**News Release**

No.: REN2408(A)

**Renesas Develops New AI Accelerator for Lightweight AI Models and Embedded Processor Technology to Enable Real-Time Processing**

*Results of Operation Verification Using an Embedded AI-MPU Prototype Announced at ISSCC 2024*

**TOKYO, Japan, February 22, 2024 ―** Renesas Electronics Corporation (TSE: 6723), a premier supplier of advanced semiconductor solutions, today announced the development of embedded processor technology that enables higher speeds and lower power consumption in microprocessor units (MPUs) that realize advanced vision AI. The newly developed technologies are as follows: (1) A dynamically reconfigurable processor (DRP)-based AI accelerator that efficiently processes lightweight AI models and (2) Heterogeneous architecture technology that enables real-time processing by cooperatively operating processor IPs, such as the CPU. Renesas produced a prototype of an embedded AI-MPU with these technologies and confirmed its high-speed and low-power-consumption operation. It achieved up to 16 times faster processing (130 TOPS) than before the introduction of these new technologies, and world-class power efficiency (up to 23.9 TOPS/W at 0.8 V supply).

Amid the recent spread of robots into factories, logistics, medical services, and stores, there is a growing need for systems that can autonomously run in real time by detecting surroundings using advanced vision AI. Since there are severe restrictions on heat generation, particularly for embedded devices, both higher performance and lower power consumption are required in AI chips. Renesas developed new technologies to meet these requirements and presented these achievements on February 21, at the International Solid-State Circuits Conference 2024 (ISSCC 2024), held between February 18 and 22, 2024 in San Francisco.

The technologies developed by Renesas are as follows:

1. **An AI accelerator that efficiently processes lightweight AI models**

As a typical technology for improving AI processing efficiency, pruning is available to omit calculations that do not significantly affect recognition accuracy. However, it is common that calculations that do not affect recognition accuracy randomly exist in AI models. This causes a difference between the parallelism of hardware processing and the randomness of pruning, which makes processing inefficient.

To solve this issue, Renesas optimized its unique DRP-based AI accelerator (DRP-AI) for pruning. By analyzing how pruning pattern characteristics and a pruning method are related to recognition accuracy in typical image recognition AI models (CNN models), we identified the hardware structure of an AI accelerator that can achieve both high recognition accuracy and an efficient pruning rate, and applied it to the DRP-AI design. In addition, software was developed to reduce the weight of AI models optimized for this DRP-AI. This software converts the random pruning model configuration into highly efficient parallel computing, resulting in higher-speed AI processing. In particular, Renesas' highly flexible pruning support technology (flexible N:M pruning technology), which can dynamically change the number of cycles in response to changes in the local pruning rate in AI models, allows for fine control of the pruning rate according to the power consumption, operating speed, and recognition accuracy required by users.

This technology reduces the number of AI model processing cycles to as little as one-sixteenth of pruning incompatible models and consumes less than one-eighth of the power.

1. **Heterogeneous architecture technology that enables real-time processing for robot control**

Robot applications require advanced vision AI processing for recognition of the surrounding environment. Meanwhile, robot motion judgment and control require detailed condition programming in response to changes in the surrounding environment, so CPU-based software processing is more suitable than AI-based processing. The challenge has been that CPUs with current embedded processors are not fully capable of controlling robots in real time. That is why Renesas introduced a dynamically reconfigurable processor (DRP), which handles complex processing, in addition to the CPU and AI accelerator (DRP-AI). This led to the development of heterogeneous architecture technology that enables higher speeds and lower power consumption in AI-MPUs by distributing and parallelizing processes appropriately.

A DRP runs an application while dynamically changing the circuit connection configuration between the arithmetic units inside the chip for each operation clock according to the processing details. Since only the necessary arithmetic circuits operate even for complex processing, lower power consumption and higher speeds are possible. For example, SLAM (Simultaneously Localization and Mapping), one of the typical robot applications, is a complex configuration that requires multiple programming processes for robot position recognition in parallel with environment recognition by vision AI processing. Renesas demonstrated operating this SLAM through instantaneous program switching with the DRP and parallel operation of the AI accelerator and CPU, resulting in about 17 times faster operation speeds and about 12 times higher operating power efficiency than the embedded CPU alone.

**Operation Verification**

Renesas created a prototype of a test chip with these technologies and confirmed that it achieved the world-class, highest power efficiency of 23.9 TOPS per watt at a normal power voltage of 0.8 V for the AI accelerator and operating power efficiency of 10 TOPS per watt for major AI models. It also proved that AI processing is possible without a fan or heat sink.

Utilizing these results helps solve heat generation due to increased power consumption, which has been one of the challenges associated with the implementation of AI chips in a variety of embedded devices such as service robots and automated guided vehicles. Significantly reducing heat generation will contribute to the spread of automation into various industries, such as the robotics and smart technology markets. These technologies will be applied to Renesas’ RZ/V series—MPUs for vision AI applications.

**About Renesas Electronics Corporation**

Renesas Electronics Corporation ([TSE: 6723](http://www.jpx.co.jp/english/)) empowers a safer, smarter and more sustainable future where technology helps make our lives easier. The leading global provider of microcontrollers, Renesas combines our expertise in embedded processing, analog, power and connectivity to deliver complete semiconductor solutions. These Winning Combinations accelerate time to market for automotive, industrial, infrastructure and IoT applications, enabling billions of connected, intelligent devices that enhance the way people work and live. Learn more at [renesas.com](http://www.renesas.com/). Follow us on [LinkedIn](https://www.linkedin.com/company/renesas/), [Facebook](https://www.facebook.com/RenesasElectronics/), [X](https://twitter.com/renesasglobal), [YouTube](https://www.youtube.com/user/RenesasPresents), and [Instagram](https://www.instagram.com/renesas_global/).

###

(Remarks) This DRP-AI technology uses a part of the results of work commissioned by the New Energy and Industrial Technology Development Organization (NEDO). All names of products or services mentioned in this press release are trademarks or registered trademarks of their respective owners. Arm, Arm Cortex are trademarks or registered trademarks of Arm Limited in the EU and other countries. All names of products or services mentioned in this press release are trademarks or registered trademarks of their respective owners.

**Media contact for further information, text and graphics or to discuss feature article opportunities:**

Alexandra Janetzko / Martin Stummer

HBI Communication Helga Bailey GmbH(PR agency), Hermann-Weinhauser-Str. 73, 81673 Munich, Germany

Tel.: +49 89 99 38 87-32 / -34

Email: alexandra\_janetzko@hbi.de / martin\_stummer@hbi.de

Web: [www.hbi.de](http://www.hbi.de/)