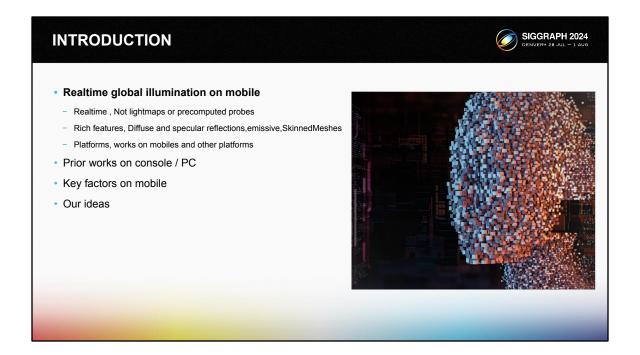


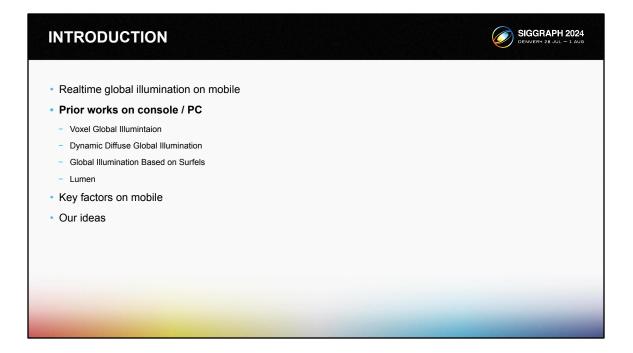
Good afternoon, everyone. I'm Shun from Tencent Games. Today, I'd like to talk about a mobile friendly global illumination solution standing out from the traditional voxelizations.



Before diving into any technical details, it might be helpful to look back at why we need to do this and what happened behind the stage.



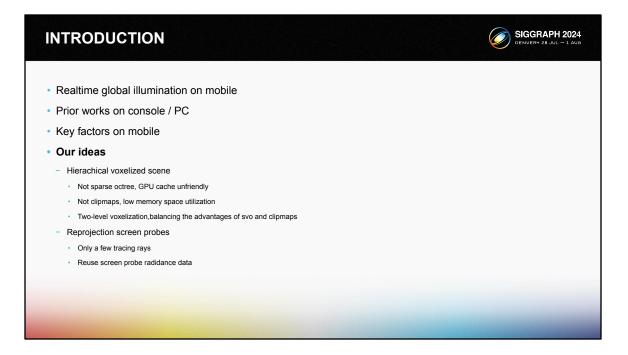
As we all know, offline baking solutions are very popular in mobile game engines. Our motivation at the very beginning was to deliver a real-time GI solution, which does not need baking and can achieve various advanced lighting effects, enabling us to create fully dynamic game scenes on today's mobile devices.



We have seen many real-time global illumination solutions shining on Console and PC. We have carefully studied their PROs and CONs and learned from them while developing our own solution. For example, the voxel-based, dynamic probe-based, and surfels-based solutions, all of them provide a simplified representation of the scene with lighting cache, and make it possible to fast sample lighting data represented by that. However, all these options either have an inherent limitation on visual effects, or seriously challenge the rendering capacity of mobile devices. For instance, voxelization based on clipmaps may cause light leakage due to the insufficient voxel precision when the shading point is in distance. Dynamic probe-based solution requires precise raytracing for dynamic generation and offset calculation to address issues like light leakage or dark spots. Surfels solutions, which are often based on Gbuffer, suffer from the lack of scene representation when the camera enters the scene for the first time. Even though Lumen does not have all these issues, it heavily relies on SDF and Meshcard, which are very performance demanding.

INTRODUCTION	SIGGRAPH 2024 DENVER* 28 JUL – 1 AUG
<ul> <li>Realtime global illumination on mobile</li> <li>Prior works on console / PC</li> <li><b>6 Key factors on mobile</b></li> <li>Peformance Optimizations</li> <li>Bandwidth</li> <li>Memory Usage</li> <li>How many gpu data for GI</li> <li>Use Device Capabilities</li> <li>Main Streaming Devices</li> <li>Battery Usage</li> <li>Energy efficent algorithms</li> <li>Our ideas</li> </ul>	

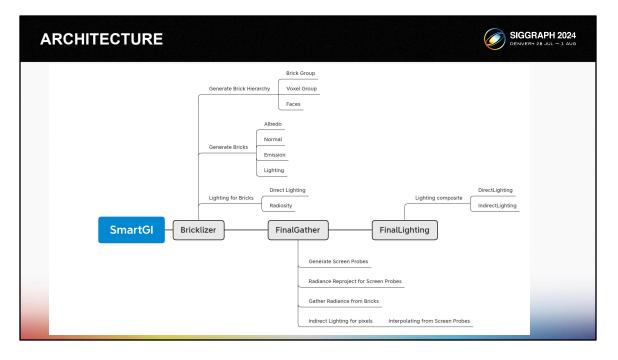
Thanks to all the analyses, we believe that the following factors are important when implementing any mobile GI, such as GPU bandwidth, memory usage, mobile hardware capabilities, and power consumption. We all know that there is still a significant gap between mobile and desktop GPU bandwidth capacities, which matters to performance and power consumption. Memory usage is also crucial; excessive memory footprint can easily lead to a crash. Although high-end mobile GPUs now support raytracing, many devices in the market are left behind. Higher power consumption means battery drainage and overheating coming much earlier, which downclocks the GPU with worse performance.



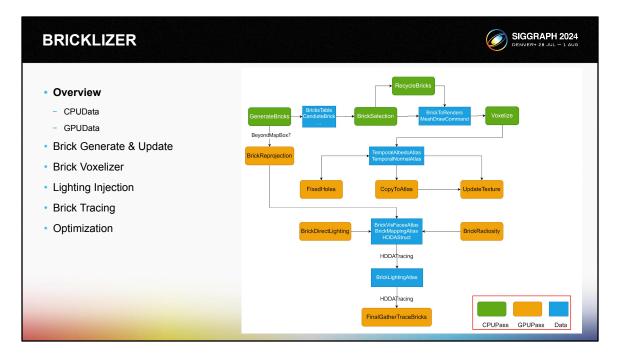
With a considerable amount of learning and struggling, we finally reached the point that a brand new voxel-based algorithm is coming up. It was built on the efficient sparse octrees and cache-friendly clipmaps to achieving a multi-level voxel-based scene representation. We also didn't forget screen probes, as what is in Lumen, to minimize raytracing workloads. Moreover, we improved it by introducing a screen probe reprojection technique, to further simplify lighting with even fewer rays.



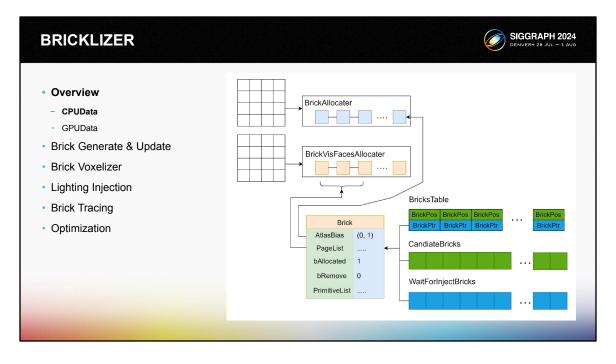
Now, let's dive into some of the most interesting implementation details.



Here is the overall architecture diagram of the solution, which consists of three main parts: Bricklizer, FinalGather, and Lighting Composition. Bricklizer is our approach to achieving hierarchical voxelization. FinalGather is the process of collecting lighting based on screen probes and interpolating lighting for each pixel shading from the radiance information of the screen probes. The final lighting composition process is relatively simple, involving the overlaying of direct and indirect lighting.



First, let's take a look at the overall process of Bricklizer. On the CPU side, we generate candidate blocks, referred to as bricks, based on the visible range. These candidate bricks are placed in a candidate queue, and a fixed number of bricks are voxelized each frame based on a certain priority. After voxelization, we have a process to repair and optimize the voxelization, such as filling holes in specific directions of the voxels and eliminating invalid faces between two voxels. Once the brick voxelization is complete, we perform lighting injection, including sampling direct light sources and reflecting light from other voxels onto the current voxel. Finally, we can use the HDDA algorithm to calculate the radiance of the screen probes.



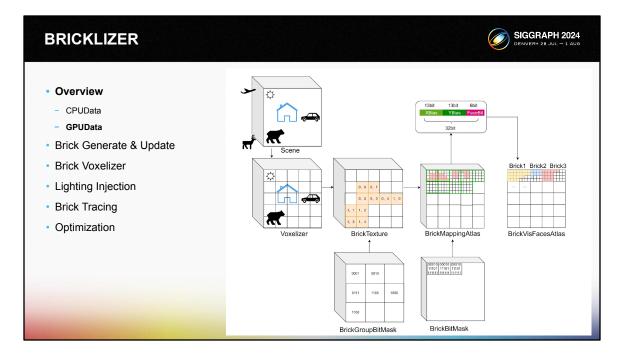
The core data structure in the CPU is illustrated as follows. During initialization, two allocators are initialized based on the given texture sizes of BrickMappingAtlas and BrickVisFacesAtlas, and an array is created by stringing together the Bias as the base elements. Some important parameters in Brick are:

AtlasBias: The Bias obtained from BrickAllocator, used for mapping to BrickMappingAtlas.

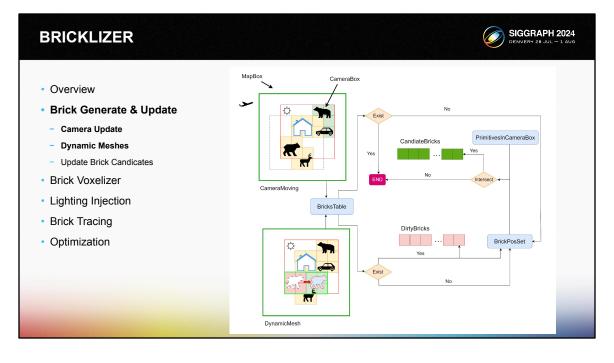
PageList: Maintains the PageBias of several Bricks in VisFacesAtlas in the form of an array.

bAllocated: Indicates whether space has been allocated and is also used to distinguish whether capture is complete.

bRemove: Bricks exist in three lists in the form of smart pointers. If a brick needs to be deleted, set this flag to 1, and then delete it from the BricksTable. The other two lists will check this flag before processing. PrimitiveList contains all the primitives intersect with the brick.

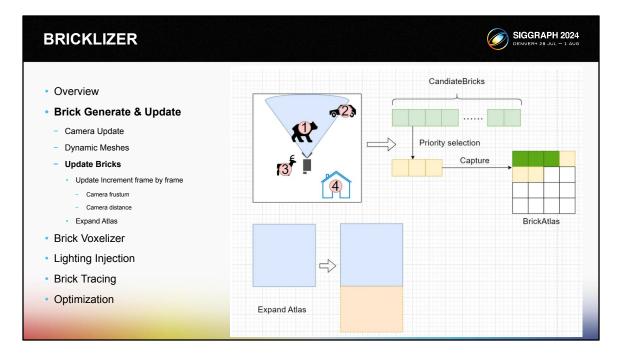


The resource organization structure in the GPU is illustrated as follows. For the 3D scene shown, BrickTexture stores the actual storage locations of each Brick with a value in BrickMappingAtlas. The default configuration is that each Brick covers a range of 4x4x4 m, with a total coverage of 512x512x128 m. BrickMappingAtlas stores data in Voxel units, where each Voxel stores a mapping pointer to the next level, composed of 32 bits. 26 bits are used to represent x and y-axis offsets, and 6 bits indicate the presence or absence of each face. The actual storage carrier for each voxel face is a 2D Atlas called BrickVisFacesAtlas, which tightly stores every valid face (where the adjacent voxel is empty or translucent). Only the last page of each Brick may contain intra-page fragments, so the space utilization is extremely high. The following two 3D Textures are auxiliary structures for HDDA Tracing. Among them, BrickGroupTexture uses 4x4x4 Bricks as a composition unit, with each grid storing a 64-bit BitMask used to indicate the presence or absence of Bricks. Similarly, BrickBitMask uses BitMask to represent the presence or absence of corresponding voxels.

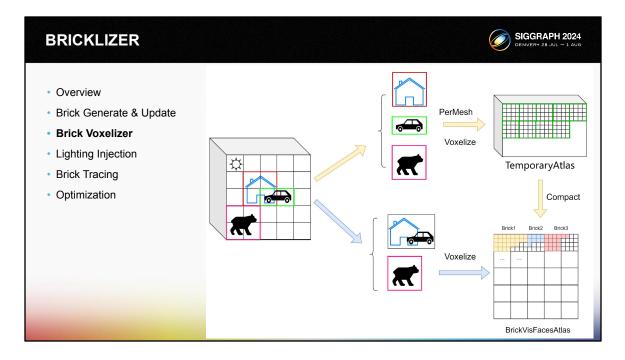


Besides full updates caused by entering the scene or BrickReset, only two changes can lead to Brick updates: one is the addition of Bricks caused by camera movement, and the other is updates caused by Mesh changes, including Mesh deletion, addition, and movement. Full updates and camera movement updates can be attributed to updates based on bounding boxes, requiring normalization of the bounding boxes to ensure that their boundaries align with Brick boundaries. Then, for each BrickPos inside, use BricksTable to determine whether the corresponding Brick exists. If not, add it to the BrickPosSet.

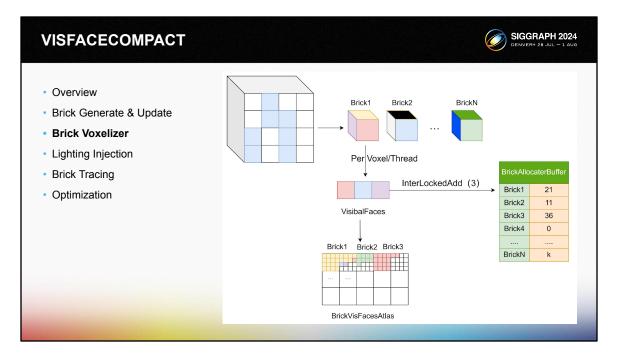
Next, obtain the updated bounding box of the Mesh, expand it to align its boundaries with Brick boundaries, and then iterate through BrickPos. If it already exists, it needs to be cleared and regenerated. If it does not exist, add it to the BrickPosSet for subsequent processing. All Bricks waiting for updates are added to the BrickPosSet, and then multithreading is used to initially cull Meshes, retaining only those intersecting with the updated bounding box. Then, multithreading is used to traverse all Bricks and Primitives. After traversal, each Brick will have its own Mesh list, and Bricks that do not intersect with Meshes can be deleted. The rest are added to CandiateBricks. In the above steps, Bricks that need to be deleted in this frame will also be added to DirtyBricks for processing.



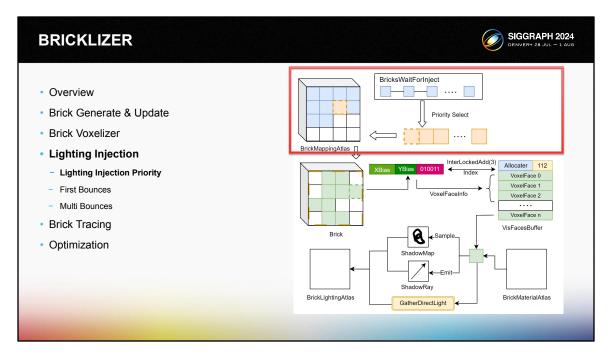
After updating the Bricks, it is necessary to select k Bricks from the CandiateBricks for updating in the current frame. The selection strategy can rely on factors such as the distance from the camera or the distance within the view frustum. As shown in the figure below, priority should be given to updating the bricks within the view frustum, followed by bricks closer to the camera. After the selection, it is essential to ensure that there are remaining Bricks and Pages in both Allocators. If there are not enough, it will trigger the recycling process or memory expansion logic. The recycling process will reclaim the space of all Bricks outside the update range. Memory expansion will double the size of BrickMappingAtlas or BrickVisFacesAtlas, and an additional pass will be required to move the original Atlas data to the corresponding locations.



The voxelization process currently has significant optimization potential and is also the most time-consuming part. In the current strategy, each brick generates a corresponding MeshDrawCommand based on its Mesh list. Each Brick then undergoes voxelization in three directions, and the results are temporarily written into a temporary 3D texture generated in the current frame for subsequent processing. When voxelizing, it is advisable to combine MultiView and voxelize in three directions simultaneously to reduce DrawCalls and other operations.

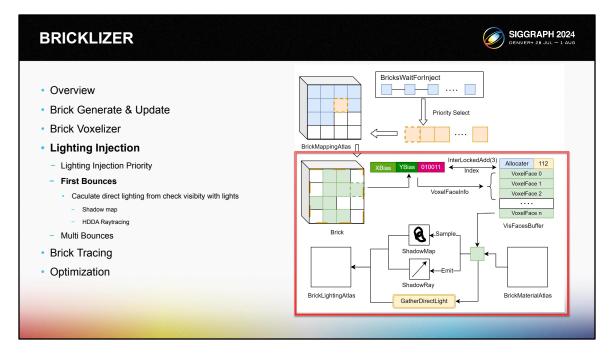


Each brick is pre-allocated with 6 to 8 pages, and each page occupies 8x8 pixels. The compression and allocation of all valid faces for voxels within a brick are handled here. The specific bias for each voxel's visface is tightly arranged within these pages, ensuring that only the last page may have unused space. If the 8 pages are insufficient, we will feedback to the CPU to request additional pages. The purpose of using large-grained pages for allocation is to facilitate efficient page recycling.



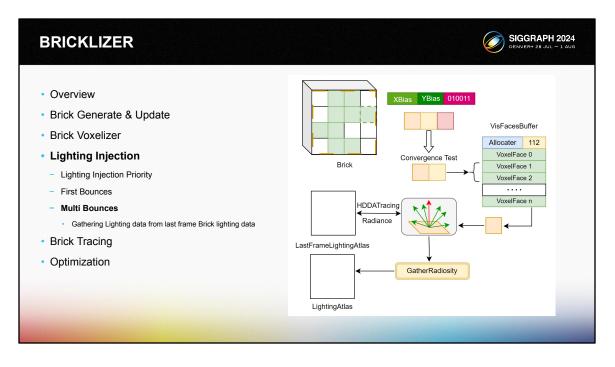
Lighting calculations can be divided into two parts for processing: direct lighting and indirect lighting.

Before calculating direct lighting for each frame, k Bricks are selected from a candidate list to undergo direct lighting computations. The selection can be based on sorting factors such as their position within the view frustum or their distance from the camera.



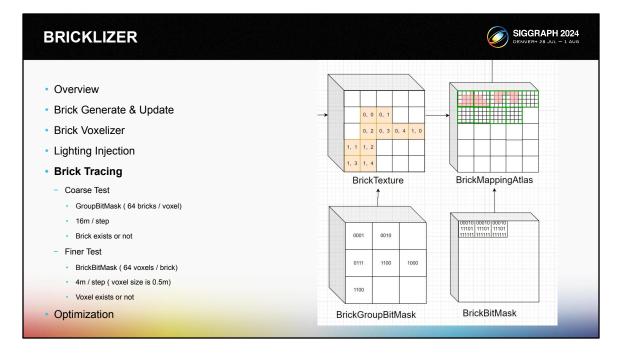
Prior to the direct lighting calculation, to maximize hardware utilization, all valid faces within a brick are compacted into a VisBuffer. As illustrated in the diagram below, the upper half depicts the compaction logic. Since the BrickMappingAtlas already stores the storage information for each face corresponding to a voxel, this value can be directly retrieved. If a voxel exists and has three valid faces, three consecutive spaces are requested from the Allocator, and the index values of these valid faces in the VisFacesAtlas are written into the buffer.

Based on the Allocator, the number of thread groups can be determined, with each thread handling one valid VoxelFace. The basic material properties are fetched from the BrickAtlas. If the ShadowMap is valid, a direct sample can be taken; otherwise, a separate ShadowRay needs to be cast to determine if the point is in shadow. Subsequently, direct lighting is calculated based on factors such as normal weights, light source type, and material information. Finally, the results are written into the LightingAtlas.



## Multi Bounces:

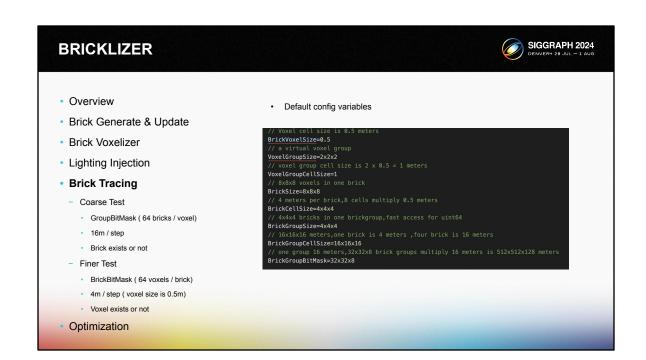
The process for indirect lighting calculations is similar to direct lighting. First, k Bricks requiring radiance updates in the current frame are identified. Then, the indices of the effective faces of these Bricks are compacted into a Buffer to facilitate subsequent GPU thread group and thread allocation. For each effective face, n rays are emitted in hemispherical directions, and the collected results are weighted and averaged to calculate the Irradiance and store it.



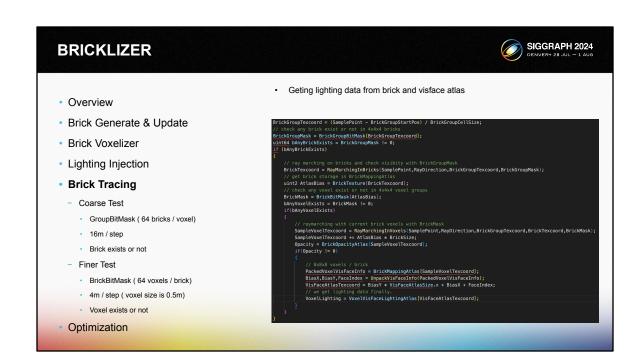
Utilizing BrickGroupBitMask and BrickBitMask, the HDDA algorithm can be implemented, enabling fast raycasting and intersection detection. The specific processing logic is as follows:

Starting Point Offset: When a ray originates from a starting point, the starting point needs to be translated outward along the ray direction to the surface of a voxel for tracing, avoiding self-intersection.

BrickGroupTracing: Based on the starting point's position, the corresponding BrickGroup is located. If the Group exists, the corresponding 64-bit BitMask is retrieved.



Here are the default config variables. The default voxelsize is 0.5 meters, there are 8x8x8 voxels in one brick, thus one brick covers the range of 4 meters. There are 4x4x4 bricks in one brickgroup, so the brickgroup covers the range of 16 meters. We have the number of 32x32x8 brickgroups , so the we can cover 512x512x128 meters.

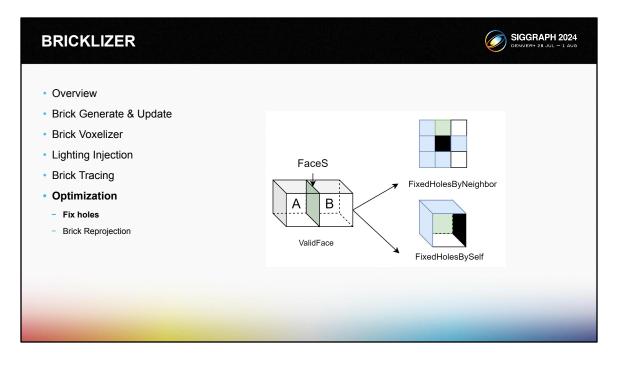


Here is the Pseudocode how we get lighting data from brick and

visfacelightingatlas.First raymarching is in bricks levels with brickgroupbitmask.Then the voxels level raymarching.Each raymarching process use 64bits to check 64 bricks or voxels is presence or not.

BRICKLIZER	SIGGRAPH 2024 DENVER+ 28 JUL - 1 AUC
<ul> <li>Overview</li> <li>Brick Generate &amp; Update</li> <li>Brick Voxelizer</li> <li>Lighting Injection</li> </ul>	<ul> <li>RayMarching with BrickMask</li> <li>[snt3 RayMarchingTeBricks(SamplePoint,SampleDir,BrickGroupTexcoord,BrickGroupMask) (stop = 0; while(stop &lt; 64)</li> </ul>
<ul> <li>Brick Tracing</li> <li>Coarse Test <ul> <li>GroupBitMask ( 64 bricks / voxel)</li> <li>16m / step</li> <li>Brick exists or not</li> </ul> </li> <li>Finer Test <ul> <li>BrickBitMask ( 64 voxels / brick)</li> <li>4m / step ( voxel size is 0.5m)</li> </ul> </li> </ul>	<pre>SampleBickTescoord = (SampleBint = BrickGroupTescoord = VFickTevupCellSize; BrickTest = BrickTescoord + (BrickGroupMask) { return SampleBickTescoord; else { SamplePoint += SampleDir * BrickCellSize; } return int3(-1,-1,-1); } </pre>
Voxel exists or not     Optimization	

the Pseudocode how we get valid brick from brickmask.and raymarching on voxel level is similar.

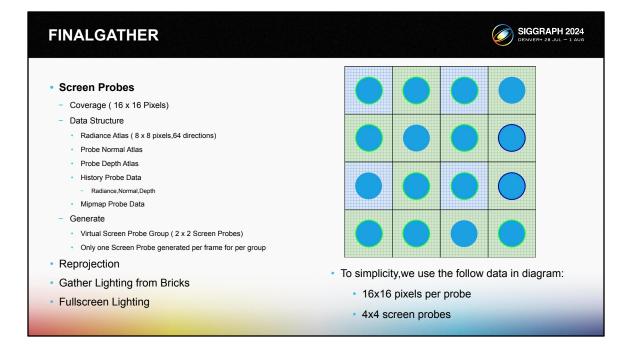


Firstly, it is necessary to identify which faces need to be repaired. To briefly explain the concept of effective faces: simply put, only observable faces are considered effective. As shown in the figure below, FaceS is the overlapping face between Voxel A and Voxel B. If B is an empty voxel or a semi-transparent voxel (i.e., Opacity < 1), FaceS can be considered visible and is thus an effective face. For each non-empty voxel, if there are effective faces and some of these faces are empty, they are added to a buffer waiting for subsequent repair processing.

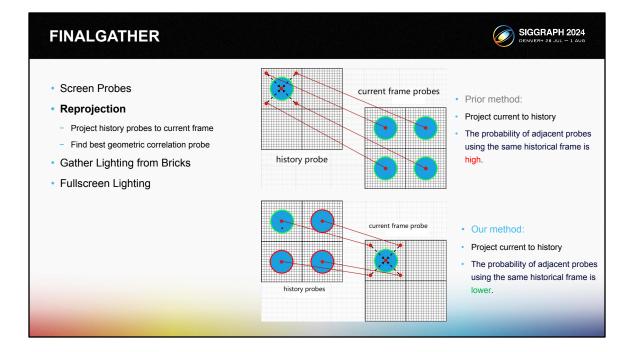
Specifically, there are two options for repair. Repairs will prioritize searching and patching within a 3x3 grid of faces based on the plane the face resides in, retrieving its Albedo and Normal for filling. Another simpler and more direct option is to directly take the weighted average of the Albedo of other faces belonging to the same voxel as this face's Albedo, and the Normal can be obtained by rotating the Normals of other faces (or simply selecting the orientation of the voxel face as the normal direction—the effect is basically correct and without light leakage).

BRICKLIZER	SIGGRAPH 2024 DENVER+ 28 JUL - 1 AUG
<ul> <li>Overview</li> <li>Brick Generate &amp; Update</li> <li>Brick Voxelizer</li> <li>Lighting Injection</li> <li>Brick Tracing</li> <li>Optimization</li> <li>Fix holes</li> <li>Brick Reprojection</li> </ul>	<text></text>

During camera movement, if it is detected in a frame that the updated bounding box exceeds the range of the mapped bounding box, the corresponding reprojection logic needs to be executed. On the CPU, all Bricks are traversed, and those that exceed the new mapped bounding box are deleted. At the same time, basic parameters such as AtlasOrigin need to be updated. On the GPU, only the corresponding BrickTexture and HDDAStruct need to be updated. An offset is calculated for the remaining Bricks and passed as a parameter to the ReprojectionPass, shifting the corresponding Bricks accordingly.



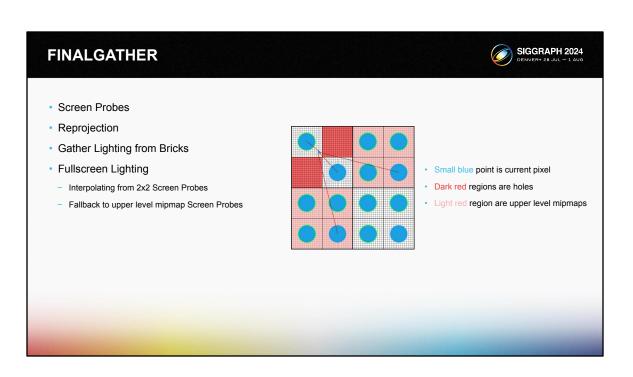
After introducing how to implement voxelization using bricklizer, we move on to the illumination sampling stage. As mentioned earlier, we utilize a screenprobe approach similar to lumen for efficient illumination sampling. As shown in the diagram, we segment the screen into 16x16 blocks, and generate a screenprobe on each block. Each screenprobe records radiance information in 64 directions. In addition to radiance information, we also generate normals and depth information for the screenprobe. Furthermore, we keep track of the screenprobe's relevant information from the previous frame. There is also mipmap data for the probes, and I will explain its purpose later. Our approach to calculating screenprobe data differs from lumen, as we group 2x2 screenprobes into a virtual group and randomly select one probe from this group for computation each frame, thus reducing the computational load to one-quarter of lumen's.



To further reduce the computational cost of calculating the radiance of screenprobes per frame, we maximize the utilization of historical screenprobe radiance data. However, there are two modes of reprojection: one is to project the current frame's screenprobes into historical frames to find neighboring screenprobes, and the other is to reproject historical frame's screenprobes into the current frame and use the geometrically closest historical frame to fill in the current frame. We found that this method results in a lower conflict rate for screenprobes compared to the former approach. For screenprobes that fail reprojection, we proceed to the next step: radiance data collection and computation.

<ul> <li>Screen Probes</li> <li>Reprojection</li> <li>Gather Lighting from Bricks</li> <li>Tracing With HDDA</li> <li>Gather Lighting Data from best voxel face</li> <li>Fullscreen Lighting</li> <li>Gather Radiance from Bricks</li> <li>Gather Radiance from Bricks</li> <li>Find Gather Radiance from Bricks</li> <li>Find nearest face of voxel around the hitpoint</li> <li>Gather Radiance from Bricks</li> <li>Find nearest face of voxel around the hitpoint</li> <li>Gather Radiance for per ray</li> <li>Idirect Lighting for pixels</li> <li>Interpolating from Screen Probes</li> </ul>	FINALGATHER	SIGGRAPH 2024 DENVER* 28 JUL - 1 AUG
	<ul> <li>Reprojection</li> <li>Gather Lighting from Bricks</li> <li>Tracing With HDDA</li> <li>Gather Lighting Data from best voxel face</li> </ul>	Generate Screen Probes Radiance Reproject for Screen Probes Send 8x8 Rays from Probe Tracing with GroupBitMask Tracing with BrickBitMask Find nearest face of voxel around the hitpoint Get the lighting data of the face Store Radiance for per ray

After determining which screenprobes require recalculation of radiance data, we cast rays into the scene using the HDDA algorithm to our voxelized representation, "Bricks," to obtain illumination data. Unlike traditional voxelization that only stores illumination data on voxels, we store separate data for each direction of the voxel. Additionally, we apply the invalid face elimination logic described earlier in the bricklizer description. This allows us to obtain different illumination data when rays hit the voxel from different directions. For example, a thin wall has different illumination on its indoor and outdoor sides.



The illumination data we calculate is based on screenprobes, but ultimately, we need illumination data for each pixel shader. The approach we use is relatively simple: we find the four screenprobes surrounding the current pixel and calculate the pixel's illumination information based on the weighted average of the geometric information. However, you may have noticed that we mentioned earlier that only one-quarter of the screenprobes in the virtual group are computed each frame, which means some screenprobes may not have actual illumination data and may fail reprojection. In such cases, the mipmap of the screenprobe comes into play. We search for appropriate screenprobes in the mipmap hierarchy and interpolate them to the current pixel point.



I am done with the detailed introduction of our algorithm, now let's take a look at how the test result in the real world may look like.

Memory				
<ul> <li>28 MB</li> <li>512 x 512 x 128</li> </ul>	Algorithm	Memory	Peformance	Coverage Area
Lighting Radiosity Off	Lumen	>280MB	-	200m
<ul> <li>0.5 ms</li> <li>Radiosity On</li> </ul>	VXGI	~50MB	~2.0ms	256m
• 1.5 ms	BRICKGI	~30MB	~2.0ms	512m
<ul> <li>Voxelize</li> <li>15 bricks</li> <li>Cached</li> <li>0 ms</li> </ul>				

Here, we have used the 2023 flagship mobile devices to test VXGI and BrickGI, while having Lumen on PC as the reference. As you can see, our BrickGI only needs less than 30MB of memory for a spherical volume with a radius of 512 meters. In contrast, Lumen and the traditional clipmap approach take more than 50MB. Furthermore, in terms of performance, we can complete all the GI calculations within 2ms.

	<ul> <li>4 level clipmaps me</li> </ul>	mory usage:		
Memory - Lumen	Texture	Size	Format	Memory
• > 282MB	AlbedoTexture	64x256x192	R32_UINT	12MB
<ul> <li>Clipmap Method</li> <li>~50MB</li> </ul>	NormalTexture	64x256x192	R32_UINT	12MB
<ul> <li>Our Bricklizer Method</li> <li>~30MB</li> </ul>	EmissiveTexture	64x256x192	R11G11B10_FLOAT	12MB
Lighting	LightingTexture	64x256x192	R11G11B10_FLOAT	12MB
Voxelize	OpacityTexture	64x64x4x32	R8	512KB
	VoxelVisBuffers	64x64x64x6	R32_UINT	3MB
	Total	1	I	51.5MB

Here is the detailed memory usage of the voxelization method based on clipmap, which is approximately 50MB.

Memory         Texture         Size         Format           - Lumen         BrickGroupBitMask         32x64x8         R32_UINT	Memory
Lumen     BrickGroupBitMask 32x64x8 R32 UINT	
BriokoroupBlandok BERGING ROLLONING	64KB
> 282MB BrickTexture 128x128x32 R16G16_UINT	2MB
Clipmap Method     Some Sector S	16KB
- Our Bricklizer Method BrickOpacityAtlas 512x1024x8 R8	2MB
• ~30MB BrickMappingAtlas 512x1024x8 R32_UINT	8MB
Lighting BrickVisFacesAlbedoAtlas 1024x1024 R32_UINT	4MB
BrickVisFacesNormalAtlas 1024x1024 R32 UINT	4MB
Voxelize BrickVisFacesLightingAtlas 1024x1024 R11G11B10_FLOAT	4MB
BrickVisFacesEmissiveAtlas 1024x1024 R11G11B10_FLOAT	4MB

Here is the detailed memory usage of our voxelization method based on the latest bricklizer technology. It consumes approximately 30MB of memory, and the scene coverage is twice that of a 4-level clipmap, reaching up to 512 meters.



Let's wrap up the advantages and possible future improvements of our algorithm.

CONCLUSION	SIGGRAPH 2024 Denver+ 28 JUL - 1 AUC
Advantages	
<ul> <li>More efficient data storage rate and lower memory usage</li> </ul>	
- Higher-precision scene representation	
<ul> <li>More friendly to the GPU cache</li> </ul>	
<ul> <li>No need with hardware ray tracing</li> </ul>	
<ul> <li>Fewer ray tracing calculations</li> </ul>	
Disadvantages	
<ul> <li>Not friendly to mirror reflection</li> </ul>	
<ul> <li>Not friendly to the changes of huge object</li> </ul>	
Future Plan	
- Reduce overdraw	
<ul> <li>Many lights injection</li> </ul>	
- Texture compression	
<ul> <li>For the Mean Squared Error Metric (MSME) evaluation of Radiosity</li> </ul>	

The advantages of the system include a more efficient data storage rate that leads to lower memory usage, enabling a higher-precision representation of the scene. Furthermore, it is designed to be more compatible with the GPU cache, eliminating the need for hardware-based ray tracing and thus reducing the number of ray tracing calculations required.

However, there are also some disadvantages, such as its inability to handle mirror reflections effectively and difficulties in adapting to changes with huge objects. Looking ahead, our future plan involves implementing strategies to reduce overdraw, incorporating multiple light injections, and compressing textures. These enhancements will be particularly useful when evaluating Radiosity using the Mean Squared Error Metric (MSME), aiming to improve both performance and accuracy.



Thank you for your time. Feel free to reach out to me if you have any further questions.