Sensory Discovery: Learning from Sensors, Learning for Sensors

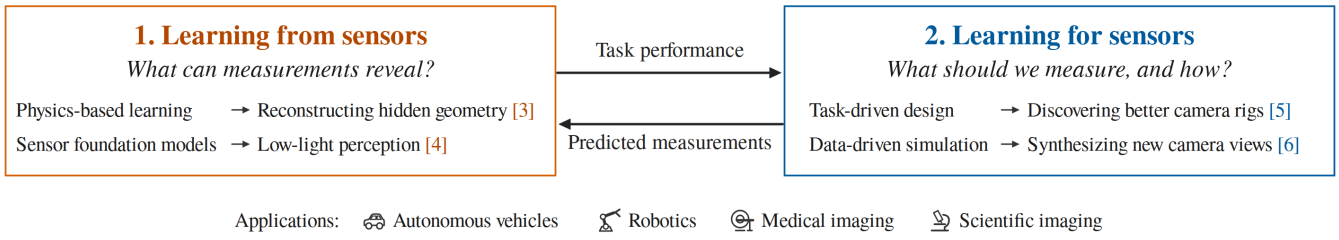
My goal is to discover new ways for machines to sense the world, expanding what they can perceive and accomplish. I work at the intersection of artificial intelligence (AI), computer vision, and sensing, with expertise in single-photon vision. Specifically, I develop methods that learn from sensors, recovering information conventional sensing pipelines discard, and methods that learn for sensors, discovering what to measure and how to measure it for a given task.

Evolution designs an animal's senses, brain, and behavior together, adapting them to the demands of its environment. Engineered sensing systems, by contrast, are often designed in stages: hardware engineers choose what to measure; signal-processing experts transform measurements into representations; AI researchers train models on these representations; and roboticists use these models to guide action. Designing these stages independently can lead to systems that discard task-relevant information and overlook new sensing strategies enabled by jointly designing sensing, perception, and action. Inspired by evolution, my group will develop Sensory Discovery: a closed loop that proposes sensing systems for a given task, predicts and interprets their measurements, evaluates downstream task performance, and uses this feedback to create better designs. This loop will both leverage well-understood sensing cues and, ultimately, discover new ones and learn how they arise through real-world interaction and experimentation.

Sensory Discovery could transform what machines perceive and how they acquire information, particularly when signals are faint, hidden, or difficult to measure. An autonomous vehicle could exploit sparse photon measurements to detect a deer crossing a dark road; a medical scanner could adapt its acquisition to better characterize a tumor deep in the brain; and a search-and-rescue robot could choose where and how to measure to locate a person beneath rubble.

But how can we discover new ways of sensing? Not only does this require "imagining" what measurements a candidate sensing system would produce and how useful those measurements would be, but, more fundamentally, it requires probing the world to discover useful sensing cues that we do not yet know exist. Prior work has taken important first steps by co-designing parts of the imaging pipeline, such as optics and reconstruction [1, 2], but largely within fixed designs and known sensing cues. My group will answer this question through two complementary research directions:

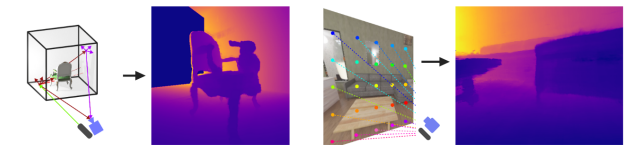
1. **Learning from Sensors**. Discovering useful sensing cues begins with the measurements we already have. Weak or indirect signals, such as indirect reflections captured by RGB cameras, lidar, and radar, can reveal hidden geometry but are often overlooked by existing methods. My group will develop physics-based learning methods that explicitly model these signals. Where the physics is difficult to model explicitly, sensor foundation models will learn transferable representations from simulated and limited real-world measurements.
2. **Learning for Sensors**. Determining what and how to measure requires searching a large, mixed discrete–continuous design space, spanning choices like modality and exposure. Evaluating candidate sensing systems is expensive: collecting real data is often impractical, while physics-based simulation can be slow or incomplete. My group will develop task-driven methods to jointly design sensing, perception, and action, supported by data-driven simulators that efficiently predict what candidate systems would measure. Through real-world interaction, these systems will learn to adapt how they sense and, ultimately, discover which cues in their environment to use.



My PhD research has established the foundations for Sensory Discovery. In learning from sensors, I showed that single-view lidar can reconstruct hidden geometry from indirect light (PlatoNeRF [3], CVPR Best Paper Finalist, top ~0.2% of submissions) and that sparse photons enable low-light perception (PhotonJEPA [4]). In learning for sensors, I demonstrated a task-driven design loop: reinforcement learning proposes imaging systems, simulation predicts measurements, and task performance guides new designs (DISeR [5]). I also showed that synthesized camera views improve perception [6]. My research, including collaborations with Meta and NVIDIA, has produced six first- or co-first-author publications at CVPR, ICCV, ECCV, ICCP, and SIGGRAPH Asia, plus three more papers in review.

Direction 1: Learning from Sensors

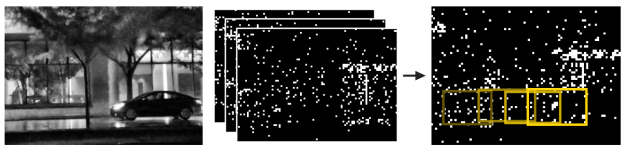
1.1 Physics-based learning



scanned lidar *flash lidar*

PlatoNeRF [3] Shoot-Bounce-3D [9]

1.2 Sensor foundation models



reference *photon frames* *tracking*

PhotonJEPA [4]

1.1 Physics-Based Learning: Reconstructing hidden geometry from indirect light

Sensors often capture information they were not designed to exploit. For example, a shadow cast onto the road or a reflection in a car door can reveal a pedestrian hidden around a corner. Such indirect cues are often treated as nuisances, yet they contain valuable information about otherwise hidden 3D geometry. My PhD research has focused on recovering this information, either by actively creating and measuring shadows with single-photon lidar or by exploiting naturally occurring shadows and reflections in RGB images.

Creating cues with single-photon lidar. Single-photon lidars can create and measure shadows that reveal geometry hidden from view via active illumination. In PlatoNeRF [3], which I conceived and led in collaboration with Meta, I proposed single-view 3D reconstruction by exploiting these generated cues. Existing methods either rely heavily on learned priors, which can hallucinate occluded geometry, or on visible shadows that are difficult to measure under high ambient light or low background albedo. My key insight was that two-bounce lidar encodes both depth through photon arrival times and occlusions through shadows. I used both as physically grounded constraints within a NeRF [7] to recover hidden geometry. PlatoNeRF reduced novel-view depth error by 25% over the strongest competing baseline.

Resolving ambiguities with physics and learned priors. Consumer lidars [8] introduce several challenges absent from PlatoNeRF, including simultaneous illumination, which mixes indirect returns, and low resolution, which limits reconstruction detail. In Shoot-Bounce-3D [9], I combined lidar physics and learned priors to disentangle mixed returns using a large simulated single-photon lidar dataset that I built and open-sourced. Our method reconstructs 3D geometry from a single measurement, including scenes with occlusions and mirrors. I recently mentored an undergraduate researcher on a project [10] where we use indirect lidar to guide an RGB generative model, grounding synthesized multiview images in physical constraints on hidden geometry. These images refine the reconstruction and add texture, demonstrating how physical signals can constrain learned priors that otherwise rely on hallucination.

Opportunistically exploiting shadows and reflections in RGB images. Indirect-light cues can also be exploited when they occur naturally, without active illumination. In our shape-from-shadow work [11], we optimized a NeRF using cast shadows observed across viewpoints, demonstrating that even a single shadow can constrain coarse hidden 3D geometry. We extended this principle in [12], treating glossy objects as indirect views into regions outside the camera's field of view and jointly recovering the reflector and hidden scene. Together, these works demonstrate that hidden geometry can be recovered from naturally occurring optical cues, without actively controlling illumination.

Future Direction (a): Perceiving the Hidden World Through Multimodal Physics. My group will investigate how complementary sensing modalities can reveal information about the world that no individual sensor can recover. Cameras, lidar, and radar interact with their environments through different physical mechanisms, providing distinct constraints on visible and occluded geometry, material properties, and motion. A central question is under what conditions combining modalities makes ambiguous information recoverable, and how to characterize what remains unobservable. Building on my work in lidar [13], radar [14], and X-ray imaging [15], I will develop physics-aware models that jointly interpret these complementary signals. As an initial project, I will extend simultaneous localization and mapping to estimate occluded geometry, even in scenes with specular surfaces, by combining RGB with indirect light measured by a consumer single-photon lidar. I will also investigate single-view reconstruction from lidar and radar, asking when their complementary measurement physics can resolve ambiguities that neither modality can resolve alone. Longer term, my goal is to enable machines to build physically grounded representations of dynamic, partially observed environments, including the geometry, motion, and uncertainty of objects beyond direct view.

1.2 Sensor Foundation Models: Perception from Raw Measurements

Modeling all relevant measurement physics is often impractical. A complementary approach is to learn task-relevant representations directly from raw sensor measurements, exploiting information that physical models may overlook.

Passive single-photon sensing illustrates this opportunity: even when measurements are too sparse to reconstruct reliable images, they may still contain enough information for perception. These sensors capture hundreds of thousands of sparse binary frames per second, making them promising for low-light, high-dynamic-range, and fast-motion vision. Many existing methods first reconstruct intensity images, but reconstruction becomes unreliable when photons are scarce. In PhotonJEPA [4], I instead developed a self-supervised method that adapts pretrained video foundation models [16] to extract semantic and motion information directly from raw photon streams using simulation-based training. By bypassing image reconstruction, PhotonJEPA enables activity recognition and object tracking at photon fluxes as low as 0.005 photons per pixel, outperforming reconstruction-first approaches in photon efficiency. These results raise a fundamental question: *how many photons does computer vision actually need?*

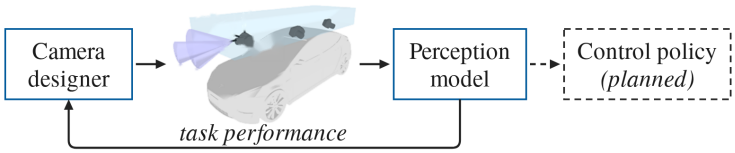
Future Direction (b): Learning to Perceive with New Sensing Systems

Near term. Building on PhotonJEPA, my group will develop foundation models that learn task-relevant representations across different raw sensor measurements. As a first step, I will extend this approach to single-photon lidar, investigating generalization across scenes, tasks, and sensing configurations using simulation and limited real-world data.

Long term. I envision foundation models that can learn to interpret unfamiliar sensing systems with minimal sensor-specific training. I will investigate models trained across diverse measurement processes, using sensor parameters and physics-based simulation to learn representations that transfer across sensing configurations. For a new sensor, a small number of calibration measurements could adapt the model to its measurement characteristics without retraining from scratch. A central question is which aspects of perception can be shared across sensors with different measurement physics, and which must be learned separately. Such models could enable general-purpose perception and evaluate the task-relevant information captured by newly designed sensing systems.

Direction 2: Learning for Sensors

2.1 Task-driven design



DISeR [5]

2.2 Data-driven simulation



source view *generated view*

Viewpoint Robustness [6]

2.1 Task-Driven Design: Discovering Better Sensing Systems

The best sensing system depends on what a machine needs to accomplish: what it measures determines what it can perceive and do. Yet designing a complete sensing system requires searching a large, mixed discrete–continuous space spanning sensor modality, placement, optics, illumination, exposure, and acquisition strategy, where evaluating each candidate can be expensive. My research has begun to explore how we can efficiently search and optimize over this space to discover sensing systems that maximize task performance under practical resource constraints.

In DISeR [5], I led the development of a framework to jointly design imaging systems and perception algorithms. We represented the space of possible imaging systems using a context-free grammar over components such as illumination, optics, sensors, and algorithms, and trained a reinforcement-learning "camera designer" to propose configurations for a downstream task. Each candidate system was instantiated in simulation, its measurements were used to train and evaluate a task-specific perception model, and the resulting task performance became the reward for the camera designer. We demonstrated the approach on depth estimation and autonomous-vehicle camera-rig design. The method rediscovered known principles such as maximizing stereo baseline, and under resource constraints identified new camera-rig designs that improved perception accuracy over industry-standard rigs by more than 6%.

I have since contributed to follow-up work on continuous lidar optimization [17] and the co-design of eye morphology and neural processing in embodied agents [18]. I am also currently extending DISeR with large-language-model (LLM)-generated imaging configurations and, with the University of Glasgow, exploring task-driven design of a wearable single-photon sensing helmet for brain tumor monitoring using photon-transport simulation.

Future Direction (c): From Sensing Design to Adaptive Sensing Discovery

Near term. My group will move beyond fixed sensor designs toward adaptive sensing policies that choose measurements based on previous observations and downstream tasks. This direction is promising because the system can *choose* which cues to create and exploit. As a first project, I will extend PlatoNeRF's fixed illumination scan to adaptively select where to illuminate, maximizing information about hidden scenes under limited acquisition budgets. I will then investigate joint design of sensing, perception, and action in embodied agents.

Long term. I aim to develop agents that autonomously discover how to acquire task-relevant information, rather than only optimize within predefined sensing design spaces. Building on LLM-generated designs, my group will investigate agents that propose sensing architectures and acquisition strategies, evaluate them through physics-based and data-driven simulation, and refine them using task-performance feedback and real-world experiments. The goal is for agents to identify and exploit useful cues that they were not explicitly told to seek – for example, discovering that indirect lidar reflections could reveal geometry hidden by a robot's gripper, avoiding the need for additional cameras.

2.2 Data-Driven Simulation: Predicting Sensor Measurements

Collecting real-world data for all possible sensor configurations is impractical. In work I conceived and led with NVIDIA [6], I developed a method that predicts how a scene would appear under previously unseen camera placements. Our model estimates scene geometry and synthesizes training views for new camera configurations, avoiding the need to recollect and relabel data for each autonomous-vehicle platform. Training on these generated views improved downstream perception by more than 14% under camera-placement changes.

Extending measurement generation beyond RGB introduces new challenges because sensing modalities encode physical information in fundamentally different forms. I recently led a project that jointly generates radar and RGB measurements by conditioning video diffusion models on scene layouts, including an initial RGB frame and object bounding boxes over time [19]. The goal of the project is to create synthetic data to train autonomous vehicle perception models. A key challenge is representation: pretrained video diffusion models operate in dense image latent spaces, whereas radar consists of sparse points. We are developing representations that preserve the physical information in radar measurements while remaining compatible with pretrained generative models.

Future Direction (d): Learning Generalizable Models of Sensing

Near term. Building on my work in novel-view synthesis and camera–radar generation, my group will develop data-driven simulators that predict what a sensor would measure under new configurations. These simulators will use known measurement physics where it is available and learn effects that are expensive or difficult to model explicitly. As an initial project, I will investigate a unified photon world model that represents a scene's underlying physical properties and predicts RGB images, event streams, photon counts, and lidar returns conditioned on sensor parameters.

Long term. An agent can exploit a cue without knowing its physics, but simulating it requires modeling how measurements arise. I aim to develop agents that learn unfamiliar sensing physics through experimentation. Using simulated sensors with hidden forward models, I will test whether agents can design experiments, infer physical principles, and revise hypotheses. Could an agent rediscover the light-transport principles underlying the rendering equation? For validation, I will test agents on unfamiliar or modified measurement physics. Ultimately, agents could incorporate newly discovered cues into data-driven simulators to evaluate and optimize sensing designs.

Towards Sensory Discovery

Together, these research directions establish a path toward machines that discover how to sense, perceive, and act. Data-driven simulators will predict what candidate sensing systems would measure, physics-based methods and sensor foundation models will interpret those measurements, and task performance will guide the next design. Ultimately, I envision systems that begin with a task rather than a fixed sensor, discover what information to acquire and how to interpret it, and, through interaction and experimentation, find and exploit new sensing cues, learning how they arise.

References

* indicates equal authorship; † indicates advised student.

[1] **T. Klinghoffer***, S. Somasundaram*, K. Tiwary*, and R. Raskar. "Physics vs. Learned Priors: Rethinking Camera and Algorithm Design for Task-Specific Imaging". In: *ICCP*. 2022.

[2] K. Tiwary, **T. Klinghoffer***, A. Young*, S. Somasundaram, N. Behari, A. Dave, B. Cheung, D.-E. Nilsson, T. Poggio, and R. Raskar. "A Roadmap for Generative Design of Visual Intelligence". In: *An MIT Exploration of Generative AI*. MIT Press, 2024.

[3] **T. Klinghoffer**, X. Xiang*, S. Somasundaram*, Y. Fan, C. Richardt, R. Raskar, and R. Ranjan. "PlatoNeRF: 3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar". In: *CVPR*. 2024. Oral; Best Paper Award Finalist.

[4] **T. Klinghoffer**, S. Somasundaram, P. Liang, and R. Raskar. "PhotonJEPA: Self-Supervised Learning in Low Light with Single-Photon Cameras". In: *Under Review*. 2026.

[5] **T. Klinghoffer***, K. Tiwary*, N. Behari, B. Agrawalla†, and R. Raskar. "DISeR: Designing Imaging Systems with Reinforcement Learning". In: *ICCV*. 2023.

[6] **T. Klinghoffer**, J. Philion, W. Chen, O. Litany, Z. Gojcic, J. Joo, R. Raskar, S. Fidler, and J. M. Alvarez. "Towards Viewpoint Robustness in Bird's Eye View Segmentation". In: *ICCV*. 2023.

[7] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. "Nerf: Representing scenes as neural radiance fields for view synthesis". In: *Communications of the ACM* 65.1 (2021), pp. 99–106.

[8] 4sense. *Apple LIDAR Demystified: SPAD, VCSEL, and Fusion*. Online; posted 1 March 2021. 2021. URL: https://4sense.medium.com/apple-lidar-demystified-spad-vcsel-and-fusion-aa9c3519d4cb.

[9] **T. Klinghoffer**, S. Somasundaram*, X. Xiang*, Y. Fan, C. Richardt, A. Dave, R. Raskar, and R. Ranjan. "Shoot-Bounce-3D: Single-Shot Occlusion-Aware 3D from Lidar by Decomposing Two-Bounce Light". In: *SIGGRAPH Asia*. 2025.

[10] E. Underwood*†, **T. Klinghoffer***, L. Collins, and R. Raskar. "Coloring Plato's Cave: Generative Single-View 3D Reconstruction from RGB and Two-Bounce Lidar". In: *Under Review*. 2026.

[11] K. Tiwary*, **T. Klinghoffer***, and R. Raskar. "Towards Learning Neural Representations from Shadows". In: *ECCV*. 2022.

[12] K. Tiwary, A. Dave, N. Behari, **T. Klinghoffer**, A. Veeraraghavan, and R. Raskar. "ORCa: Glossy Objects as Radiance-Field Cameras". In: *CVPR*. 2023.

[13] N. Behari, A. Young, S. Somasundaram, **T. Klinghoffer**, A. Dave, and R. Raskar. "Blurred LiDAR for Sharper 3D: Robust Handheld 3D Scanning with Diffuse LiDAR and RGB". In: *CVPR*. 2025. Highlight.

[14] N. Sortur†, J. Goodwin, P. Patel, L. E. Martinez Jr., **T. Klinghoffer**, R. S. Caceres, and R. Walters. "Radar2Shape: 3D Shape Reconstruction from High-Frequency Radar using Multiresolution Signed Distance Functions". Under review.

[15] D. Gilo†, **T. Klinghoffer**, and O. Litany. "EPI-NAF: Enhancing Neural Attenuation Fields for Limited-Angle CT with Epipolar Consistency Conditions". In: *ISBI*. 2025.

[16] M. Assran, A. Bardes, D. Fan, Q. Garrido, R. Howes, M. Muckley, A. Rizvi, C. Roberts, K. Sinha, A. Zholus, et al. "V-jepa 2: Self-supervised video models enable understanding, prediction and planning". In: *arXiv preprint arXiv:2506.09985* (2025).

[17] N. Behari, A. Young, **T. Klinghoffer**, A. Dave, and R. Raskar. "Task-Driven Implicit Representations for Automated Design of LiDAR Systems". In: *CVPR*. 2026. Highlight.

[18] K. Tiwary, A. Young, Z. Tasneem, **T. Klinghoffer**, A. Dave, T. Poggio, D.-E. Nilsson, B. Cheung, and R. Raskar. "What if Eye. . . ? Computationally Recreating Vision Evolution". In: *Science Advances* (2025).

[19] **T. Klinghoffer***, M. Keidar*, T. Borreda, and O. Litany. "RadarDream: Joint RGB–Radar Driving Video Generation Across Visibility Conditions". In: *Under Review*. 2026.