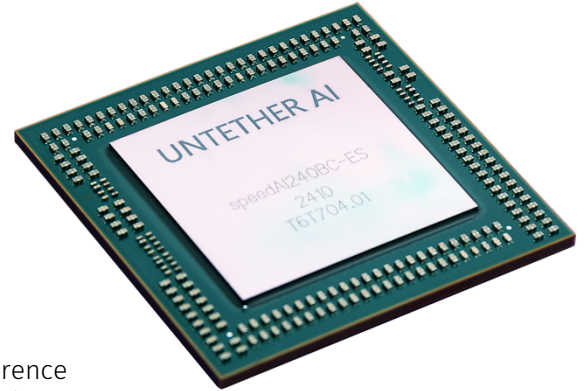


# Unmatched efficiency and scalability with AI inference accelerators



## Overview

The speedAI240™ device is designed for real-time deep learning inference applications. This second generation at-memory architecture is optimized for CNNs, attention networks, and recommendation systems.

## Key Highlights

### Efficiency Boosts Performance

The speedAI240 maximizes performance by dedicating every square millimeter to inference, achieving industry-leading 30 TFLOPS/W in 2000 FP8 TFLOPS.

### Flexibility at Scale

With over 1,400 RISC-V cores, it offers coarse-grained control over 370,000 processing elements, interconnected by a high-speed NOC for smooth data flow.

### Precision and Accuracy

Leveraging 8-bit floating point data types (FP8) and a custom network-on-chip, speedAI devices minimize data movement costs while ensuring accuracy.

### Scalable Architecture

External DRAM supports large models with two LPDDR5 x64 interfaces, while four PCIe Gen5 interfaces facilitate chip-to-chip and card-to-card interconnects for building nodes capable of handling even the largest language models swiftly.

## Target Applications

### Vision/Object Detection

Smart Retail and City

AgTech

Aerospace and Defense

### Finance, Enterprise Datacenter

Speech-to-text

Sentiment analysis

Chatbots

### Automotive

Advanced driver assistance systems (ADAS)

In-Vehicle Experience (IVE)

## Specifications

Package	Dimensions
Package dimensions	40x40 mm
Interconnect interface	PCIe Gen5 x16 Supports 160 GB/s total PCIe bandwidth
External Memory	Two LPDDR5 x64 interfaces, up to 64 GB capacity Offers 100 GB/s LPDDR5 bandwidth
On-Chip Memory	238 MB on-chip SRAM Up to 1 PB/s of on-chip memory bandwidth

## Performance

All estimates use the Untether AI FP8 data type.

Network	Single Chip
ResNet50	80,000
BERT-Large 384	10,000
YoloV7	5,000

For more information: Visit: [www.untether.ai/](http://www.untether.ai/)  
Contact: [www.untether.ai/about/contact/](http://www.untether.ai/about/contact/)

### Notice

THE INFORMATION DISCLOSED TO YOU HEREIN (THE "MATERIALS") IS PROVIDED SOLELY FOR THE SELECTION AND USE OF UNTETHER AI'S PRODUCTS. TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, MATERIALS ARE MADE AVAILABLE "AS IS". UNTETHER AI MAKES NO REPRESENTATIONS OR WARRANTIES, WHATSOEVER WITH RESPECT TO THE MATERIALS OR THE PRODUCTS, INCLUDING BUT NOT LIMITED TO REPRESENTATIONS OR WARRANTIES OF MERCHANTABILITY; SECURITY; RELIABILITY; ACCURACY; QUALITY; INTEGRATION; FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, OR TITLE; THAT THE INFORMATION PROVIDED IN THIS MATERIAL IS SUITABLE FOR ANY PURPOSE; NOR THAT THE IMPLEMENTATION OF SUCH INFORMATION WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS, OR OTHER RIGHTS. WITHOUT LIMITING THE GENERALITY OF THE FOREGOING, UNTETHER AI EXPRESSLY DISCLAIMS ANY REPRESENTATION, CONDITION, OR WARRANTY THAT ANY INFORMATION PROVIDED TO YOU HEREUNDER, CAN OR SHOULD BE RELIED UPON BY YOU FOR ANY PURPOSE WHATSOEVER. UNTETHER AI DISCLAIMS ANY AND ALL LIABILITY RELATED TO THIS MATERIAL AND WILL NOT BE LIABLE FOR ANY LOSSES OR DAMAGE CAUSED BY RELIANCE ON THE INFORMATION IN THIS MATERIAL.

No license, either expressed or implied, is granted for any intellectual property rights of Untether AI or any third party through the information in this Material. Untether AI shall not be liable (whether in contract or tort, including negligence, or under any other theory of liability) for any loss or damage of any kind or nature related to, arising under, or in connection with, the Materials (including your use of the Materials), including for any direct, indirect, special, incidental, or consequential loss or damage (including loss of data, profits, goodwill, or any type of loss or damage suffered as a result of any action brought by a third party) even if such damage or loss was reasonably foreseeable or Untether AI had been advised of the possibility of the same. Untether AI assumes no obligation to correct any errors contained in the Materials or to notify you of updates to the Materials or to any products. You may not reproduce, modify, distribute, or publicly display the Materials without Untether AI's prior written consent. You should obtain the latest relevant Material before placing orders and should verify that such information is current and complete. All orders are subject to Untether AI's contract which outlines any applicable terms and conditions for a product.

### Trademarks

Untether AI, tsunAI, speedAI, imAI, gine SDK are trademarks and/or registered trademarks of Untether AI Corporation in the U.S and other countries. Other company names may be trademarks of the respective companies.