WHITE PAPER

A Vision for Ambient Computing

Enabling ambient computing to reach its full potential

Contents

Overview

Embedding computers in our clothing, walls, and streets has been a vision for decades in the making. Ambiq was founded to make this vision — ambient computing — happen. This white paper discusses the key trends driving ambient computing, including the chip and system-on-a-chip (SoC) technology that underpins it. The broader semiconductor and microcontroller (MCU) markets are reviewed to give a sense of where things stand today. The future of ambient computing is then discussed, focusing on technology trends relating to semiconductor manufacturing, SoC architectures, neural networks, and wearables. This white paper then concludes with a short discussion of Ambiq's roadmap and its alignment with ambient computing.

Market Outlook

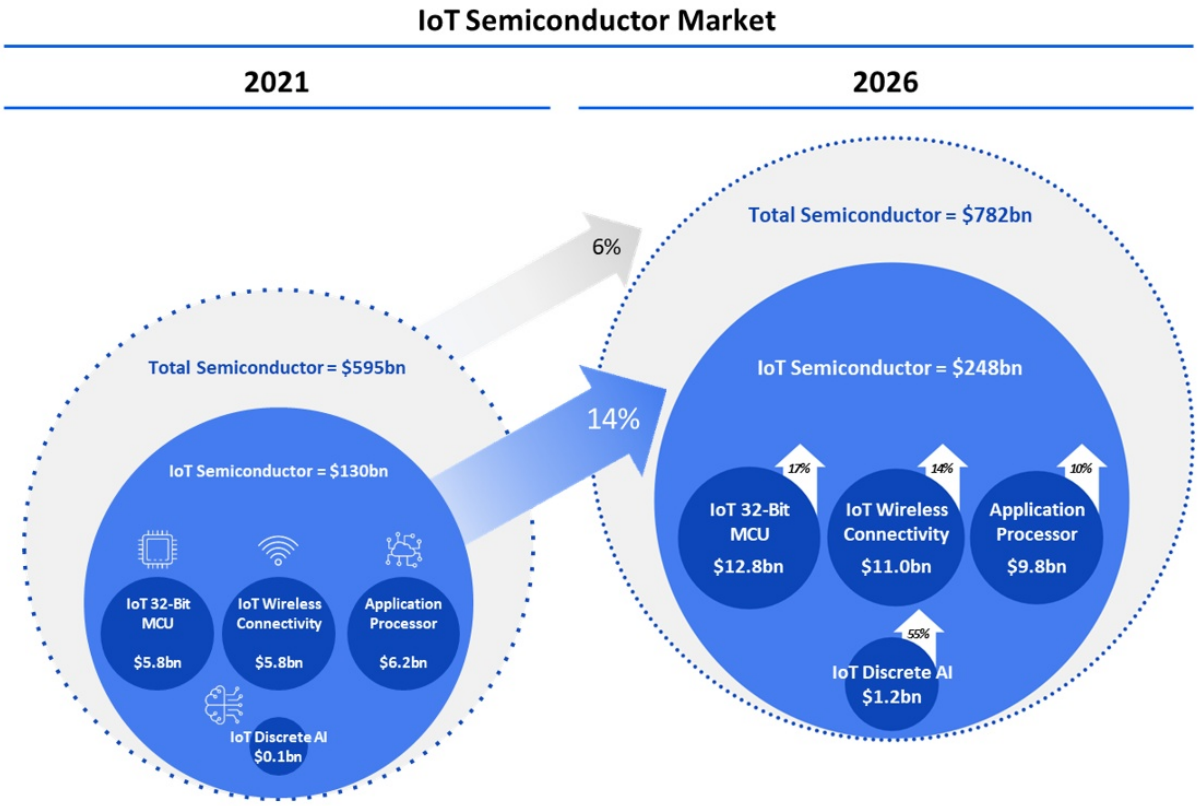
The current rapid deterioration in the global economy and weakening consumer demand due to inflation and rising interest rates, compounded with the prolonged war between Ukraine and Russia, the COVID lockdown in China, and the social unrest in Iran, Africa, and Latin America, make it difficult to stay optimistic about the economic outlook in the next few years. The semiconductor industry, which experienced the most extreme growth spike due to chip, material, and labor shortages during the pandemic years, is not insulated from this downturn as economic conditions are expected to worsen into 2023.

Since June this year, Industry analysts, including Gartner, IDC, and WSTS, revised their forecast, predicting worldwide semiconductor revenue to decline in 2022 and 2023. However, sentiments among world economists aren't all gloom and doom about the next few years.

The IMF updated its forecasts in October, predicting the global growth to slow from 6.0% in 2021 to 3.2% in 2022 and 2.7% in 2023. Even though they will likely be the weakest growth since 2001, they are still considered positive. Despite persistent fears of a slowdown next year, the positive market reaction to the weaker-than-expected October consumer price index (CPI data) seemed to indicate a transition to a new phase of inflation dynamic analysis - the likelihood of the delivery of sustained disinflation in 2023. The general perception, or more adequately hope, of inflation, possibly peaking and then encountering disinflationary headwinds (after the October CPI report), produced a downward shift in the U.S. interest rate. More, with accelerated digital transformation, while consolidating resources and tightening spending, manufacturers of consumer electronics, hardware, and software vendors look for continuous emergence of disruptive companies with transformative solutions to survive the upcoming downturn and come out on top from a shallow recovery in late 2023 and early 2024.

Overall Size of the MCU Market

Figure 1-1: Large and Growing TAM



Addressing the Fastest-Growing Segments of IoT Semiconductors

- The broader semiconductor market is growing at a CAGR of 6%, while the **IoT SEMICONDUCTOR MARKET IS GROWING AT A RAPID CAGR OF 14%**
- Within this market, **AMBIQ OPERATES IN THE FASTEST GROWING SEGMENTS, WITH EACH EXPECTED TO DOUBLE BY 2026**: 32-Bit MCUs, Wireless Connectivity and Application Processors
- The AI-enabled segment of the semiconductor market is growing at a CAGR of 20%, while the **IoT AI-ENABLED SEGMENT IS GROWING AT A RAPID CAGR OF 43%**
- Ambiq's existing and in-development **IoT AI-CENTRIC PRODUCTS** will enable us to address IoT Discrete A.I. applications, a market projected to **GROW AT 55% CAGR**

The number of Internet of Things (IoT) devices worldwide is forecast to almost triple from 9.7 billion in 2020 to more than 29 billion IoT devices in 2030. In 2030, the highest number of IoT devices will be found in China and India, with around 5 billion consumer devices. IoT devices

are used in all industry verticals and consumer markets, with the consumer segment accounting for approximately 60 percent of all IoT-connected devices in 2020. This share is projected to stay at this level over the next ten years.

Major industry verticals with more than 100 million connected IoT devices are electricity, gas, steam and A/C, water supply and waste management, retail and wholesale, transportation and storage, and government. Overall the number of IoT devices across all industry verticals is forecast to grow to more than eight billion by 2030.

The most important use case for IoT devices in the consumer segment is consumer internet and media devices such as smartphones, where the number of IoT devices is forecast to grow to more than 17 billion by 2030. Other use cases with more than one billion IoT devices by 2030 are connected (autonomous) vehicles, I.T. infrastructure, asset tracking and monitoring, and smart grids.

The Ambient Computing Vision

From visions of smart dust in the early 2000s to the emergence of the Internet of Things (IoT) in the early 2010s, ambient computing has been gaining momentum for several decades. These visions invariably show smart homes and buildings that anticipate their inhabitants' every need.

A smart doorbell recognizes your child walking home from the bus stop and unlocks the front door. Your watch alerts you that your child is walking in the door so that you can greet them. Your watch also detects that you're cold and asks whether you want to raise the temperature in the house by a few degrees. And the possibilities from there are endless.

This Vision Is Happening Slowly But Surely

As we engage with customers and partners and map out the future for Ambiq®, we see this ambient computing vision happening NOW. Quite simply, all of the key hardware pieces exist. Homes are becoming smart with connected security, safety, climate, and personal control/sensing systems.

But There Are Gaps to Be Filled

The hardware pieces exist, but ambient computing has not yet reached its full potential. The disparate hardware and software ecosystems are not yet fully integrated.

At the same time, the hardware ecosystem is not fully capable yet. Sensors consume too much power, and typical ambient computing system-on-chips (SoCs) do not have enough horsepower. The confluence of semiconductor, SoC, and neural network trends and Ambiq's innovation are solving this problem. We describe key trends and Ambiq's plans in the following discussion.

Semiconductor Technology Trends

Ambient computing requires a full spectrum of silicon options covering a wide range of costs, power, and performance. A commensurately wide range of semiconductor manufacturing options is available, from 40nm down to advanced nodes that increasingly include FinFET-based 12nm, 6nm, and beyond. We see key trends, including:

- **Moore's Law is alive and well (at least for ambient computing):** High-volume ambient computing products that integrate digital circuits, nonvolatile memory, and analog/R.F. circuits have migrated from 90nm and 65nm down to 40nm and 22nm. 12nm FinFET-based process nodes are available now with low leakage transistors, low leakage SRAM bitcells, and even early nonvolatile virtual memory (NVM) options. Ambiq is working closely with TSMC as an early adopter of this 12nm technology to help drive the next generation of SPOT-enabled SoCs with ever-lower voltages (e.g., <0.4V). Ambiq is also prepared to jump quickly to 6nm low-power technology variants, especially as memory and processing needs continue to rise. Mobile computing and datacenter wafer volume have helped the amortization of leading node semiconductor equipment, making trailing node cost (e.g., 22nm and 12nm) more accessible for ambient computing use cases. We are optimistic about that trend continuing (for the right markets.)

- **Performance is vastly increasing:** More advanced nodes are inherently faster, even at lower voltages. 22nm ambient computing parts are capable of hundreds of MHz operation. The introduction of the FinFET (with its nearly ideal sub-threshold slope) at 12nm will enable even faster frequencies – a trend that will continue as we leap to 6nm.

- **Gates are getting less expensive – in cost and power:** Scaling to 22nm, 12nm, and beyond also means that transistors and digital gates are less costly and can thus be used a bit more freely. This has important architectural implications that we'll explore in the next section on SoC trends. These smaller transistors (and the lower voltages that can be targeted) are also more energy efficient. When combined with innovative technologies like Ambiq's SPOT platform, huge gains in power efficiency are ahead.

- **A single process node won't be enough:** The needs of ambient computing are diverse, and there is no single cost/integration/power/performance point that works for all use cases. We will see the full spectrum of nodes from 40nm down to 12nm and beyond used across applications, and we will even see different process nodes used within a single package. A 12nm chip may be ideal for complex digital functions, while various cost, power, and performance requirements may force R.F., analog, and memory tasks to older nodes. Expect to see system-in-package solutions to address the increasingly diverse needs of ambient computing.

Architecture and SoC Trends

Moore's Law is alive and well, but innovation may be even more impactful at an SoC level. We see the following key trends for ambient computing:

- **SoC offerings will be varied:** A single SoC (or even a single SoC family) is NOT sufficient for the breadth of ambient computing. A full spectrum of parts is required, from tiny parts with a few hundred kB of memory and limited integration to multi-core SoCs with M.B. of memory. As discussed later, Ambiq's portfolio is growing yearly to cover this full spectrum.
- **SoCs will be increasingly heterogeneous, especially at the high end:** Moore's Law, and the "less expensive" transistors that come with it, mean that we can throw a large number of gates at a given problem. In fact, we can build entire sub-systems to attack each class of problem. Mobile phones with graphics sub-systems, audio sub-systems, communications sub-systems, etc., have been doing this for years. Low-cost MCUs will increasingly have multiple cores, including a host MCU, a GPU, an NPU, DSPs for radio stack processing, and various digital and analog peripherals. Even the simplest and most cost-constrained applications can leverage these complex SoCs. It should also be noted that the value of any given processor core in a heterogeneous system is minimized, so the particular processor core in use (Arm®, Tensilica, CEVA, RISC-V, etc.) will be less critical. Common software ecosystems (e.g., TensorFlow Lite Micro) will instead become more important.
- **SoC solutions will have much more memory at the high end:** 22nm parts offer 5MB+ of on-chip memory, and 12nm parts offer 10MB+ of memory at a reasonable price point. This will be disruptive for memory-intensive applications like graphics and neural networks/A.I. (see the next section). However, this explosion in memory size combined with increasingly leaky memory bitcells will create major leakage power problems. Ambiq is introducing important innovations starting in Apollo5 and will continue in subsequent families.
- **Memory options will become more creative and heterogenous:** Ambient computing chips typically offer NVM for code storage and SRAM for data/code that is both written and read (e.g., caches, stack storage, frame buffer storage, etc.) As ambient computing cases evolve, and in particular, as neural network and graphics needs evolve, this division of labor will become more nuanced. Large on-chip SRAM (see the previous point about larger memories) can be leaky and occupy a significant die area. New SRAM-like NVM options promise to address the leakage and die area issues by trading short retention time (e.g., seconds or minutes) for write speed. These memory options could appear as caches or memories targeted at "always on" memories that are read and written frequently, like frame buffers and neural network activation buffers. The need for

significant memory, intended for frame buffers and neural networks, may also be partially met by integrated in-package off-chip memories. High-speed and low voltage/power interfaces can manage the overhead of going off-chip while giving better cost trade-offs for large memories.

- **Sensors and sensor interfaces will take big strides forward in power efficiency:** Compute is only one slice of the performance and energy efficiency pie. Expect to see advances in sensors like microphones and image sensors. These sensors MUST be "always on," but CANNOT do so at compromised performance. Ambiq is working to solve this challenge. For example, we are currently developing an "always listening" audio system capable of capturing 48kHz audio over an 80dB+ dynamic range at <30uW (including the microphone.) In recent years, we have seen the emergence of "smart sensors" that integrate not just the sensor but also compute and even neural network functions. This trend will continue, especially as ultra-efficient compute options, like Ambiq's SPOT platform, proliferate.
- **Get ready for huge power/performance improvements:** These trends add to the disruptive performance and energy efficiency increases. This will enable unprecedented new solutions, which we discuss in the next two sections.

Neural Network Trends

AI-powered features are ubiquitous in this market and will rapidly become a must-have for wearables. Consumers have come to expect their devices to deduce their activities and accurately track them, monitor their vitals and provide valuable feedback, and interact with them using voice-driven interfaces. Technological advances in power efficiency, compute capabilities, and algorithms will continue to increase these features' practicality, accuracy, and sophistication while enabling new ones.

Currently, the classes of wearable A.I. features are largely dictated by the type of sensor data typically available on wearable devices: audio, gyroscopic, and accelerometer sensors, proximity sensors, and health sensors. This implies that the number of A.I. tasks in this space is in the dozens, which tends to fall into audio classification, time series, and computer vision problems. There is still much to do with the available data, including more sophisticated context detection and accurate inferences. Beyond that, novel sensors such as millimeter-wave radar and multi-spectral photo detectors will likely become widely available. A.I. will be used to extract meaningful inferences from their fuzzy real-world data.

The relatively small number of model types and the massive scale of endpoint deployment means that A.I. developers can amortize the effort of hyper-optimizing A.I. features across millions of devices. We believe this will lead to the continued use of generalized A.I. frameworks such as Tensorflow to produce models but a de-emphasis on the use of general framework-based runtimes in favor of "hand coding" the deployed production versions of these models on endpoint devices.

Endpoint A.I. is constrained by power, compute, and memory. NPUs help by adding computational capacity and reducing inference latency (with a concomitant reduction in power). Still, it is important to note that this represents only a fraction of the total energy cost of the feature. Power-efficient sensor sampling, compute, and memory are all equally, if not more, important.

Even with all the recent technological advances, some A.I. features are still out of reach for IoT endpoint devices. For example, speech transcription and translation are still relegated to large devices. These types of tasks currently require an order of magnitude more memory than is typically available on endpoint devices. We believe two trends will eventually render these tasks practical: rapid algorithmic advances in model density and compression and hardware advances allowing larger endpoint memories. We're excited to participate in both trends.

Wearable Trends

Improvements in SoCs and manufacturing, along with an explosion in neural networks, mean that ambient computing systems will be VASTLY more capable in the next 1-3 years. This will enable an entirely new generation of wearables. In particular, we see that:

- **Health is king, and the on-wrist doctor is coming:** We have worked hard at Ambiq to deliver constant improvements in performance and energy efficiency (which will continue). Our customers have used this increased capability to evolve simple step-counting wearables into on-wrist health devices. This trend will continue toward an on-wrist doctor with the emergence of more efficient sensors and a massive increase in on-chip neural network capabilities.
- **Context awareness will explode on the wrist:** Though mobile phones are getting better at detecting user context (driving? parked? walking?), wearables can give a far more intimate view of context with the help of efficient sensors and on-chip neural network capabilities. Instead of just "walking" or "running," the next generation in wearables will detect how the user feels (happy? sad? healthy? cold?) and be capable of serving up useful services (raise the temperature, schedule a doctor appointment, suggest a workout, etc.)
- **WearOS-based wearables with "Fitbit-grade" battery life is coming:** The massive increase in performance and power efficiency that we are enabling at Ambiq will bring about Android/Linux products capable of 1+ week battery life. More on this in the next section.
- **All of this is coming to more than just the wrist:** Every function on the wrist (context detection, health monitoring, and more) will also come to the ear and smart hearables. More interestingly, the massive increase in processing and neural network capabilities will enable low-latency hearing enhancement, innovations in noise cancellation, and unprecedented high-fidelity music experience.

Ambiq's Roadmap to Ambient Computing

Ambiq was founded in 2010 to enable ambient intelligence (Ambiq = Ambient I.Q. = Ambient Intelligence) with the world's most energy-efficient semiconductor solutions. We remain committed to that with an innovative roadmap that includes the following:

- An expanding Apollo family from the low end (e.g., Apollo4 Nano) up to the high end (Apollo5, Apollo6, and beyond) will enable the next generation of wearables
- A new family, Atomiq, will enable hearables and audio products capable of hearing enhancement, context awareness, active noise cancellation (ANC,) and high-fidelity audio
- Explorations of Linux/Android-grade processors in an "Apollo-like" power footprint

To make all of this happen, we focus our research and development investments on a few key areas:

1. taking SPOT to new levels of energy efficiency and performance at 12nm, 6nm, and beyond
2. developing A.I. hardware and software at the highest energy efficiency
3. developing the world's most energy-efficient sensing interfaces to feed that A.I. hardware and software, and
4. developing the world's most energy-efficient connectivity.

About Ambiq

Ambiq's mission is to enable intelligent devices everywhere by developing the lowest-power semiconductor solutions to drive a more energy-efficient, sustainable, and data-driven world. Ambiq is a pioneer of ultra-low-power semiconductor solutions based on the proprietary and patented Subthreshold Power Optimized Technology (SPOT®) platform. SPOT provides a game-changing, multi-fold improvement in energy efficiency for our end customers' electronic products. Ambiq has helped leading manufacturers worldwide develop products that run for weeks (rather than days) on a single charge while delivering a maximum feature set in compact industrial designs. Ambiq's goal is to take Artificial Intelligence (A.I.), where it has never gone before in mobile and portable devices, using Ambiq's advanced ultra-low power system on chip (SoC) solutions. Ambiq has shipped more than 190 million units as of April 2022. For more information, visit www.ambiq.com.

6500 River Place Boulevard, Building 7, Suite 200, Austin, TX 78730
www.ambiq.com
sales@ambiq.com
+1 (512) 879-2850

A-SOCAPG-WPNA01EN v1.1
February 2023