

Accelerating Edge AI Development in Real-Time Control Apps

Sponsored by Texas Instruments: NPU-equipped MCUs open the door to optimized edge AI in systems ranging from wearable health monitors to physical AI in humanoid robots.

Much of the electronics world today obsesses over artificial intelligence (AI) in big system-on-chips (SoCs) and complex, multichip designs. But what may be overlooked are far-reaching applications that can be enabled inside smaller chips like microcontrollers, making both industrial devices and consumer devices more intelligent and more efficient.

The reason? Primarily response times, power consumption, performance, development complexity, memory footprint, and the cost to transform data into the real-time decisions required with AI capabilities. Fortunately, there are strong indicators that's all changing.

Consider Texas Instruments' MCUs with integrated TinyEngine neural processing units (NPUs). These dedicated hardware accelerators optimize deep-learning inference operations to reduce latency and improve energy efficiency when processing at the AI edge. With this innovation, engineers are now able to deploy intelligence almost anywhere.

The idea behind the TinyEngine NPU is to execute computations required by neural networks in parallel to the primary CPU running application code. It addresses key design constraints that have prevented widespread adoption of embedded AI by utilizing 120X less energy per inference and 90X lower latency compared to software-based AI.

By executing machine-learning algorithms in parallel to the primary CPU,

real-time processing of neural-network models can take place on resource-constrained devices. This optimizes deep-learning inference latency and power consumption when processing at the edge, eliminating the round-trip latency of cloud-based inferencing for increased system responsiveness.

TI plans to integrate the TinyEngine NPU across its entire microcontroller portfolio (Fig. 1). It will lead to expansion of edge AI capabilities into devices such as portable battery-powered products, medical wearables, as well as personal electronics and industrial equipment that were previously unable to support meaningful AI workloads.



1. TI is integrating the TinyEngine NPU across its entire microcontroller portfolio, including general-purpose and high-performance real-time MCUs. The NPU is a specialized hardware accelerator designed to perform parallel machine-learning inference alongside the main Arm Cortex-M CPU, offering high performance in energy-constrained and real-time applications (Source: TI)

Transforming Real-Time Control Applications

The need for real-time monitoring and control has become essential to help minimize downtime, reduce energy consumption, and improve overall reliability in motor systems for industrial automation applications. Machine monitoring helps avoid downtime scenarios by providing real-time insights. To do so, motor-control applications in appliances, robotics, and industrial systems increasingly call for intelligent features like adaptive control and predictive maintenance.

Meeting this growing demand for intelligent real-time control applications, TI's AM13Ex MCUs with integrated TinyEngine NPUs add advanced capabilities without compromising performance. They enable designers to implement sophisticated motor control and AI features simultaneously without external components, lowering bill-of-materials costs by up to 30%.

They also facilitate the ability to maintain precise real-time control loops for up to four motors while [delivering AI processing 20x faster than software](#) — a vital advantage when designing appliances, drones, humanoid robotics, or HVAC applications.

In addition, TI's real-time-control-optimized MCUs offer high-performance analog, control, and digital peripheral integration, support ambient temperature ranges from -40 to 105°C , and operate with a 3.3-V supply voltage.

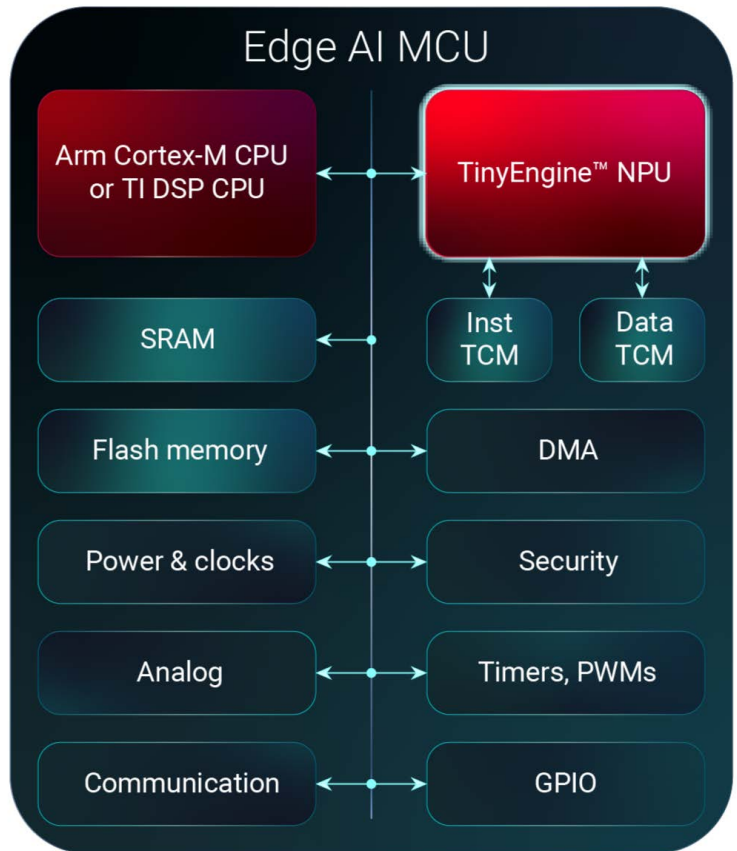
[The AM13E230x MCU, for example, enables predictive fault detection and adaptive control algorithms in real-time control applications](#) by combining an Arm Cortex-M33 CPU with a TI TinyEngine NPU in a single device (Fig. 2). The MCU supports such features as predictive fault detection, adaptive control algorithms, anomaly detection and intelligent load balancing, where cost, size and power considerations have traditionally limited the use of edge AI.

Advanced Intelligence at Your Fingertips

It's not easy to create a new tool that's dramatically different from its predecessors and does the same job better. Addressing that issue, designers can get started on their edge AI designs quickly with TI's royalty-free [CCStudio Edge AI Studio](#), a collection of graphical and command line tools designed to accelerate edge AI development.

With the CCStudio Edge AI Studio development environment, you can go from start to finish using an integrated workflow covering data collection and labeling, feature extraction, neural-network model selection and tuning, and model compilation and deployment to target hardware.

Edge AI Studio supports AI-accelerated devices such as processors featuring a C7 NPU or microcontrollers integrat-



2. Shown is a simplified block diagram of a TI edge AI MCU with an integrated TinyEngine NPU. (Source: TI)

ing a TinyEngine, as well as AI-supported devices without a dedicated accelerator. The platform offers 60+ code examples and application-specific reference designs (arc fault detection, motor fault prediction).

It also supports industry-standard frameworks like PyTorch, an open-source machine-learning library widely used for developing and training neural networks. Trained models are automatically converted into optimized software libraries without manual coding.

Conclusion

TI's TinyEngine NPU can unlock edge AI acceleration in a vast number of embedded systems. It thus helps accelerate the adoption of edge AI in any electronic device, from real-time monitoring in wearable health monitors and home circuit breakers to physical AI in humanoid robots.

Priced under US\$1 in 1,000-unit quantities, MCUs such as the MSPM0G5187 can reduce system and operating costs by offering an affordable alternative to other processor architectures.