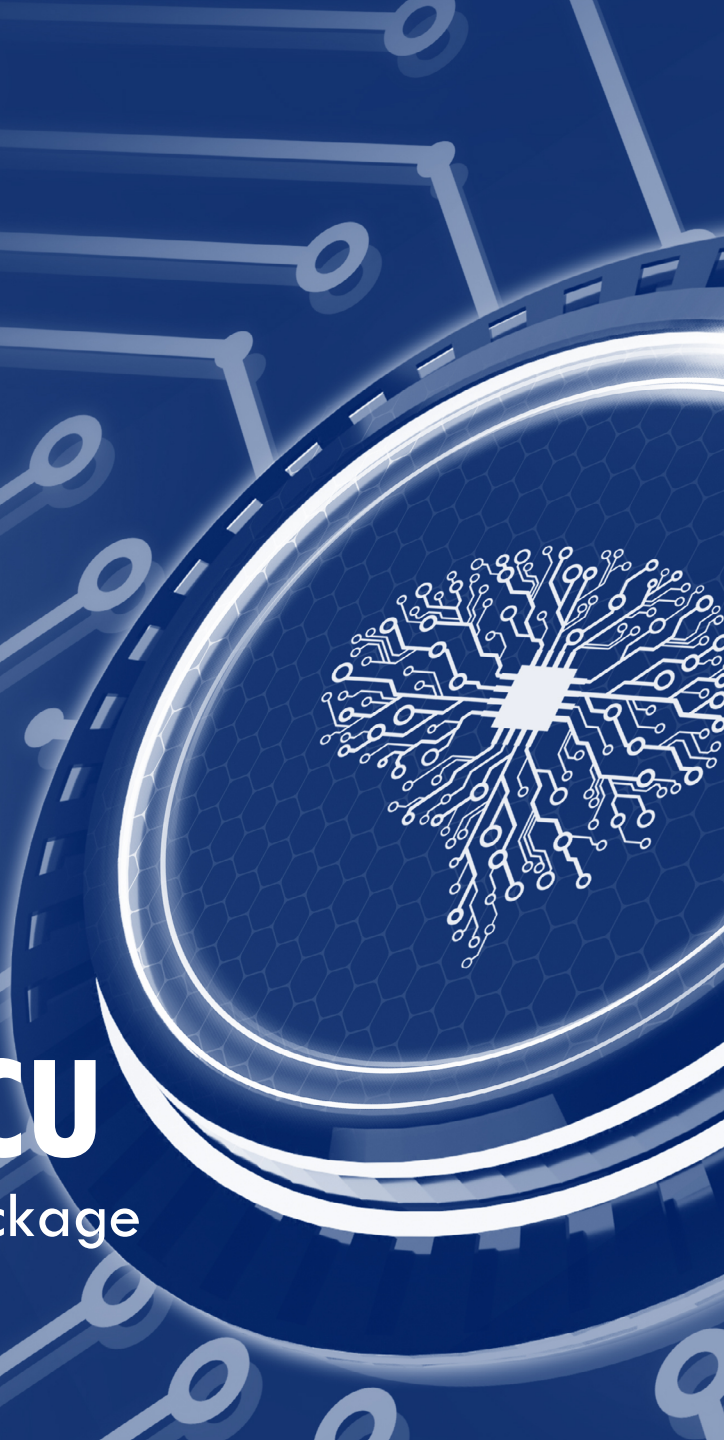
ALIF SEMICONDUCTOR



Technical Dive Into the Alif E1C MCU

Enhanced AI/ML capability in a small package

Alif Semiconductor has released their latest high-capability, small form factor, embedded microcontroller unit (MCU). The E1C, part of the Ensemble® product line, is designed to leverage artificial intelligence (AI) and machine learning (ML) processes, making it suitable for performing advanced adaptive control decisions, yet small and energy-efficient enough to serve the wearable electronics market.

Image credit: Adobe Stocks

Background on AI/ML

Localized AI/ML techniques are currently being used in endpoint devices to optimize processes and deal with large datasets without heavy dependence on the cloud. As sensor technology becomes more cost effective, more data can be collected, and the need for smarter ways to rapidly handle this data becomes essential. Useful for determining everything from detecting and classifying abnormal health markers for an individual to separating out confounding factors, virtually every industry will use AI/ML in the upcoming years.

AI/ML Development and Applications

At the core of AI/ML are pattern recognition and logical decision making capability. What makes these tools useful for the modern world is their ability to reduce the noise and isolate the signal; they can find patterns that would be difficult or time consuming for a human to find using statistical methods alone.

While often used interchangeably, there are a few distinct properties that separate AI from ML. In some ways, ML can be thought

of as a subset of AI, but there is more to this definition. Another way to think of the difference between AI and ML is by examining what information is evaluated by each system.

ML is geared towards making predictions and control decisions based on large sets of data. Its primary function is pattern recognition, making it useful for finding correlations between sensor data, image recognition routines, and other complex data operations. ML is about examining big data.

AI can handle big data sets, but it is also about evaluating user input and handling human interfacing. Instead of just hard facts from the dataset, it may also evaluate input selected by human operators. One of the big driving forces for AI development is in Natural Language Processing (NLP) and Large Language Models (LLM) that can make decisions based on the user's native language and manner of speaking. Overall, AI is about examining data from people, be them operators, technicians or members of the general public, and using this information in conjunction with big data sets.

Current Challenges is AI/ ML Programming

In order to perform all of the high-level calculations required by AI/ML, fast processors are required. However, high-speed calculations come with a price; they will consume more power and the battery will discharge faster in mobile or wearable applications, requiring larger and/or more expensive batteries. Furthermore, higher power consumption means more heat is dissipated by the device which requires large heat sinks or expensive active cooling solutions leading to a larger footprint, limiting versatility and potential applications.

Alif Semiconductor E1C

The E1C provides a good compromise between high-speed computing, suitable for AI/ML applications, without the need for bulky space requirements or high power usage. Equipped with an Arm® Cortex®-M55 core and several very low power states, the E1C can provide computing capability while using power efficiently — all packed into a small footprint. Designed for mobile or wearable electronic applications and intelligent Internet of Things (IoT) devices, it can handle complex data analysis and fit on a wristband.

Photo credit: Adobe Stocks

E1C Performance Metrics

At the heart of the E1C is an Arm Cortex-M55 core, with additional AI/ML acceleration through an Arm Ethos™ Neural Processing Unit (NPU) and large on-chip memory, with the capability to expand the memory externally if required.

Photo credit: Adobe Stocks

Arm Cortex -M55 Core

The Arm Cortex-M55 CPU Core is capable of running at speeds of up to 160 MHz while generating 4.4 CoreMarks™ per MHz, making it ideal for fast calculations in embedded systems. Additionally the CPU core makes use of the Helium™ Vector Processing Extension, which is designed to accelerate ML and digital signal processing (DSP) applications, such as audio data analysis or vibrational data analysis. Furthermore, this CPU core also has the ability to handle floating point math, increasing accuracy of calculations.

Arm Ethos NPU

Besides the M55 Core, the E1C also includes an Arm Ethos-U55 NPU. The NPU reduces AI/ML inference times thanks to its ability to process up to 46 Giga Operations Per Second (GOPs). During each clock cycle, the NPU is capable of performing 128 Multiply-Accumulate (MAC) operations in parallel, rather than time-consuming serial operations as most legacy MCU must do. Combining this NPU with the Cortex-M55 Core, the system is up to two orders of magnitude faster than the Cortex-M55 Core alone while resolving inferences for voice recognition and object detection for example.

The Cortex-M55 CPU is typically run with a Real Time Operating System (RTOS), which is essential for deterministic applications. The AI/ML portion of the application is compiled by Arm's Vela compiler that divides the ML workload between the CPU and the NPU, with usually 95% or more of the workload landing on the NPU. This frees up the Cortex -M55 CPU core to either sleep during the ML inference computation to save energy, or the CPU can do other useful work during the inference. The high speed processing means low latency and quick, but well-informed control decisions are possible.

Memory and Memory Management

One time-saving component to the E1C is its smart memory use. There are up to 1.9 MB of high endurance, non-volatile MRAM and also up to 2.0 MB of SRAM embedded on the chip. The SRAM can be accessed immediately by the CPU and NPU, as there are no wait states to slow down operation. External to the E1C, additional SRAM or Flash memory can be added, which is controlled by a high-speed OctalSPI interface that supports HyperBus Protocol execute in place (XIP), enabling code to be run directly on the external memory rather than being loaded into the internal RAM before execution.

Low Power Consumption With aiPM

One of the major advantages of the E1C is its low power consumption made possible through Alif's proprietary Autonomous Intelligent Power Management — aiPM™— technology designed to extend battery life for wireless sensor networks, IoT, mobile and wearable applications.

aiPM makes power savings possible using dynamic control of current flow to match immediate processing demands, and then shut down that current flow as demand drops to any of six different power domains on the chip. Ultimately, this means powering up only the portions of the chip that are needed, when needed. aiPM is executed in hardware so the burden of power management is removed from the software developer. Of course, the software can configure aiPM in advance to behave as desired while running a given use case, and software can override aiPM hardware if needed.

Power Modes

An extension of aiPM's function is to manage system-level power modes that account for the use of processors, peripherals, clocks, power regulation, and optional retention of SRAM data content in the lowest power mode for rapid restart. The five system-level modes are: go, ready, idle, standby and stop.

- Go: In the go mode, the processor and peripherals are powered on and running at any frequency up to their maximum as required. This mode typically uses the most power but with a capable CPU core like E1C has, the work can be done quickly before moving to a lower power mode.

- Ready: In the ready mode, the core is powered but its clock is gated off. The peripherals are available to run autonomously if required. Ready mode enables the CPU to wake and return to go mode very quickly. The CPU can draw as little as 22uA per MHz of operation while running the benchmark CoreMark™ code.

- Idle: When idle, the CPU core is powered off, but the peripherals can still run. The E1C has a programmable Event Controller that enables peripherals to collect and then transfer data between themselves and to/from memory in a specific sequence, all without CPU intervention to save power while acquiring data. Even the DMA controllers can be active in idle mode to streamline this data movement. Idle mode uses less power than ready, but it takes a little longer to reach the go mode when processing is required.

- Standby: In standby, the CPU core and most of the peripherals are powered off, though some can run within the lowest power region of the chip to enable wakeup events from outside signals and internal sources including serial data interfaces, timers, counters, and analog signal comparators. Standby mode creates a great balance between power consumption and having many useful ways to wake and return to go mode. Wake time from standby is just 4 µs.

- Stop: When the stop mode is active, all but a few vital heartbeat systems are powered off. What remains optionally powered on are a real-time counter, an event counter/timer, an analog comparator, and I/O pins, any of which can wake up the CPU core. There's also an option to retain the contents of some or all of the SRAM in stop mode for rapid restarts after waking up. Stop mode has only 700 µA leakage current, meaning the chip can remain in stop mode for a long time and not drain the battery. Even in this extreme low-power condition, it only takes 400 µs for the CPU core to wake and return to go mode.

All the power conditioning and management is handled inside the chip. A single power source may be applied from 1.7V to 4.2V, then internal DC-DC converters and LDOs generate multiple internal voltage rails for operation.

Small Form Factor and Packaging

One of the major advantages of the E1C is its extremely compact form factor. Measuring only 3.9 mm x 3.9 mm for the 90-bump Wafer Level Chip Scale Package (WLCSP90), this chip is capable of being embedded virtually anywhere that sensor data must be collected and processed.

Also available is the 120-bump, Fine Pitch Ball Array package (FPBA120) at 6.6 mm x 6.6 mm that can be routed on a circuit board using just four layers to reduce the total cost and complexity of an end product.

Finally, for the most rugged applications, a 64 pin Thin Quad Flat Package (TQFP64) will help the chip withstand heavy vibration and other industrial challenges.

Connectability

The E1C offers a wide variety of interfacing options with a highly selectable multiplexing system to get the most use of peripherals and pins in small packages. Interfaces include USB-HS, MIPI-DSI display interface, I3C sensor interface, many channels of UARTs, SPIs, I2Cs. Microphones can be connected via PDM or I2S, where I2S can also drive stereo sound output. A low-power camera can connect by a dedicated 8-bit interface. Finally, high-precision analog signals can be sensed and stimulated with SAR and sigma-delta ADCs, DACs, and analog comparators.

Alif Semiconductor will make available the E1C-DK development kit in 4Q2024, enabling users to rapidly get started in their embedded electronics projects.

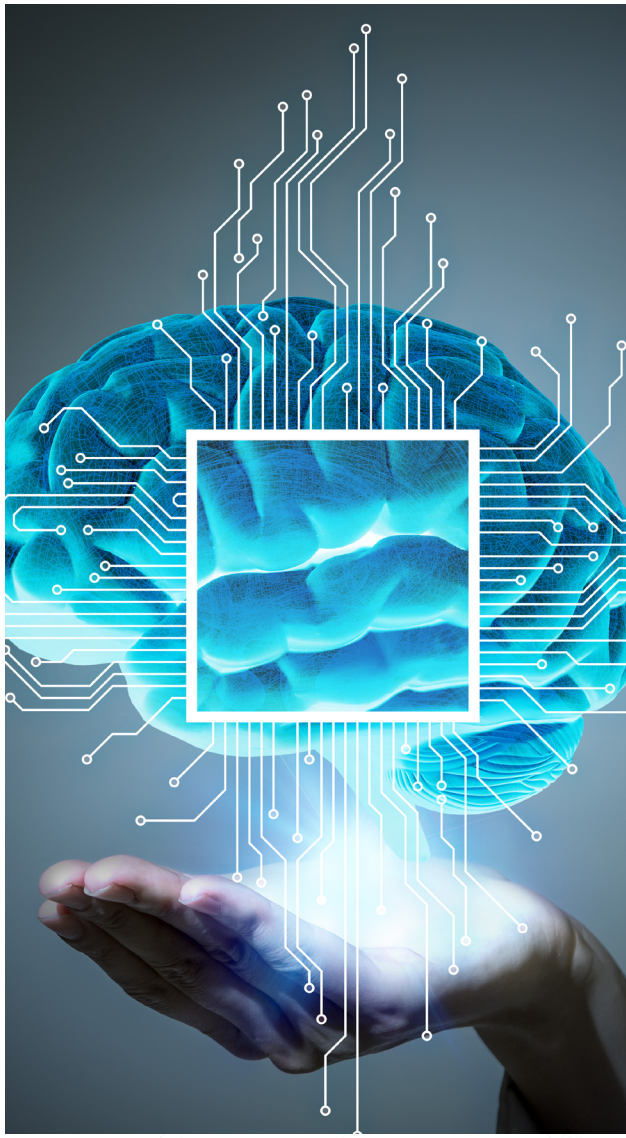


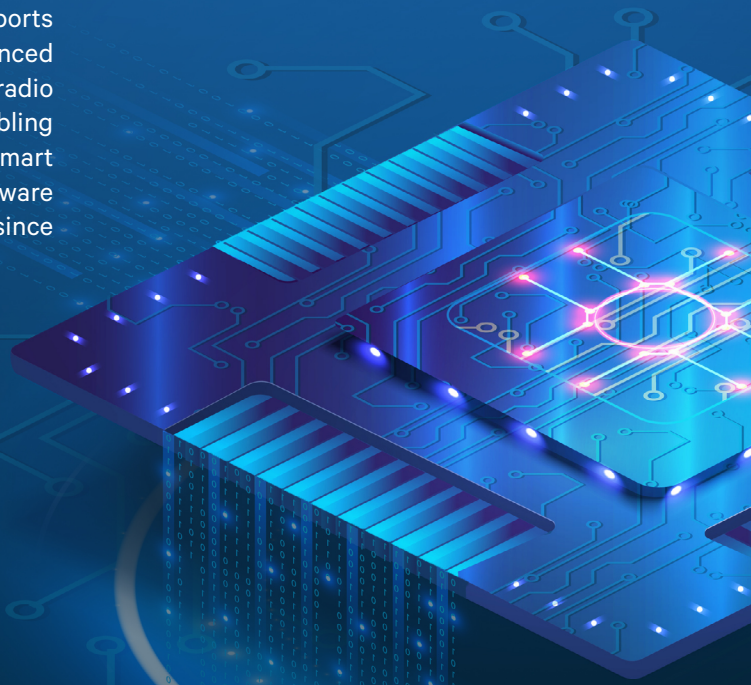
Photo credit: Adobe Stocks

Final Thoughts

While there are many MCUs on the market, the E1C stands out among them due to its extremely efficient AI/ML processing in a tiny footprint with very low power usage compared to other MCUs.

Get started with evaluation of the E1C MCU and then rapidly onto development of your own end product by using Alif Semiconductor's E1C-DK development kit. This kit provides access to all E1C MCU signals, enables power measurement, and includes sensors such as four microphones and an Inertial Measurement Unit for motion capture. There are also means to connect a camera and a display panel. All of this, and a library of software drivers and example code from Alif, enables fast evaluation and development of AI/ML-enabled applications.

For wireless applications, Alif offers the Balletto™ B1 wireless MCU that has the same resources as E1C MCUs but adds a Bluetooth BLE 5.3 radio that supports LE Audio for music and AI/ML-enhanced hearing assistance applications. This radio also supports the 802.15.4 protocol enabling Thead and Matter for smart home and smart building applications. It's an easy software porting job between E1C and B1 MCUs since they share the same architecture.



ALIF SEMICONDUCTOR

For more information on Alif Semiconductor's solutions or application-specific inquiries, please visit our website.