting environments [Vogl 2010] clipped to the hemisphere using luminance AHD, HHD, and HLSH (Section 2.2); the reference is numerical integration of the input environment map. Each IrradZ format has the lowest error for some environments; there is no single best choice. For lightmaps, where the normals are usually oriented towards the hemisphere normal, the visual difference is usually slight, with AHD and HHD having higher contrast; however, the runtime solves for HHD and HLSH make them an attractive choice for some use cases.

and can be trivially blended, but has no physical analogue and cannot represent high-frequency directional lighting. In practice, HHD provides the best balance of error and runtime properties for our production lightmaps, but different use cases may see different conclusions.

3.4 Directional Encodings as a Level of Detail

HHD and HLSH can also be used as a lighting level-of-detail scheme for non-lightmapped geometry, where the coefficients are generated from volumetric SH in the vertex shader and then evaluated in the pixel shader. This is particularly useful for lower-end platforms, since fetching and evaluating high-order spherical harmonics in pixel shaders can be expensive (particularly when incorporating a hemisphere or cone multiply), while evaluating these compact representations is very inexpensive.

To compute the coefficients, we use either the solve from linear SH for HHD (Section 3.1.2) after a quadratic-to-linear hemisphere multiply (Section 5.1) or the direct solve to irradiance-convolved HLSH (Section 5.2); in both cases, we perform the solve in the global frame for luminance SH after computing the RGB irradiance in the vertex normal direction. As vertex shader outputs, we compute I_z and $\mathbf{d}' = \text{Luminance}\left(\frac{C_d}{I_z}\right)\mathbf{d}$; then, in the pixel shader, we reconstruct $C_a = I_z\,(1 - \mathbf{n}_v \cdot \mathbf{d}')$. This interpolation scheme minimizes the number of interpolants, preserves the tangent-space I_z as an invariant, and pushes any cancellation in the direction vector into the ambient term, which mimics the behavior in the solves. A slightly more accurate option would be to interpolate $\text{Luminance}(I_z) \times \mathbf{d}'$ rather than $\mathbf{d}'$, dividing again in the pixel shader; in practice, we have seen little visual difference and prefer the reduced pixel shader cost.

4 ENCODING IRRADZ FROM SPHERICAL HARMONICS

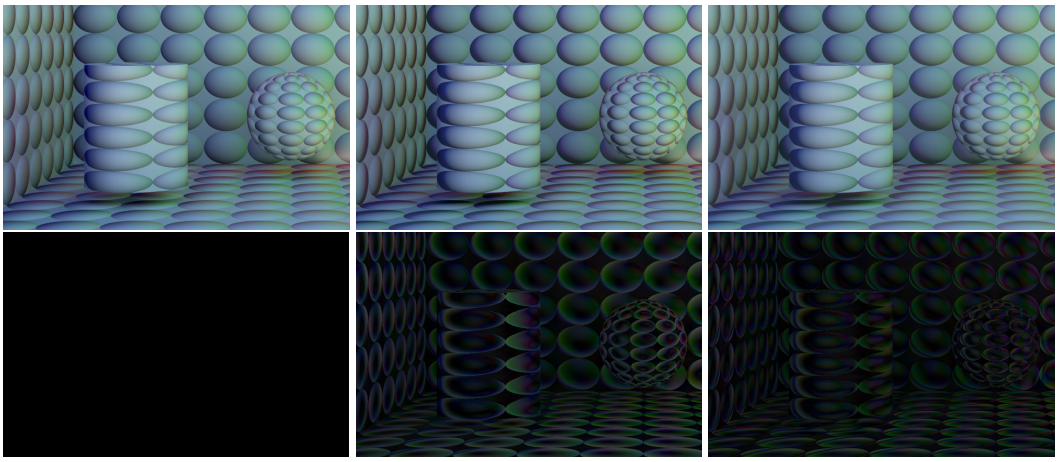
The AHD, HHD, and HLSH lighting models are used to represent directional irradiance, which means that directional irradiance data is required to encode them. Dense sampling of irradiance over the hemisphere during a bake generally incurs infeasible memory usage and computational cost; instead, as in prior work [Ishmukhametov 2011; Sloan and Silvennoinen 2018], we encode to (quadratic) spherical harmonics, perform the closed-form irradiance convolution, and then solve for the encoded IrradZ coefficients that best represent the spherical harmonic irradiance.

Past encoding methods for AHD [Iwanicki 2013; Sloan and Silvennoinen 2018] have used spherical error or taken the optimal linear direction from the spherical harmonic as the highlight direction $\mathbf{d}$, both of which reduce reconstruction accuracy (Figure 6); the optimal linear direction's l_1^0 coefficient is polluted by the hemisphere light that necessarily results from integration of SH over the hemisphere, significantly increasing error, and using the optimal linear direction also neglects energy in higher bands (which can cause the optimal $\mathbf{d}$ to vary in both ϕ and θ).

In this section, we set out a framework for nonlinear solves, applicable to both AHD and HHD,[2] that correctly account for hemispherical error and search for the optimal highlight direction on the hemisphere. In our solves, we enforce the IrradZ constraint, which, in addition to ensuring exact reconstruction for local Z, biases the error distribution to be lower around the hemisphere normal, better matching the typical distribution of normals in a normal map. The IrradZ value used in the constraint is preferably given by sampled scalar irradiance, but computing it from the quadratic SH is a reasonable alternative.

It is important to note that, since we are encoding for evaluation with normal maps, we are inherently encoding a nonphysical quantity for which there is no ground truth. Normal maps approximate local geometric detail in a way that, for physically accurate results, must be incorporated into the path tracing as actual geometric detail (for example, to model inter-reflection between opposing normals), which is decidedly non-trivial. In prior work, Schüssler et al. [2017] proposed

[2]HLSH may be solved to directly using the method described in Section 3.2.2



RGB RMSE: 3.795, 4.874, 5.159 RGB RMSE: 1.514, 1.805, 1.597

(a) Reference (lightmap encoded using quadratic SH)

(b) RGB AHD using the optimal linear SH direction and w_a from the hemispherical Gram.

(c) RGB AHD using a searched-for direction and w_a from the hemispherical Gram.

Fig. 6. In prior work [Iwanicki 2013; Sloan and Silvennoinen 2018], the optimal linear direction and spherical error were used for encoding AHD parameters. This yields significantly higher error compared to the hemispherical solve; the highlight direction is too close to the hemisphere normal and the ambient weight is too low, so grazing normals are excessively darkened. By comparison, using the correct hemispherical solve (Section 4) produces a much closer result to the target quadratic SH. Images are lightmaps encoded using The Baking Lab [Pettineo 2024].

a model for path tracing normal maps in a well-defined and energy-conserving way; however, it is not clear how that model could be applied when considering the outgoing radiance from an arbitrary, runtime-applied normal map. We instead restrict ourselves to a simpler problem: we consider the incident radiance on a hemisphere surrounding the geometric normal as received during the bake and then aim to encode the irradiance that an arbitrary normal would receive from that lighting environment, independent of any interreflection. In effect, this means that the SH encoded in the bake *is* our reference, and that, save using higher-order SH that have increasingly little irradiance contribution, there is no other ground truth.

4.1 Hemispherical Least-Squares Encoding

Encoding irradiance in hemispherical lighting models usually entails taking source spherical harmonic radiance data produced in an offline bake, performing an irradiance convolution, and then encoding to the coefficients of the lighting model of choice. Often neglected is the fact that when encoding irradiance to be evaluated over the hemisphere we must take care to minimize reconstruction error over *only* the hemisphere, rather than the entire sphere; disregarding error in the lower hemisphere (knowing those directions will never be sampled) often significantly decreases error for the upper hemisphere.[3] The L2 error over the hemisphere Ω^+ for the SH representation of

[3]Note that when performing the cosine convolution for irradiance it is important that the lower hemisphere remain zero (i.e. not discarded), since otherwise invalid data will be included in the convolution; hemispherical error should only be used after the convolution.

two functions $A(\omega) = a \cdot \mathrm{Y}(\omega)$ and $B(\omega) = b \cdot \mathrm{Y}(\omega)$ is given by

$$E = \int_{\Omega^+} \left(a \cdot \mathrm{Y}(\omega) - b \cdot \mathrm{Y}(\omega)\right)^2 \mathrm{d}\omega \tag{23}$$

$$= \int_{\Omega^+} \left((a-b) \cdot \mathrm{Y}(\omega)\right)^2 \mathrm{d}\omega, \tag{24}$$

where $\mathrm{Y}(\omega)$ is the vector of SH basis functions evaluated in the direction ω.

To simplify the error integral, we use the Gram matrix $\mathbf{G} = \int_{\Omega^+} A(\omega)B(\omega)^T \mathrm{d}\omega$; given that

$$\int \big(a \cdot A(s)\big)\big(b \cdot B(s)\big)\mathrm{d}s = a^T \mathbf{G} b, \tag{25}$$

$$\therefore E = (a-b)^T \, \mathbf{G} \, (a-b) \, . \tag{26}$$

In this general form, solving for b yields the standard least-squares solution $b = \mathbf{G}^{-1}a$.

The Gram matrix can be computed for various basis functions and integration domains; $\mathbf{G}$ for quadratic spherical harmonics over the hemisphere is given in Appendix B, and over the sphere is $\mathbf{I}$.

4.2 Nonlinear Solves for IrradZ AHD and HHD

The spherical harmonic representation L_{sh} of an IrradZ AHD or HHD function is

$$\begin{aligned} L_{sh} &= I_z \left(w_a \mathrm{sh}_{amb} + \frac{\pi(1-w_a)}{\mathbf{d}_z} \mathrm{sh}_{dir} \right) \\ &= C_a \mathrm{sh}_{amb} + \pi C_d sh_{dir}. \end{aligned} \tag{27}$$

I_z is the irradiance in the hemisphere normal $\int_{\Omega^+} \mathrm{R}(\omega) \frac{\cos\omega}{\pi} \mathrm{d}\omega$. When encoding from a hemispherical radiance signal projected into spherical harmonics, I_z is equivalently given by

$$I(\mathbf{n}) = \frac{2}{\sqrt{3\pi}} l_1 \cdot \mathbf{n}_{sh}, \tag{28}$$

where $\mathbf{n}_{sh} = [-\mathbf{n}_y, \mathbf{n}_z, -\mathbf{n}_x]$ and $\mathbf{n}$ is the hemisphere normal; alternatively, I_z can be approximated by the irradiance spherical harmonic. For AHD, $\mathrm{sh}_{amb} = [2\sqrt{\pi}, 0, 0, 0, \dots]$ is an ambient light in spherical harmonics, while for HHD it is a hemisphere light (Appendix A.1); for both, $\mathrm{sh}_{dir} = \mathrm{Y}(\mathbf{d})$ is the vector of SH basis functions evaluated in the direction $\mathbf{d}$ (equivalent to a directional light in spherical harmonics). Radiance reconstruction of $L(s)$ for a direction s is given by

$$R(s) = L_{sh} \cdot \mathrm{Y}\,(s)\,, \tag{29}$$

and irradiance reconstruction by

$$I(s) = \mathrm{SHConvCos}(\mathrm{L_{sh}}) \cdot \mathrm{Y}\,(s)\,, \tag{30}$$

where $\mathrm{SHConvCos}(l)$ applies a normalized cosine convolution (per-band scales of $[1, \frac{2}{3}, \frac{1}{4}, \dots]$) to map the input radiance l to irradiance.

Solving for both the highlight direction $\mathbf{d}$ and ambient weight w_a in combination is a nonlinear optimization due to the IrradZ-preserving scale of $\frac{1-w_a}{\mathbf{d}_z}$ on the directional light. In practice, we perform a golden section search for $\mathbf{d}_z$ along the ϕ given by the optimal linear direction, solving for w_a analytically and computing the error at each step. This neglects variation in ϕ due to energy in higher bands; if desired, performing a golden section search for ϕ with the search for $\mathbf{d}_z$ in an inner loop results in a slight decrease in error, though the difference is usually minor (e.g. less than 7% RMSE) given the presence of the $w_a \in [0, 1]$ constraint.

4.2.1 Direct Solve for w_a. Given $\mathbf{d}$, it is possible to directly solve for the ambient weight w_a. We want to minimize the reconstruction error E over the hemisphere Ω^+:

$$\begin{aligned} E &= \int_{\Omega^+} \Big(\mathrm{Y}(\omega)\cdot\big(w_a a + (1-w_a)d\big) - \mathrm{Y}(\omega)\cdot i\Big)^2 \mathrm{d}\omega \\ &= (w_a(a-d)+d-i)^T\,\mathbf{G}\,(w_a(a-d)+d-i)\,, \end{aligned} \tag{31}$$

where $a = \mathrm{SHConvCos}(\mathrm{sh}_{amb})$ is the irradiance from the unit-intensity ambient or hemi light, $d = \frac{\pi}{\mathbf{d}_z}\mathrm{SHConvCos}(\mathrm{sh}_{dir})$ is the irradiance from the directional light in the tangent-space highlight direction $\mathbf{d}$, $i = \frac{1}{L_z}\mathrm{SHConvCos}(L_{sh})$ is the normalized target irradiance, and $\mathbf{G}$ is the hemispherical Gram matrix (Section 4.1, Appendix B). Differentiating the error function E with respect to w_a and setting equal to zero, we have

$$0 = (a-d)^T\mathbf{G}(w_a(a-d)+d-i) + (w_a(a-d)+d-i)^T\mathbf{G}(a-d). \tag{32}$$

Expanding and using the symmetry in $\mathbf{G}$, the closed form expression for w_a is

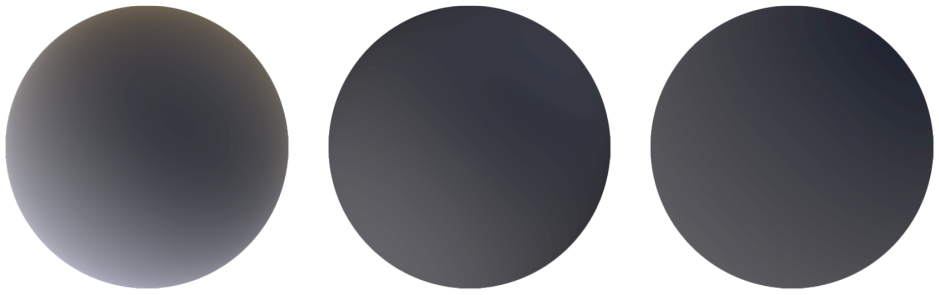
$$w_a = \frac{(a-d)^T\mathbf{G}(i-d)}{(a-d)^T\mathbf{G}(a-d)}. \tag{33}$$

5 HEMISPHERICAL OCCLUSION FOR SH REPRESENTING SPHERICAL DATA

Thus far we have dealt with purely hemispherical radiance data in tangent space. In practice, it is common to have spherical radiance; this can occur if lightmap baking is done in world-space on curved surfaces, but is more usually encountered at runtime, where SH radiance data is sampled from a volumetric data structure or simply used as distant environment lighting. Evaluation of irradiance from a spherical radiance function uses the entire hemisphere of radiance around the normal being queried; for normal maps, this results in light leaking from below the surface hemisphere (Figure 2). A more accurate approach is to multiply the radiance function by a hemisphere in the vertex normal or an occlusion cone, perform the irradiance convolution, and then evaluate, adding self-bounce (Appendix A) as appropriate.

In tangent space, zeroing out the lower hemisphere—or, equivalently, multiplying by an upper hemisphere—can be done in a least-squares manner by reprojecting the source spherical harmonic over the hemisphere (i.e. $\int_{\Omega^+}\mathrm{Y}(\omega)(\mathrm{Y}(\omega)\cdot i)\mathrm{d}\omega$). This reduces to multiplication with the hemispherical Gram matrix (Appendix B), where the SH vector i after hemisphere multiplication is given by $\mathbf{G}i$. Note that each application of a hemisphere multiply is effectively a stronger enforcement that the lower hemisphere is zero, and, in the absence of an infinite-order SH expansion, is lossy to the contents of the upper hemisphere; a hemisphere multiply in spherical harmonics should ideally only be applied if the input data is known not to be hemispherical.

For arbitrary $\mathbf{n}$ and arbitrary-order SH, performing a hemisphere multiply generally means first rotating the SH into $\mathbf{n}$-aligned tangent space, projecting, then rotating back, which is a fairly expensive operation [Hable 2014]. If we restrict ourselves to only the linear SH coefficients after projection, however, we show that the intermediate rotations can be avoided. Using this, it is possible to either produce a linear spherical harmonic that can be used as input for a direct HHD solve (Section 3.1.2) or, more directly, to produce HLSH coefficients optimally representing the surface hemisphere. While the encoding to HLSH from quadratic SH is inherently lossy, it's relatively inexpensive, exactly preserves irradiance in the hemisphere normal, and provides much higher accuracy for grazing normals than using unoccluded SH (Figure 7). These multiplies can additionally be extended to represent irradiance after occlusion by an arbitrary cone (Section 5.2.1), which is particularly useful for convolving a radiance environment with an AO visibility cone [Jimenez et al. 2020] (Figure 8).



(a) Unclipped SH (b) Hemisphere-Clipped SH (c) HLSH Hemi-Multiply

Fig. 7. Using irradiance from an SH that has not been clipped to the surface plane hemisphere results in light leakage for grazing normals (left). Correct occlusion (center) clips radiance under the surface hemisphere, and a hemisphere-multiply solve to HLSH (right) is a reasonable approximation for this occlusion. The source lighting is a quadratic SH probe from a production map.

5.1 Oriented Hemisphere Multiply for Quadratic to Linear SH

By decomposing the hemisphere-projected linear SH p_{sh}, where p_{sh} in tangent space is given by $\mathbf{G}_{\Omega^+} l$ and l is our source quadratic SH, we derive an exact and relatively inexpensive method for oriented hemisphere projection. From the hemispherical Gram matrix (Appendix B), the L0 band is given by $p_0 = \frac{l_0}{2} + \frac{\sqrt{3}}{4} l_1^0$; with rotation, this is

$$p_0 = \frac{l_0}{2} + \frac{\sqrt{3}}{4}(\mathbf{n}_{sh} \cdot l_1). \tag{34}$$

For p_1, the normal-aligned $[l_1^{-1\prime}, 0, l_1^{1\prime}]$ components are given by $l_1 - (\mathbf{n}_{sh} \cdot l_1)\mathbf{n}_{sh}$. The L1 contribution from the L2 band p_2 is given by rotating the input SH l_2 band into tangent space, forming the L1 vector $[l_2^{-1\prime}, 0, l_2^{1\prime}]$, and rotating back to the original coordinate space; using Hable's method for spherical harmonic rotation [Hable 2014], p_2 reduces to

$$\begin{bmatrix} (2\mathbf{n}_y^2 - 1)\mathbf{n}_x & -(2\mathbf{n}_y^2 - 1)\mathbf{n}_z & \sqrt{3}\mathbf{n}_y\mathbf{n}_z^2 & -2\mathbf{n}_x\mathbf{n}_y\mathbf{n}_z & (\mathbf{n}_x^2 - \mathbf{n}_y^2 + 1)\mathbf{n}_y \\ -2\mathbf{n}_x\mathbf{n}_y\mathbf{n}_z & (2\mathbf{n}_z^2 - 1)\mathbf{n}_y & \sqrt{3}(1 - \mathbf{n}_z^2)\mathbf{n}_z & (2\mathbf{n}_z^2 - 1)\mathbf{n}_x & (\mathbf{n}_y^2 - \mathbf{n}_x^2)\mathbf{n}_x \\ (2\mathbf{n}_x^2 - 1)\mathbf{n}_y & -2\mathbf{n}_x\mathbf{n}_y\mathbf{n}_z & \sqrt{3}\mathbf{n}_x\mathbf{n}_z^2 & -(2\mathbf{n}_x^2 - 1)\mathbf{n}_z & (\mathbf{n}_x^2 - \mathbf{n}_y^2 - 1)\mathbf{n}_x \end{bmatrix} \begin{bmatrix} l_2^{-2} \\ l_2^{-1} \\ l_2^0 \\ l_2^1 \\ l_2^2 \end{bmatrix}. \tag{35}$$

The normal-aligned $l_1^{0\prime}$ is given by $\frac{\sqrt{3\pi}}{2} I_z$ after projection as per Equation (28). In combination, $[p_1^{-1}, p_1^0, p_1^1]$ is given by

$$\begin{bmatrix} p_1^{-1} \\ p_1^0 \\ p_1^1 \end{bmatrix} = \frac{1}{2}\left(\begin{bmatrix} l_1^{-1} \\ l_1^0 \\ l_1^1 \end{bmatrix} - (\mathbf{n}_{sh} \cdot l_1)\mathbf{n}_{sh} \right) + \frac{\sqrt{3\pi}}{2} I_z \mathbf{n}_{sh} + \frac{3\sqrt{5}}{16} p_2. \tag{36}$$

5.2 HLSH Hemisphere Multiplies for Volumetric Lighting

This same approach can be used to directly encode quadratic radiance SH to hemisphere-occluded, irradiance-convolved HLSH in an oriented frame. In tangent space, the calculation can be combined

into a single tangent-space matrix multiplication:[4]

$$\begin{bmatrix} C_a \\ c_x \\ c_y \end{bmatrix} = \begin{bmatrix} \frac{1}{4\sqrt{\pi}} & 0 & \frac{2203}{8192\sqrt{3\pi}} & 0 & 0 & 0 & -\frac{7\sqrt{5}}{128\sqrt{\pi}} & 0 & 0 \\ 0 & 0 & 0 & -\frac{2573}{4096\sqrt{3\pi}} & 0 & 0 & 0 & -\frac{11\sqrt{15}}{128\sqrt{\pi}} & 0 \\ 0 & -\frac{2573}{4096\sqrt{3\pi}} & 0 & 0 & 0 & -\frac{11\sqrt{15}}{128\sqrt{\pi}} & 0 & 0 & 0 \end{bmatrix} sh_{rad}, \quad (37)$$

To apply this encoding in world space, we reuse our p_2 components from Equation 35 as the tangent-space l_2^{-1} and l_2^1 components, along with the zonal L2 (quadratic band) coefficient p_{zl2} given by

$$p_{zl2} = \sqrt{\frac{4\pi}{5}} \left(Y_2(\mathbf{n}) \cdot l_2 \right) \quad (38)$$

where $\mathrm{Y}_2(\mathbf{n}_v)$ is the quadratic spherical harmonic band evaluated in the hemisphere normal $\mathbf{n}_v$ and l_2 is the L2 coefficient vector of the source radiance function.[5] In combination, we have a linear spherical harmonic $[p_0, p_1^{-1}, p_1^0, p_1^1]$ with coefficients

$$C_a = \frac{1}{4\sqrt{\pi}} l_0 + \frac{2203}{8192\sqrt{3\pi}} (\mathbf{n}_{sh} \cdot l_1) - \frac{7\sqrt{5}}{128\sqrt{\pi}} p_{zl2}, \quad (39)$$

$$\begin{bmatrix} -c_y \\ c_z \\ -c_x \end{bmatrix} = \frac{2573}{4096\sqrt{3\pi}} \left(l_1 - (\mathbf{n}_{sh} \cdot l_1)\, \mathbf{n}_{sh} \right) + \frac{11\sqrt{15}}{128\sqrt{\pi}} p_2 + (I_z - C_a)\mathbf{n}_{sh}, \quad (40)$$

where $I_z = \mathrm{Y}(\mathbf{n}_v) \cdot \mathrm{SHConvCos}(\mathrm{sh}_{rad})$ is the evaluated irradiance in the hemisphere normal from the source radiance spherical harmonic. Irradiance reconstruction is then given by $I(\mathbf{n}) = C_a + c_x \mathbf{n}_x + c_y \mathbf{n}_y + c_z \mathbf{n}_z$.

If high-roughness indirect specular is approximated as a cosine lobe, this same HLSH-encoded irradiance function can be used for fetching the irradiance in the reflection direction; occlusion is particularly important in this case since the reflection lobe is likely to intersect with the hemisphere.

5.2.1 Cone Multiplies for Volumetric SH. This method can be further generalized to account for a visibility cone oriented along the hemisphere normal with half-angle α. Rather than using the hemispherical Gram matrix $\mathbf{G}_{\Omega^+}$, multiplication with a cone in the hemisphere normal can be done by multiplying the source radiance SH by $\mathbf{G}_{\mathrm{cone}} = \int_0^{2\pi} \int_0^{\alpha} (\mathrm{Y}(\theta, \phi) \sin\theta)\, \mathrm{d}\theta \mathrm{d}\phi$, convolving with the irradiance kernel, and then solving to HLSH as per Section 3.2.2.[6]

While this method requires that the cone normal is aligned with the hemisphere normal, HLSH reconstruction is well behaved for directions away from the cone center; as such, this can be used to model a visibility cone for ambient occlusion by taking the cone center as the hemisphere normal. Although doing so fails to account for the clipping of the cone by the surface hemisphere in certain cases, this still provides significantly more accurate results than no occlusion and is relatively inexpensive.

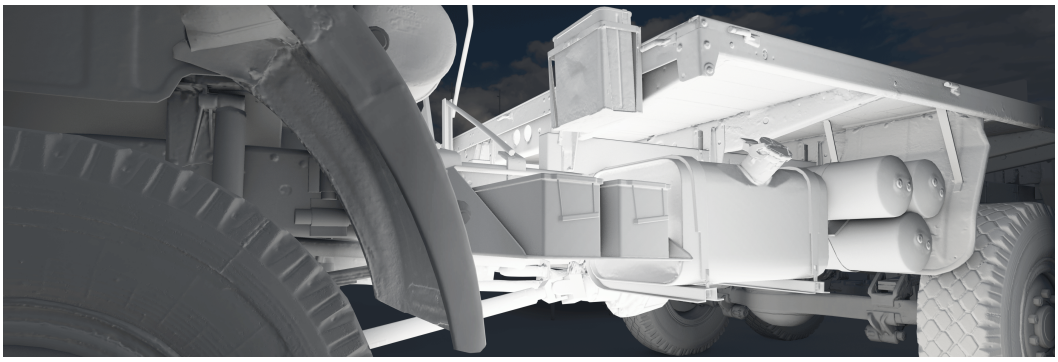
6 CONCLUSIONS AND FUTURE WORK

We have presented a range of novel techniques related to IrradZ-constrained hemispherical lighting. We introduced the HHD and HLSH reconstruction models, each of which carries advantages over AHD for runtime use in level-of-detail and blending, showed how to solve for them from linear spherical harmonics, and compared their accuracy and visual characteristics against AHD. We

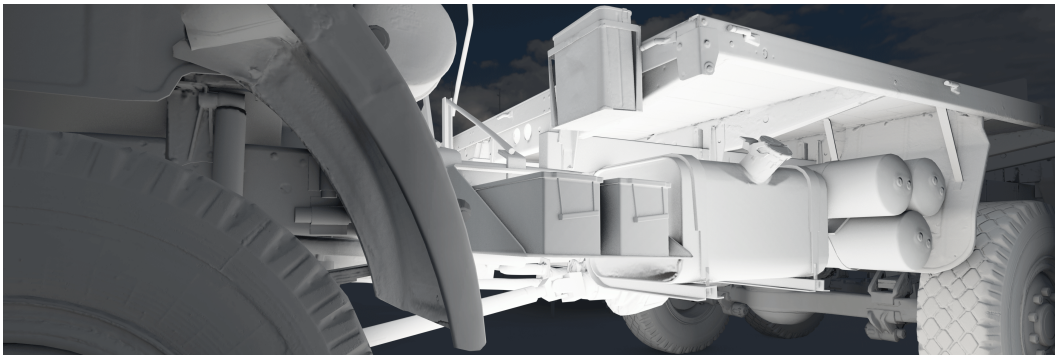
[4] These coefficients were generated by applying an order-4 SH hemisphere multiply, convolving with the diagonal-matrix irradiance kernel, and then solving to HLSH.

[5] $\mathrm{Y}_2(\mathbf{n})$ will have been computed as part of the irradiance reconstruction for I_z and so can be reused.

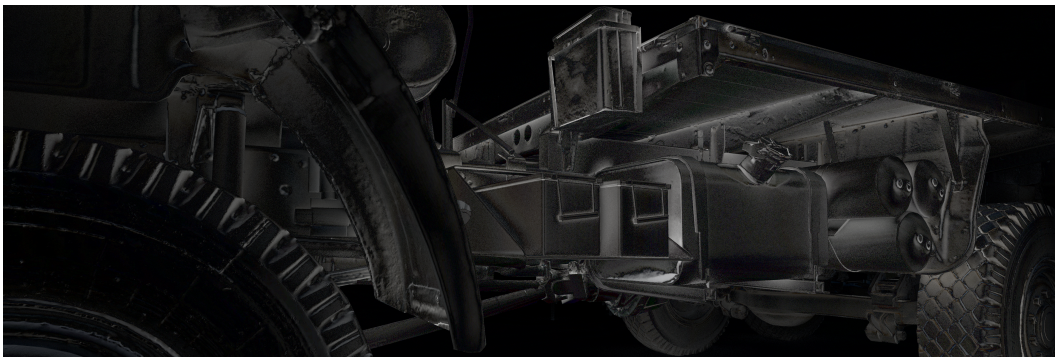
[6] We provide a sample implementation at https://www.shadertoy.com/view/lcVSDh.



(a) Ambient Occlusion Only



(b) Cone Occlusion via HLSH Projection



(c) Difference (in linear space)

Fig. 8. Compared to quadratic SH irradiance evaluation scaled by an AO term (top), evaluation of HLSH produced via multiplication of quadratic SH with a runtime-generated AO visibility cone (bottom) (Section 5.2.1) produces far more visually convincing results and avoids light leakage. Images are secondary diffuse irradiance.

then outlined the nonlinear solve from spherical harmonics for AHD and HHD, correcting errors in prior work and achieving higher quality and more efficient bakes. Finally, we showed how world-space hemisphere multiplies and hemispherical lighting formats can be used to inexpensively add hemispherical or visibility cone occlusion to SH data representing spherical radiance.

Many of these techniques have been used in recent commercial video games. Both the AHD and, more recently, HHD model have been extensively used in the form of IrradZ lightmaps, with the improved nonlinear solve algorithm used in the lightmap generation for both. HHD has also been used, in conjunction with the quadratic-to-linear hemisphere multiply, as a runtime per-vertex level-of-detail scheme for volumetric SH data.

In the future, it would be interesting to further investigate quantization of IrradZ values. We assume the IrradZ parameterization in general due to its local-Z irradiance interpolation, and its use in production on multiple titles. However, while it is clear that linear interpolation of scalar irradiance is desirable, it is not necessarily the case that w_a and $\mathbf{d}$ are the optimal encoded parameters for representing the hemispherical lighting variation when considering quantization and the parameter distribution for the different models. For example, the distribution of optimal w_a values for HHD is shifted higher than that for AHD, and as such the highlight directions are pushed closer to the horizon, indicating that a nonlinear encoding scheme for each may reduce error for quantized values. Additionally, while IrradZ interpolation is generally well-behaved, encoding $C_d\mathbf{d}$ directly (along with I_z) would further improve interpolation behavior. These investigations are left as future work.

ACKNOWLEDGMENTS

The authors thank Michał Iwanicki, Adrien Dubouchet, and Liam Fike for their help in preparing this paper. Naty Hoffman read early drafts and gave helpful feedback.

REFERENCES

James F. Blinn. 1978. Simulation of Wrinkled Surfaces. *SIGGRAPH Comput. Graph.* 12, 3 (aug 1978), 286–292.

John Carmack, Michael Abrash, John Cash, et al. 1996. Quake.

John Carmack, Robert A. Duffy, Jim Dosé, et al. 1999. Quake III: Arena.

Hao Chen. 2008. Lighting and Materials of Halo 3. In *Game Developers Conference*.

Pascal Gautron, Jaroslav Krivánek, Sumanta N Pattanaik, and Kadi Bouatouch. 2004. A Novel Hemispherical Basis for Accurate and Efficient Rendering. *Rendering Techniques* (2004), 321–330.

Ralf Habel, Bogdan Mustata, and Michael Wimmer. 2008. Efficient Spherical Harmonics Lighting with the Preetham Skylight Model. In *Eurographics 2008 - Short Papers*.

Ralf Habel and Michael Wimmer. 2010. Efficient Irradiance Normal Mapping. In *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (Washington, D.C.) *(I3D '10)*. Association for Computing Machinery, New York, NY, USA, 189–195.

John Hable. 2014. Simple and Fast Spherical Harmonic Rotation. http://filmicworlds.com/blog/simple-and-fast-spherical-harmonic-rotation/

JT Hooker. 2016. Volumetric Global Illumination at Treyarch. In *SIGGRAPH Course: Advances in Real-Time Rendering in 3D Graphics and Games*.

Denis Ishmukhametov. 2011. Efficient Irradiance Normal Mapping. (2011). https://www.cg.tuwien.ac.at/research/publications/2011/Ishmukhametov_Denis_2011_EIN/

Michał Iwanicki. 2013. Lighting Technology of the Last of Us. In *ACM SIGGRAPH Talks*.

Michał Iwanicki and Peter-Pike Sloan. 2017. Ambient Dice. In *"Eurographics Symposium on Rendering - Experimental Ideas & Implementations"*, Matthias Zwicker and Pedro Sander (Eds.). The Eurographics Association.

Jorge Jimenez, Xian-Chun Wu, Angelo Pesce, and Adrian Jarabo. 2020. *Practical Real-Time Strategies for Accurate Indirect Occlusion*. Technical Report. Activision.

Jorge Jimenez, Xian-Chun Wu, Angelo Pesce, Adrián Jarabo, and Activision Blizzard. 2016. Practical Real-Time Strategies for Accurate Indirect Occlusion. https://api.semanticscholar.org/CorpusID:222131256

Sam Martin. 2011. Enlighten Real-Time Radiosity. (08 2011). https://enlighten.atlassian.net/wiki/spaces/SDK310/pages/904999344/Directional+irradiance

Gary McTaggart. 2004. Half-Life 2 Source Shading. In *Game Developers Conference*.

Quirin Meyer, Jochen Süßmuth, Gerd Sußner, Marc Stamminger, and Günther Greiner. 2010. On Floating-Point Normal Vectors. *Computer Graphics Forum* 29, 4 (2010), 1405–1409. https://doi.org/10.1111/j.1467-8659.2010.01737.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2010.01737.x

David Neubelt and Matt Pettineo. 2015. Advanced Lighting R&D at Ready At Dawn Studios. In *SIGGRAPH Course: Physically Based Shading in Theory and Practice*.

M. J. Pettineo. 2024. The Baking Lab. https://github.com/TheRealMJP/BakingLab.

Ravi Ramamoorthi and Pat Hanrahan. 2001. An efficient representation for irradiance environment maps. In *SIGGRAPH 2001 Conference Proceedings, August 12–17, 2001, Los Angeles, CA*, ACM (Ed.). ACM Press, pub-ACM:adr, 497–500.

Vincent Schüssler, Eric Heitz, Johannes Hanika, and Carsten Dachsbacher. 2017. Microfacet-based normal mapping for robust Monte Carlo path tracing. *ACM Trans. Graph.* 36, 6, Article 205 (nov 2017), 12 pages. https://doi.org/10.1145/3130800.3130806

Ari Silvennoinen and Peter-Pike Sloan. 2021. Moving Basis Decomposition for Precomputed Light Transport. *Computer Graphics Forum* 40, 4 (2021).

Peter-Pike Sloan and Ari Silvennoinen. 2018. Directional Lightmap Encoding Insights. In *SIGGRAPH Asia Technical Briefs*. Article 12. https://www.ppsloan.org/publications/lmap-sabrief18.pdf

Peter-Pike Sloan and Ari Silvennoinen. 2020. Precomputed Lighting Advances in Call of Duty: Modern Warfare. In *SIGGRAPH Course: Advances in Real-Time Rendering in Games*.

A James Stewart and Michael S Langer. 1996. Towards accurate recovery of shape from shading under diffuse lighting. In *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 411–418.

Bernhard Vogl. 2010. https://dativ.at/lightprobes/

A IRRADZ SELF-BOUNCE

Combining the IrradZ I_z value with a hemispherical model enables reasoning about missing data from self-bounce—the energy reflected by the surface onto itself—not captured in the bake. The fraction of the incident radiance received by a normal $\mathbf{n}$ on a hemisphere oriented around $\mathbf{n}_v$ is given by a hemisphere light $\frac{1}{2}(1 + \mathbf{n} \cdot \mathbf{n}_v)$; therefore, the fraction bounced off the surface is given by $1 - \frac{1}{2}(1 + \mathbf{n}_v \cdot \mathbf{n}) = \frac{1}{2}(1 - \mathbf{n}_v \cdot \mathbf{n})$. The diffuse reflected radiance is then given by a hemisphere colored by the albedo α times I_z in the anti-normal direction:

$$\alpha I_z \frac{1 - \mathbf{n}_v \cdot \mathbf{n}}{2}. \tag{41}$$

This can be inexpensively evaluated at the pixel normal and added to the result. This technique does not put any constraints on the method of evaluating the indirect irradiance and requires only that I_z is known; Geomerics [Martin 2011] independently derived and used this self-bounce term for their similarly-irradiance-preserving directional lighting model. Self-bounce has also been explored in determining shape from shading [Stewart and Langer 1996] and in ambient occlusion [Jimenez et al. 2020].

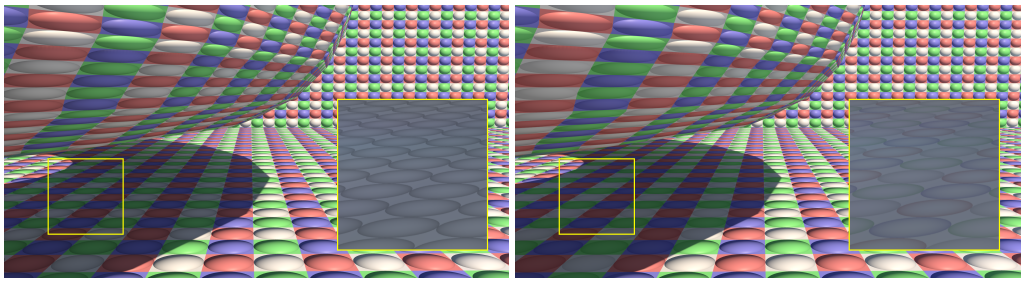
Hemisphere self-bounce is much higher frequency than the incident lighting at a surface, compactly and inexpensively reconstructing detail and self-illumination from a lower-resolution bake (Figure 9). In interactive applications, material texture frequency is usually much higher than the lightmap frequency, and so self-bounce reconstructs indirect illumination detail that would be impractical to directly encode.

A.1 Hemisphere Light Self-Bounce

A derivation for hemisphere self-bounce can be also be produced by examining the SH projections. Hemispheres have non-zero SH projections for odd degree polynomials above linear, while the clamped cosine has only even degrees above quadratic. This means that when computing the outgoing radiance for a Lambertian surface only linear SH are used.

A hemisphere of unit intensity oriented in $\mathbf{h}$ in SH has coefficients

$$\left[\sqrt{\pi}, \frac{-\sqrt{3\pi}\mathbf{h}_y}{2}, \frac{\sqrt{3\pi}\mathbf{h}_z}{2}, \frac{-\sqrt{3\pi}\mathbf{h}_x}{2}\right]. \tag{42}$$



(a) Self-Bounce Off (b) Self-Bounce On

Fig. 9. Total lighting from an HHD lightmap, with indirect lighting shown in a cut-out. Self-bounce smooths the lighting, conserves energy, and adds subtle detail to high-albedo normal-mapped surfaces.

Convolution with a normalized cosine results in

$$\left[\sqrt{\pi}, \frac{-\sqrt{3\pi}\mathbf{h}_y}{3}, \frac{\sqrt{3\pi}\mathbf{h}_z}{3}, \frac{-\sqrt{3\pi}\mathbf{h}_x}{3}\right]. \tag{43}$$

Evaluating linear SH in the direction $\mathbf{n}$ results in a vector

$$\left[\frac{1}{2\sqrt{\pi}}, \frac{-\sqrt{3}\mathbf{n}_y}{2\sqrt{\pi}}, \frac{\sqrt{3}\mathbf{n}_z}{2\sqrt{\pi}}, \frac{-\sqrt{3}\mathbf{n}_x}{2\sqrt{\pi}}\right]. \tag{44}$$

The dot product of the SH vectors can be scaled by the intensity times the albedo A to compute outgoing radiance:

$$\frac{A}{2} + \frac{A(\mathbf{h}\cdot\mathbf{n})}{2} = A\frac{1+(\mathbf{h}\cdot\mathbf{n})}{2}. \tag{45}$$

A second hemisphere with direction $-\mathbf{h}$ and intensity B results in the equation

$$A\frac{1+(\mathbf{h}\cdot\mathbf{n})}{2} + B\frac{1-(\mathbf{h}\cdot\mathbf{n})}{2}. \tag{46}$$

B HEMISPHERICAL GRAM MATRIX

The Gram matrix for spherical harmonics when integrated over the sphere is $\mathbf{I}$, and is sparse and symmetric over the hemisphere. For SH through the quadratics, it is given by:

$$\mathbf{G} = \begin{bmatrix}
\frac{1}{2} & 0 & \frac{\sqrt{3}}{4} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{2} & 0 & 0 & 0 & \frac{3\sqrt{5}}{16} & 0 & 0 & 0 \\
\frac{\sqrt{3}}{4} & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{\sqrt{15}}{16} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{3\sqrt{5}}{16} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\
0 & \frac{3\sqrt{5}}{15} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\
0 & 0 & \frac{\sqrt{15}}{16} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\
0 & 0 & 0 & \frac{3\sqrt{5}}{16} & 0 & 0 & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2}
\end{bmatrix} \tag{47}$$

C DERIVATION OF DIRECT SOLVE FOR HHD FROM LINEAR SPHERICAL HARMONICS

From the definition of our scaled basis $\mathbf{b}$ in Equation 8, it follows that

$$\begin{bmatrix} \mathbf{b}_c \\ \mathbf{b}_x \\ \mathbf{b}_y \\ \mathbf{b}_z \end{bmatrix} = \begin{bmatrix} 2C_a + C_d \\ C_d \mathbf{d}_x \\ C_d \mathbf{d}_y \\ C_a + C_d \mathbf{d}_z \end{bmatrix} = \begin{bmatrix} 2C_a + C_d \\ C_d \mathbf{d}_x \\ C_d \mathbf{d}_y \\ I_z \end{bmatrix}, \tag{48}$$

Using the identities $C_a = I_z w_a$ and $C_d = \frac{I_z(1-w_a)}{\mathbf{d}_z}$ (from Equations (4) and (5)), we have:

$$C_d = \frac{I_z - C_a}{\mathbf{d_z}}, \tag{49}$$

$$\therefore \mathbf{b}_c = 2C_a + \frac{I_z - C_a}{\mathbf{d}_z}. \tag{50}$$

From Equation 48 we can define the highlight direction $\mathbf{d}$ in terms of $\mathbf{b}$, C_a, and C_d:

$$\mathbf{d} = \frac{[\mathbf{b}_x, \mathbf{b}_y, \mathbf{b}_z - C_a]}{C_d}; \tag{51}$$

since $\mathbf{d}$ is a unit vector, this is equivalent to

$$\mathbf{d} = \frac{[\mathbf{b}_x, \mathbf{b}_y, \mathbf{b}_z - C_a]}{\sqrt{(\mathbf{b}_x)^2 + (\mathbf{b}_y)^2 + (\mathbf{b}_z - C_a)^2}}. \tag{52}$$

Substituting $\mathbf{d}$ into Equation 50 yields

$$\begin{aligned} \mathbf{b}_c &= 2C_a + \frac{(\mathbf{b}_z - C_a)\sqrt{\mathbf{b}_x^2 + \mathbf{b}_y^2 + (\mathbf{b}_z - C_a)^2}}{(\mathbf{b}_z - C_a)} \\ &= 2C_a + \sqrt{\mathbf{b}_x^2 + \mathbf{b}_y^2 + (\mathbf{b}_z - C_a)^2}. \end{aligned} \tag{53}$$

This is a quadratic in C_a for which the negative root gives the solution:

$$(\mathbf{b}_c - 2C_a)^2 = \mathbf{b}_x^2 + \mathbf{b}_y^2 + (\mathbf{b}_z - C_a)^2, \tag{54}$$

$$3C_a^2 + (2\mathbf{b}_z - 4\mathbf{b}_c)C_a + (\mathbf{b}_c^2 - (\mathbf{b}_x^2 + \mathbf{b}_y^2 + \mathbf{b}_z^2)) = 0, \tag{55}$$

$$C_a = \frac{(2\mathbf{b}_c - I_z) - \sqrt{(2\mathbf{b}_c - I_z)^2 - 3(\mathbf{b}_c^2 - \|\mathbf{b}_{xyz}\|^2)}}{3}. \tag{56}$$