Evolution of AI for ADAS/AD Applications

Authors:

Ahmed Sadek, VP Engineering

Tom Toma, Director Product Marketing

Table of Contents

Introduction

What is the best way for automotive developers to introduce artificial intelligence (AI) to their products? Is it by making incremental changes within the constraints of traditional, stack-oriented architectures? Or by moving to larger, end-to-end networks designed from the ground up to grow and scale?

As described in our companion paper, "Steering Towards Heterogeneous Compute for ADAS," automakers and OEMs are deploying generative AI-based assistants. Those assistants run large multi-modal models (LMMs) on system-on-chips (SoCs) designed by Qualcomm Technologies for heterogeneous compute. At the same time, it's becoming clear to developers that traditional, stack-oriented architectures – with heavy, human-defined interfaces and rule-based planning – won't suffice for automated driving at Level 2+ and up. There lie problems such as system fragmentation, complex maintenance and limited scalability.

But the customer functions for automated driving are well suited to AI planner and end-to-end (E2E) networks that take advantage of the general knowledge in foundational models. The transition to E2E networks can significantly enhance system integration and performance, simplify maintenance and allow scalability of automotive systems.

This paper examines the role of AI in advanced driver assistance systems (ADAS) and the integration of AI models with the Qualcomm® AI Hub. Automotive developers can see the advantages of using E2E networks with generative AI and evaluate how they can make the transition in their own companies.

The Potential for AI in ADAS

Sensor fusion, an integral aspect of ADAS, relies on AI algorithms to combine data from multiple sensors, including cameras, radar and lidar. As a result, the ADAS creates a comprehensive and precise representation of the vehicle's surroundings. The system can then make context-informed adaptations based on a rich understanding of the environment.

Of course, the human driver is the biggest beneficiary of the context-informed adaptations capabilities of AI in ADAS. The system analyzes the sensor data and applies predefined rules or machine learning models to respond in real time by adjusting vehicle speed, applying brakes or providing steering assistance. AI algorithms enable the ADAS to respond swiftly and effectively to potential hazards, resulting in enhanced driver confidence.

Looking ahead, AI in ADAS is poised to drive the evolution of automated driving systems. With advanced perception algorithms, route planning capabilities and precise control mechanisms, AI empowers vehicles to navigate and interact with the environment.

Customer Functions

Consider the customer functions for automated driving, as shown in Figure 1:

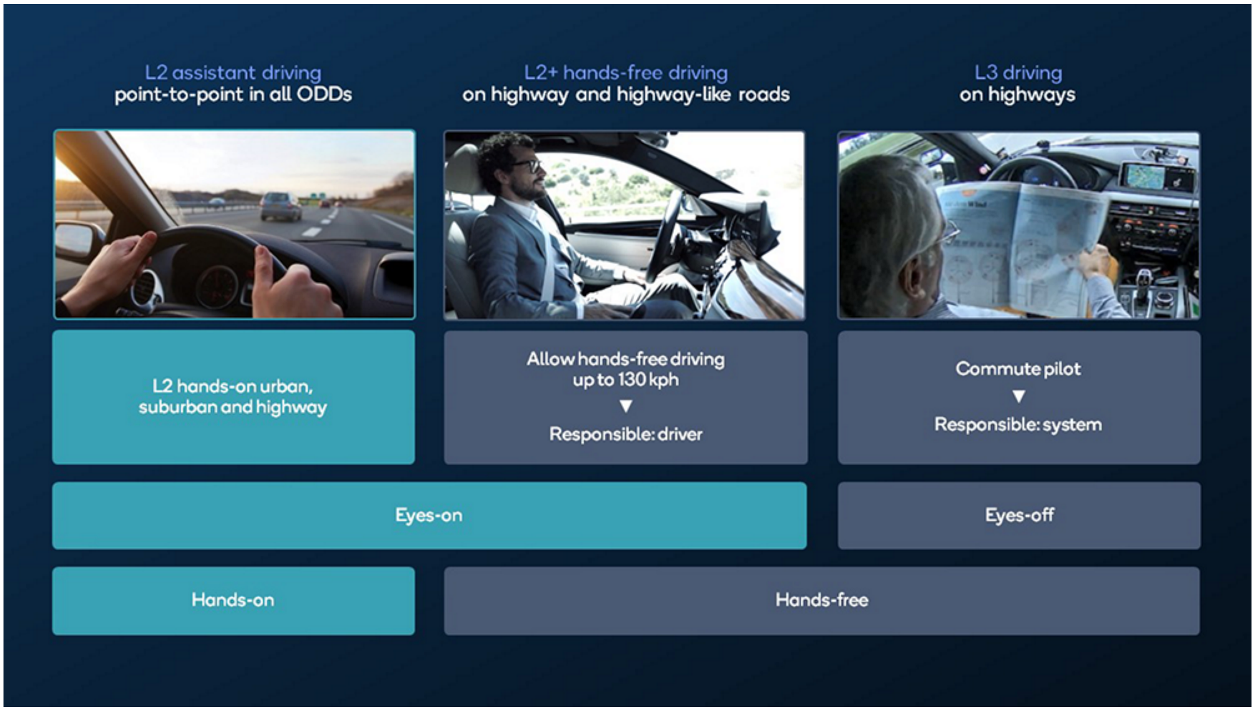

Figure 1: An overview of driving experiences, with emphasis on interaction with the driver

Basic levels: Include emergency braking, lane departure warning and similar event-driven features.

Level 2: Adds comfort functions like adaptive cruise control and requires the driver to keep hands on the steering wheel, supported by multiple radars and cameras.

Level 2+: Introduces partial automation like traffic jam assist and lane change assist, with the driver needing to remain alert.

Level 3: Allows the vehicle to handle certain tasks independently but requires the driver to be ready to intervene.

Levels 4 and 5: Progresses toward high and full automation, where the vehicle can operate autonomously within specific scenarios or entirely on its own.

Parking and non-driving functions: Include advanced parking aids and remote surveillance capabilities to monitor the vehicle's surroundings and protect it from tampering.

Driver and occupant monitoring: Utilizes in-cabin sensors to ensure driver alertness and passenger safety.

OEM functions: Focus on advancing ADAS through research and development, including data collection, additional sensors and event recording for continuous improvement.

Over time, developers will build AI into more of those functions and levels. But they must do so in the context of a system architecture.

Current Approaches to System Architecture for Automated Driving

The traditional architecture for automated driving is modular and stack-oriented. It follows the sequence of perceive-plan-act.

- **Perceive:** The perception block breaks down into camera perception, radar perception, late object level or frame level sensor fusion. The resulting inference is based on an environment model (EM) at a given instant, followed by a prediction block that provides an intention and trajectory prediction for the next several seconds.
- **Plan:** The planning block is then responsible for searching and optimizing a set of drive policies that ensure smooth, collision-free trajectory.
- **Act:** The policies are fed into the motion control block.

The APIs among all the modules are human-defined and specified, as shown in Figure 2.

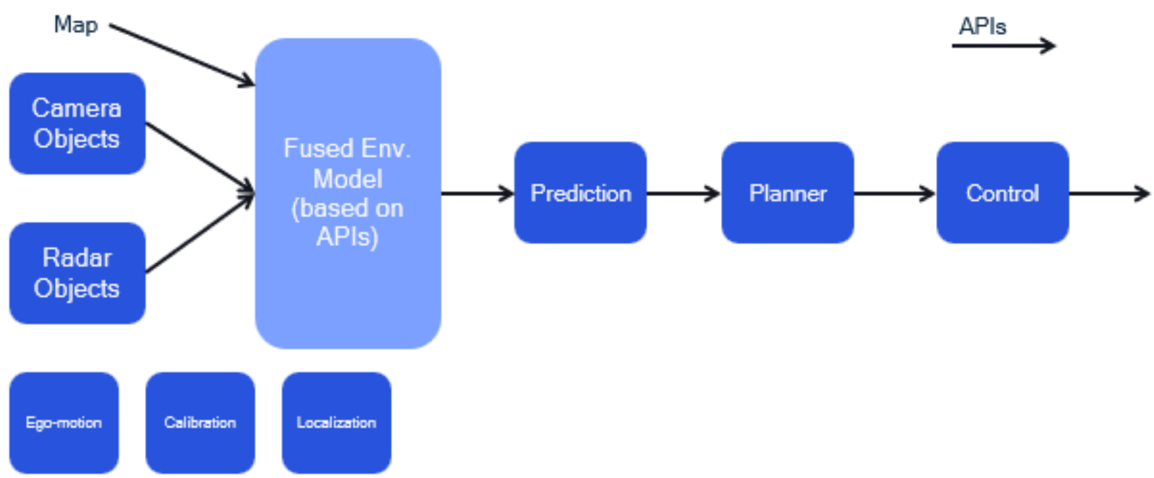

Figure 2: Existing system architecture example

The modular nature of the architecture makes it easy to see how the stack would perform in different scenarios. That, in turn, makes it easier to identify and resolve defects reported from fleet vehicles.

Shortcomings of the Traditional Architecture

But the architecture has downsides.

- The biggest disadvantage of a modular architecture is that per-sensor perception and late object fusion suffer from high latency in validating tracked objects.
- Whether because of size or because of the speed difference between the object and the ego vehicle, the architecture makes it difficult to track objects that span several sensors.
- Late object fusion requires highly accurate association among the objects/detections received from the different sensors. But that association is prone to errors in human-defined algorithms, especially for corner cases (for example, at intersections).

Moreover, the human-defined interface of the traditional architecture is problematic. It risks being over-defined in the effort to include all known scenarios, or being under-defined because some corner cases are not in the data set or are difficult to define. That affects the ability of the ADAS to achieve a human-like driving experience. It also complicates such offline tasks as data annotation, the reprocessing and simulation pipeline, interface documentation and maintenance across different trims and generations.

Figure 3 shows examples of those shortcomings.

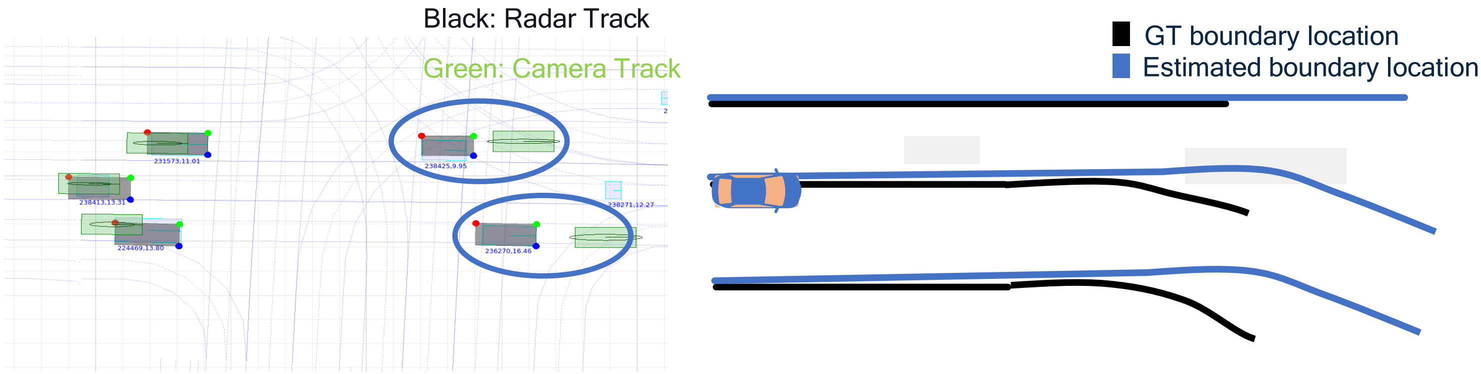

Figure 3: Examples of challenges with existing architectures

Evolution to Hybrid Approaches

To overcome the problems of modular design, hybrid design approaches use AI to consolidate some blocks and replace the human-level interface with AI-defined interfaces (or neural network features).

Modern architectures consolidate modular designs of 10-20 blocks into a few large blocks: notably, a low-level perception (LLP) block, an AI planner and a motion control block as shown in Figure 4.

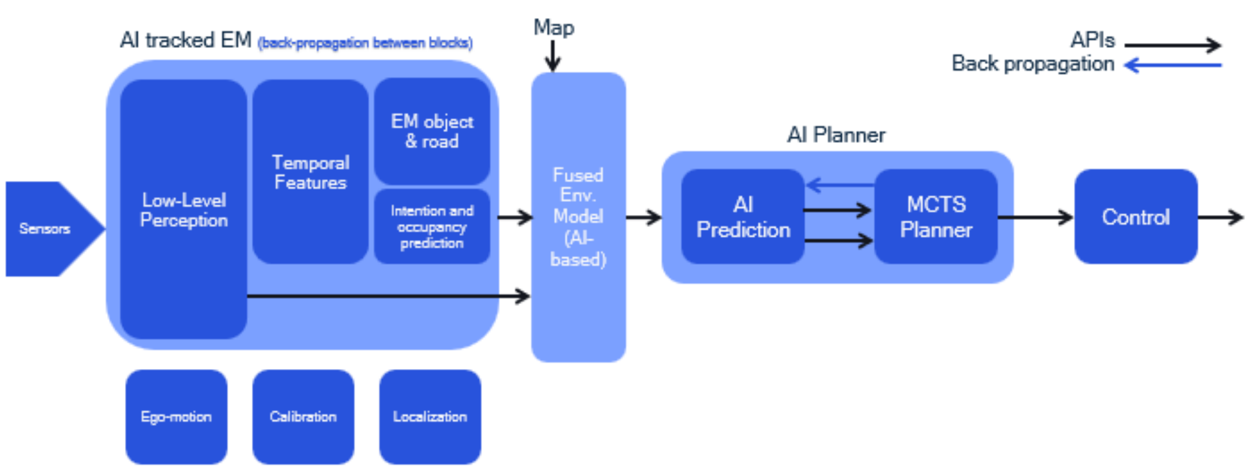

Figure 4: Modern, consolidated architecture with redundancy

The LLP block performs low-level fusion of raw sensor inputs, which reduce the need for human-defined interfaces and association within the perception block.

Low-level perception has different architecture options including bird's-eye view (BEV). BEV is an important baseline that offers a natural way to represent all the sensors in a feature space and fuse them together. BEV is known for high compute and bandwidth requirements, which developers can meet using approaches such as:

- smart grid respesenation that trades off requirements of nearby high accuracy with long-range, lower localization accuracy;
- the use of dense BEV representation near the ego vehicle and direct regression from 2D to 3D features (pull-from-2d approaches); and
- the use of priors to adaptively sample/query the BEV grid.

Compound AI, generative AI and E2E networks

The modern architecture allows developers to leverage transformers and integrate AI in an E2E network, instead of continuing to add AI in slices, one optimization at a time. The larger, E2E approach lets them take advantage of networks designed from the start to scale and grow. In essence, the E2E networks enabled by the modern architecture represent compound AI, integrating driver-specific information into ADAS with the help of generative AI.

That leap forward has benefits not only for developers but also for consumers. It is designed to help improve vehicle safety by personalizing the driver assistance. It can lead to increased driver satisfaction by making the driving experience more intuitive and responsive. Figure 5 depicts the role of E2E AI in automotive development.

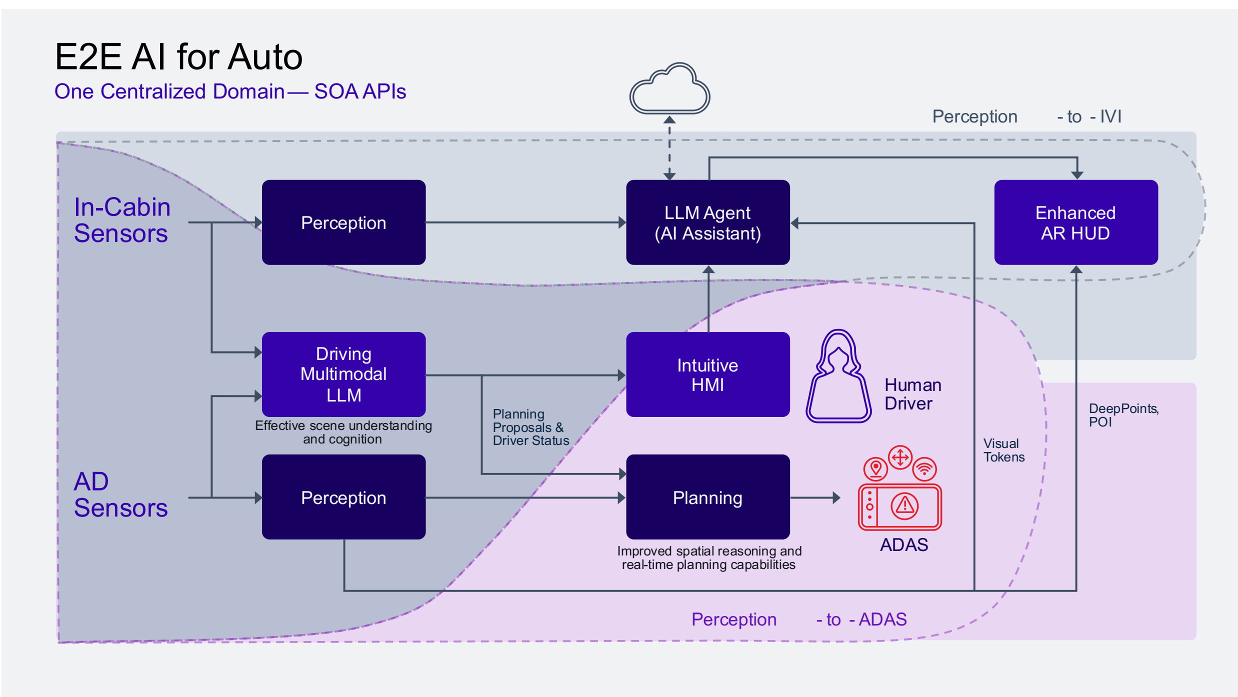

Figure 5: E2E AI for automotive

The upper third of the figure represents the factors inside the car that affect AI; the bottom third, the factors outside the car affecting AI. The middle represents how E2E AI uses internal and external factors as context to provide intuitive information to the driver. (Ultimately, that is supplemented by information from the cloud, but much of it takes place on the device.) There is also the enhanced information provided as information from both internal and external domains is combined.

As compound AI evolves in automotive technology, it will support safety and efficiency in multiple areas:

- **Scenario simulation –** Generative AI can simulate countless driving scenarios and outcomes based on real-time environmental data and historical driver behavior. This capability allows ADAS to make context-informed adaptations about when to intervene and how to alert the driver, enhancing the driving experience.
- **Customized, in-car assistance –** Using natural language processing and voice generation, generative AI can provide dynamic communication with the driver, offering guidance and alerts in a natural, conversational manner that adjusts to the driver's behavior and preferences.
- **Predictive behavior modeling –** Generative AI coupled with large vision models (LVMs) can predict driver behavior based on past patterns, enabling ADAS to anticipate potentially unsafe actions (like sudden lane changes or hard braking) and take preemptive actions to mitigate risks.
- **Enhanced learning capabilities –** As generative AI continues to learn from a wide array of driver interactions and scenarios, it can continuously improve the ADAS algorithms, making them smarter and more attuned to individual drivers over time.
- **Dynamic content creation:** For long drives, generative AI can create personalized content such as audio books or music playlists. The content not only caters to the driver's tastes but also responds to their current level of alertness, helping to keep the driver engaged and alert.

The Qualcomm AI Hub

While compound AI running on E2E networks in modern architectures sets the stage for the next wave of automotive innovation, what about the models? Where will developers get them? How will developers optimize them for in-vehicle chipsets, then test and deploy them?

The Qualcomm AI Hub is a developer-first, online platform designed to simplify on-device AI development for Snapdragon® and Qualcomm® platforms. Developers can use the Qualcomm AI Hub to benchmark and conduct early-stage development of models for automotive applications in areas such as object detection and perception. They can choose from a range of AI models optimized and validated by Qualcomm Technologies, Inc. or bring their own AI models (BYOM). That enables them to work with their preferred frameworks and models, which the Qualcomm AI Hub then compiles and optimizes for Snapdragon and Qualcomm platforms. As a result, allows developers to deploy and benchmark models faster on those platforms.

Against a specific device or chipset, the Qualcomm AI Hub can automatically convert models from Pytorch or ONNX for deployment on TensorFlow Lite, ONNX RT and Qualcomm AI Engine Direct. Even with testing on locally hosted cloud devices, the workflow takes less than five minutes and requires only a few lines of code. AI models optimized through the Qualcomm AI Hub can achieve up to four times faster inferencing performance.

Conclusion

This white paper has explored the significant role that AI plays in automated driving, including how AI algorithms enable vehicles to perceive the environment, plan and act.

Automotive technology is evolving beyond the constraints of traditional, stack-oriented architecture to a modern architecture marked by E2E networks and compound AI. Developers who choose to pivot to E2E networks can enhance system integration and performance, simplify maintenance and upgrades, and facilitate scalability of their products. Without adopting E2E networks in ADAS and the user experience, developers may face increased challenges such as system fragmentation, complex maintenance, and limited scalability.

The transition is an important step for automotive developers. Combined with the Qualcomm AI Hub, the pivot to E2E networks offers them the opportunity to shorten time to commercialization and realize the full potential of on-device AI.