Gleanmer: A 6 mW SoC for Real-Time 3D Gaussian Occupancy Mapping

Zih-Sing Fu*, **Peter Zhi Xuan Li***, **Sertac Karaman**, **Vivienne Sze**
Massachusetts Institute of Technology, ***Equal contribution***

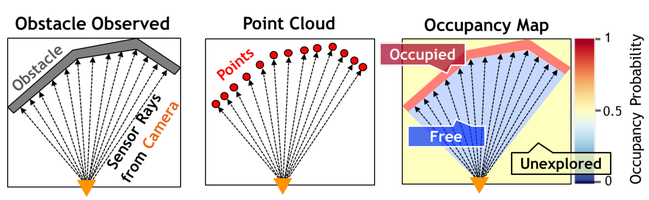
LEAN

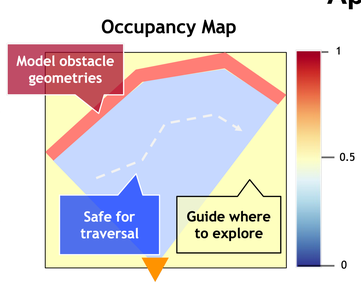
The **first fabricated SoC** for real-time **3D Gaussian occupancy mapping at 6 mW**, achieved by **co-designing the GMMap algorithm with dedicated hardware**, enabling **spatial perception** on AR/VR headsets, drones, and insect-scale robots.

1. Motivation & Applications

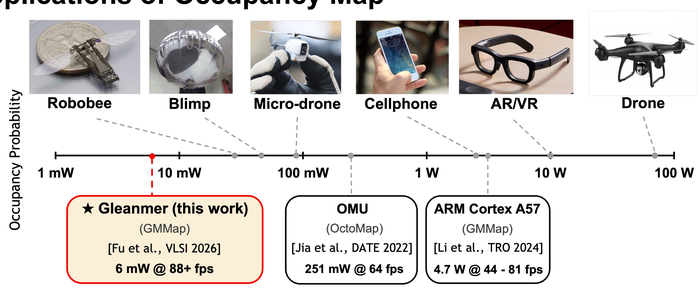
What is an Occupancy Map?

Mapping refers to the process of building a **spatial representation** of an environment from sensor rays.





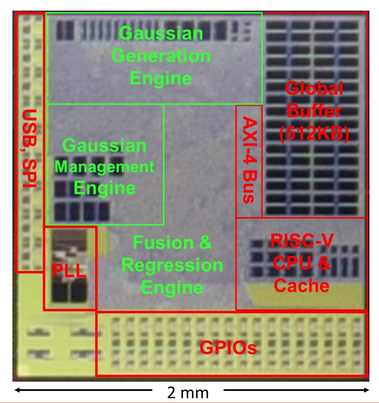
Applications of Occupancy Map



Occupancy maps classify the 3D environment into **occupied, free, and unexplored regions** → essential for a wide range of spatial perception tasks.

Occupied regions tells AR/VR headsets where real surfaces are so virtual objects render correctly, **free regions** enable safe robot navigation, and **unexplored regions** guide robots toward regions worth exploring.

Gleanmer: 6mW SoC for 3D Mapping

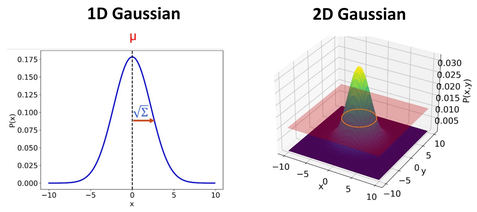


Technology	16 nm
Area	4mm²
Logic Gates	3.2 million
SRAM	622 KB
Frequency	125 MHz (construction) 250 MHz (query)
Supply Voltage	0.85 V (construction) 0.95 V (query)
Construction Throughput	88 – 331 fps
Query Throughput	540K – 1320K cps*
Construction Power	4 – 6 mW
Query Power	6 mW

* cps: coordinate per second

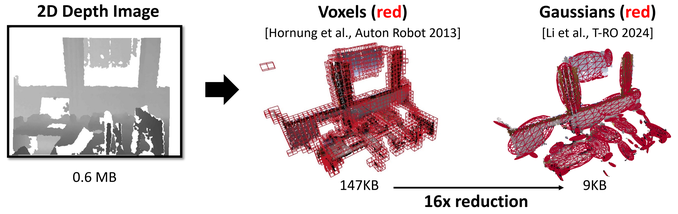
2. Background: Gaussians

What is a Gaussian Distribution?

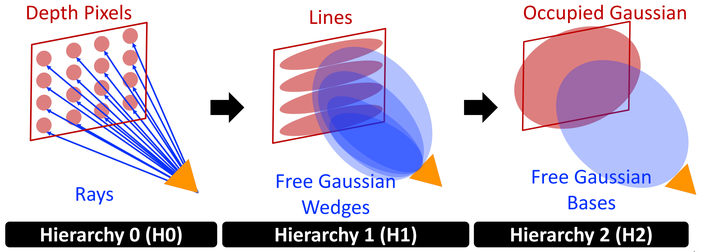


Probabilistic distribution parametrized by mean (**μ**) and covariance matrix (**Σ**). Iso-surface is visualized as an **ellipsoid**.

Gaussians Compactly Represent 3D Geometry



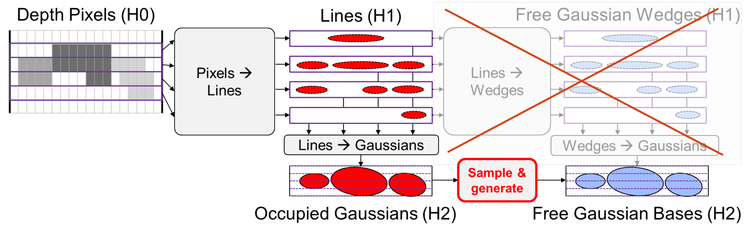
Gaussian Construction [Li et al., T-RO 2024]: H0 to H2



Decreasing arithmetic intensity; increasing compactness & memory access

3. Co-Design Contributions

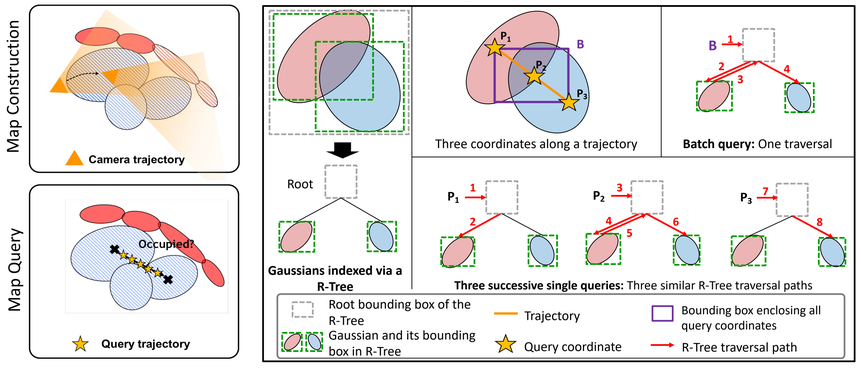
Efficient Hierarchy Conversion (H0 to H2)



Eliminating wedges (H1) by generating free Gaussian bases directly **from occupied Gaussians (H2)** reduces map construction energy by up to **63%**.

Amplified Optimizations at Gaussian Level (H2)

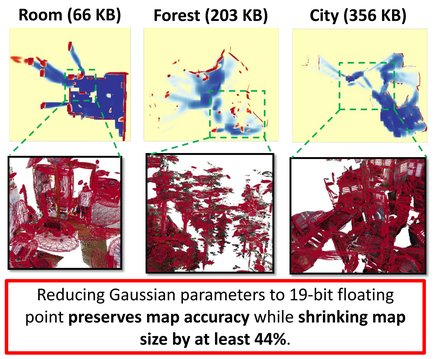
Cache Spatially Nearby Gaussians | **Reusing Gaussians across Batch Query**



Cache spatially nearby Gaussians: Up to **7×** construction and **9×** query throughput.
Batch 16 queries into one R-tree traversal: Up to **81%** lower query energy, **<2%** area overhead.

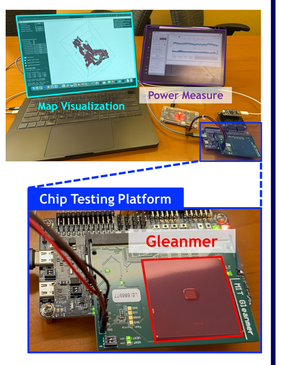
4. Results & Live Demonstration

Map Visualization



Reducing Gaussian parameters to 19-bit floating point **preserves map accuracy** while **shrinking map size by at least 44%**.

Live Demonstration



Comparison with Other Hardware

	NVIDIA Jetson TX2 (Cortex A57 + 256-core GPU)	OMU [Jia et al., DATE 2022]	Gleanmer (This work)
Mapping Framework	GMMap [Li et al., T-RO 2024]	OctoMap [Hornung et al., Auton Robot 2013]	GMMap [Li et al., T-RO 2024]
Map Accuracy	96% – 99%	93% – 97%	96% – 99%
Map Construction Throughput	44 – 81 fps	61 – 64 fps	**88 – 331 fps**
Map Query Throughput	500 – 800K cps	Not Reported	**540K – 1320K cps**
Average Map Construction Power	4.7 W	251 mW	**4.5 mW**
Average Map Query Power	2.0 W	Not Reported	**5.7 mW**

With comparable accuracy, Gleanmer (16 nm) consumes at least **341× less power than Jetson TX2** (16nm) and **44× less than OMU** (12 nm).

Sponsors: MathWorks Fellowship, Amazon, NSF CPS 2400541, Intel USP. **Acknowledgement**: Berkeley EECS for SophiaLake chip testing platform.