Triple Convergence of AI, Connectivity and Compute in Next-Generation Edge Devices

Copyright © Counterpoint Research
All rights reserved

2026

Executive Summary: The Dawn of the AI-Native Edge

For the last decade, the IoT industry has been stuck in a system assembly loop. We have essentially been duct-taping standalone processors to separate wireless modules and external AI accelerators, creating a fragmented hardware legacy that is starting to show its age. This modular approach served us well in the early days, but it resulted in a rigid architectural dependency on the cloud. The unrelenting tether to distant data centers has become a liability because of prohibitive latency, rising bandwidth costs, and the constant concern of data privacy.

As AI moves into mission-critical and cost-sensitive territory, we are seeing a fundamental shift toward AI-native connected processing solutions. We are finally moving past the era where AI is treated as an expensive guest on the board. Instead, we are seeing next-generation wireless standards and high-performance neural processing co-designed on a single die. This is not just about shrinking components, but a technological inflection point driven by ultra-efficient TinyML models and unified memory architectures.

The Catalyst: Convergence in Action

Leading this transition is the Synaptics® SYN765x, which is essentially the first solution to treat Wi-Fi 7 and Edge AI processing as a single, cohesive unit. By putting a dedicated NPU right next to a tri-band radio, it transforms the wireless link from a passive data pipe into an active sensing layer. We are now at a point where a device can handle human presence detection or acoustic intelligence locally without needing a single camera lens or a round trip to the cloud.

Strategic Value: Beyond the Spec Sheet

From an analyst's perspective, the move to a converged architecture like the SYN765x is not just a win for engineering; it is a massive shift in product economics. Wi-Fi 7 makes it future-proof, as 54% of the global broadband CPE shipments are expected to be on Wi-Fi 7 by 2030[1]:

- **Economic Efficiency:** We are seeing BoM reductions of up to 25% and PCB footprints dropping below 100 mm² with the SYN765x. In high-volume manufacturing, those numbers are the difference between a niche product and a market leader.

- **Autonomous Performance:** Slashed latency of less than 10ms finally makes local, split-second decision-making viable for industrial situations or for retail use cases where zero-lag customer experience is critical.

- **Privacy by Design:** By processing sensitive data locally, OEMs can sidestep the massive compliance hurdles of GDPR and the EU AI Act, as the raw data simply might be restricted from leaving the silicon.

---

[1] Source: Counterpoint Research Wi-Fi Devices services

© Counterpoint Research 2026

- **Development Agility:** The integration tax, which refers to the six to nine months spent fighting firmware compatibility, is effectively eliminated by a unified software stack.

The Path Forward

While modular architectures will likely survive as a niche insurance policy for long-cycle industrial assets that require manual hardware swaps, the integrated MCU has become the definitive choice for anyone looking to scale. Today, the job description for an architect has changed as it is no longer about building a system from parts, but about selecting the right integrated platform that can support the entire lifecycle of a truly intelligent, autonomous edge system.

© Counterpoint Research 2026

Table of Contents

© Counterpoint Research 2026

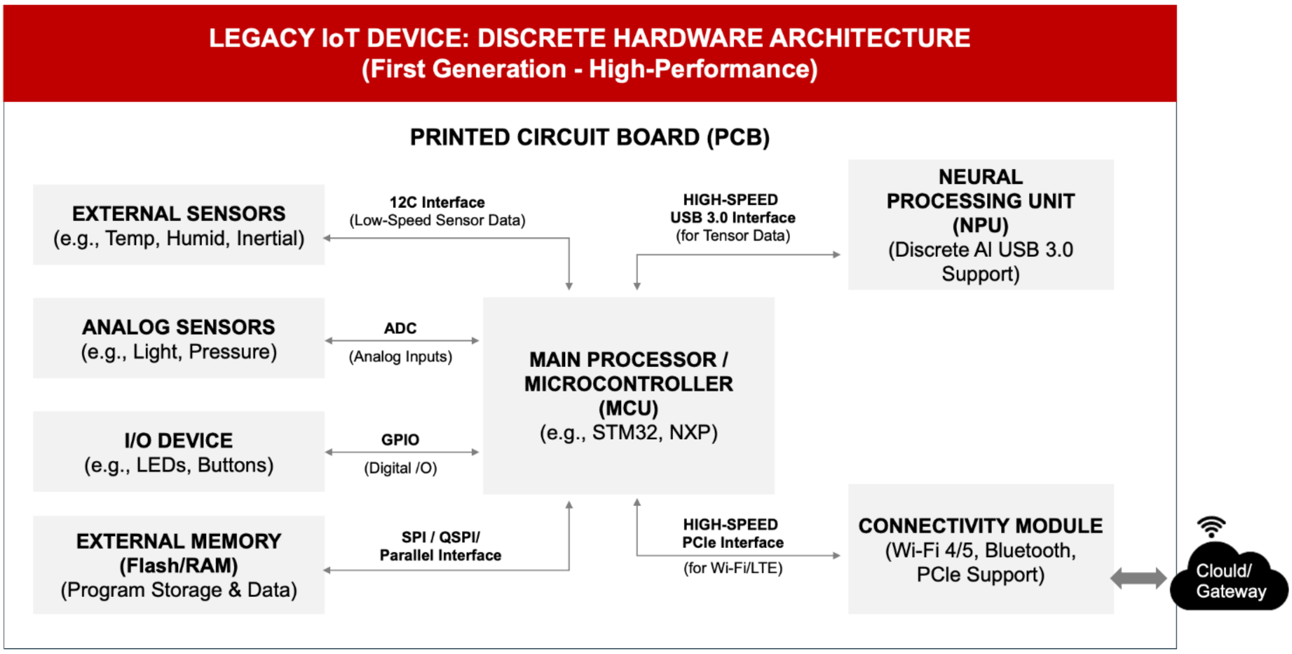
THE PROBLEM

Fragmented Legacy: The High Cost of Discrete Systems

The first generation of IoT architecture was built primarily around modularity, and hence the discrete hardware set-up. These early devices typically combined a standalone microcontroller or a processor, NPU, and connectivity module function as separate physical components on a PCB, typically communicating via standard interfaces like SPI, I2C, or PCIe.

Since these edge devices lacked the necessary local compute power, the architecture relied heavily on distant data centers for data processing and intelligence. This dependency created a **bandwidth bottleneck** that introduced significant network latency, high bandwidth costs, and persistent concerns regarding the security of sensitive data in transit. While this modular approach offered initial flexibility, it introduced systemic inefficiencies and a tethered reliance on connectivity that now hinder the industry's shift toward true autonomy.

A good example of the fragmented architecture is a first-generation smart home camera where the processor was NXP's i.MX series or Texas Instruments' Sitara with an external NPU like Intel Movidius stick, and a separate Wi-Fi module from Expressif Systems. The block diagram below illustrates this generic, multi-vendor legacy architecture.



Source: Counterpoint Research

Complex Component Integration

Complex integration of various components is a significant bottleneck for developers. Original Equipment Manufacturers (OEMs) were responsible for stitching together multiple hardware components from various suppliers. This required managing firmware compatibility across different architectures and ensuring interoperability across a fragmented vendor ecosystem.

In practice, this manual integration for the discrete set-up extended development cycles by six to nine months[2]. Identifying compatible components, architecture design and PCB layout itself

[2] Source: Counterpoint Research OEM Survey

© Counterpoint Research 2026

takes around two months, with prototyping requiring another couple of months. However, the most resource-consuming effort is towards software and driver development, which can take up to six months. Toolchain friction needs to be overcome as using different vendor SDKs (e.g., one for the NPU and another for Wi-Fi) requires significant glue code.

Industry sources indicate that engineering overhead for maintaining these disparate software stacks often consumes 30%-40% of the total research[2] and development budget, diverting resources away from actual product innovation.

Inefficient Power Consumption and Latency Constraints

Power consumption in legacy systems is inherently suboptimal as the data must move across multiple physical traces and interfaces between the processor, the radio, and external memory. Every time a bit of data crosses a chip-to-chip interface, it incurs an energy penalty, i.e. uses additional power.

Latency remains a critical constraint for any application requiring immediate action. Cloud-based inference introduces a round-trip delay that includes data packetization, transmission over a network, cloud processing, and the return command.

Standard cloud round-trips typically range from 50 to 500 milliseconds depending on a number of factors such as model complexity, connectivity protocols, and hardware. This delay due to cloud round-trips is unacceptable for mission-critical environments such as industrial robotics or autonomous safety systems.

Cost Structure Mismatch with Mass Market

Finally, the cost structure of modular designs was fundamentally misaligned with mass-market scaling. Each additional discrete component, such as the AI accelerator, wireless module and the power management IC, adds to the BoM.

- **Hardware Costs:** Individual components are expensive compared to one integrated component. Besides, using multiple discrete chips requires larger PCBs and more complex power delivery networks, which further drive up the costs.
- **Operational Expenses:** Beyond hardware, cloud-dependent models introduce ongoing opex through data ingress and egress fees. Shifting to a hybrid or pure edge model can result in significant cost reductions in opex.
- **Scalability Gap:** While a high BoM might be acceptable for a premium industrial gateway, it is untenable for a high-volume consumer or low-value industrial sensors.

As IoT deployments expanded into cost-sensitive and latency-critical environments, this fragmented architecture became increasingly unsustainable. The industry started to look for integrated solutions as AI use cases started to expand and make way in IoT subsystems. The quest for integrated solutions does not mean every use case would use an integrated chip or hardware, but in a large number of mass-market products, the integrated chips are likely to be the first choice.

© Counterpoint Research 2026

THE SOLUTION

Technology Inflection Point: The Rise of the AI-Native Silicon

A fundamental shift is occurring in how we define the brain of an IoT device. Historically, microcontrollers were simple logic gates, while connectivity and AI were treated as expensive, external additions. We are now entering the era of the **AI-native connected SoC**, a new class of silicon that marks the first time high-performance neural processing and next-generation wireless standards have been co-designed on a single die.

In this new architecture, AI is no longer a guest on the chip, but it is a permanent resident. Unlike legacy MCUs that might struggle to run even basic machine learning tasks, an AI-native MCU features a heterogeneous compute engine specifically built for inference. This includes a dedicated NPU that sits directly alongside the CPU and the radio.

The shift toward convergence is primarily driven by three technological inflections occurring simultaneously. The changes are not just hardware- or silicon-related but also include AI models and memory architecture.

1. **Efficient AI Models**

   The AI models have become significantly more efficient with the use of techniques such as quantization, pruning, and model distillation. These techniques have drastically reduced compute and memory requirements, making on-device inference practical even in the most resource-constrained environments.

   As a result, Edge AI no longer requires a power-hungry Linux gateway; it can now run on a microcontroller for many use cases. For example, the TinyML movement has produced specialized models like TinyYOLO for object detection, or Keyword Spotting (KWS) networks that fit into less than 256 KB of RAM. By using 8-bit or even 4-bit quantization, developers can shrink a neural network by 75% with negligible loss in accuracy, allowing a smart doorbell to distinguish between a person and a passing car entirely on its local processor.

2. **Silicon Purpose-Built for AI**

   At the silicon level, the era of the general-purpose processor is giving way to specialized architecture designed specifically to handle the mathematical rigor of neural networks. For decades, edge devices relied on standard central processing units to manage logic, but these chips were never optimized for the parallel processing required by AI. Today, NPUs and dedicated hardware accelerators are being integrated directly into low-power SoCs, marking a fundamental shift in semiconductor design. This transition is evident in platforms such as Qualcomm's AI-enabled edge SoCs and Synaptics' integrated AI-connectivity platforms, where compute, inference and wireless subsystems are co-designed rather than added as discrete components.

   This industry transition has led to the categorization of AI-native silicon into distinct performance tiers to meet diverse application needs. At the high end of the spectrum, Synaptics' VS680 edge SoC now features multi-core processors paired with integrated NPUs capable of delivering nearly 8 TOPS. This level of localized compute power allows for sophisticated multimodal tasks such as 4K video analytics, real-time object tracking and multi-stream encoding to occur entirely on-device without any cloud intervention.

   At the more resource-constrained end of the market, a new generation of AI-enabled microcontrollers is emerging. Vendors such as STMicroelectronics and Ambiq Micro are

© Counterpoint Research 2026

integrating micro-NPUs and DSP extensions directly into ultra-low-power MCUs, while Espressif Systems combines AI inference capabilities with Wi-Fi and Bluetooth connectivity in a single chip. These chips often utilize advanced processor cores paired with specialized micro-NPUs to provide what the industry calls intelligent gearing. This architectural approach allows a chip to remain in an ultra-low-power, always-on sensing state while monitoring for specific triggers, such as a voice command or a structural vibration. Once an event is detected, the silicon can instantly shift into a high-performance inference mode to process the data. By matching the compute power to the immediate environmental demand, these purpose-built chips achieve a level of efficiency that makes sophisticated AI practical for even the smallest battery-powered sensors.

3. Unified Design and Optimized Memory

In parallel with silicon specialization, advancements in memory architecture have pushed the boundaries of performance-per-watt to new heights. The edge designs are increasingly moving away from traditional architectures where data travels long distances between the processor and external memory. Instead, the industry is shifting toward Compute-in-Memory and tightly coupled SRAM designs that integrate memory directly into the silicon fabric. This is a critical evolution because in Edge AI, the energy cost of moving data often exceeds the energy required for the actual computation. By keeping data local to the processing cores, these unified designs slash latency and power consumption, enabling complex neural networks to run on devices that previously lacked the thermal or battery capacity to support them.

Furthermore, a maturing software ecosystem has emerged to support these edge-native deployments, effectively bridging the historical gap between high-level data science and low-level embedded engineering. The development platforms and automated toolchains now allow engineering teams to convert pre-trained models from standard frameworks like TensorFlow Lite or ONNX into highly optimized C code in minutes. This streamlined workflow enables ready-to-use AI deployment, where sophisticated models can be ported to resource-constrained hardware without requiring manual, assembly-level optimization.

The convergence of these architectural and software factors marks a technological inflection point for the industry. Edge devices are no longer mere data collectors that serve as passive conduits to the cloud; they have evolved into intelligent, autonomous systems capable of understanding their environment and taking immediate action. This shift represents the transition from connected things to truly intelligent systems that provide value at the point of interaction or data generation.

Vendor Spotlight: Synaptics® SYN765x

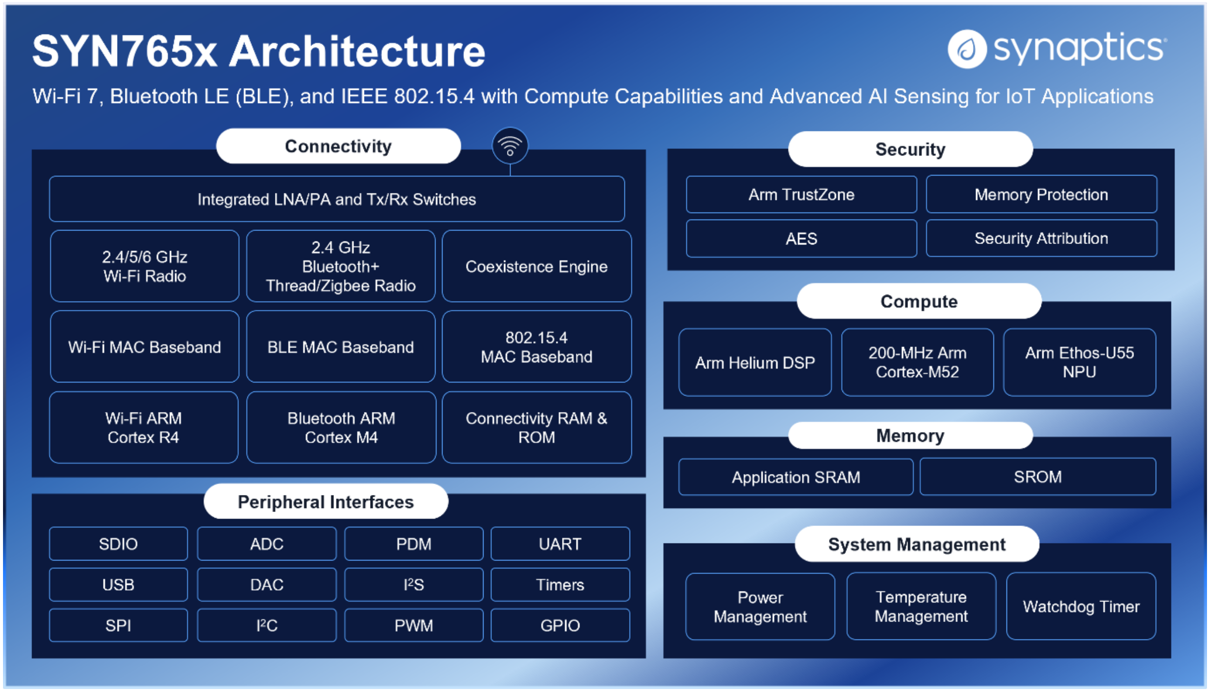
Synaptics has established a clear lead in this category with the **SYN765x**, the industry's first AI-native Wi-Fi 7 connectivity and compute-ready SoC. This SoC represents a radical departure from traditional triple-combo chips. It integrates tri-band Wi-Fi 7, Bluetooth 6.0 and 802.15.4 (Matter/Thread) with an Arm Cortex-M52 core and an Arm Ethos-U55 NPU.

What makes the SYN765x a category-defining product is its ability to operate in a standalone configuration. In many designs, this single chip can replace the need for an external application processor entirely. It leverages its integrated AI to transform standard wireless signals into high-fidelity sensors:

- **AI-Native Wi-Fi Sensing:** The SYN765x uses on-chip machine learning to analyze Channel State Information (CSI), enabling it to detect presence and track motion without a camera.

© Counterpoint Research 2026

- **Bluetooth 6.0 Channel Sounding:** It provides centimeter-level distance measurements, offering a low-power alternative to expensive mmWave radar or ultra-wideband (UWB) technologies.
- **On-Device Audio Intelligence:** The integrated NPU handles sound event detection and keyword spotting locally, ensuring that private conversations never leave the device.



© Counterpoint Research 2026

THE TANGIBLE VALUE

Benefits of Convergence

The move to a converged architecture delivers substantial advantages across the entire product lifecycle:

- **Cost Efficiency and Market Scale**

  Integration drastically reduces the physical complexity of the device. By utilizing a single-chip solution like the **Synaptics® SYN765x**, OEMs can reduce their Bill of Materials (BoM) by up to 25%. These savings extend beyond the silicon to smaller PCB footprints and simplified power delivery networks further lower manufacturing costs, making sophisticated AI viable for high-volume consumer and low-value industrial sensors that were previously price-prohibitive.

- **Power Optimization and Always-on Intelligence**

  Integrated designs generally exhibit lower power consumption due to reduced interconnect losses and optimized power management across all functional blocks on a single substrate. The power consumption difference between discrete and integrated systems is substantial, with integrated solutions typically offering 15% to 40% better power efficiency than discrete multi-chip implementations. In always-on use cases like smart thermostats or industrial sensors, this allows for sophisticated AI processing within a milliwatt power budget. For battery-powered devices, this represents a massive efficiency gain.

- **Real-Time Responsiveness**

  Latency is minimized because inference occurs inches away from the data source. By moving processing to the integrated edge, latency is slashed to 5-20 milliseconds, providing the near-instantaneous responsiveness required for real-time control and safety-critical operations. For example, using Wi-Fi sensing on an integrated SoC allows a device to detect a person's presence and trigger a light or an alarm in milliseconds, without the 200 to 500 milliseconds delay inherent in cloud processing.

- **Compact Form Factors**

  Higher integration enables smaller, more aesthetically pleasing designs. This is critical for consumer wearables and sleek home automation devices where board space is at a premium.

- **Privacy by Design and Regulatory Compliance**

  Privacy has become a primary driver for edge-based intelligence. Processing sensitive data locally, whether it is biometric markers for a door lock or patient vitals in a hospital, ensures that raw, personal information never leaves the physical device.

  This '**Privacy by Design'** approach aligns with increasingly strict global regulations, such as the EU AI Act and GDPR. For example, a smart healthcare monitor can now analyze a patient's heart rate and only transmit a high-level emergency alert to the cloud, keeping the raw medical data encrypted and secure on the device itself. This minimizes the risk of large-scale data breaches and simplifies the compliance burden for manufacturers.

© Counterpoint Research 2026

Difference Between Cloud-First IoT and AI-Native Edge IoT

Metric	Cloud-First IoT (Legacy)	SYN765x AI-Native Edge	Why it Matters
Latency	$200ms - 2s+$ (Network dependent)	<10ms (Real-time)	Essential for Automotive Safety (detecting breathing/movement instantly).
Power Profile	High: Constant radio transmission to stream raw data.	Ultra-Low: Radio stays dormant; NPU processes data at mW levels.	Enables Retail Sensors to run for years on a single coin-cell battery.
Bandwidth	Saturated: Continuous streaming of raw audio/Wi-Fi packets.	Minimal: Only sends a few bytes when an event is detected.	Reduces infrastructure costs for Smart Buildings with 1,000+ nodes.
Data Privacy	Vulnerable: Raw data (audio/ sensing) is transmitted and stored.	Secure-by-Design: Raw data is purged after on-chip inference.	Guarantees GDPR/CCPA Compliance for retail and public spaces.
Reliability	Network Dependent: Fails if Wi-Fi/ Cellular signal drops.	Autonomous: Sensing and alerts work even during network outages.	Critical for Structural Fault Detection in remote or shielded areas.

Connectivity as an Intelligent Layer

In traditional IoT architecture, connectivity was treated strictly as a transport layer. Its sole responsibility was to move data packets from the device to the cloud, acting as a passive pipe. However, in the converged model, connectivity becomes an active participant in the intelligence stack, evolving from a simple communication tool into a sophisticated sensor and data filter. The radio is now the sensor!

From Communication to Contextual Sensing

Latest wireless technologies, particularly Wi-Fi 7 and Bluetooth 6.0, are being utilized for environmental awareness far beyond data transmission. By analyzing Channel State Information (the subtle ways a wireless signal reflects off objects and people), integrated chips can now “see” without cameras.

The Synaptics® SYN765x is a primary example of this shift. Because it houses an NPU directly alongside its Wi-Fi 7 radio, it can process these signal variations in real time to detect human presence, count occupants in a room, or even monitor the breathing patterns of a patient. Similarly, the introduction of Bluetooth 6.0 Channel Sounding allows for centimeter-level distance measurements. This enables secure, proximity-based access control and high-

© Counterpoint Research 2026

precision asset tracking that is resistant to signal spoofing attacks, all without the need for expensive, dedicated radar hardware.

Role of Intelligent Data Filtering

Connectivity also acts as the gatekeeper for the cloud, serving as a critical data filter. In the old model, devices streamed raw, noisy data over the network, leading to high bandwidth costs and cloud bloat.

In a converged architecture, the edge device uses its local intelligence to preprocess data and selectively transmit only meaningful insights. For instance, a smart industrial vibration sensor doesn't need to report normal status every second. Instead, it only activates the high-bandwidth radio link when its local AI identifies a specific harmonic pattern that signals an impending bearing failure. This approach reduces cellular or Wi-Fi data usage by over 90%, significantly lowering operational costs and extending the battery life of remote installations.

Orchestration and the Software-Defined Edge

Integrated connectivity is the backbone of device orchestration and lifecycle management. Since the radio and the AI-accelerated processor are co-designed on a single platform like Synaptics Astra, over-the-air (OTA) updates become seamless and secure.

This allows the hardware to be software-defined. A manufacturer can deploy a fleet of generic sensors today and, via an OTA update, push a new AI model that turns those sensors into specialized glass-break detectors or Wi-Fi-based motion trackers. Connectivity is no longer just a way to send a message, but it is the intelligent interface that allows a device to adapt to its environment and evolve its capabilities long after it has been installed.

Simplifying the Development Journey

The most significant advantage of this convergence is how it fundamentally redefines the development lifecycle. In the legacy system assembly model, engineering teams often spent the first six to nine months of a project just solving low-level hardware friction. This included debugging complex timing issues between a standalone processor and a separate radio, or manually optimizing drivers to ensure an external NPU could communicate with the rest of the system without crashing. These hidden costs, often called the **integration tax**, frequently drained R&D budgets before a single line of application-specific code was even written.

By moving to a converged platform supported by a unified software stack, developers can effectively bypass these foundational hurdles. The development environments now provide high-level abstractions that treat the NPU, CPU and connectivity as a single entity. This allows teams to move straight to application-level innovation. Instead of writing low-level firmware for data movement, they can use pre-optimized frameworks to deploy sophisticated AI models such as those for acoustic fingerprinting in industrial maintenance or Wi-Fi-based presence detection in smart buildings.

This transition marks a critical shift from system assembly to platform adoption. In the assembly era, a company's value was often tied to its ability to make disparate parts work together, while in the platform era, that plumbing is handled by the silicon and its accompanying software stack. This empowers companies to refocus their best engineering talent on the user experience and the unique, brand-defining features that actually move the needle in the market. Ultimately, it allows for a fail-fast, innovate-faster approach, where new AI-driven capabilities can be prototyped, tested and deployed at a pace that was previously impossible in the world of embedded hardware.

© Counterpoint Research 2026

Counterpoint

APPLICATIONS & SELECTION OF PATH

Application Heatmap

The impact of convergence is best understood through real-world applications. In smart homes and appliances, integrated platforms enable always-on voice recognition, vision-based interaction, and predictive automation without relying on cloud processing. This improves responsiveness and enhances the overall user experience.

In industrial environments, edge AI enables predictive maintenance and machine vision directly on-site. This reduces downtime, improves operational efficiency and minimizes reliance on centralized infrastructure. Retail and smart spaces benefit from contextual awareness through technologies such as Wi-Fi sensing. This enables occupancy detection, behavior analysis and energy optimization without the need for additional hardware. In healthcare and wearables, local processing supports continuous monitoring and real-time insights while preserving data privacy.

Across these verticals, the common theme is clear – convergence enables intelligence to be embedded at the point of interaction, transforming both functionality and economics. However, beyond these established categories, the integration of AI-native connectivity like the **Synaptics® SYN765x** is unlocking a new class of **Physical AI** applications that move beyond traditional boundaries. Below are some of the applications that can be powered by an AI-native connected MCU like the SYN765x

SYN765x AI-Native Connected MCU Industry Applications & Advantages

Industry	Application	Legacy Method	SYN765X Advantage	Key Technical Value
Smart Retail	**Anonymous Heat Mapping**	Optical Cameras (High bandwidth, Privacy risks)	**Wi-Fi Signal Sensing:** Tracks dwell time and foot traffic via signal disruption.	**Privacy-by-Design:** 100% anonymous; zero image capture; battery-powered edge processing.
Automotive	**Child Presence Detection**	Weight Sensors (Tricked by cargo); Cameras (Blind spots)	**Respiration Monitoring:** Detects micro-movements (breathing) through blankets and seats.	**Life-Sign Logic:** Differentiates living beings from inanimate objects via Wi-Fi sensing.
Infrastructure	**Structural Fault Detection**	Periodic Manual Inspections (Reactive)	**Micro-Vibration Analysis:** Monitors signal shifts caused by material fatigue in beams/walls.	**Continuous** Monitoring: Early warning system for bridges/buildings before cracks are visible.
Industrial	**Acoustic maintenance**	Scheduled Maintenance / Human "Ear"	**Acoustic Fingerprinting:** NPU identifies the specific "hiss" or "grind" of failing parts.	**Predictive Uptime:** Triggers local alerts 48+ hours before mechanical failure occurs.
Healthcare	**Non-Invasive Fall Detection**	Wearables (Pendants/ Watches) or Cameras	**CSI-Based Motion Tracking:** Analyzes Channel State Information (CSI) to detect the distinct "signature" of a fall.	**High Compliance:** Works in private areas (e.g. bathrooms) where cameras are forbidden; no wearable required.
Smart Home	**Secure Sound Event Detection**	Always-on Microphones (Cloud-streaming)	**Local Audio Inference:** NPU identifies window breaks, smoke alarms, or a baby's cry locally.	**Edge Privacy:** Audio never leaves the device; eliminates latency and "always-listening" privacy concerns.

Source: Counterpoint Research

© Counterpoint Research 2026

Counterpoint

In addition to the above use cases, there are many other interesting applications where an MCU like the SYN765x can be used:

1. Non-Invasive Sleep Diagnostics and Wellness

The Challenge: Traditional sleep tracking requires wearable devices that can be uncomfortable or need frequent charging, often leading to inconsistent data.

The SYN765x Solution: By integrating the SYN765x into a smart pillow or bedside lamp, manufacturers can offer a completely **contactless biometric solution**. Utilizing advanced **Wi-Fi sensing**, the chip monitors micro-movements of the chest wall to track respiration rates and restlessness. The integrated **on-chip NPU** processes this raw signal data locally, ensuring user privacy while identifying complex patterns indicative of sleep apnea or restless leg syndrome (RLS) without the need for cloud-based analysis.

2. Precision Livestock Management in Smart Agriculture

The Challenge: Monitoring animal health in sprawling, metal-heavy barn environments is notoriously difficult due to signal interference and the sheer scale of the facilities.

The SYN765x Solution: This system utilizes **Bluetooth 6.0 markers** on cattle paired with SYN765x-enabled hubs. By leveraging **high-accuracy channel sounding**, the hubs can pinpoint cow locations with centimeter-level precision. This allows farmers to identify early signs of illness, such as huddling or decreased movement toward feeding troughs. To bridge the distance in large agricultural complexes, the **Wi-Fi 7 backhaul** provides a robust, high-bandwidth link between hubs at ranges up to 200 m, ensuring a seamless data fabric across the entire operation.

3. Frictionless Retail: Thru-Glass Gesture Interaction

The Challenge: Optical systems like LIDAR or cameras struggle with the glare and reflections of storefront glass, and they often raise significant consumer privacy concerns.

The SYN765x Solution: Retailers can transform standard windows into **interactive digital catalogs** using SYN765x Wi-Fi sensing. Because Wi-Fi signals penetrate glass without the refractive interference that plagues cameras, the system can accurately detect hand gestures and swipes in real time. The **integrated MCU** processes these "air gestures" at the edge, delivering a zero-lag, intuitive browsing experience that allows shoppers to interact with the brand 24/7 without ever stepping inside.

4. Predictive Infrastructure Health: Volumetric Leak Detection

The Challenge: Most water damage is only discovered after it has caused structural rot or mold, as slow leaks behind drywall are invisible to the naked eye.

The SYN765x Solution: By deploying SYN765x nodes in a mesh configuration within a building's framework, property managers can utilize **RF signal attenuation** for preventative maintenance. Since water is a high absorber of **5GHz and 6GHz frequencies**, the system can detect minute signal drop-offs at specific 3D coordinates. An AI-powered hub can then triangulate the moisture buildup, alerting the owner to a slow leak months before a physical stain appears, and effectively shifting home maintenance from reactive repair to **predictive preservation**.

© Counterpoint Research 2026

Choosing the Right Path: Strategic Frameworks for Integration

Transitioning AI to the edge involves a delicate balancing act. Unlike the nearly infinite resources of the cloud, edge devices operate within strict physical limits for power, memory and heat. This has led to the rise of **Intelligent Gearing**, a strategy where developers choose between the highest possible accuracy and the lowest possible power draw depending on the immediate environmental demand.

For instance, a wearable device might use a highly compressed, ultra-low-power model for continuous background monitoring, switching to a robust, high-performance model only when an anomaly is detected. This tiered approach allows for a new generation of autonomous systems that can operate for years on a single battery while remaining capable of sophisticated reasoning.

While the industry is moving toward **Integrated MCUs** as the primary choice for mass-market intelligence to solve these power and cost constraints, certain specialized scenarios still warrant a **Discrete or Modular** strategy. The following framework identifies the ideal use cases for each approach.

Architectural Decision Framework

Integrated MCU (e.g. SYN765x)		Discrete Architecture
Optimization (Cost, Power, Size)	Primary Goal	Flexibility & Customization
High Mass-market scaling, lower unit cost	Production Volume	**Low to Mid** Higher per-unit cost
Ultra-Compact Ideal for smart sensors, smart homes, etc.	Form Factor	**Standard/ Large** Industrial gateways, sleek consumer tech
Ultra-Low Best for battery-operated "Always on"	Power Profile	**High/ Mains** Power efficiency secondary to performance
Standardized Retail, home, common industrial zones	Deployment Environment	**Extreme/ Niche** Harsh environments, specialized RF
Standard (3-10 years) Aligns with consumer/enterprise refresh cycles	Lifecycle Requirement	**Extended (10-20 years)** Allows for individual radio or NPU upgrades
Low Single-point certification, pre-integrated	Regulatory Risk	**High** Multiple components, separate certifications.

Source: Counterpoint Research

These considerations do not negate the benefits of convergence, but they define the new boundaries of IoT design. For the vast majority of consumer, retail and industrial applications, the gains in **BoM, power and latency** far outweigh the flexibility lost from modularity. The goal of the architect is no longer to build a system from parts, but to select the integrated platform that best aligns with the product's intended lifecycle.

© Counterpoint Research 2026

Strategic Implications for the Ecosystem

The transition to a converged architecture isn’t just a hardware upgrade; it is a fundamental shift in the business of building and deploying IoT. For **OEMs**, this evolution significantly simplifies the physical product development process. At the same time, it introduces a much more critical decision at the start of the design cycle. Platform selection is no longer about picking the cheapest microcontroller; it has become a high-stakes strategic priority. The choice of a platform like Synaptics Astra or the SYN765x determines the long-term performance, cost structure and scalability of an entire product line. OEMs must now look for silicon-plus-software ecosystems that allow them to push over-the-air updates and new AI models to devices already in the field, ensuring they don't become obsolete the moment a new AI standard is released.

Network operators are also finding themselves at a crossroads. As connectivity becomes an intelligent layer, the value of providing a simple data pipe is rapidly commoditizing. The real opportunity now lies in moving up the value chain to offer sophisticated services like automated device orchestration, fleet-wide edge analytics, and secure lifecycle management. By leveraging the local processing power of AI-native MCUs, operators can provide intelligence-as-a-service, where they manage the health and security of millions of autonomous nodes rather than just billing for data packets.

Module vendors face the most urgent need to adapt. The days of selling a standalone Wi-Fi or Bluetooth module are numbered as customers demand more comprehensive, ready-to-compute solutions. To remain relevant, these vendors are increasingly integrating AI capabilities directly into their modules, transforming them into mini-systems that can handle vision, voice and sensing out of the box. Across the entire industry, the ability to differentiate is moving up the stack – success no longer depends on who has the best individual components, but on who can deliver the most integrated, intelligent and scalable end-to-end solution.

Looking toward the end of this decade, we are moving into a world of truly autonomous edge systems. We are moving past the era where a device simply detects and reports to an era where it understands and adapts.

The next frontier involves the rise of **ultra-efficient AI models** and techniques like **federated learning**. This will allow a fleet of devices such as smart streetlights or industrial sensors to learn from their local environments and share those insights with each other to improve accuracy, all without ever sending raw, private data back to a central server. This local learning capability will make devices increasingly self-sufficient, allowing them to recalibrate their sensors or optimize their power consumption in real time based on local conditions.

In this future, the role of the **cloud** will undergo a permanent shift. It will no longer be the brain responsible for every micro-decision, but it will evolve into a coordinator and aggregator. The cloud will handle long-term trend analytics, global fleet coordination and massive data storage, while the real-time, mission-critical intelligence resides firmly at the edge. This shift will only further reinforce the necessity of integrated architecture. As we move toward these intelligent, self-sufficient systems, the hardware that powers them must be as converged and capable as the software it runs.

© Counterpoint Research 2026

GLOSSARY OF TERMS

- **Bluetooth:** A short-range wireless technology standard used for exchanging data between fixed and mobile devices. It is a cornerstone for low-power, short-distance IoT connectivity.
- **BoM (Bill of Materials):** A comprehensive list of raw materials, assemblies and components required to manufacture a product. In the IoT industry, it represents the total hardware cost of a device.
- **CSI (Channel State Information):** Information describing how a signal propagates from a transmitter to a receiver. In AIoT, CSI is often used for gesture recognition and occupancy sensing without the need for cameras.
- **GDPR (General Data Protection Regulation):** A legal framework that sets guidelines for the collection and processing of personal information from individuals in the European Union. It is a primary driver for moving data processing to the edge to ensure privacy.
- **MCU (Microcontroller Unit):** A small computer on a single integrated circuit designed for specific, low-power control tasks. Unlike more powerful processors, MCUs typically manage simple input/output functions.
- **MPU (Microprocessor Unit):** A highly integrated circuit that performs the functions of a central processing unit. MPUs are more powerful than MCUs and are capable of running complex operating systems like Linux.
- **NPU (Neural Processing Unit):** A specialized circuit designed to accelerate machine learning algorithms by performing massive parallel mathematical operations. It allows edge devices to handle complex AI tasks locally and efficiently.
- **OEM (Original Equipment Manufacturer):** A company that produces parts and equipment that may be marketed or integrated by another manufacturer. In this context, OEMs are the primary builders of hardware devices.
- **ONNX (Open Neural Network Exchange):** An open-source format for AI models that allows them to be transferred between different software frameworks and hardware accelerators. It provides a common language for deploying deep learning models across diverse hardware.
- **OTA (Over-The-Air):** A method of delivering new software, firmware or configuration settings to devices wirelessly. This is critical for maintaining security and updating functionality in deployed IoT systems.
- **PCB (Printed Circuit Board):** A structural board used to mechanically support and electrically connect electronic components using conductive pathways. It is the physical foundation of any electronic device.
- **SDK (Software Development Kit):** A collection of software tools and libraries used by developers to create applications for specific hardware platforms. It bridges the gap between hardware capabilities and the final software application.
- **SoC (System on a Chip):** An integrated circuit that integrates all necessary components of a computer or electronic system into a single chip. In AIoT, SoCs

© Counterpoint Research 2026

combine processing, AI acceleration and connectivity to reduce device size and power consumption.

- **SRAM (Static Random-Access Memory):** A type of fast, power-efficient semiconductor memory used to store data temporarily. It is often used as a high-speed internal cache for AI processing on edge devices.
- **TensorFlow:** An end-to-end open-source platform for machine learning developed by Google. It includes specialized versions, such as TensorFlow Lite, designed for deployment on resource-constrained edge devices.
- **TinyML (Tiny Machine Learning):** A field of study in machine learning and embedded systems that enables AI models to run on extremely low-power hardware. It focuses on bringing intelligence to the smallest devices at the very edge of the network.
- **TinyYOLO (Tiny You Only Look Once):** A simplified, highly efficient version of the YOLO real-time object detection system. It is optimized to perform rapid visual recognition on devices with limited processing power.
- **TOPS (Tera Operations Per Second):** A metric used to measure the performance of AI accelerators and NPUs. It indicates how many trillions of mathematical operations a chip can perform in one second.
- **UWB (Ultra-Wideband):** A short-range, low-power radio technology that uses high bandwidth for precise location tracking and spatial awareness. It is being increasingly used in IoT for secure access control and high-accuracy positioning.
- **WiFi (Wireless Fidelity):** A family of wireless network protocols based on the IEEE 802.11 standards. It provides the high-speed local area networking necessary for most IoT devices to connect to the internet.

© Counterpoint Research 2026

Authors, Copyright, User Agreement and Other General Information

Mohit Agrawal
Research Director
mohit.agrawal@counterpointresearch.com

COUNTERPOINT TECHNOLOGY MARKET RESEARCH

Hong Kong | USA | South Korea | India | UK | Argentina | China | Taiwan | Japan

info@counterpointresearch.com

©2026 Counterpoint Technology Market Research. This research report is prepared for the exclusive use of Counterpoint Technology Market Research clients and may not be reproduced in whole or in part or in any form or manner to others outside your organization without the express prior written consent of Counterpoint Technology Market Research. Receipt and/or review of this document constitutes your agreement not to reproduce, display, modify, distribute, transmit or disclose to others outside your organization the contents, opinions, conclusions or information contained in the report. All trademarks displayed in this report are owned opinions, conclusions or information contained in the report. All trademarks displayed in this report are owned by Counterpoint Technology Market Research and may not be used without prior written consent.

COUNTERPOINT RESEARCH

Beijing | Boston | Hong Kong | New Delhi | London | Seoul | Shanghai | Shenzhen | Taipei | Tokyo

info@counterpointresearch.com

Contact Us:

www.counterpointresearch.com

info@counterpointresearch.com

@CounterPointTR

@Counterpoint Research

Counterpoint