ZDNet Korea news print

200 billion pre-order 'jackpot' in Europe... Native Fabless Soteria "Mass production of 4nm at Samsung this year"

[Interview] CEO Kim Jong-man "Targeting the Customized HPC Accelerator Niche"... Plans for mass production of 2nm chips next year

Inari Input l :2024/02/26 14:04 RevisionModified: 2024/02/28 00:06 Semiconductor Display

"Soteria is a CSSP (Customer Specific Standard Product) fabless company that designs and mass-produces on-demand semiconductor chips with pre-customer acquisition, pre-order contracts, and development funding. It's because we're looking for a stable and clear business model."

Soteria, a Korean fabless startup, is an ultra-low-power, high-performance computing (HPC) accelerator company founded in 2018. While three AI semiconductor startups, including Furiosa AI, which has recently been making headlines in Korea, focus on AI accelerators aimed at ultra-large data centers in order to compete with NVIDIA, Soteria is differentiated in that it provides customized HPC accelerators targeting large and small data centers.

In other words, the business strategy is to secure customers and markets stably by targeting niche markets.

Kim Jong-man, CEO of Soteria (Photo = ZDNet Korea)

Soteria's focus is on ultra-low-power HPC accelerator ASICs (application-specific semiconductor systems) and Compute ExpressLink (CXL)-based near data processing (NDP) memory solutions.

Soteria plans to tape out the 0.3V low-power HPC accelerator 'Artemis' by the end of April this year and put wafers into the Samsung Electronics Foundry's 4-nanometer (mn) process at the end of this year, and mass-produce it in October. The NDP memory 'Ecclesia' is scheduled to be designed and verified this year and tape-out in 2026.

The reason why Soteria was able to do a single tapeout without the multi-project wafer (MPW), which is the usual process of making a prototype before mass production, is that it received the first development subsidy of $1.7 million (about 2.3 billion won) from customers in Europe and North America, and signed a supply contract worth 200 billion won and received some upfront payment.

Major customers include Argo, a British blockchain company, and ACME, a Swiss data center business. This is the result of the efforts of CEO Kim Jong-man, who founded Soteria, to run and sell all over Europe since 2021. As a result, Soteria was able to immediately begin mass production, which is unusual in the semiconductor startup industry.

CEO Kim said, "Of course, technology is important in the semiconductor business, but technology marketing and mass production development capabilities are more

important for startups. We believe that there are products only when there are customers, and it is true technology that meets customer requirements," he said, emphasizing that "Soteria is a true ASIC company that receives orders from customers and produces competitive chips that meet customer needs and value chains through collaboration."

Kim Jong-man graduated from Seoul National University's Department of Electrical Engineering and worked as a senior researcher at LG Electronics. After that, he returned to his studies and earned a master's degree in electrical engineering and a Ph.D. in computer science from Pennsylvania State University, and was a professor in the Department of Electrical and Computer Engineering at Georgia Tech. Currently, there are about 20 people working on Soteria. Among them, 80% of the developers are from Samsung Electronics, and 10 are veterans with more than 20 years of experience in semiconductor work.

The following is a Q&A with CEO Kim Jong-man.

Q. I'm curious about the development plan (roadmap) of the Soteria chip.

"Soteria is currently working on two projects. Artemis, a 0.3V low-power HPC accelerator, is scheduled to be taped out at the end of April this year and mass-produced in October at Samsung Electronics' 4nm process. In Korea, Samsung Electronics' 4nm single tape-out is proud to be the first of its kind for Soteria. The chip has secured more than 200 billion won in demand from North American and European companies. In addition, the CXL-based NDP memory 'Ecclesia', which is aimed at the small and medium-sized data center market, is expected to be designed and verified this year and tape-out in 2026.

As for future plans, the company plans to tape out Artemis with a next-generation 2nm process by the end of 2025 and target mass production in 2026. As for the foundry, Samsung Electronics is the leading company. The third-generation AI neuromorphic semiconductor (NPU) is expected to deliver samples in 2025."

Q. Why is Artemis, an ultra-low-power HPC accelerator, targeting eco-friendly data centers?

"Normally, 4nm processes use a voltage of 0.75 volts (V), but Soteria's HPC accelerator 'Artemis' uses 0.3V to achieve ultra-low power. We developed the chip independently without using the library of Arm and Synopsis to drive the chip at ultra-low power.

0.3V Artemis is expected to be in high demand for eco-friendly data centers that use immersion cooling. Recently, generative AI has been using a lot of data, which consumes a lot of power in data centers. That's why new data centers from Meta, Google, Microsoft, and others are building immersion cooling, which reduces power by 40 percent."

Q. You said that you supply customized chips to the medium and small data center market?

"Soberly, we decided that competing in the hyperscale data center market that NVIDIA is targeting would be difficult to win in the short to medium term. Our strategy is to

target a large number of U.S. financial companies on the NASDAQ and mid-to-large data center cloud markets to supply customized accelerator chips.

What differentiates us from competitors is that we directly receive algorithms, specifications, protocols, and workloads from our customers, collaborate with them, design optimal chips, and mass-produce and supply CSSP (Customer Specific Standard Productor) that is excellent in price competitiveness, power efficiency, and maintenance. As a start-up company, the process of securing development costs and upfront funding from customers is arduous and must be rigorously verified. In this process, the dedication and teamwork of all employees is making this year even brighter."

Q. There have been discussions with Intel about foundry cooperation, but why do you use Samsung?

"Apparently, there were discussions not only with Intel but also with TSMC on the possibility of mass production of large volumes. However, there were some risks in terms of schedule and experience, and as it was our first product, we decided that we could not succeed in mass production without a really long partnership. Most of our employees are from Samsung, and we have been able to work closely with Samsung compared to overseas foundries, and our development is nearing completion, and many partners evaluate our products on a par with our competitors. We are looking forward to collaborating with Samsung not only for 4nm but also for the next 2nm products."