Lidar and Autonomous Cars

Simon Verghese

Director, Sensor Systems
Waymo

5 June 2024, ISSW

WAYMO

2009-2020 Tech Feasibility

2009 GoogleX project

2015 1st driverless

2020

2024

Millions

miles on public roads

Billions

miles in simulation

13+ states

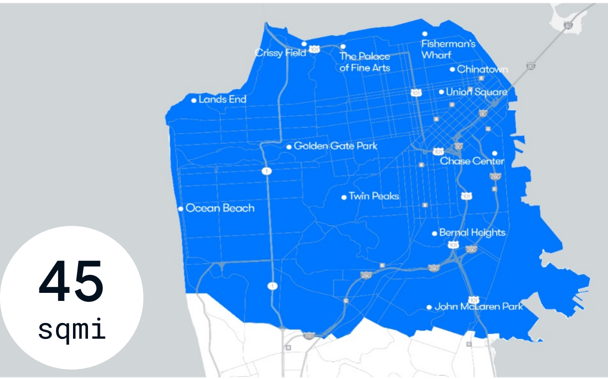
across the USA

2020-2024 Commercial Ops

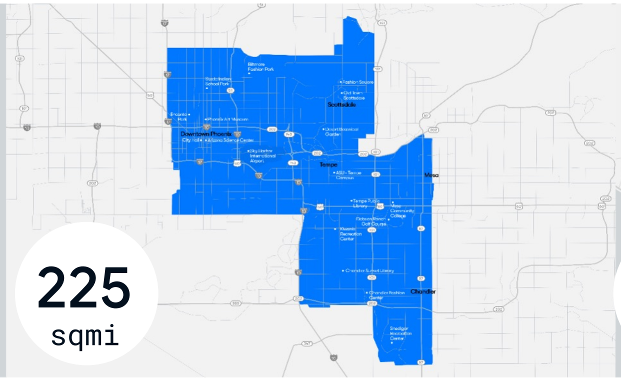
2009 — 2015 1st driverless — 2020 Commercial Ops — 2024

10s of thousands

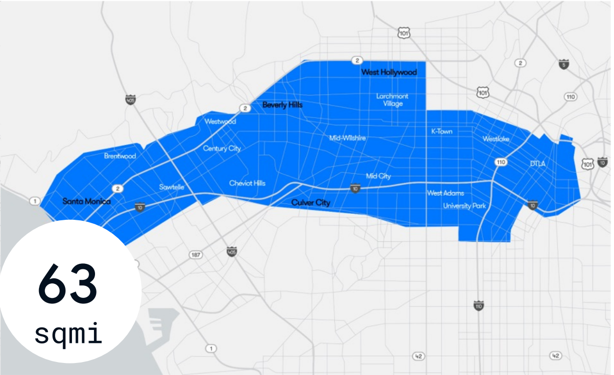
Paid public rides a week



San Francisco



Phoenix



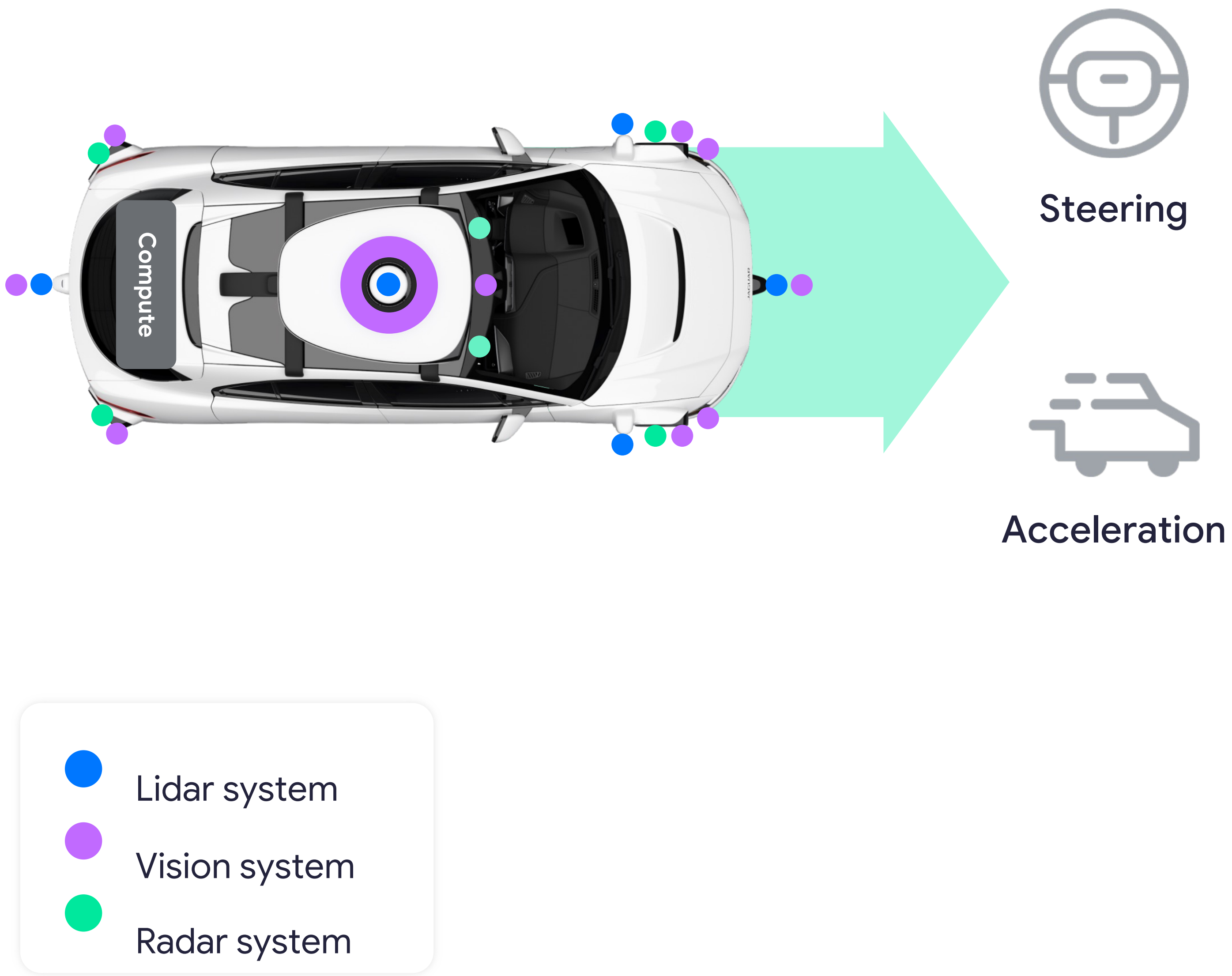
Los Angeles

Next Up

LA

ATX

Where's the difficulty?



Steering

Acceleration

● Lidar system
● Vision system
● Radar system

Safety-critical application

Extremely low tolerance to errors

Real Time Latency Requirements

Typically want decisions at >=10Hz

Complex high-dimensional input

O(10) different sensors

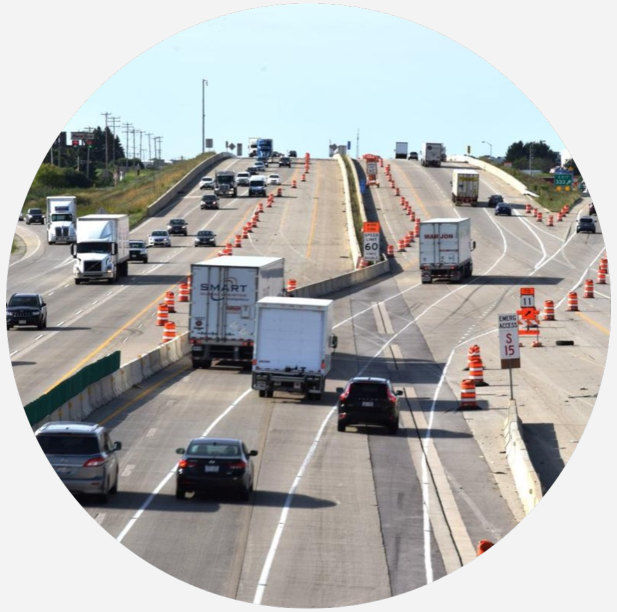
100s of millions sensor readings per second

Multiple seconds of context often needed

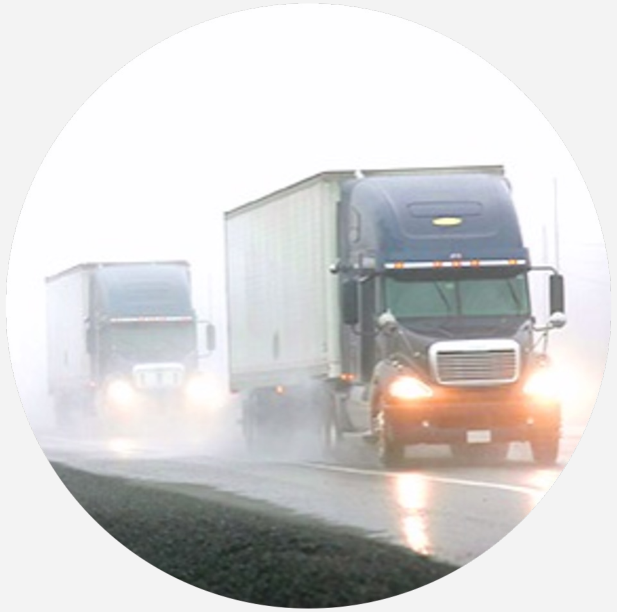
Diverse Operating Domains

Dense Urban

Suburban



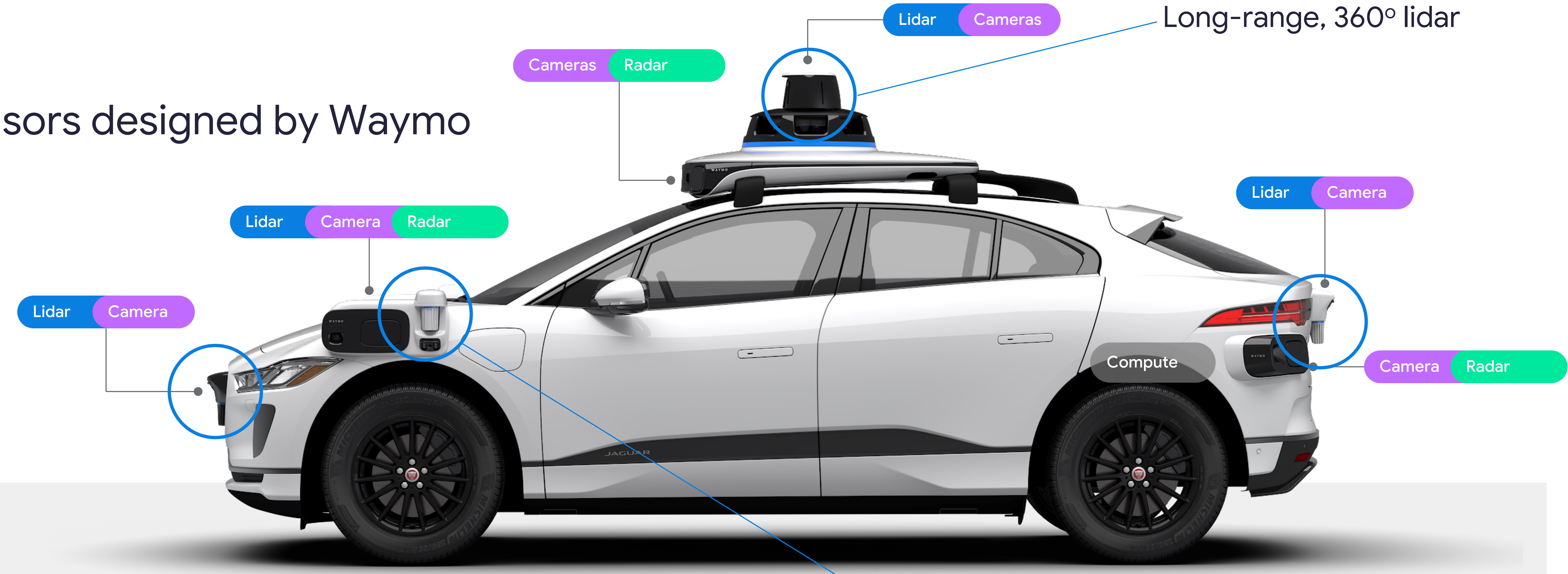
Freeway



Weather / Season /
Time of Day

A 3D Sensor Suite for Autonomous Driving

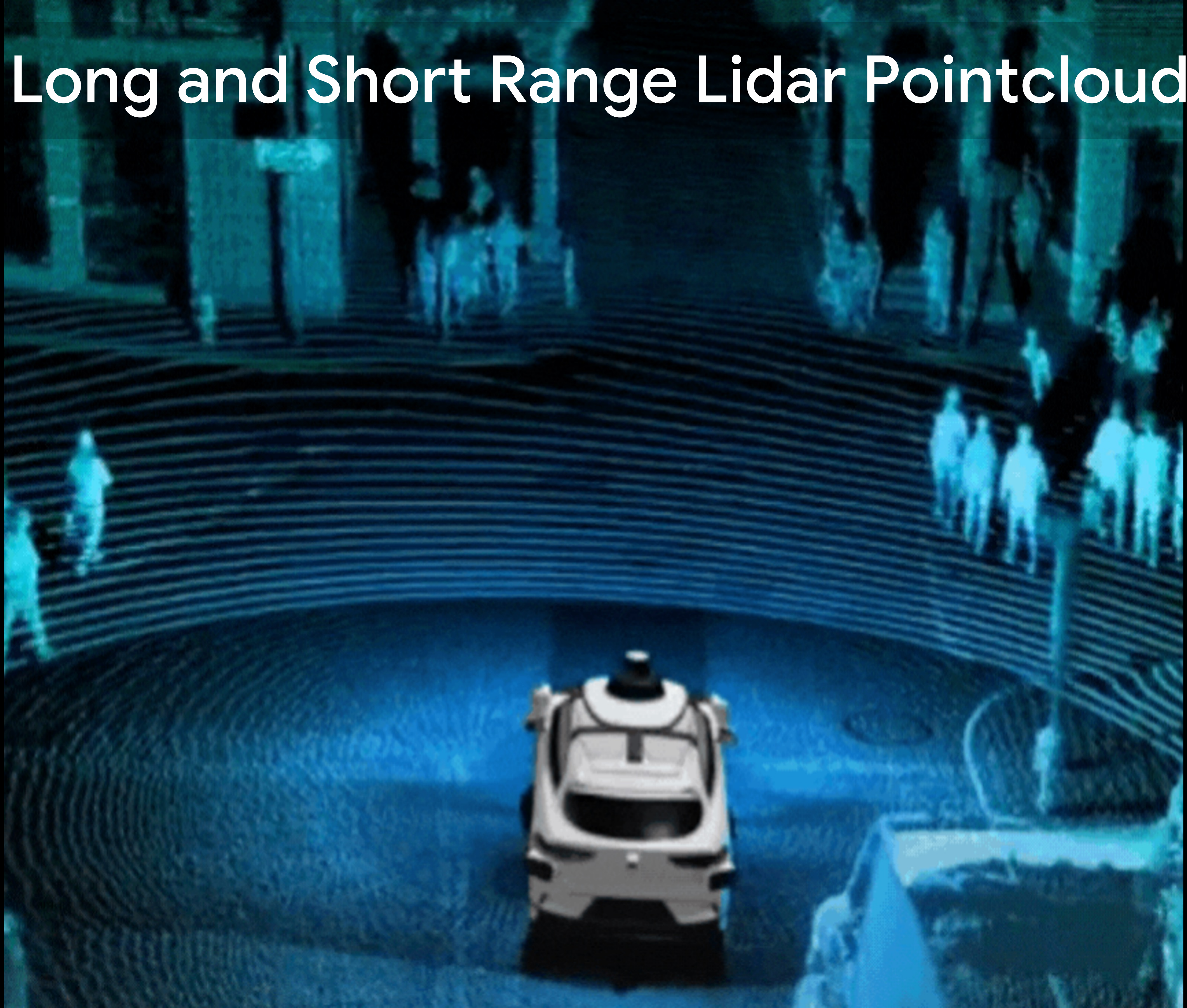
Sensors designed by Waymo



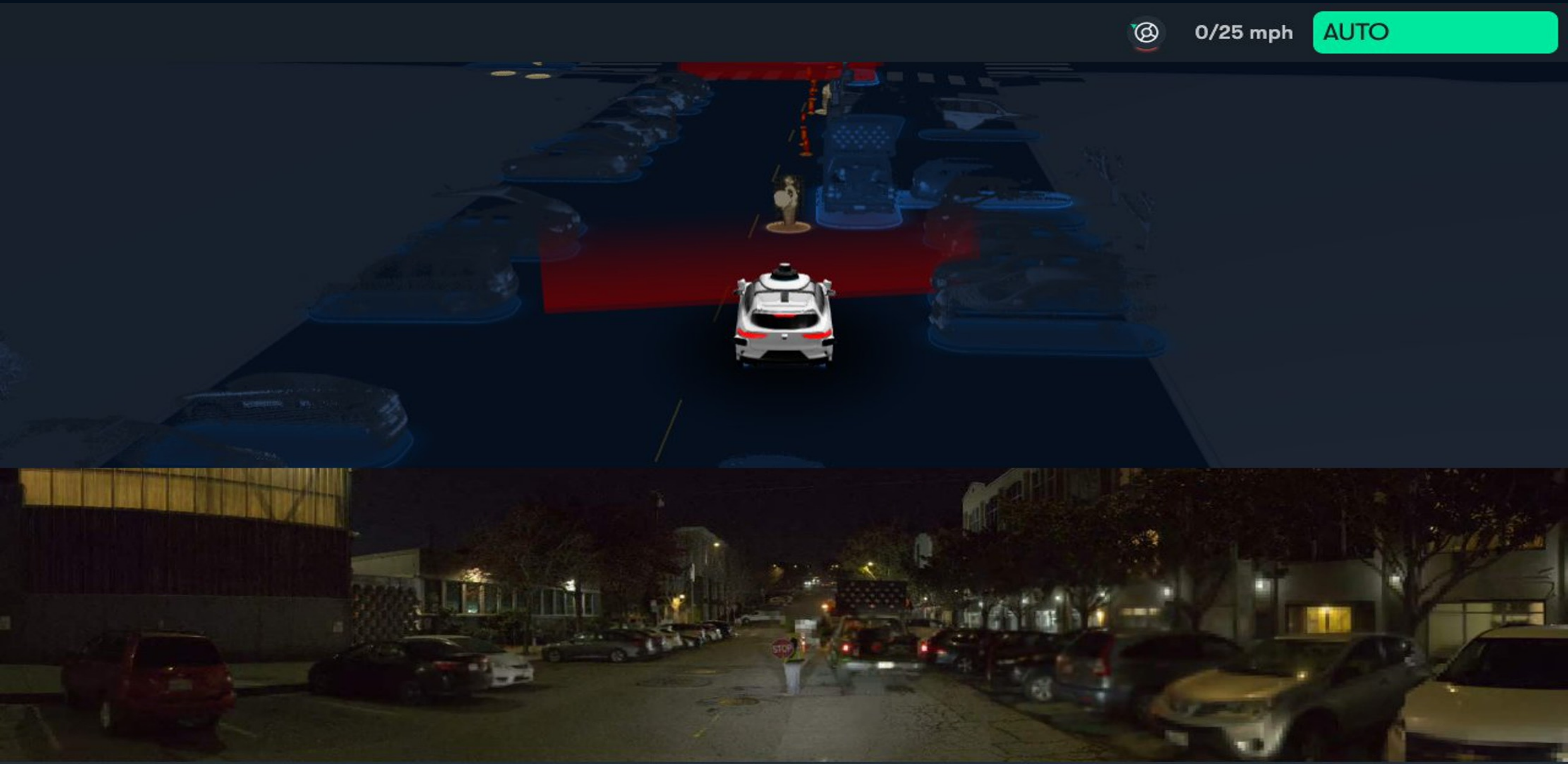
Currently using pulsed time-of-flight lidars

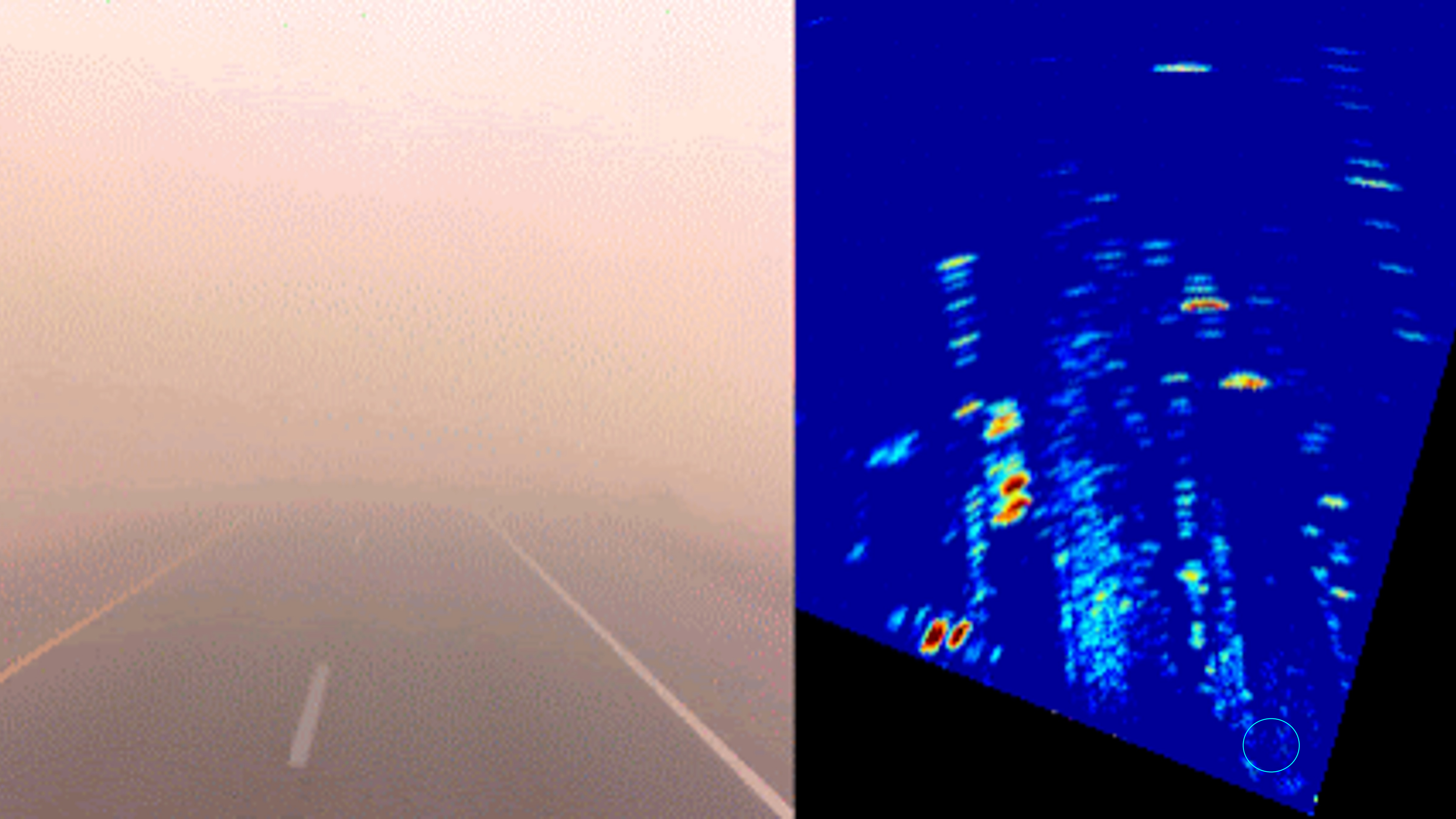
Short-range perimeter lidar (x4)

Long and Short Range Lidar Pointcloud



Complex Multi-agent Environments

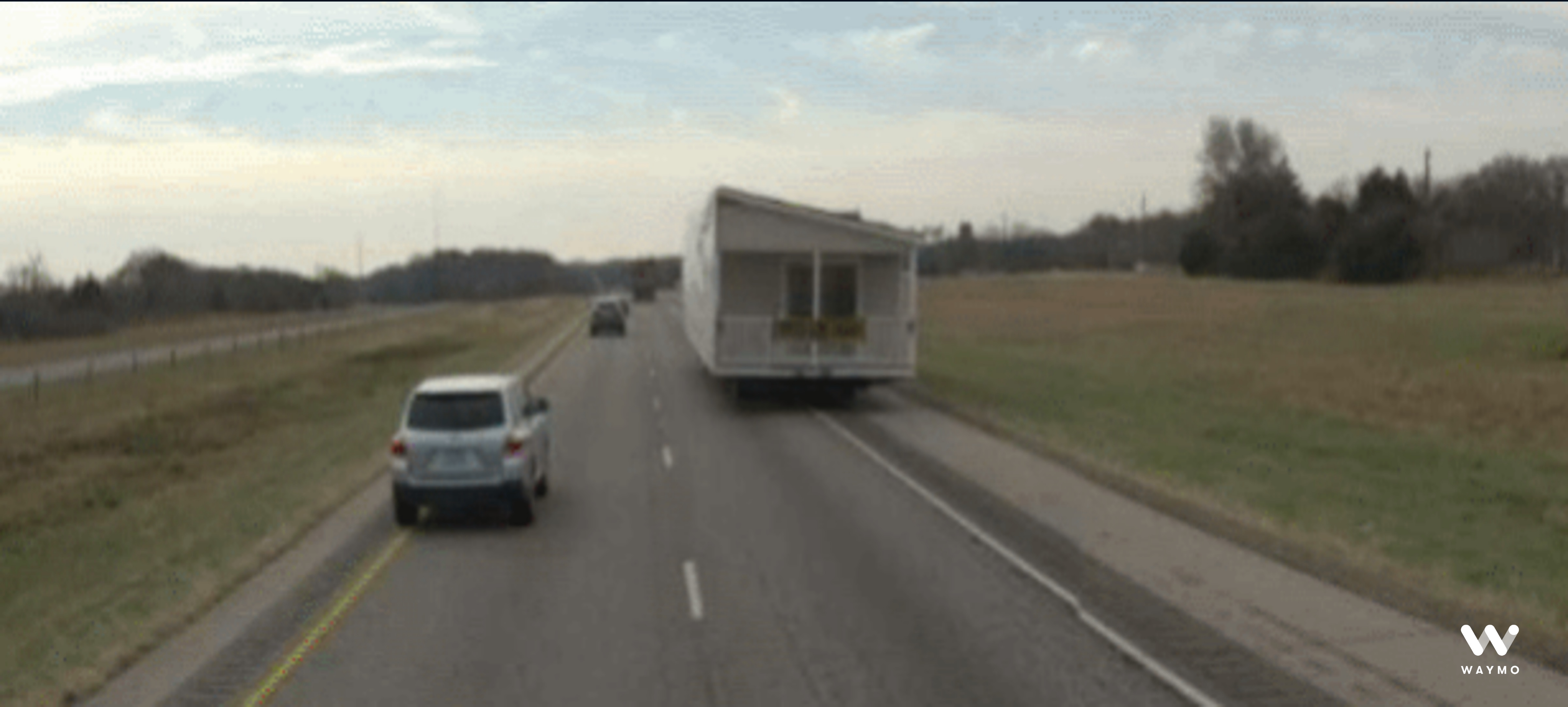




Rare Events

A BBQ grill falling from the back of a pickup truck at a speed of 64 mph

Rare Events: House in Lane

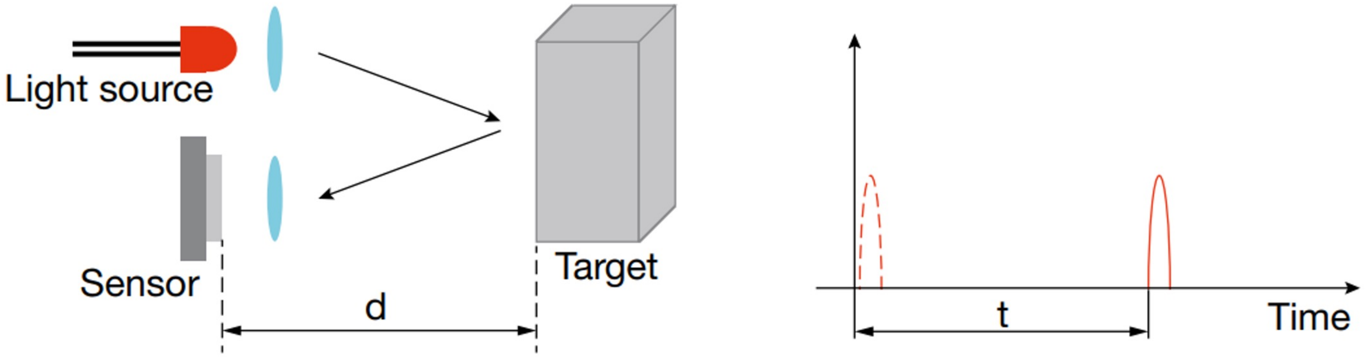


Rare Events: Construction worker walking across highway with sign

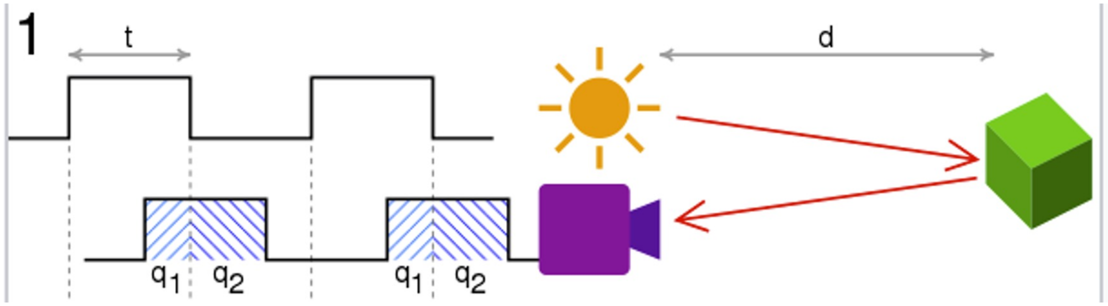
WAYMO

Automotive Lidar Technologies

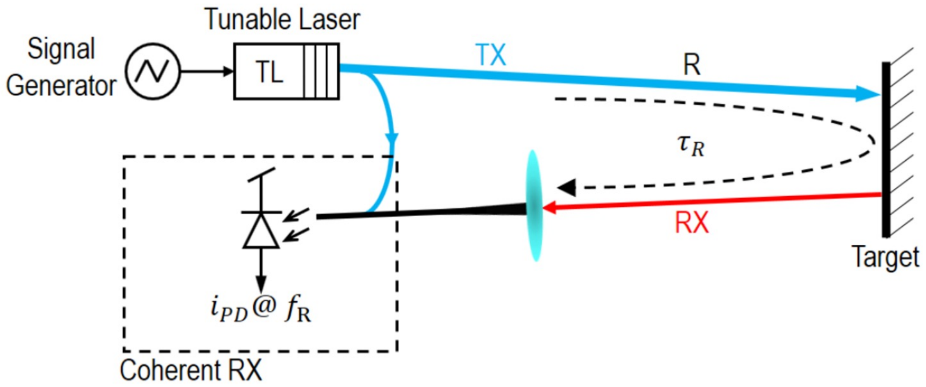
- **Time of Flight (ToF)**
  - **Direct (pulsed)**
  - Indirect (Modulated)
- Coherent Lidar (FMCW, PSK)



Osram Application Note No. AN106 "Time-of-flight (ToF) measurement using pulse lasers"



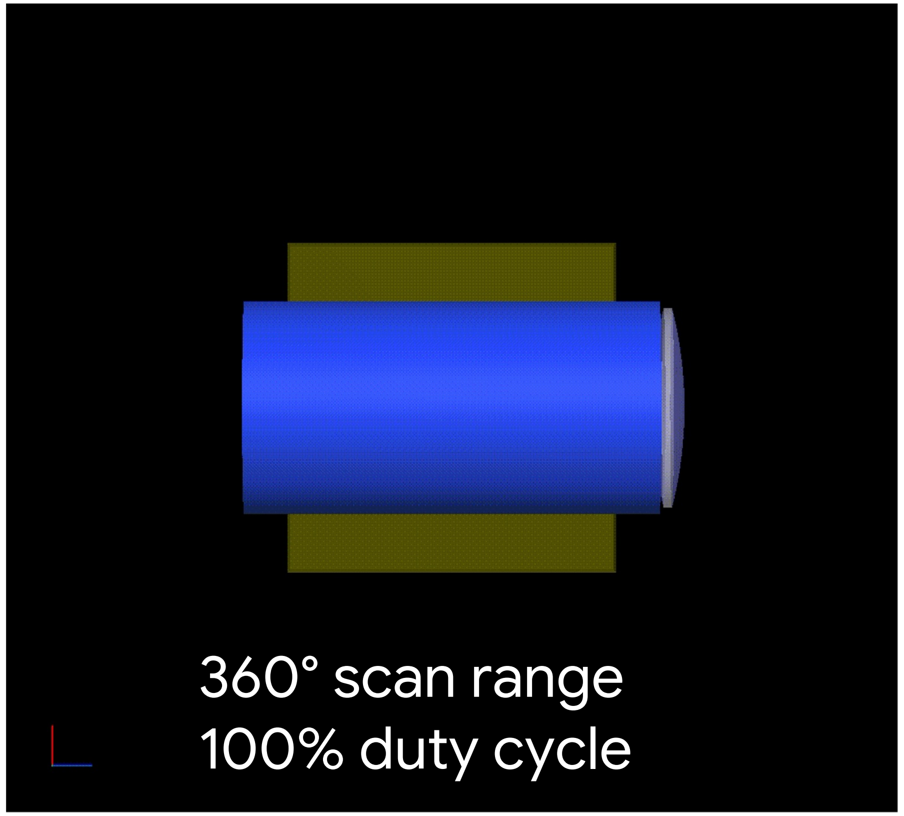
https://en.wikipedia.org/wiki/Time-of-flight_camera



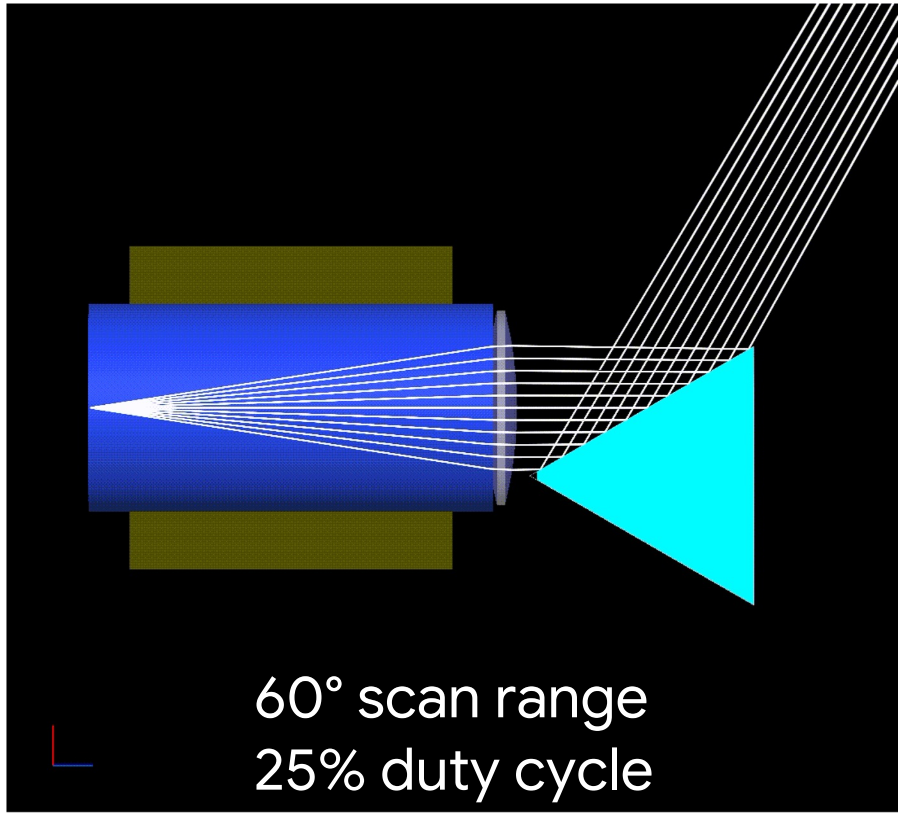
Baghmisheh, B., Ph.D Thesis, Berkeley 2016, "Chip-Scale Lidar"

Mechanical scanning

Most efficient use of optoelectronic semiconductors across azimuth

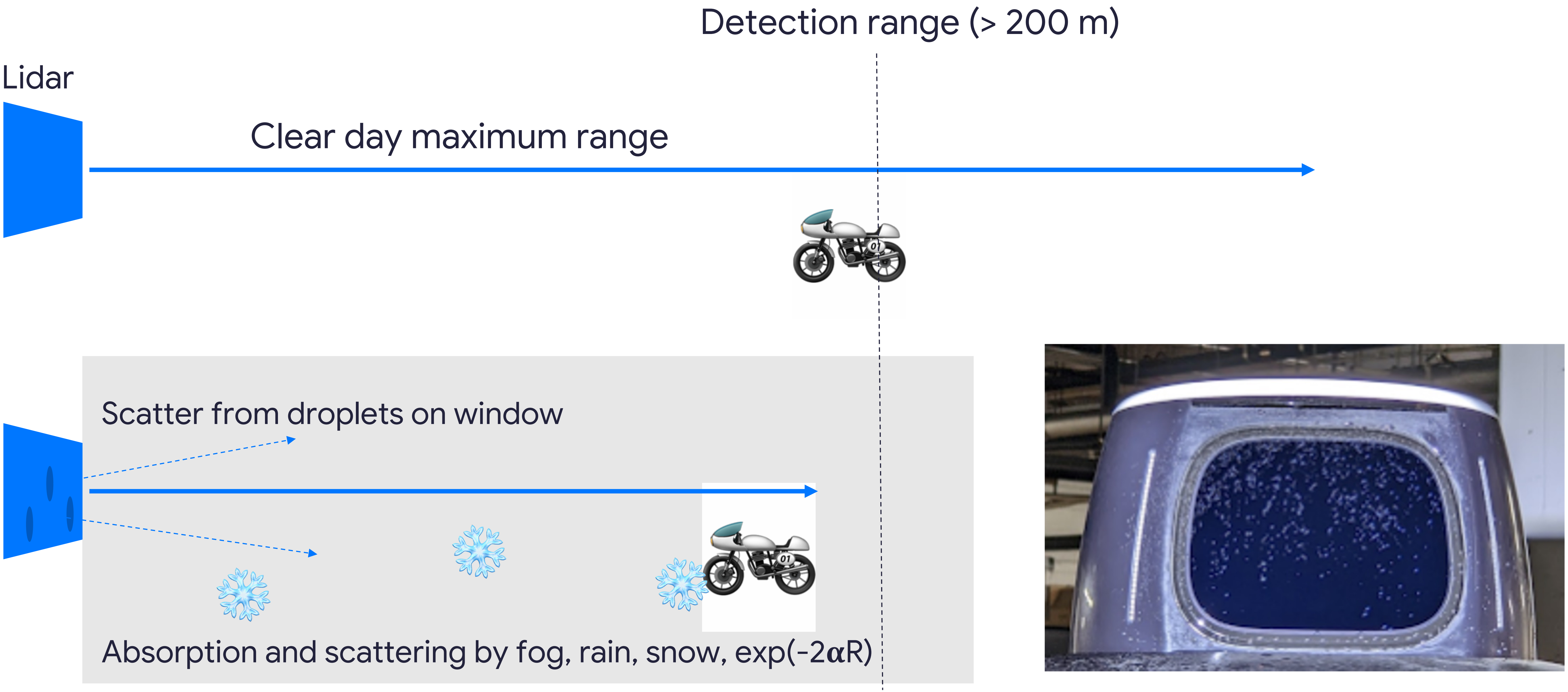


Highest FOV/$ ratio
Fewer optical surfaces
Wireless power and data links



Wired power and data connection
Simpler thermal management
Better car integration and industrial design

SNR Margin for weather - related signal loss



Detection range (> 200 m)

Lidar

Clear day maximum range

Scatter from droplets on window

Absorption and scattering by fog, rain, snow, $\exp(-2\alpha R)$

High Dynamic Range Receivers

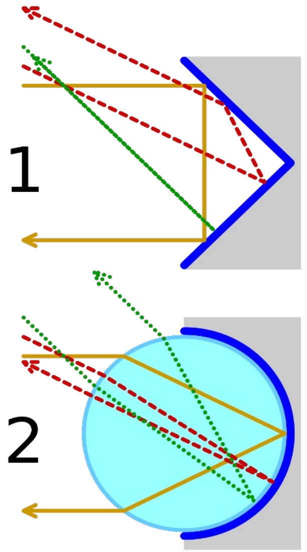
Automotive environment has a high dynamic range of returned lidar signals

Weak Diffuse Targets

Tires, asphalt, etc. (< 5% reflective)

Retroreflectors

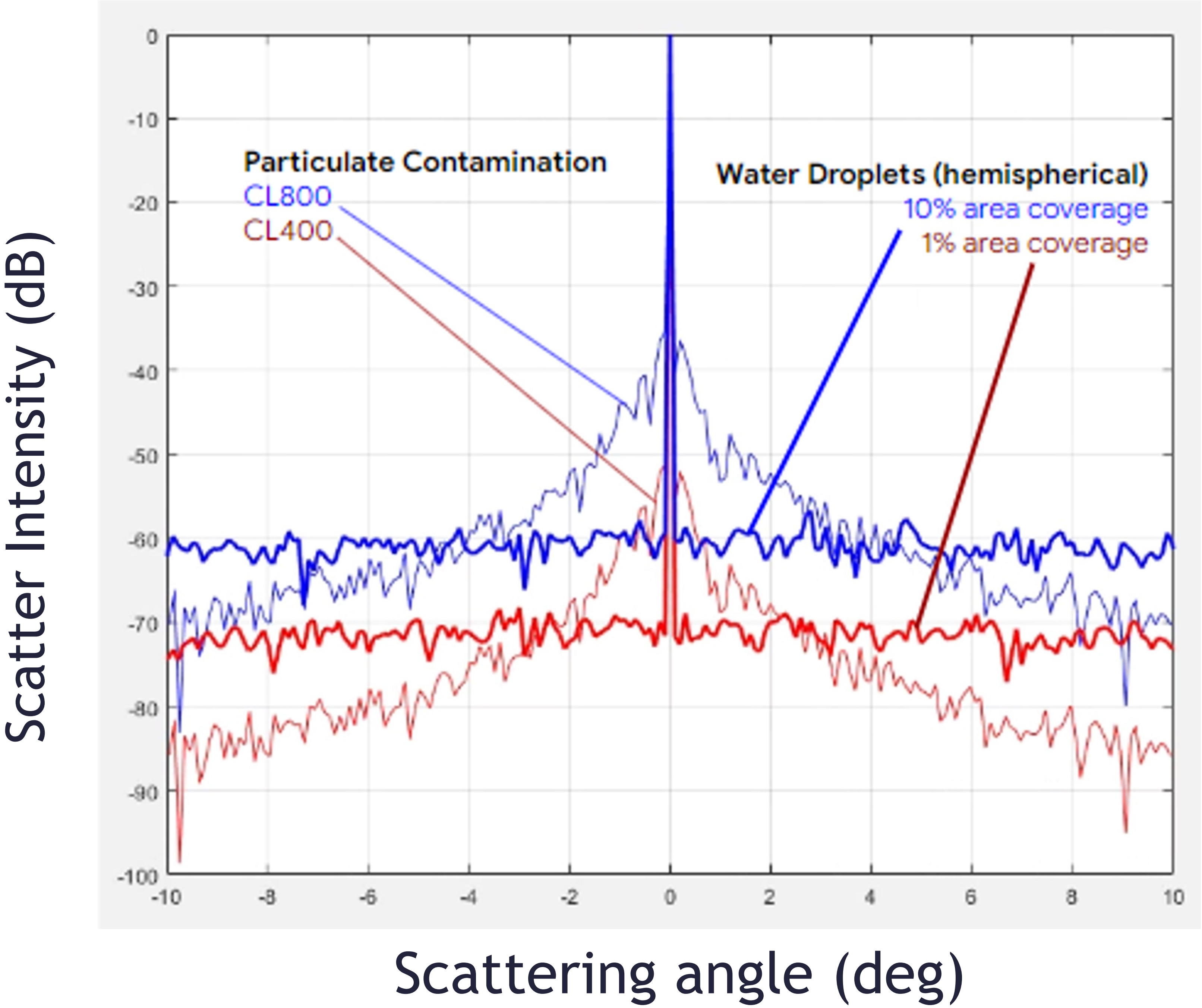
Lane markings, license plates, road signs (10 to 30+ dB gain)



Images from: https://en.wikipedia.org/wiki/Retroreflector Creative Commons

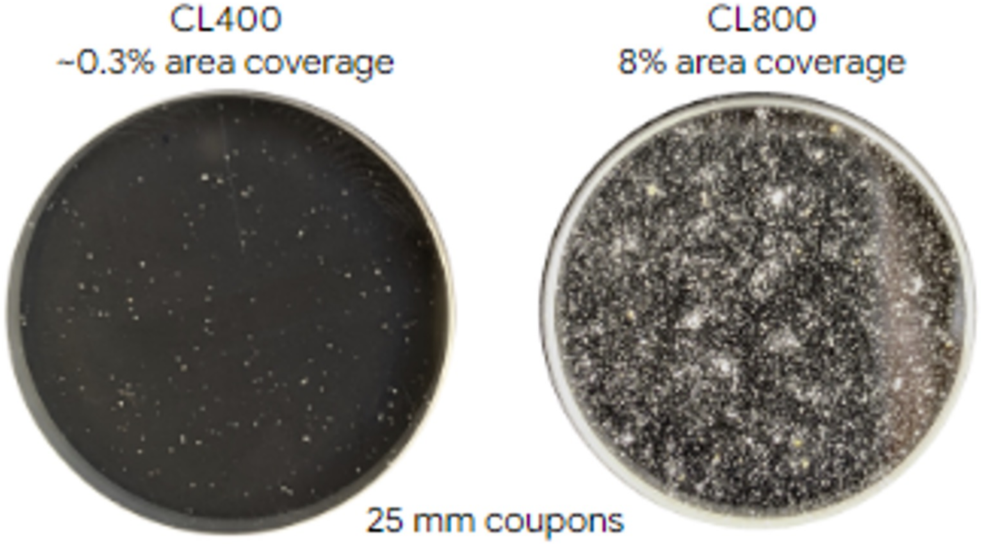
Fouled Windows: Scatter and Crosstalk

Particles scatter light into adjacent channels



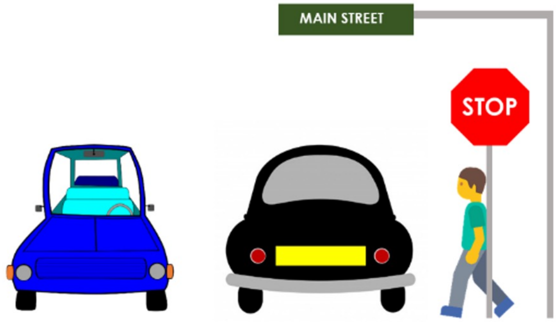
Particulate Cleanliness Level (CL)

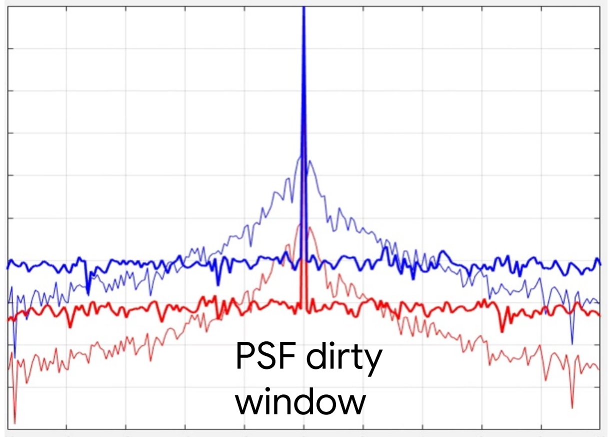
Ref: MIL-1246C / IEST-STD-1246C



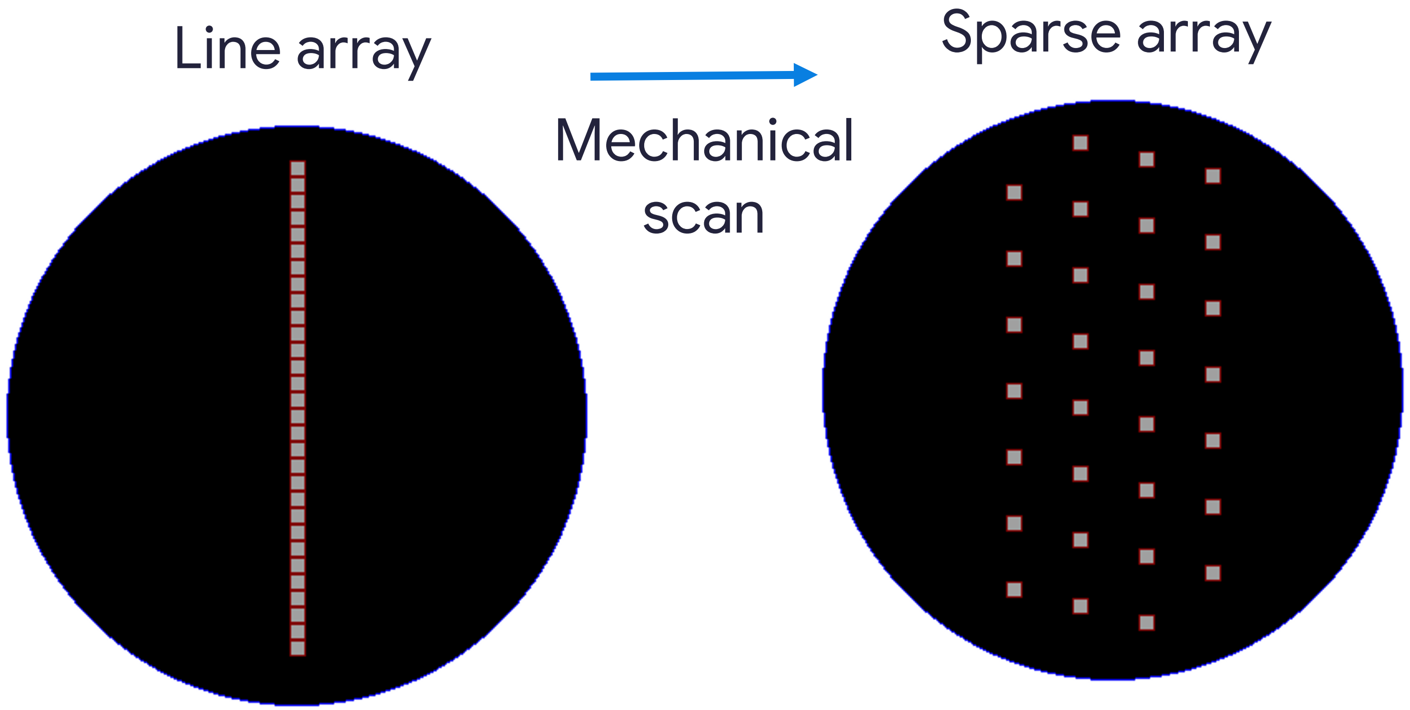
R. H. Shepard, C. Salsbury, “Predicting Stray Light Artifacts in Direct Detection Lidars,” OSA Conference on Lasers and Electro-Optics (CLEO), Optical Technologies for Autonomous Cars and Mobility I, AF1M.2 (2020).

Fouled Windows - Sparse Array vs. Line Scanner

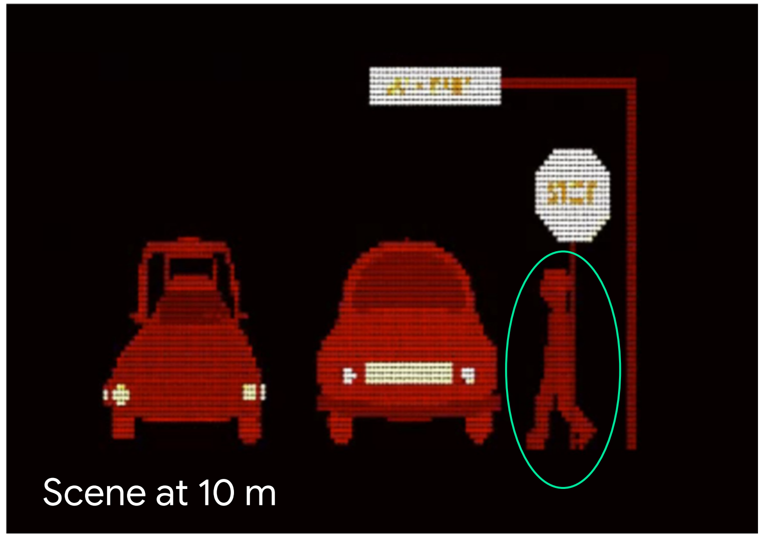




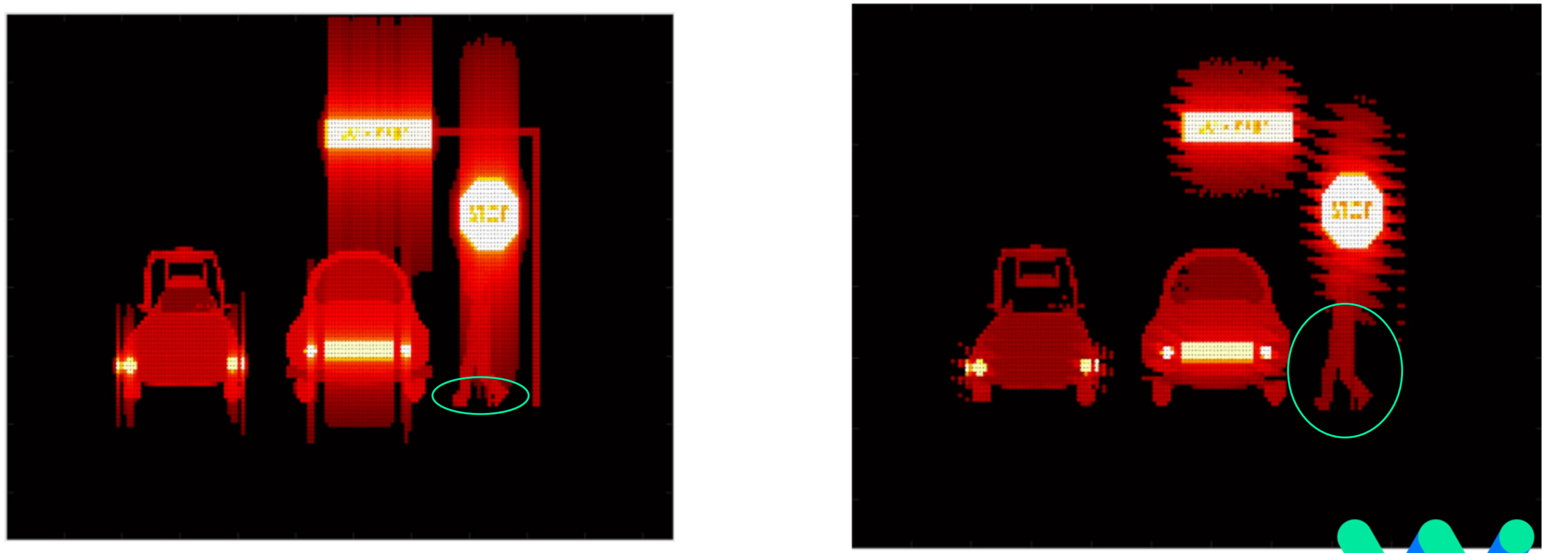
⊗ (convolve)



Clean Window



Lidar point clouds with a contaminated window

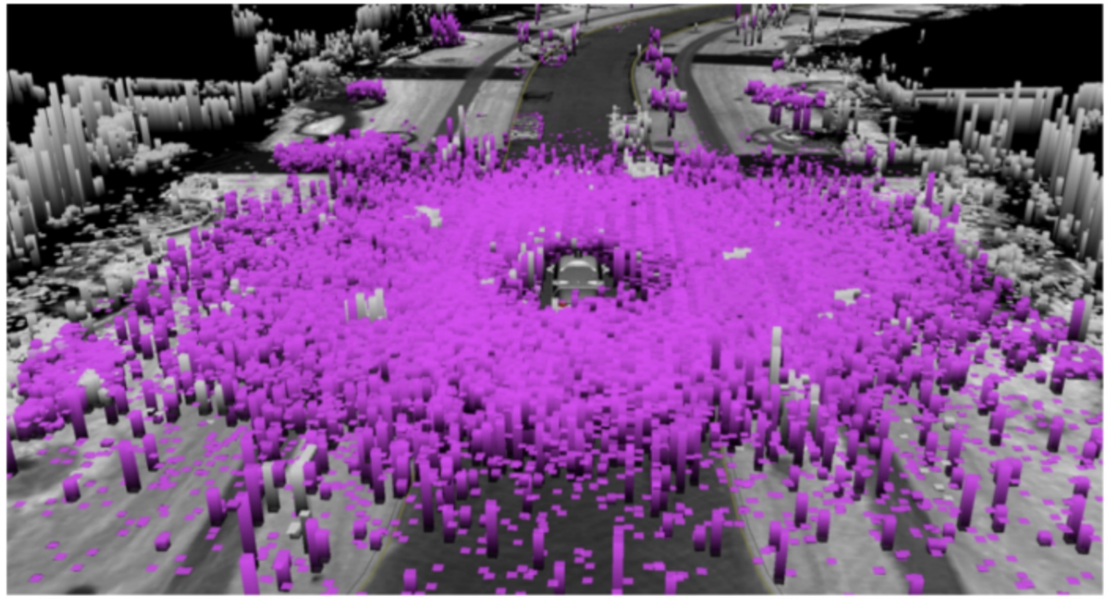


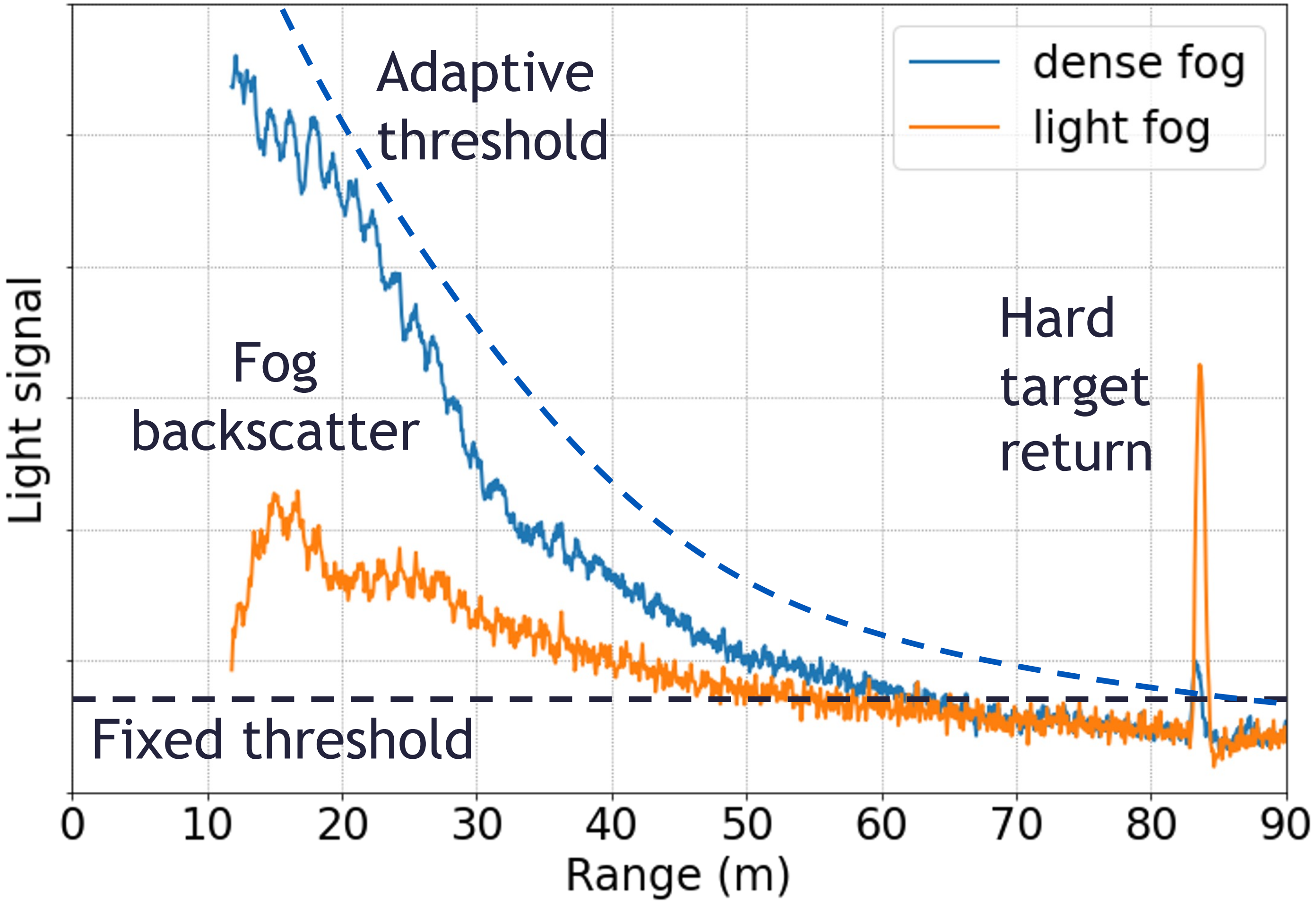
J. Dunphy, “Integrated Lidar Sensors for L4 Autonomous Vehicles”, Symposium on VLSI Technologies and Circuits (2023)

Atmospheric attenuation – Full Waveform Sampling

Infer weather attenuation and separate hard targets from fog/dust/rain/snow

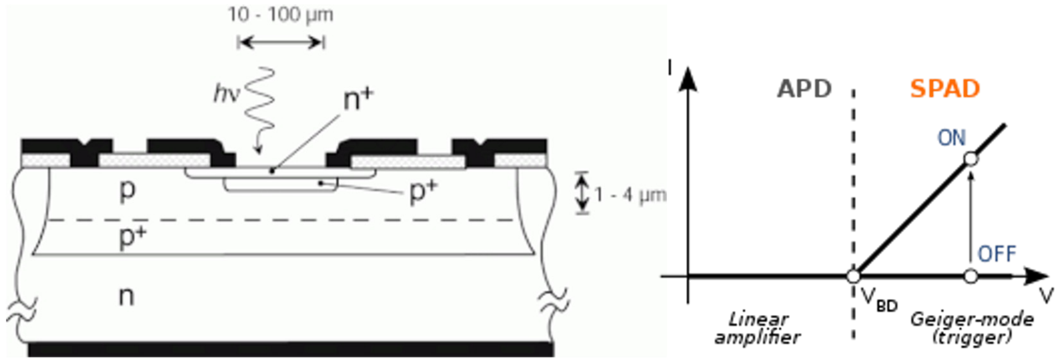
Spurious detections from droplets (purple) can be separated from real targets (gray) by adaptive thresholding





Silicon GmAPD Trends

(SPAD – Single-photon avalanche diode)



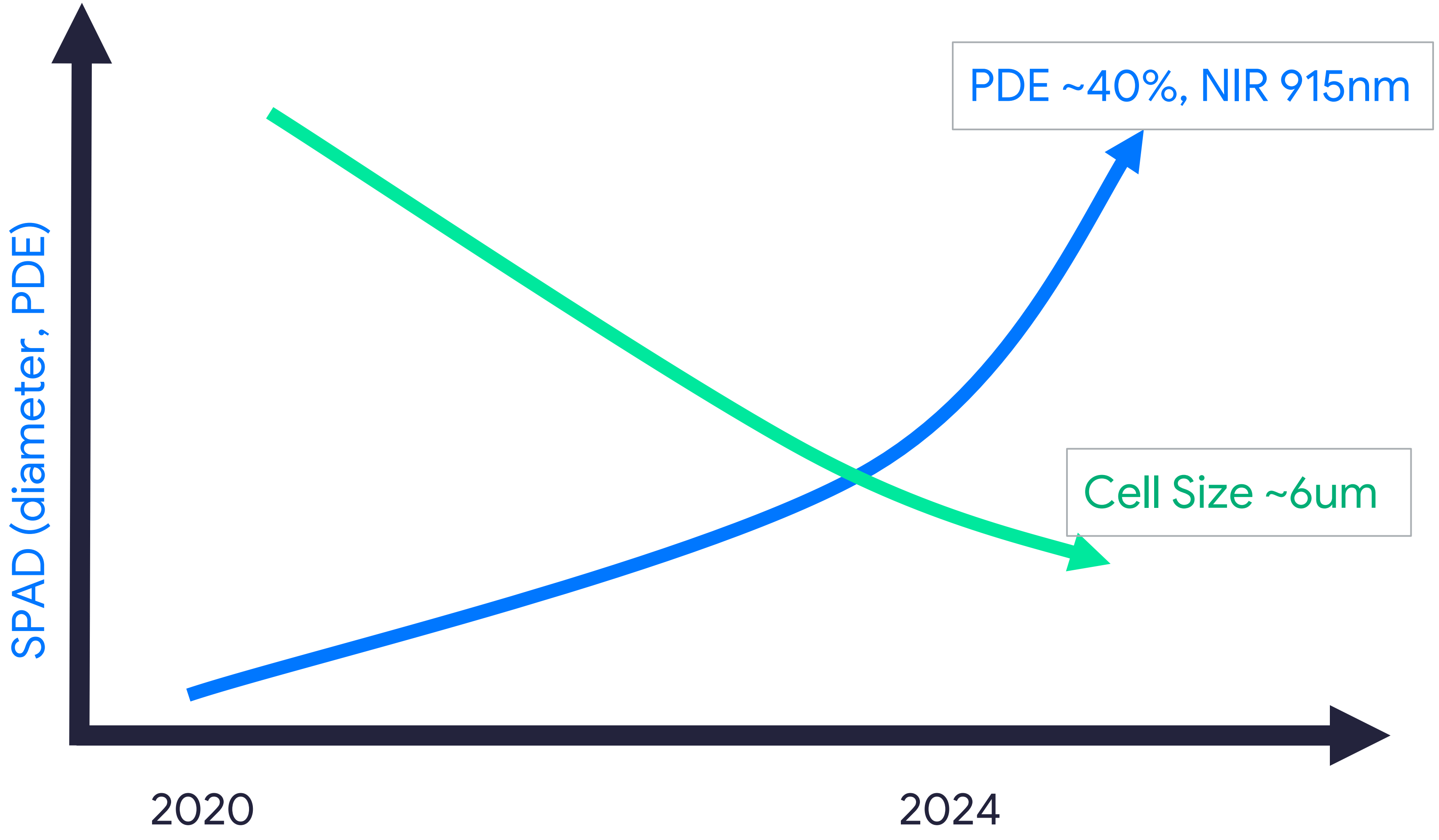
https://en.wikipedia.org/wiki/Single-photon_avalanche_diode

Photon Detection Efficiency (PDE)

- PDE has advanced (past 5 years):
  - 5% => 10% => 20% => 40% (915nm)*
- Increasing PDE improves signal-to-noise of lidar detection

SPAD Size

- Diameter is shrinking (past 10 years)
  - 50um => 25um => 15um => 6um*
- Denser SPAD array gives higher dynamic range per const-diameter macropixel



*Fujisaki et al, “A back-illuminated 6um SPAD depth sensor with PDE 36.5% at 940nm via combination of dual diffraction structure and 2x2 on-chip lens,” Symposium on VLSI Technologies and Circuits (2023) [Sony]

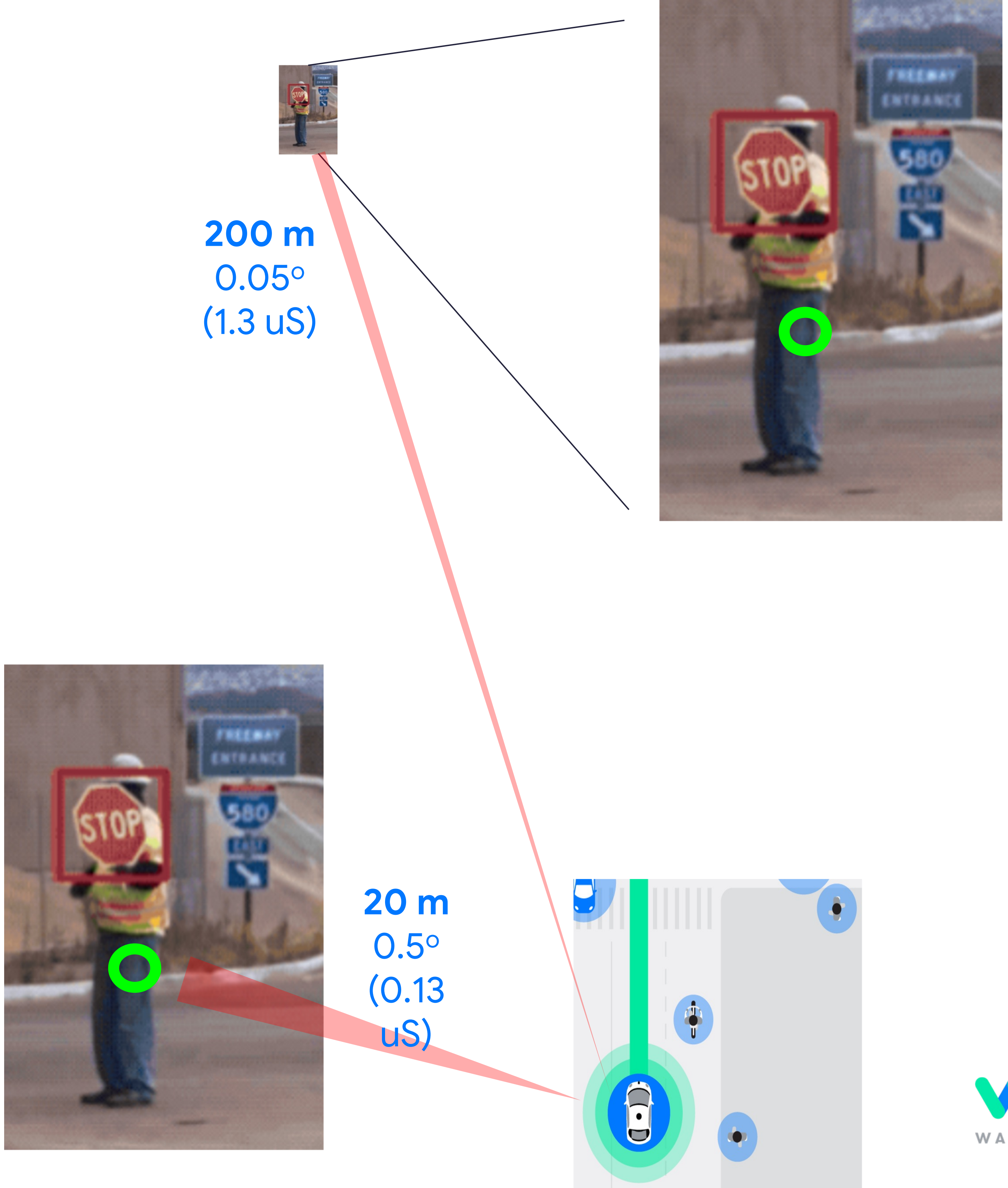
Short vs long range lidars

Number of SPADs per IFOV

	Short range	Long range	Ratio
Range (m)	20	200	10:1
Azimuth resolution (17 cm cross-range at max range)	0.5°	0.05° (0.9 mrad)	1:10
Listening time for max range (us)	0.13 μs	1.3 μs	10:1
Time per IFOV (10 Hz spin rate)	139 μs	14 μs	1:10
Time per IFOV / Listening time = shots per IFOV	900	9	1:100
SPADs per IFOV (channel)	10	1000	100:1
Optimal accumulation approach	Digital SPAD	Analog SiPM	

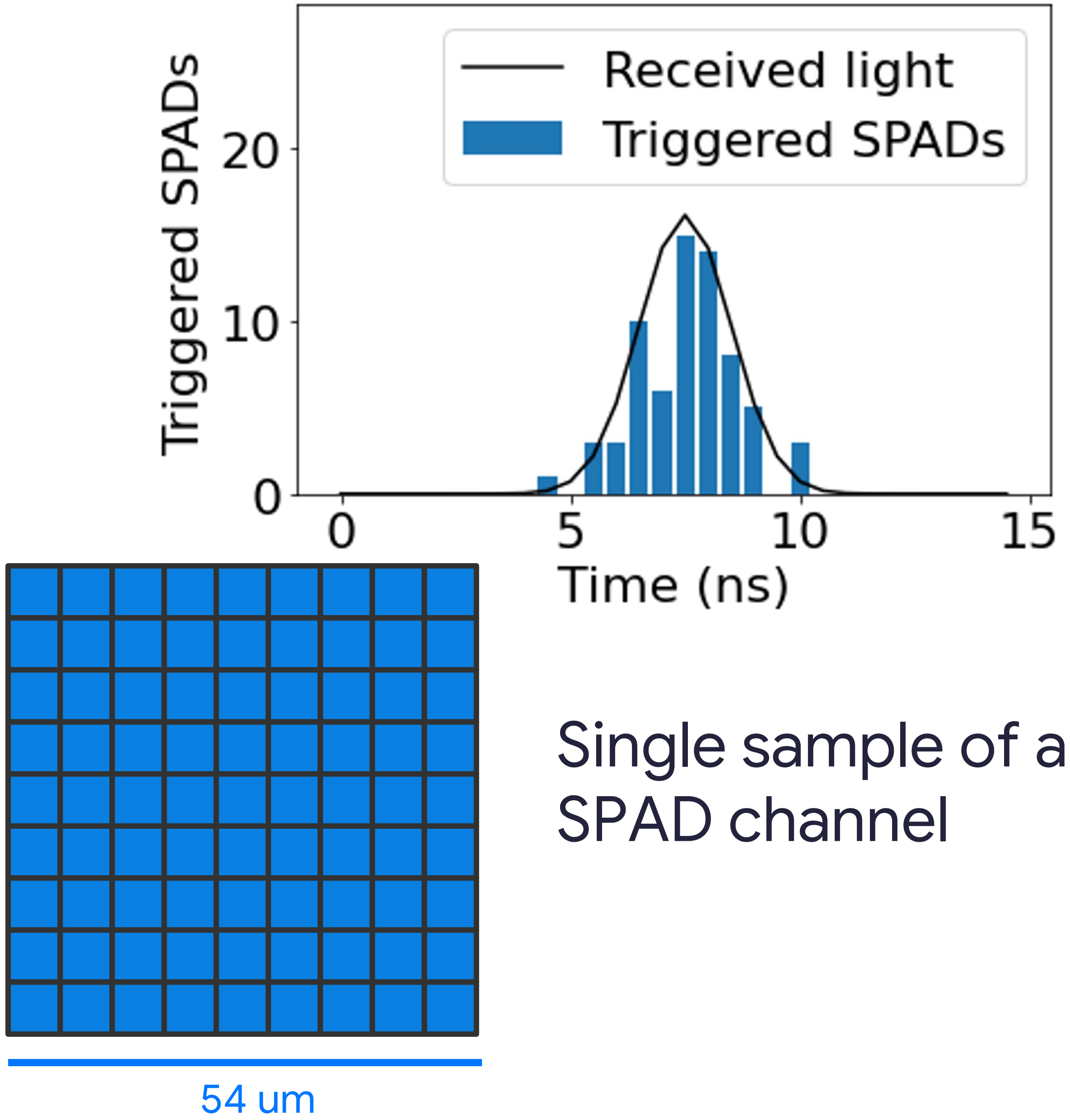
J. Dunphy, “Integrated Lidar Sensors for L4 Autonomous Vehicles”, Symposium on VLSI Technologies and Circuits (2023)

SiPM – Silicon Photo Multiplier (aka MPPC – Multi Pixel Photon Counter)

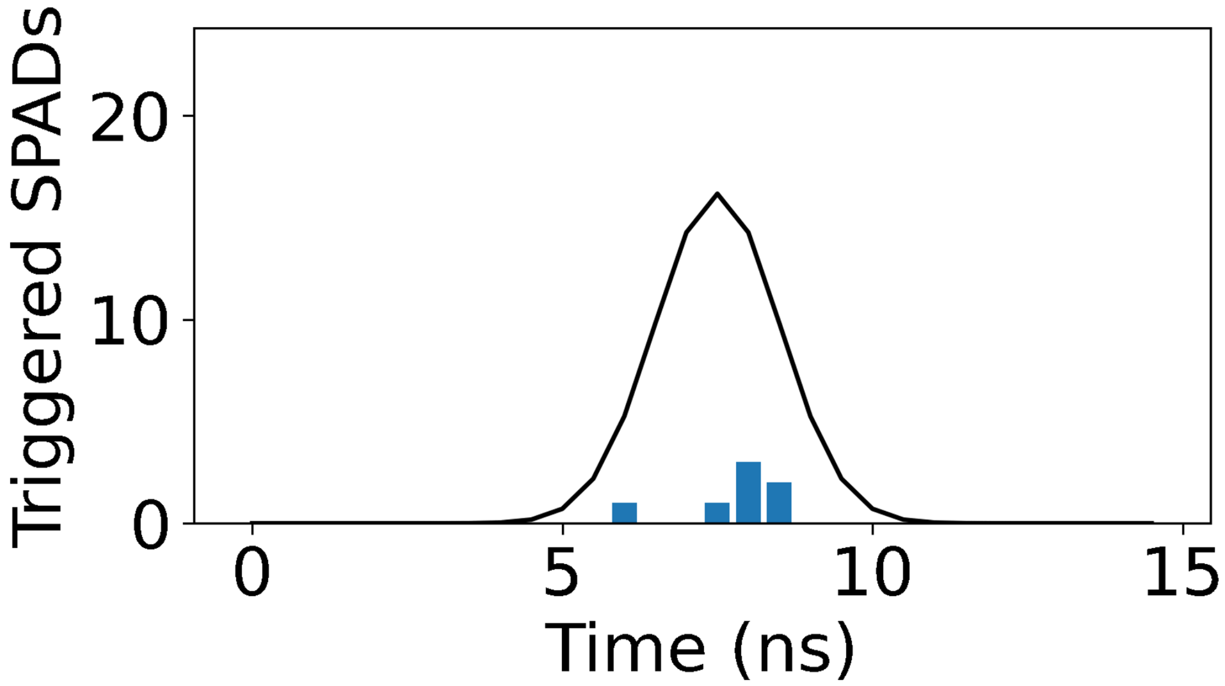


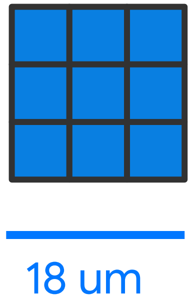
Accumulating multiple laser pulses

If you have time, take multiple shots to reduce SPADs per pixel



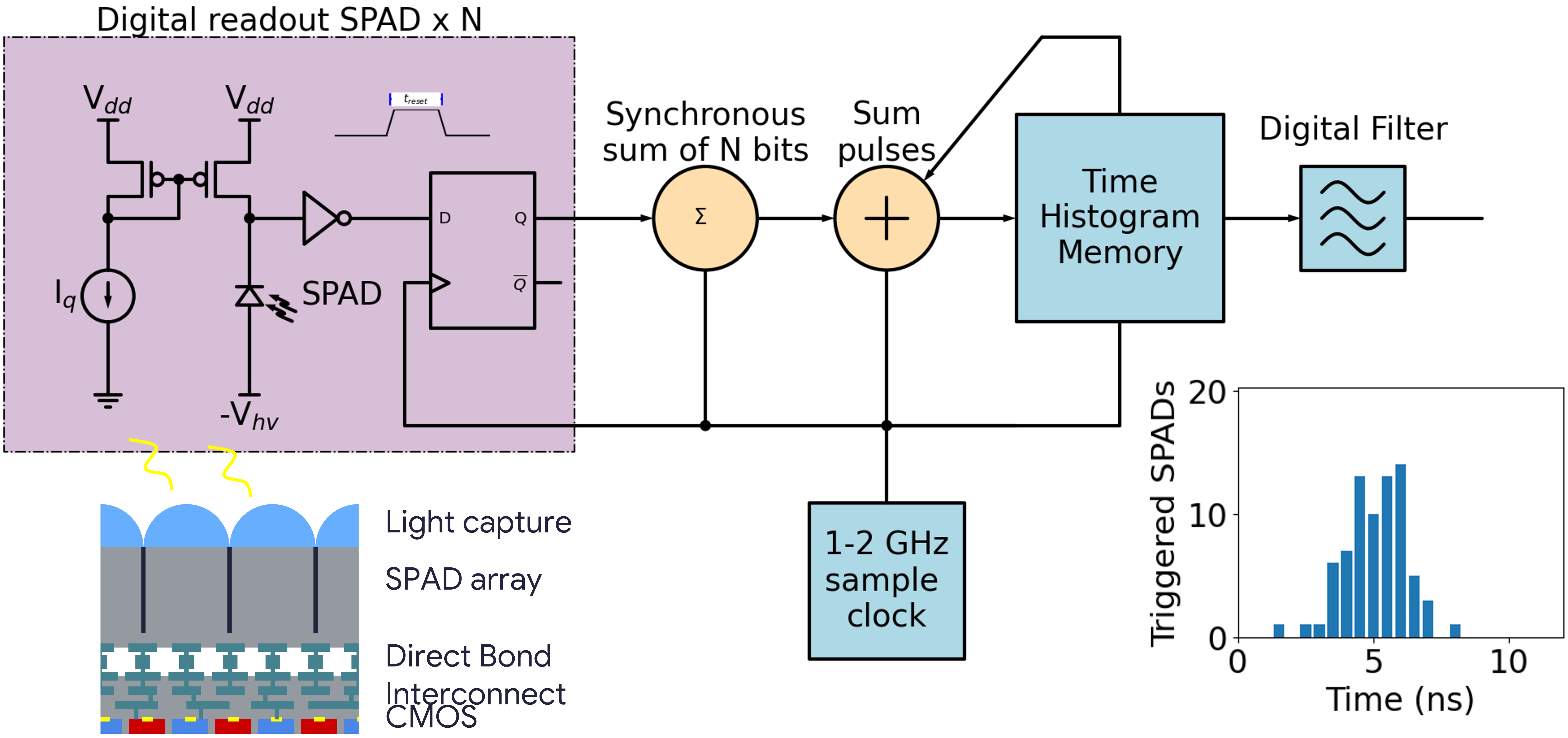
Single sample of an 81 SPAD channel



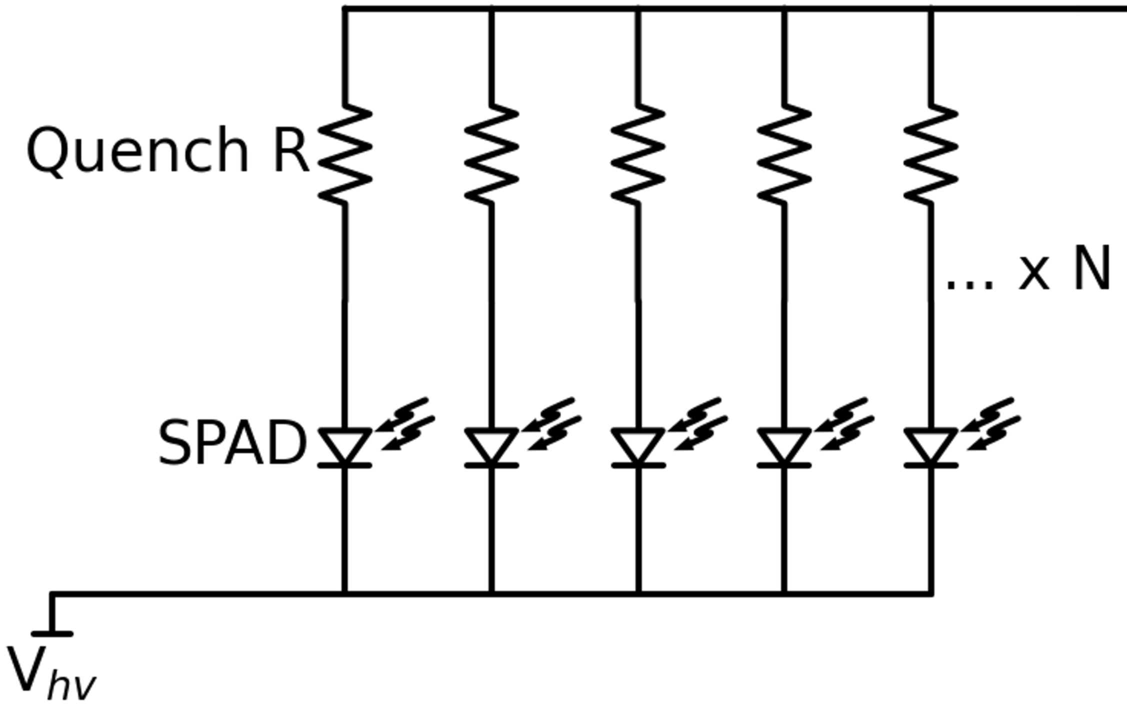


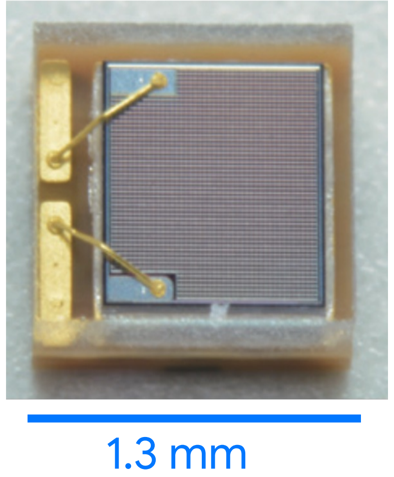
Sum of 9 stacked histograms of a 9 SPAD channel.

Monolithic Digital Accumulation of SPAD Signal



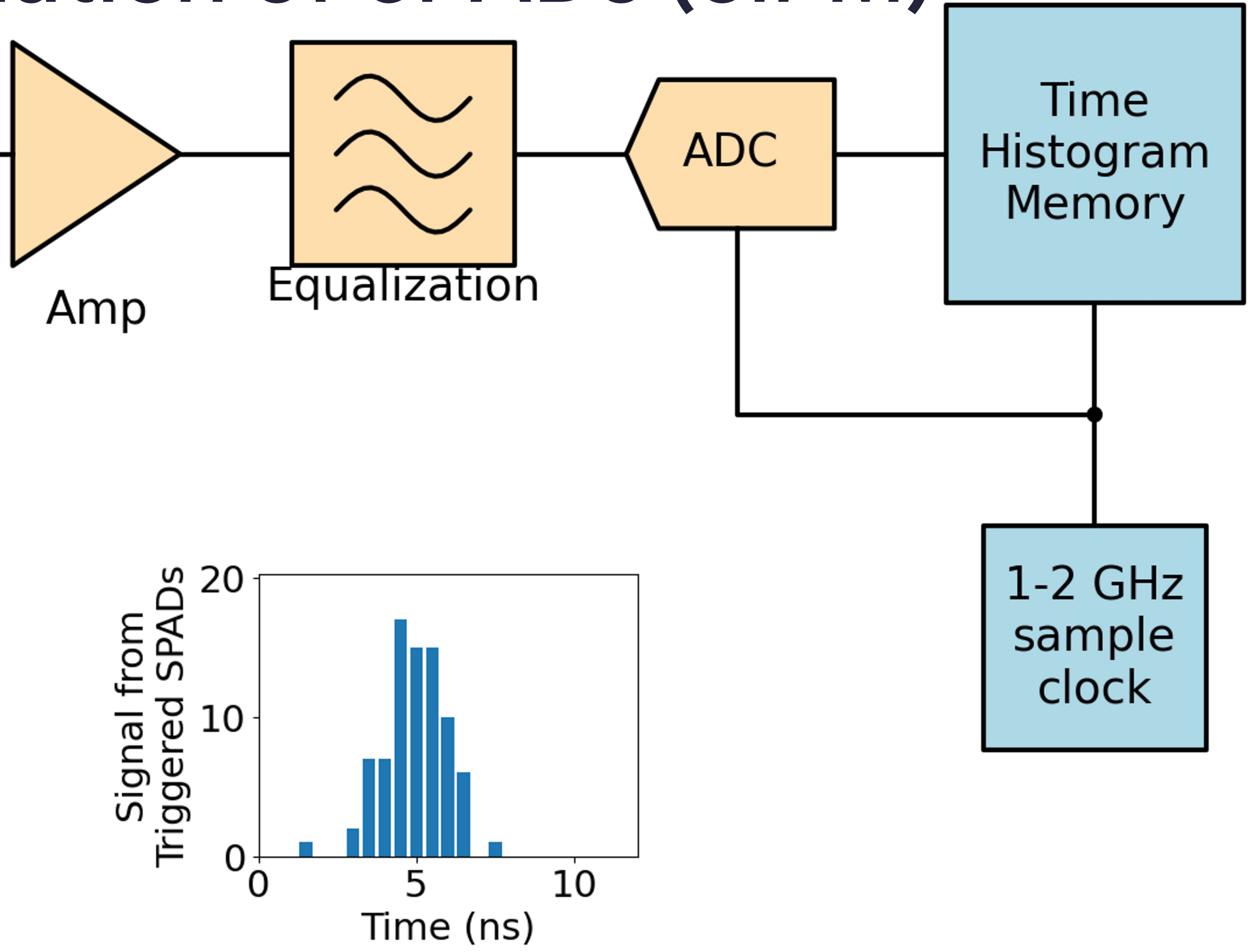
Analog Accumulation of SPADs (SiPM)





Hamamatsu
S25639-1325PS
2120 SPADs (25 um)
(example)

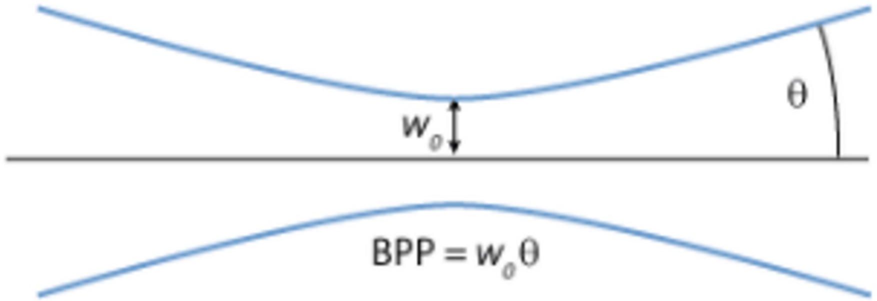
SiPM Sensor


Amp
Equalization
ADC
Time Histogram Memory
1-2 GHz sample clock

Readout electronics

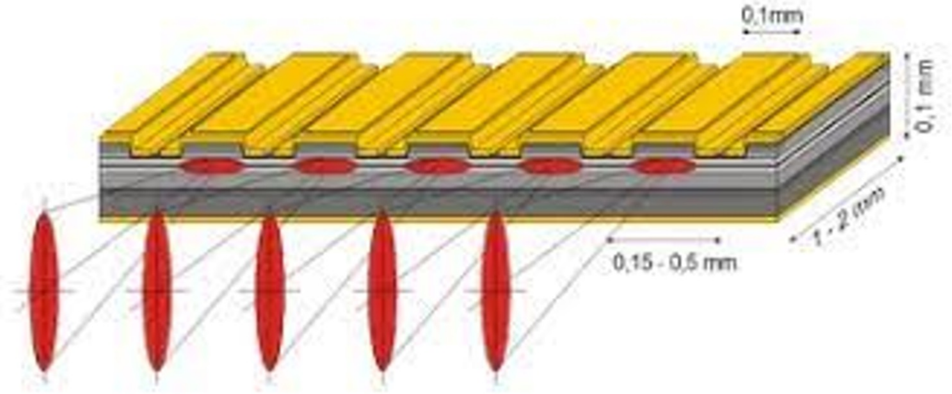
Multimode Laser Diode Properties

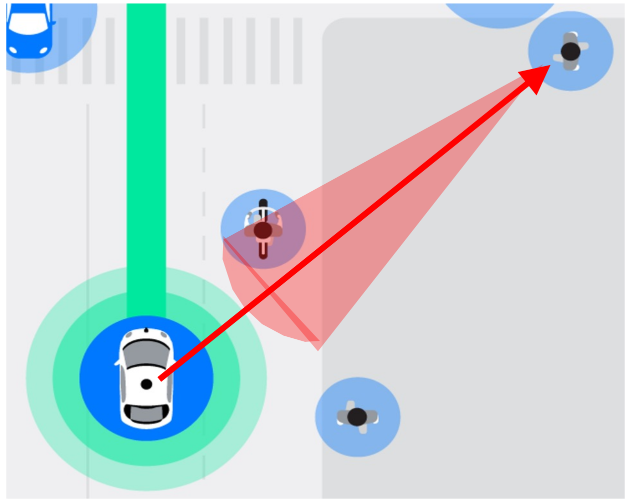
- **Beam Parameter Product (BPP)**
  - ○ BPP = $\boldsymbol{\omega} \cdot \boldsymbol{\theta}$ = (beam-waist) x (half-angle)
  - ○ BPP = $M^2 \cdot (\boldsymbol{\lambda}/\boldsymbol{\pi})$ for nearly gaussian beam

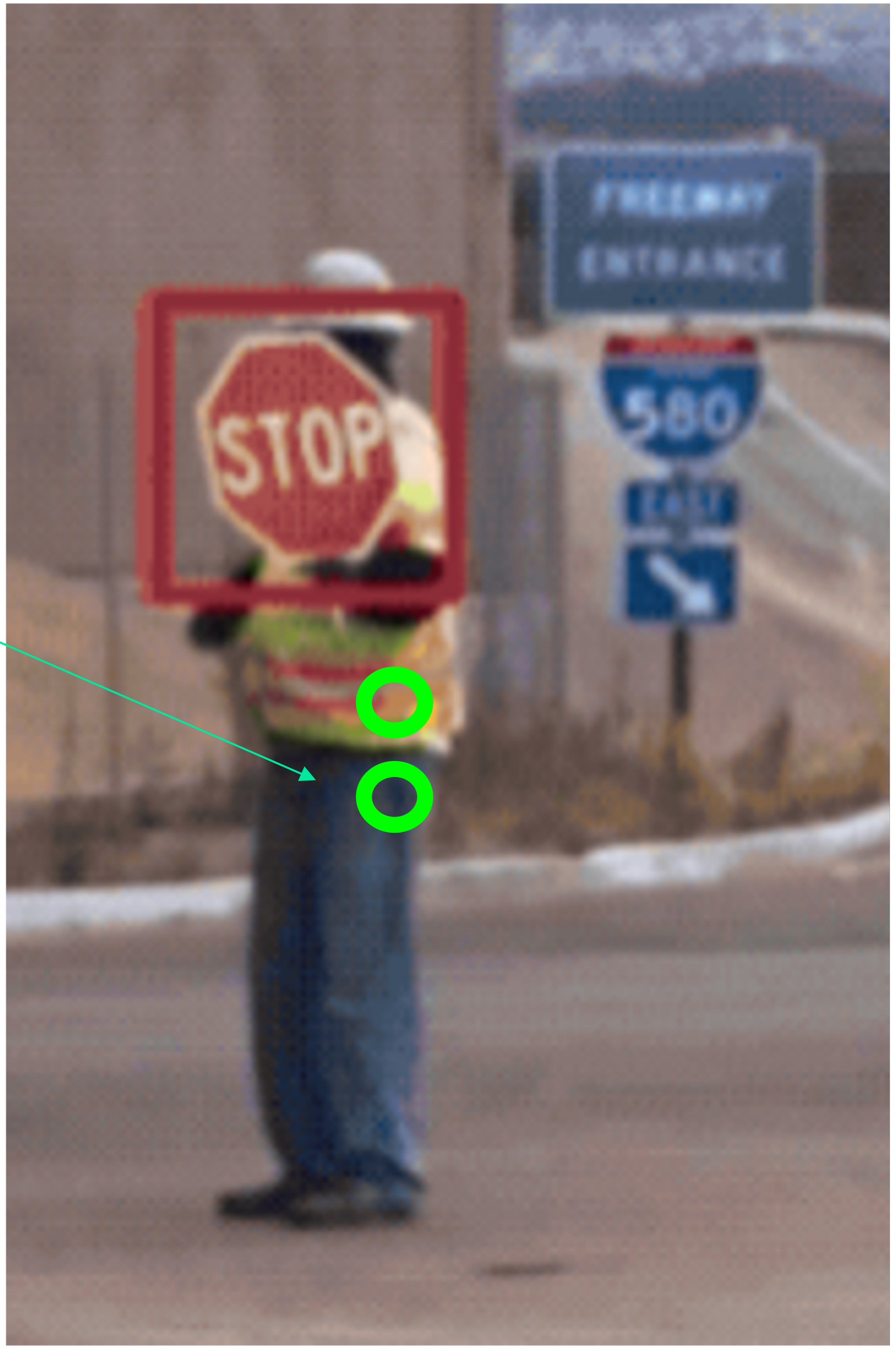


- **Returned Laser Power [W]**

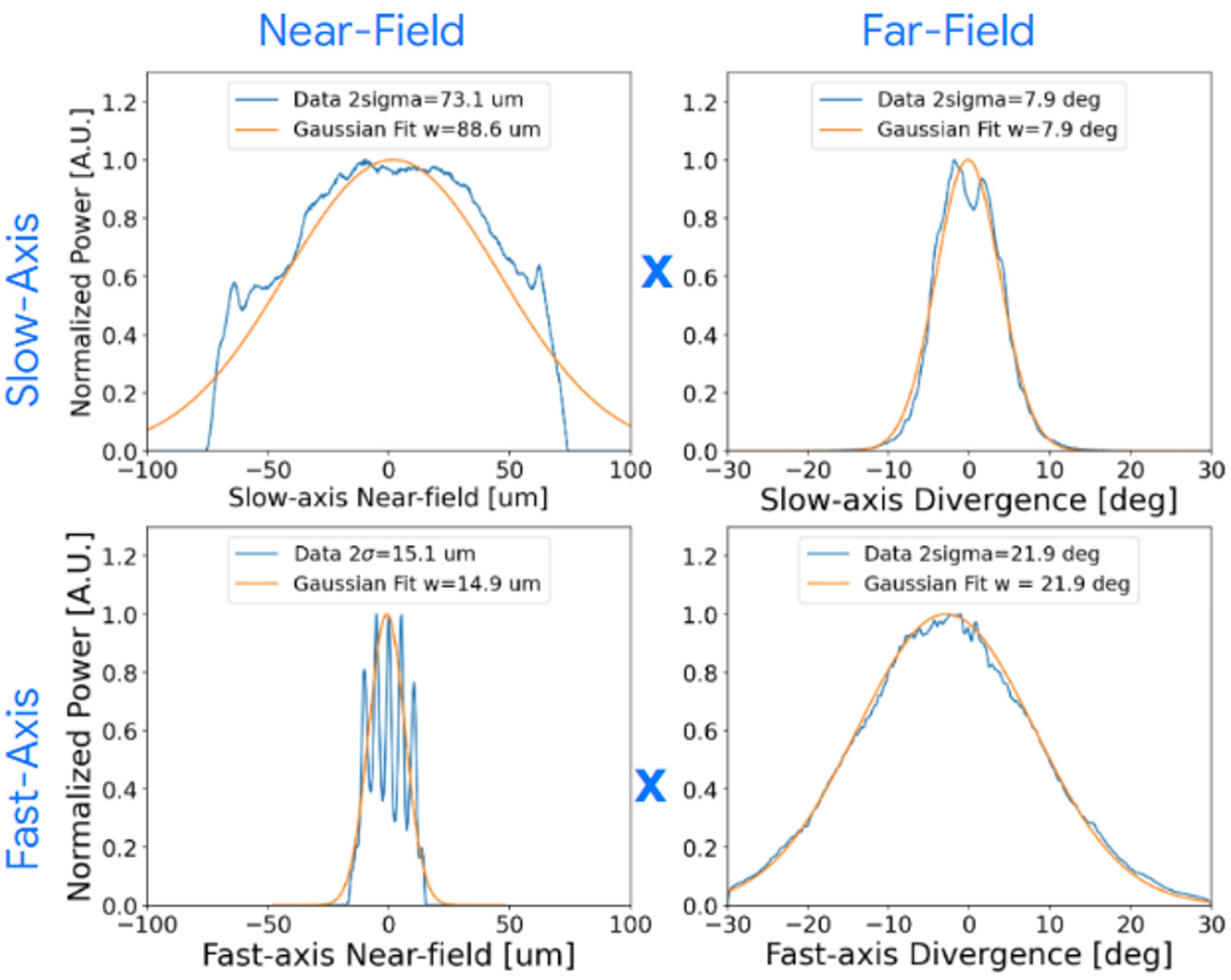
$$P_{sig} = \rho_{target}\left(\frac{A_{lens}}{\pi R^2}\right)P_{tx}$$







Power scaling: More junctions and higher current

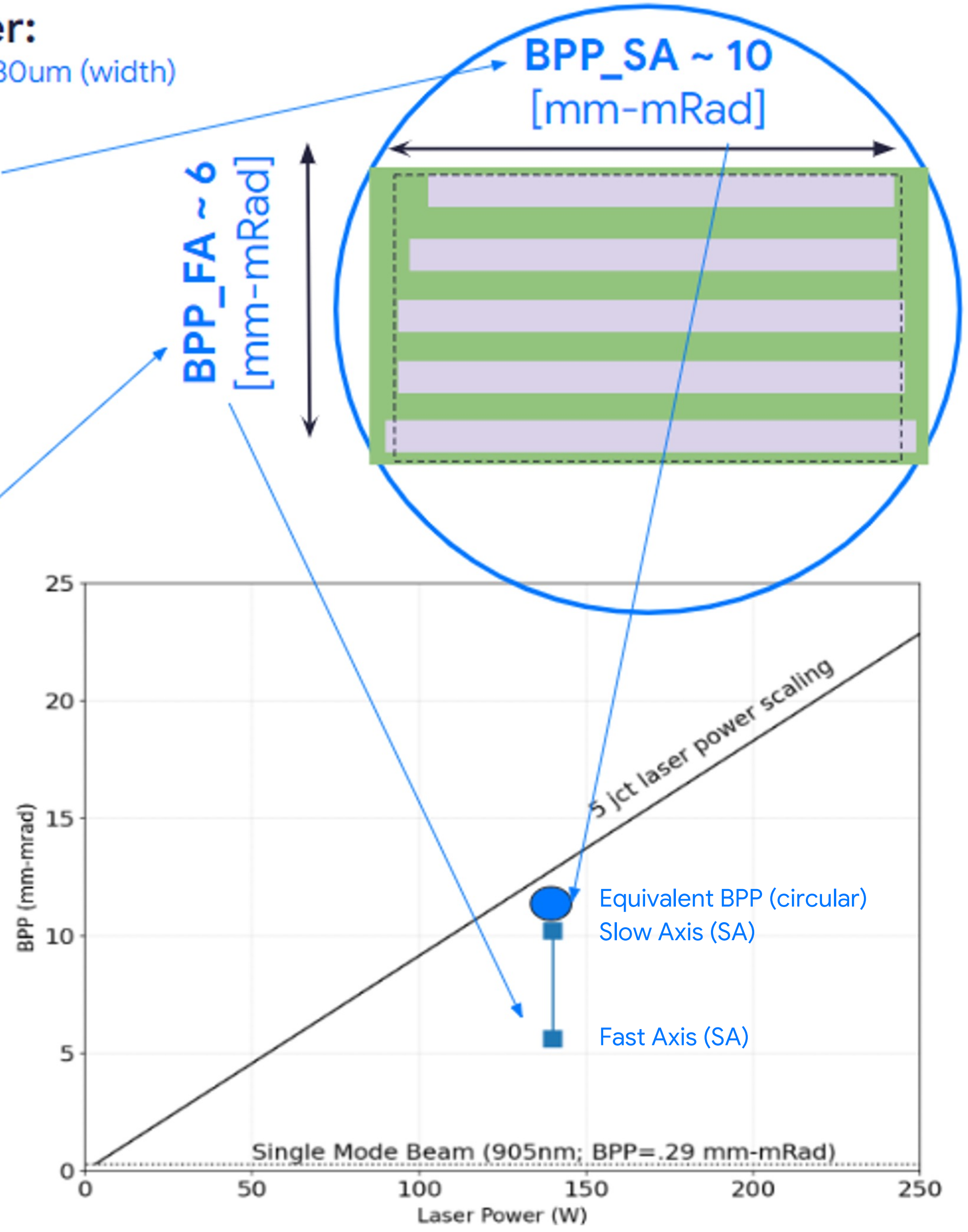


130 um

5 stack Laser:
20um (height) x 130um (width)
905nm

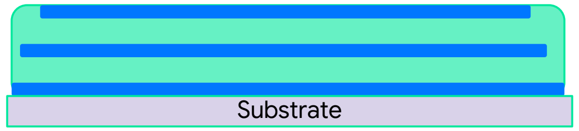
BPP_SA
73um x 7.9deg
~10 [mm-mRad]

BPP_FA
15um x 21.9deg
~6 [mm-mRad]

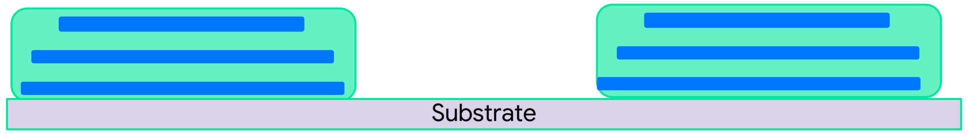


Beam Quality vs Power

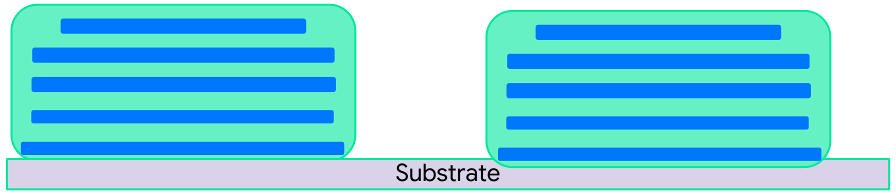
- **2015** 200um wide, 120W
  3 junction, ~1 mm long
  Single Emitter



- **2020** 130um wide, 60W
  3 junction
  8ch Bar (300um pitch)



- **2024** 130um wide, 140W
  5 junction
  8 ch Bar (200um pitch)

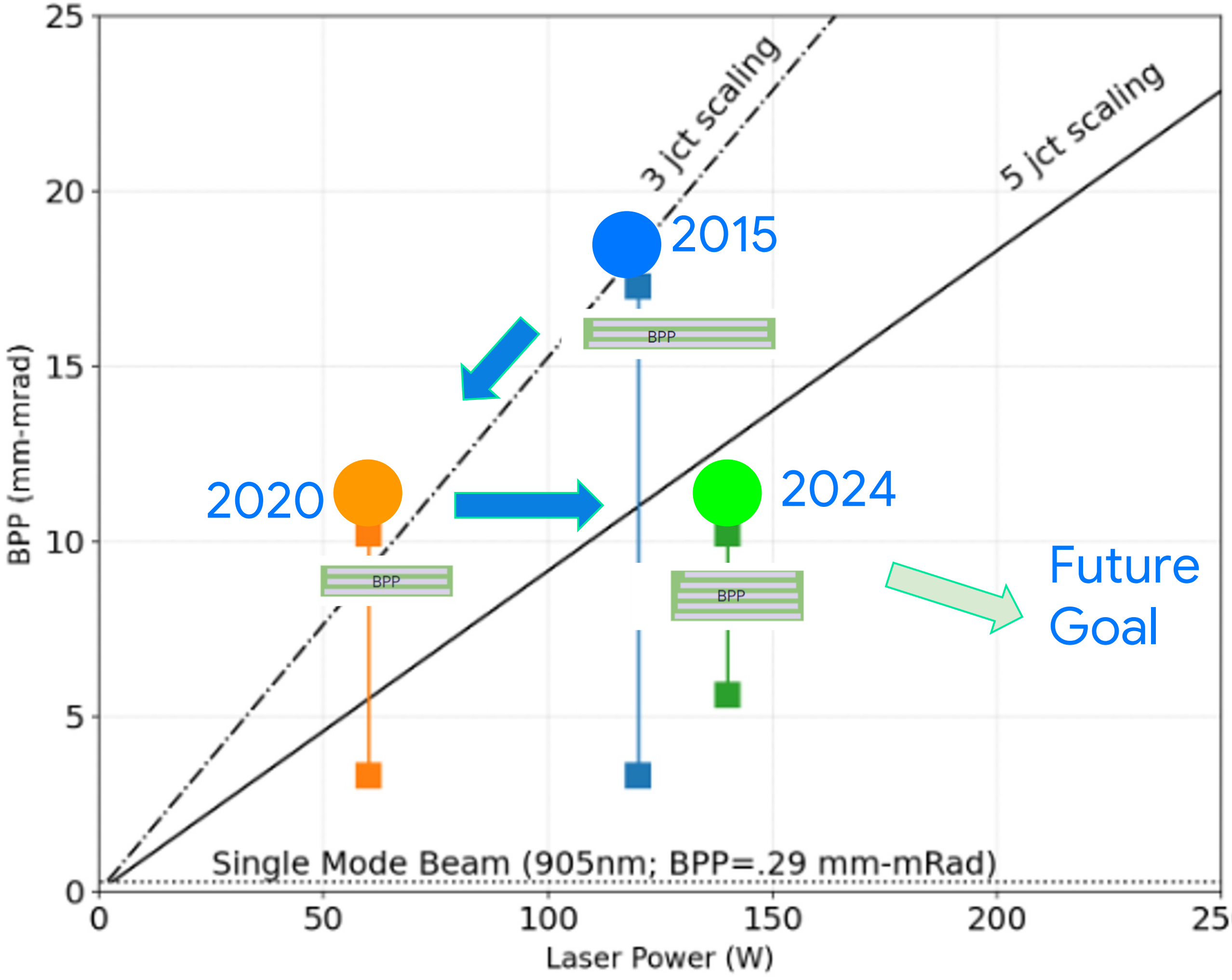


Laser diode improvements:

- Fast-Axis BPP – OK
- Turn-on Voltage (Vto) – OK
- Slope Efficiency – much better

Increasing drive current

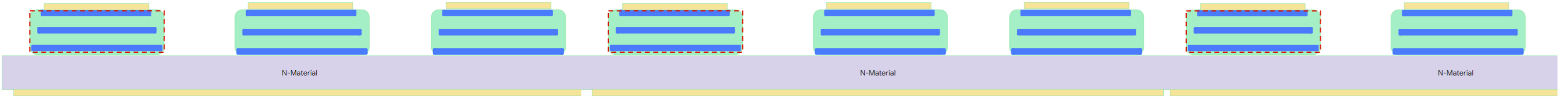
- Higher Power
- Good wall-plug efficiency with more junctions



Addressable Emitters in Arrays

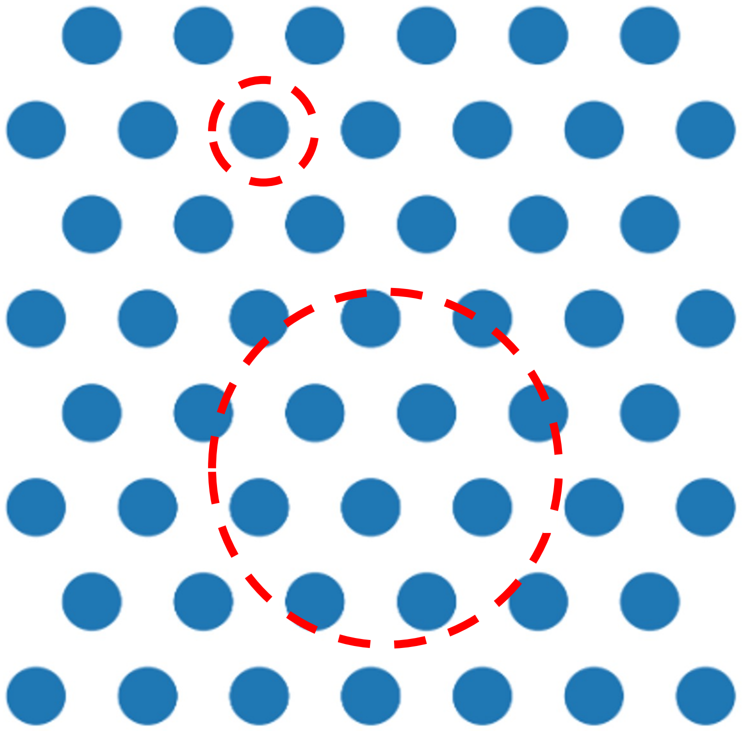
Edge-Emitters in linear arrays

- Base Cell:
  - P-contact (addressable)
  - Width x heigh
  - Pitch: 200um pitch



VCSEL 2D (sub)arrays

- Base Cell options:
  - single or group of VCSELS
  - Common-cathode or separate cathode contact



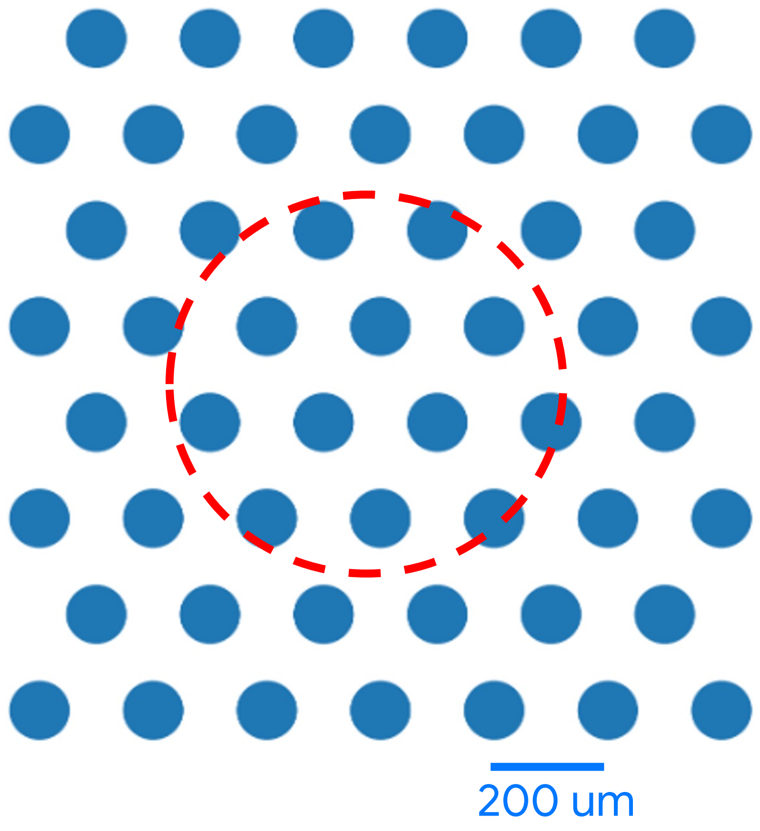
VCSEL Trends

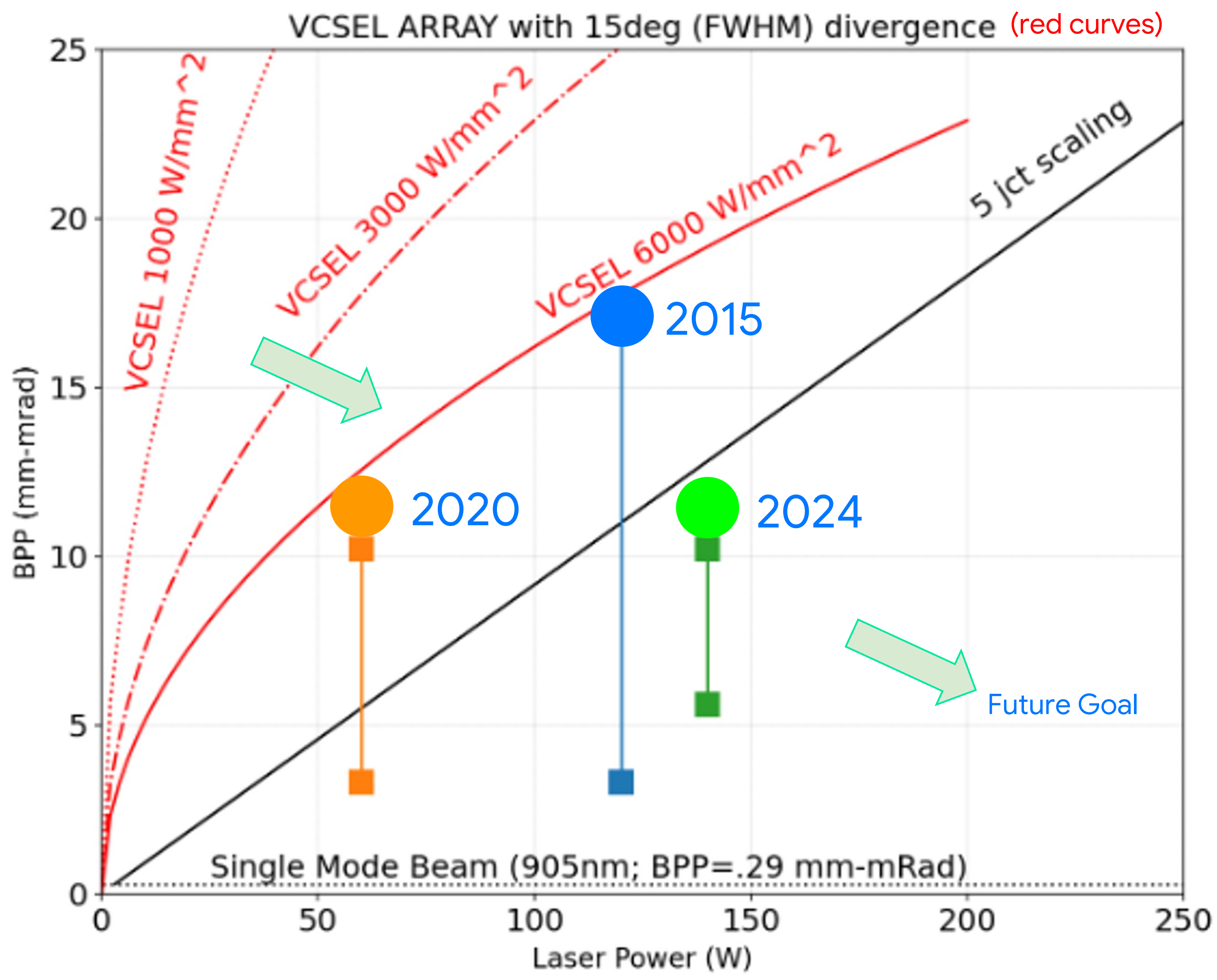
Power

- more junctions (no BPP penalty)
- Increased Diameter
- Increased Fill-factor
- scales as Power/area * subarray_radius^2

BPP (Beam Parameter Product)

- Assume 15 deg (FWHM) divergence
- Scales as radius (sqrt(laser_power))





David A. Schleuning et al, “Lidar for autonomous vehicles: trends in lasers and detectors” Photonics West 2024 (12867-1)

Ingredients for a Scalable ML Stack

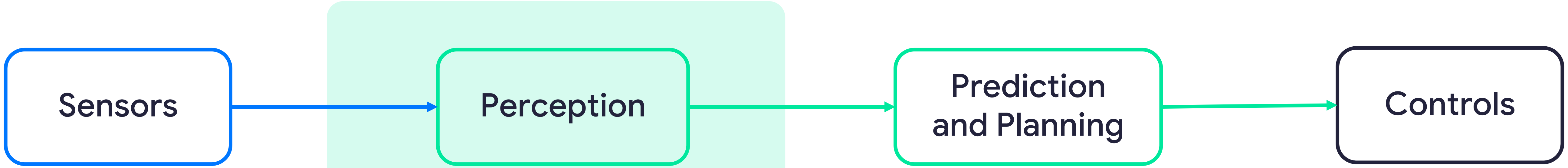
WAYMO

- **Lidar Detection Models**
- Behavior Prediction
- Auto-labeling
- Leveraging LLM / VLM Knowledge

ML – Machine Learning
LLM/VLM – Large Language Model / Visual Language Model : LLMs that understand text and images

Credit: D. Anguelov, *Ben Taskar Memorial Lec*ture (youtube Feb 2023)

High Level System Diagram



- Lidars
- Radars
- Cameras

Aggregate to
Bird’s Eye View (BEV)

Intermediate Outputs

- 3d object boxes, heading, human keypoints, attributes...
- occupancy probability grids

Pros:

- Improved generalization
- Plan validation capability
- Useful for simulation testing
- Data compression (Behavior-Prediction data needs >> Perception)

Cons:

- Requires feature design, can grow complex
- Needs labeling

Can we have an efficient Transformer architecture For Bird’s Eye View processing?

Improve on traditional solutions:

Dense convolution

○ quadratic in BEV range

Sparse convolution

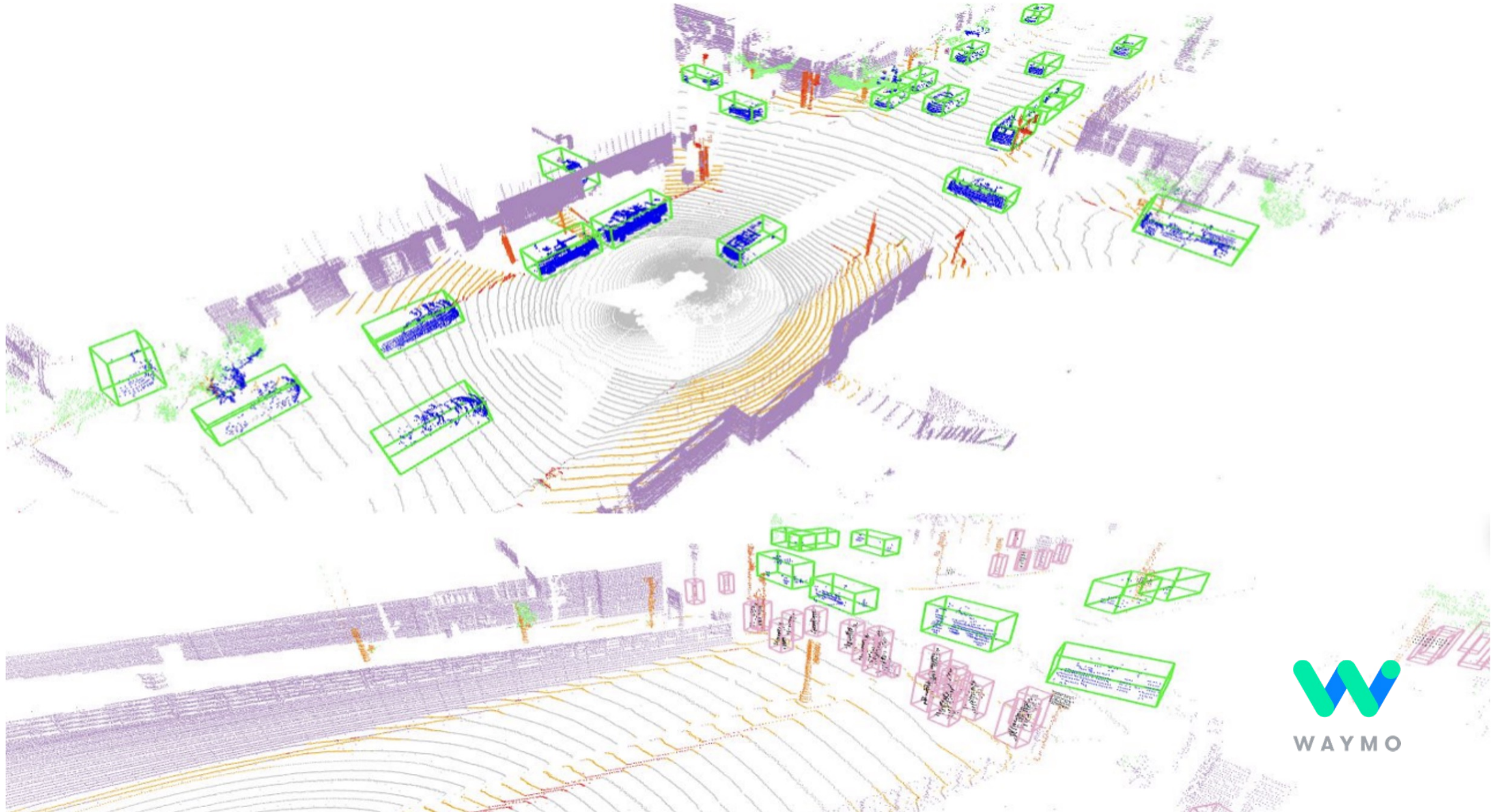
○ incompatible with efficient TPU utilization

○ unable to pass information between ‘point islands’

SWFormer: Sparse Window Transformer for 3D Object Detection in Point Clouds

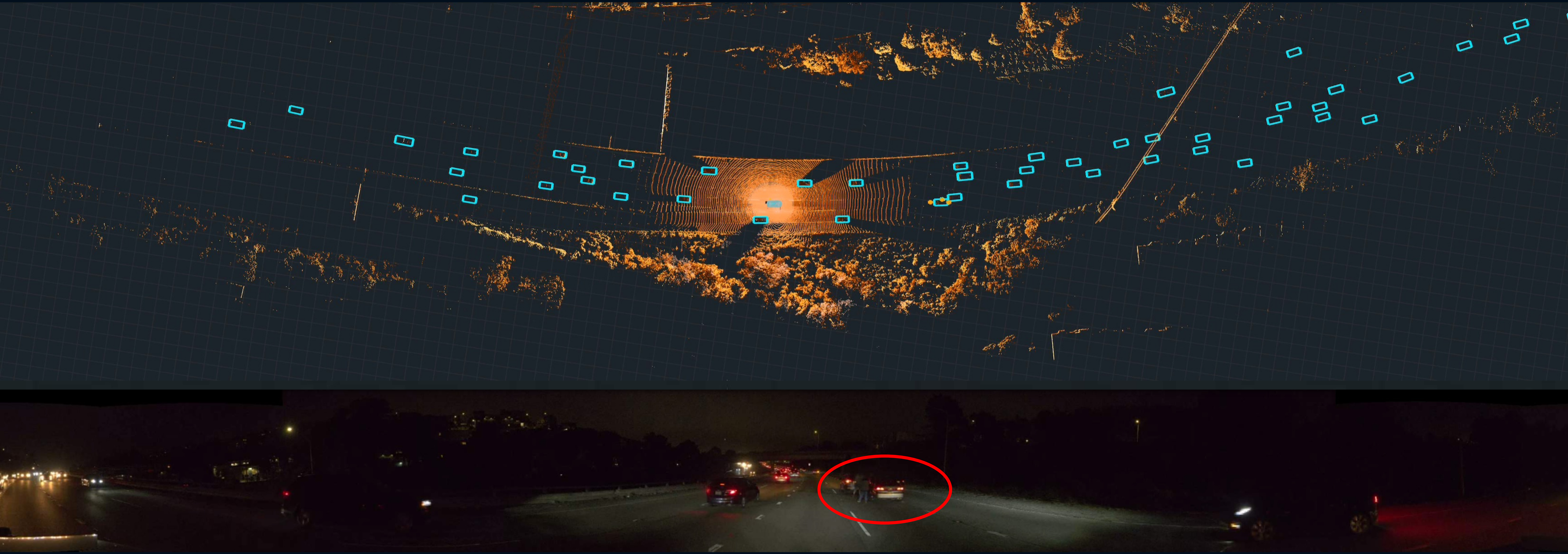
Pei Sun, Mingxing Tan, Weiyue Wang, Chenxi Liu, Fei Xia,
Zhaoqi Leng, and Dragomir Anguelov

ECCV 2022



>200m Detection on Freeway: Night-time

Accident happening in front of the Waymo car

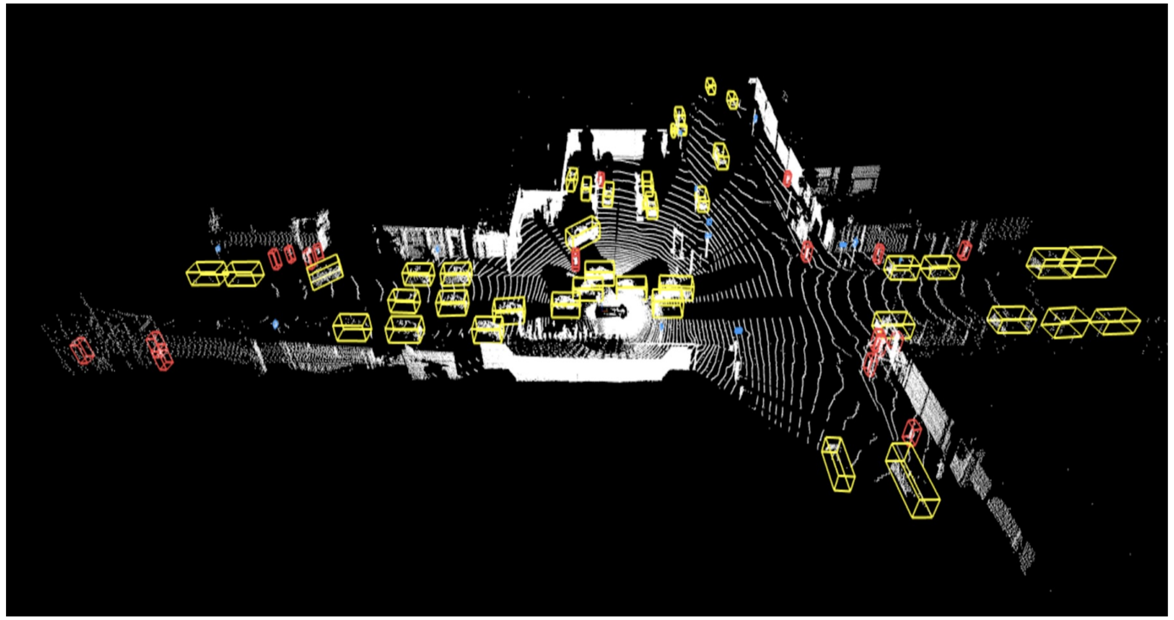


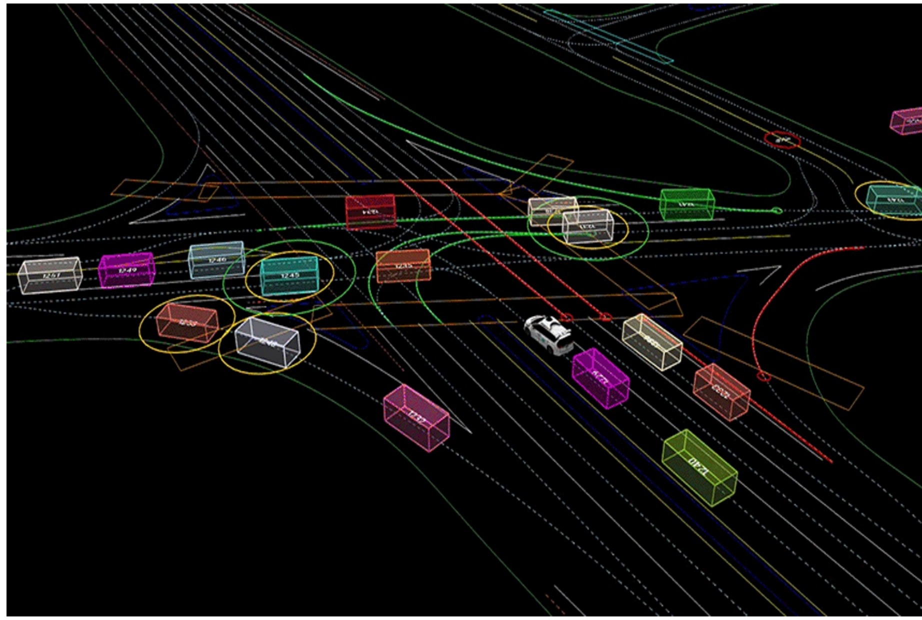
Reference panorama image. For visualization only.

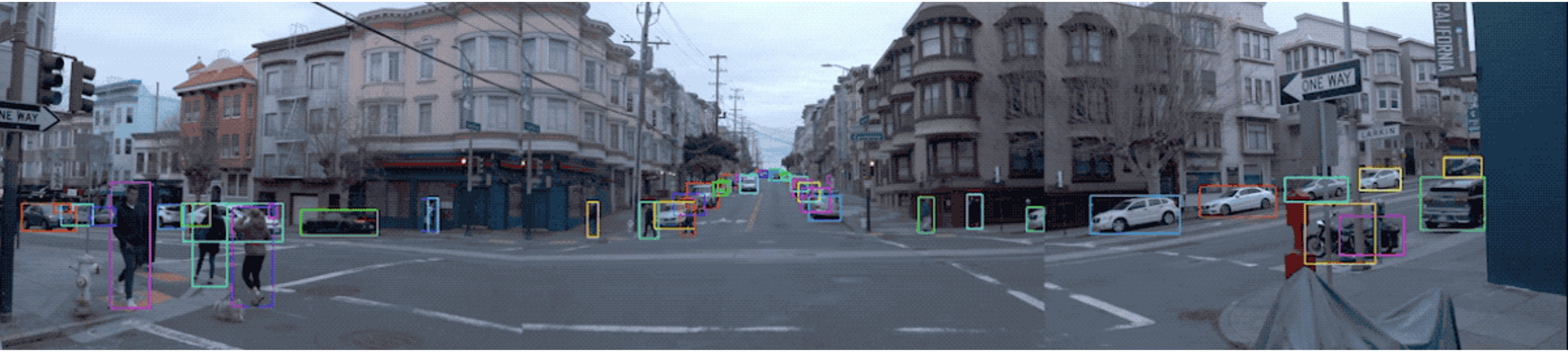
WAYMO Open Dataset

www.waymo.com/open

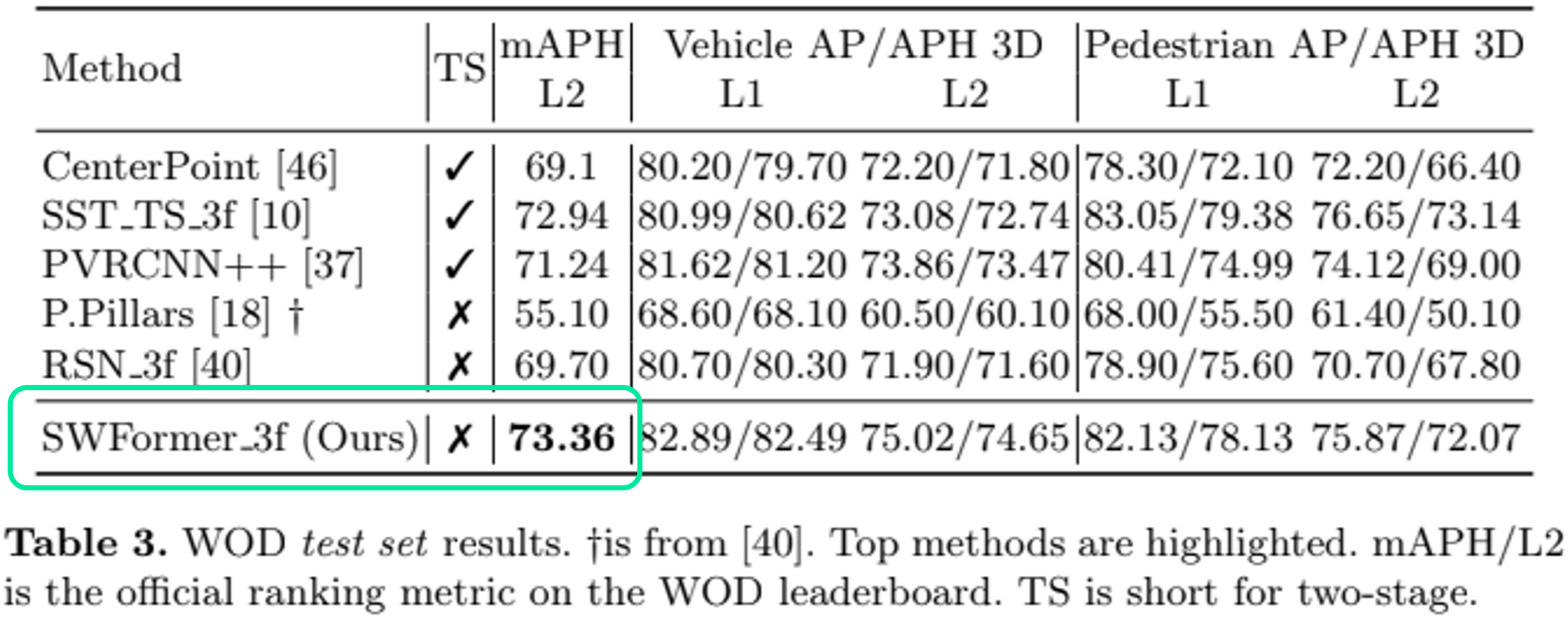
- >100k real world scenes
- Camera and lidar data
- Diverse and high quality ground-truth labels







Waymo Open Dataset Test Set Results



Method	TS	mAPH	Vehicle AP/APH 3D		Pedestrian AP/APH 3D	
		L2	L1	L2	L1	L2
CenterPoint [46]	✓	69.1	80.20/79.70	72.20/71.80	78.30/72.10	72.20/66.40
SST_TS_3f [10]	✓	72.94	80.99/80.62	73.08/72.74	83.05/79.38	76.65/73.14
PVRCNN++ [37]	✓	71.24	81.62/81.20	73.86/73.47	80.41/74.99	74.12/69.00
P.Pillars [18] †	✗	55.10	68.60/68.10	60.50/60.10	68.00/55.50	61.40/50.10
RSN_3f [40]	✗	69.70	80.70/80.30	71.90/71.60	78.90/75.60	70.70/67.80
SWFormer_3f (Ours)	✗	**73.36**	82.89/82.49	75.02/74.65	82.13/78.13	75.87/72.07

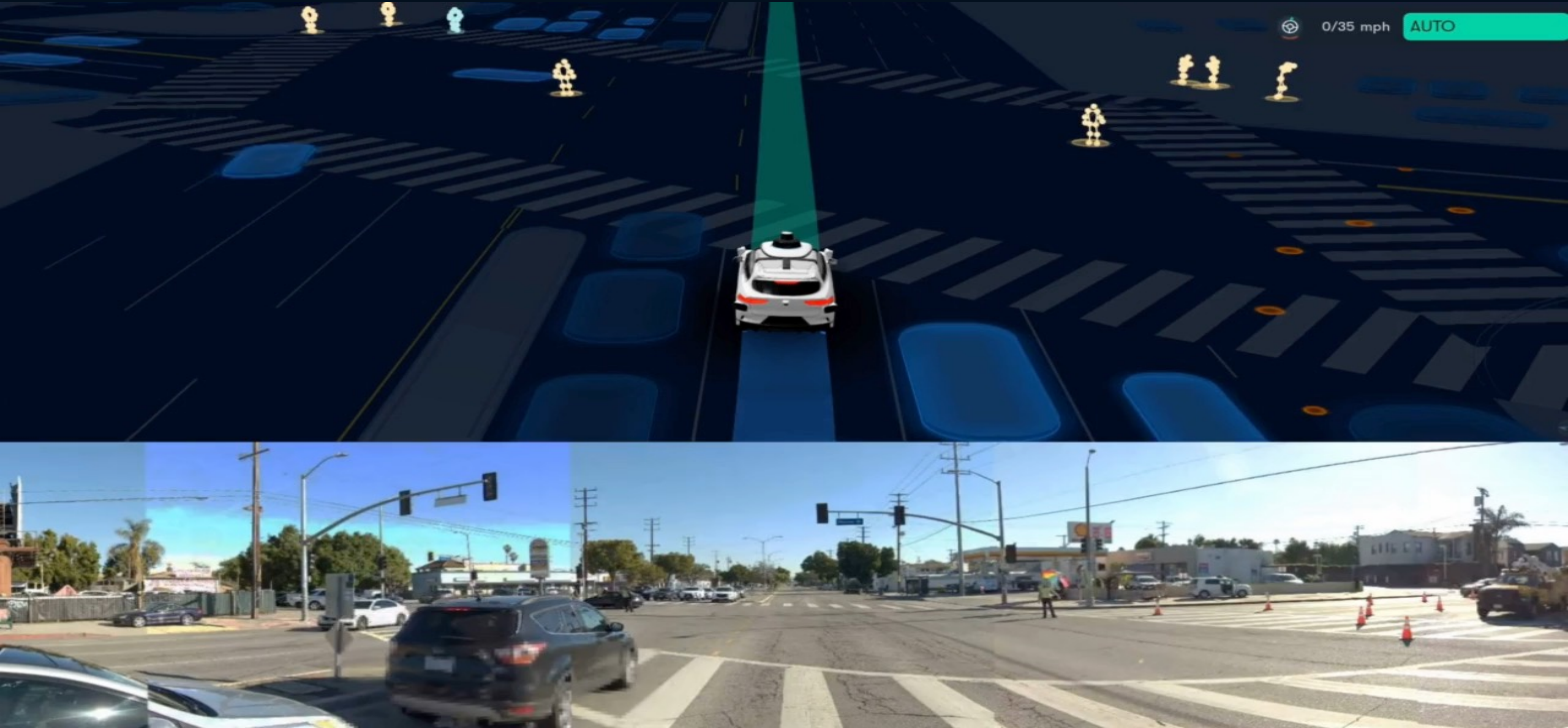
Table 3. WOD *test set* results. †is from [40]. Top methods are highlighted. mAPH/L2 is the official ranking metric on the WOD leaderboard. TS is short for two-stage.

mAPH – mean Average Precision (mAP) weighted by heading accuracy (mAPH) is the official metric for Waymo evaluation
mAPH/L2 – official metric calculated for hardest edge cases in the Waymo Open Dataset. 100% is perfect precision-recall

Pedestrian Keypoints

Multi-modal Pose Estimation with 2D Weak Supervision in Autonomous Driving, Jingxiao Zheng, Xinwei Shi, Alexander Gorban, Junhua Mao et al., 2021

HUM3DIL: Semi-Supervised Multi-modal 3D Human Pose Estimation, Andrei Zanfir, Mihai Zanfir, Alexander Gorban et al, CoRL 2022

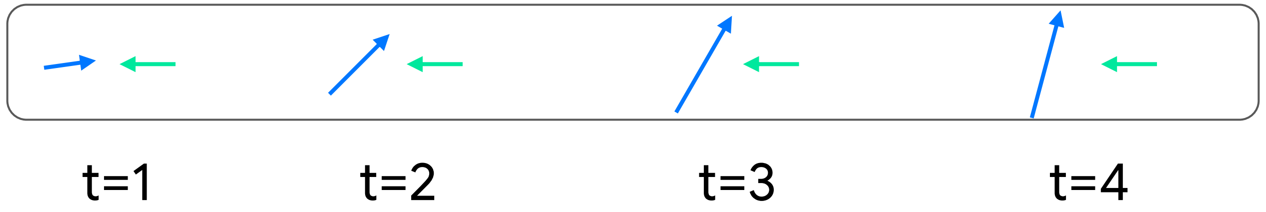


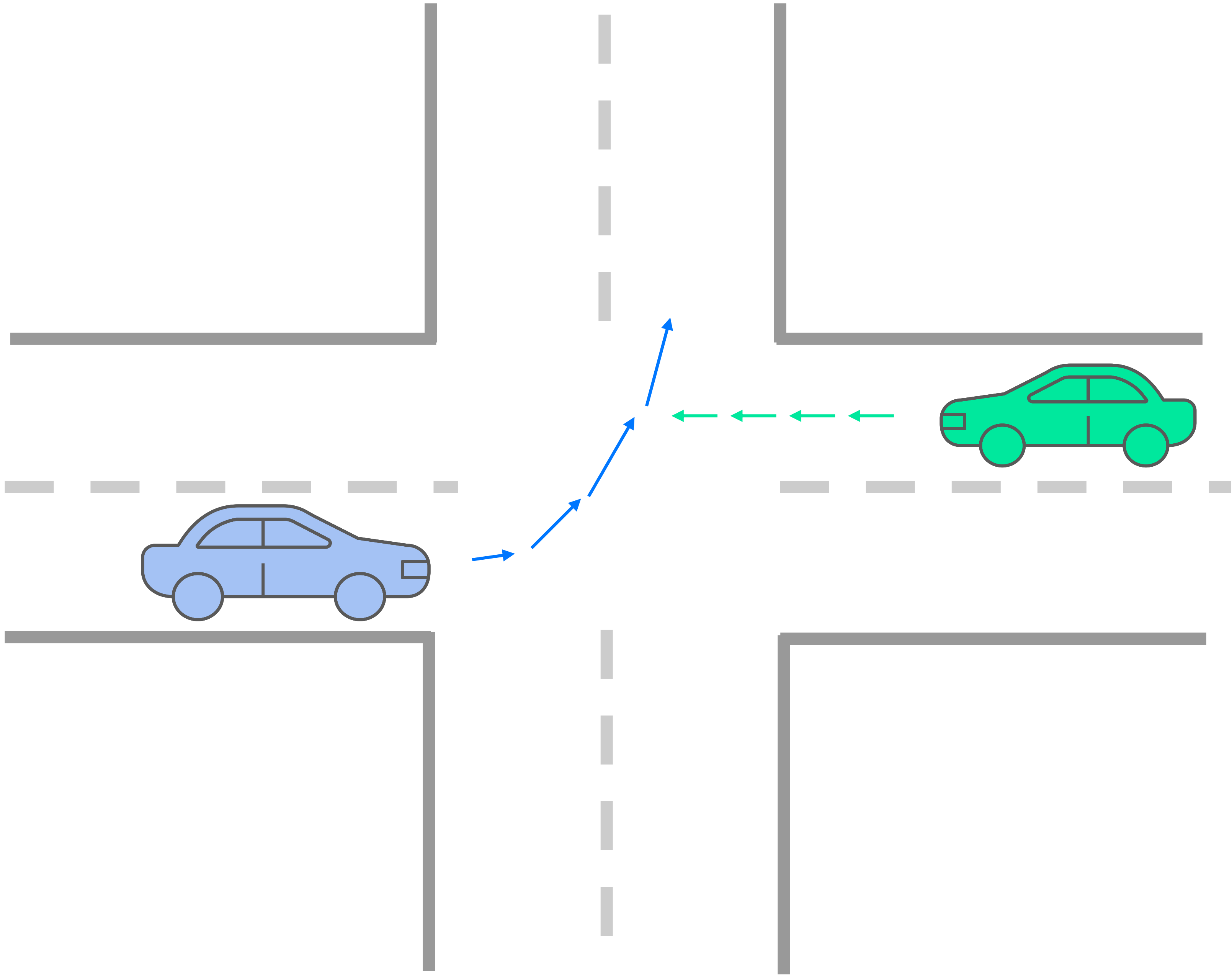
Driving is a Conversation

LLM-inspired model:

- Discrete motion tokens
- Causal cross-attention decoder

Motion token sequence:





MotionLM: Multi-Agent Motion Forecasting as Language Modeling

Ari Seff, Brian Cera, Dian Chen, Mason Ng, Aurick Zhou, Nigamaa Nayakanti, Khaled S. Refaat, Rami Al-Rfou, Benjamin Sapp

ICCV 2023

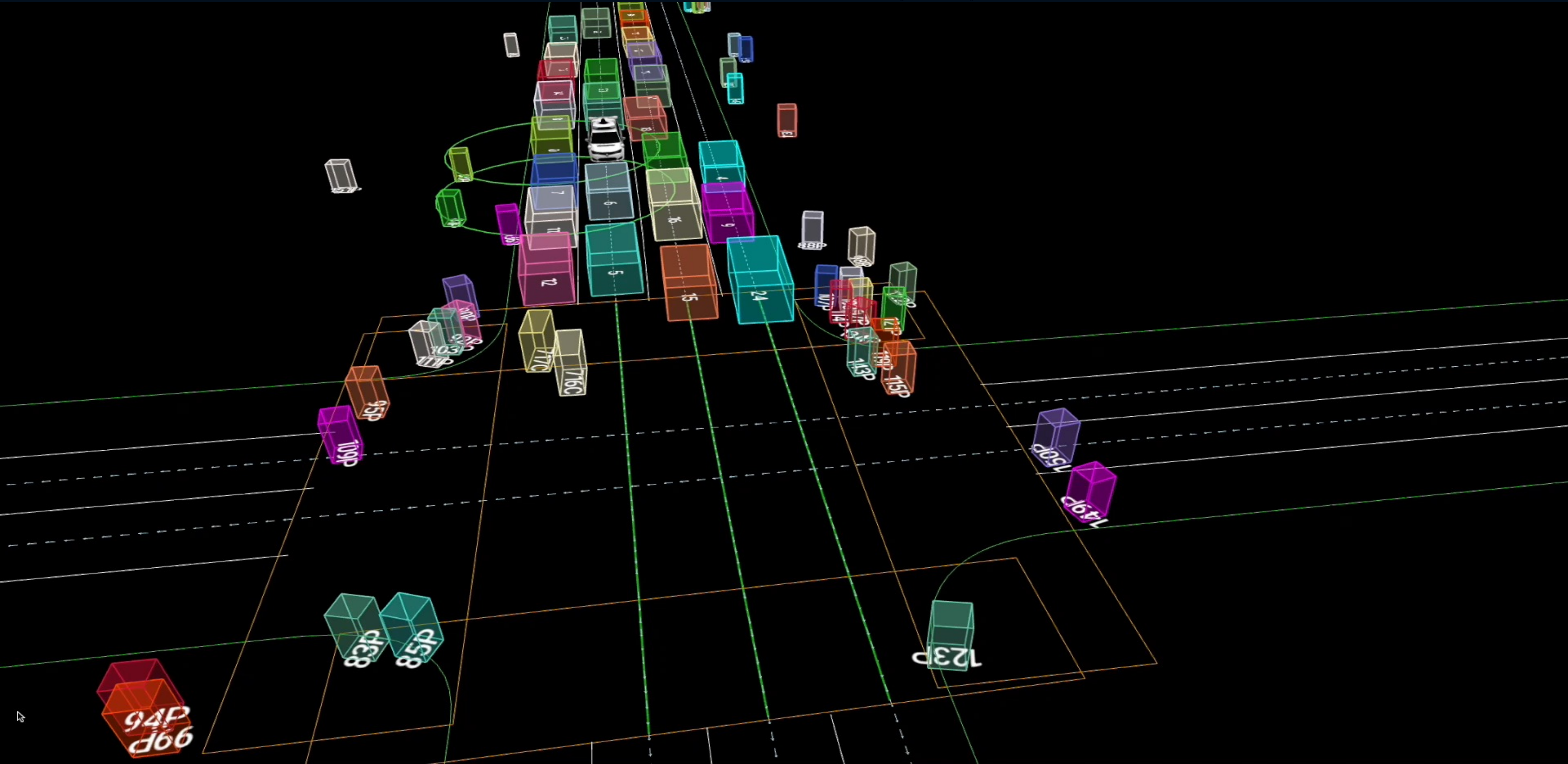
Ingredients for a Scalable ML Stack

- Lidar Detection Models
- Behavior Prediction
- **Auto-labeling**
- Leveraging VLM / LLM Knowledge

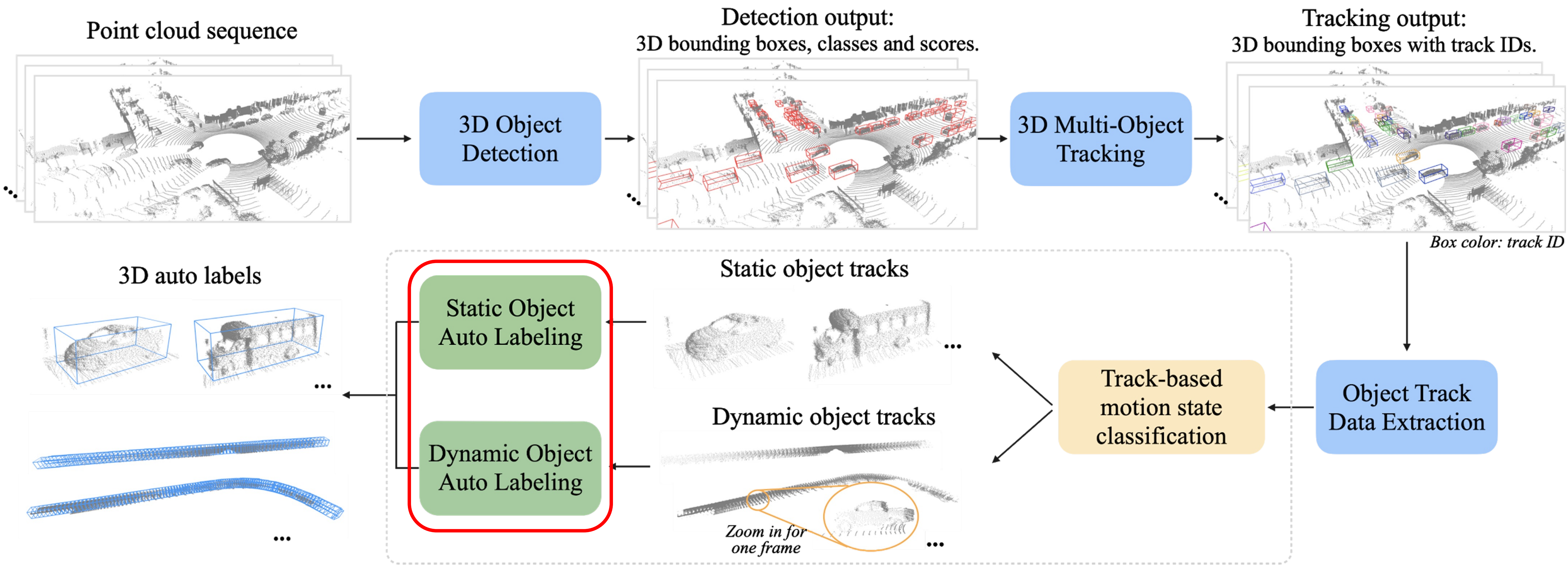
Credit: D. Anguelov, *Ben Taskar Memorial Lecture* (youtube Feb 2023)

Scaling up training: Auto labeling

Offboard 3D Object Detection from Point Cloud Sequences,
Charles Qi, Yin Zhou, Mahyar Najibi, Pei Sun, Khoa Vo, Boyand Deng,
Dragomir Anguelov, CVPR 2021



Object-centric offboard 3D detection pipeline

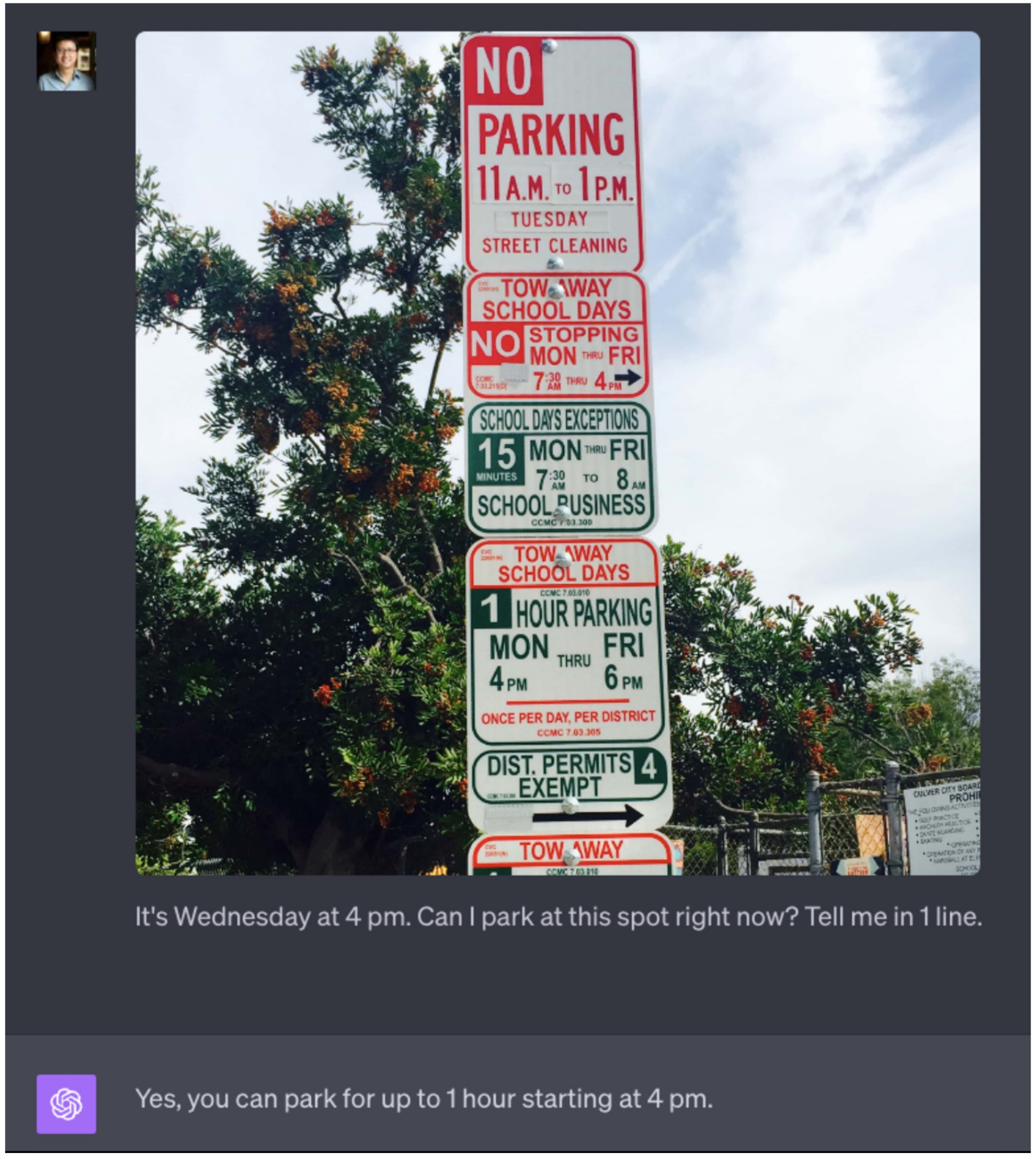


Ingredients for a Scalable ML Stack

- Lidar Detection Models
- Behavior Prediction
- Auto-labeling
- **Leveraging VLM / LLM Knowledge**

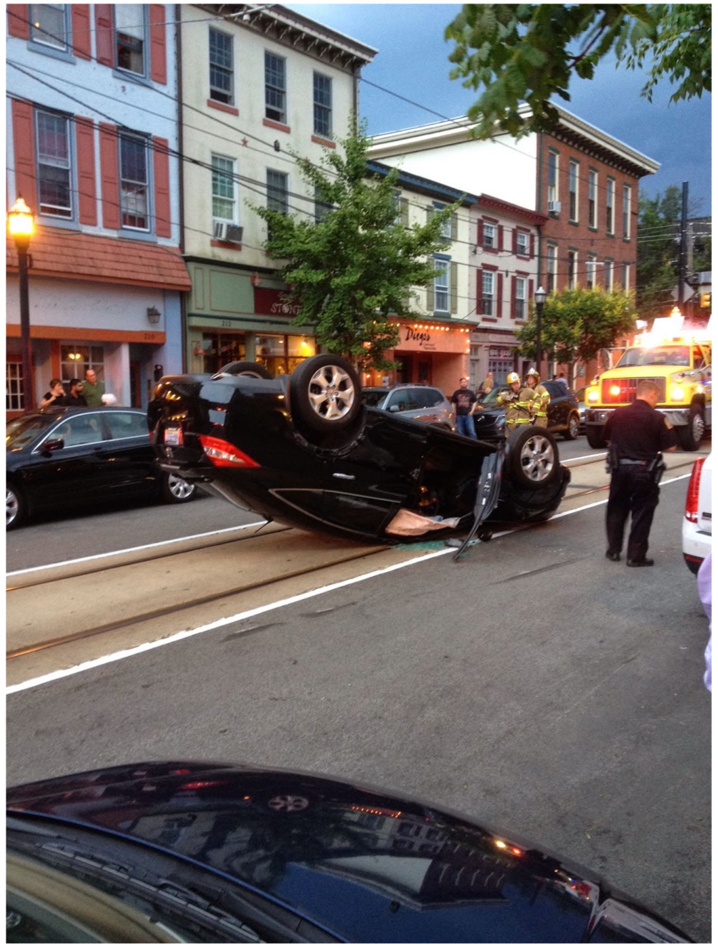
Credit: D. Anguelov, *Ben Taskar Memorial Lecture* (youtube Feb 2023)

We want to leverage the knowledge and reasoning capabilities in LLM/VLM



It's Wednesday at 4 pm. Can I park at this spot right now? Tell me in 1 line.

Yes, you can park for up to 1 hour starting at 4 pm.

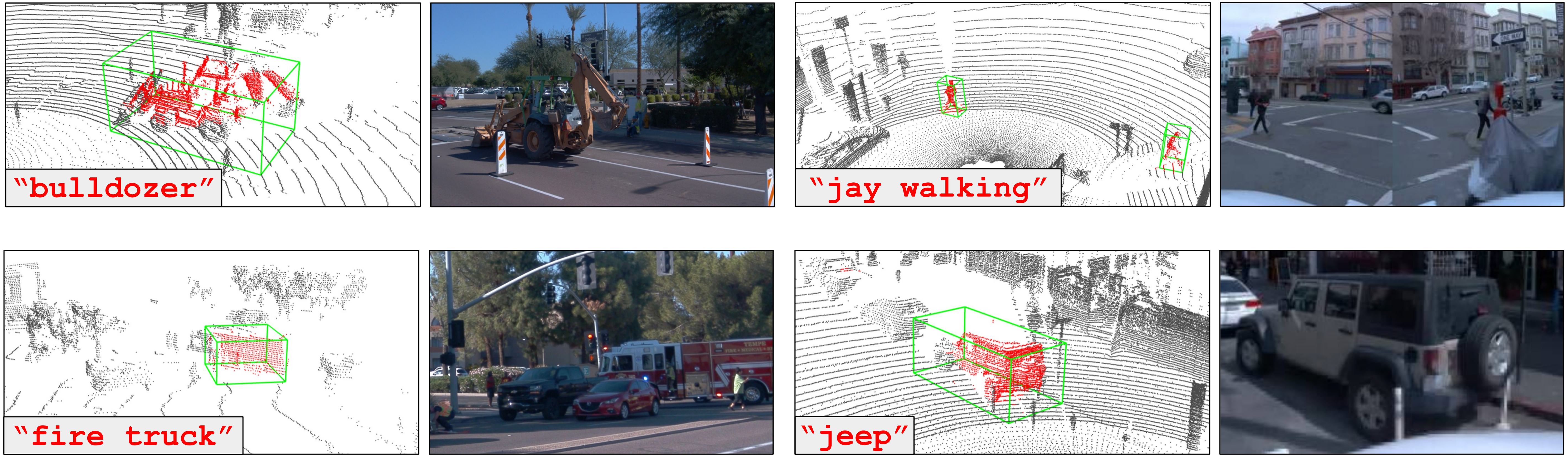


I am driving on the street, this is what I see from the front window. Tell me what's happening and what I should pay special attention to. And tell me what I should do and specify the reasons.

LLM/VLM: You are driving on the street and see a car flipped over on its side. There are firefighters and police at the scene. You should pay special attention to the traffic around you and slow down. You should also be aware of the possibility of further accidents. You should call 911 and report the accident. You should also stay out of the way of the emergency vehicles.

LLM/VLM – Large Language Model / Visual Language Model : LLMs that understand text and images

How can we leverage 2D foundation models to train 3D open-vocabulary models?



Unsupervised 3D Perception with 2D Vision-Language Distillation for Autonomous Driving (UP-VL)

Mahyar Najibi*, Jingwei Ji*, Yin Zhou, Charles R. Qi, Xinchen Yan, Scott Ettinger and Dragomir Anguelov, ICCV’23

Summary

Challenges for lidars in autonomous driving
- High dynamic range scenes
- Weather: window fouling, Mie scattering, impairment estimation

Detectors
- SPADs: PDE 5% => 20% (with path to ~40% at 915nm)
- SiPMs vs SPAD array tradeoffs

Laser diodes
- Broad-area lasers: 60 => 140W (2.3x better with same beam quality)
- VCSEL arrays: steadily improving W/mm^2

Machine Learning and sensor processing
- Sparse point cloud processing with transformers capture nonlocal context in a top-down view
- Large Language Models are inspiring new simulation techniques and helping with rare events

Thank you

waymo.com/careers

waymo.com/open *(dataset)*