Ara240 Discrete Neural Processing Unit (DNPU)

ARA240

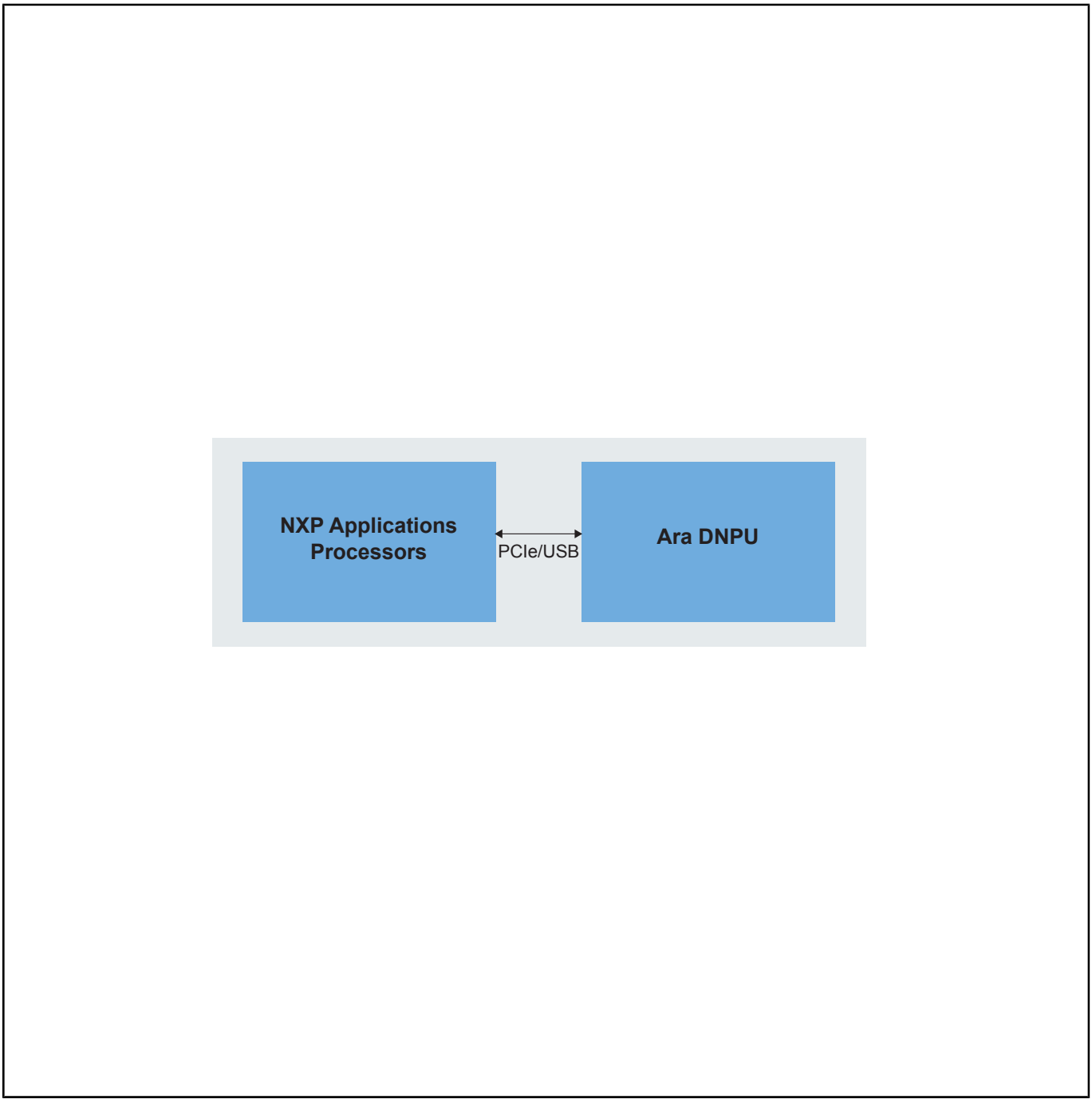
Preproduction

This page contains information on a preproduction product. Specifications and information herein are subject to change without notice. For additional information please contact your sales representative.

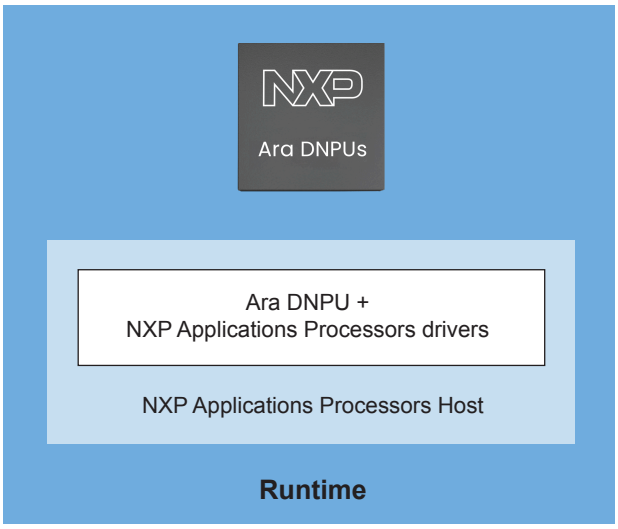
Last Updated: Feb 16, 2026

The Ara240 Discrete Neural Processing Unit (DNPU) enables real-time generative AI, large language models (LLMs) and vision language models (VLMs) execution on AI-enabled compute and embedded systems, delivering low-latency, lower operational costs and enhanced data privacy. Its innovative architecture combines balanced compute, large on-chip memory and high off-chip bandwidth to efficiently execute large models.

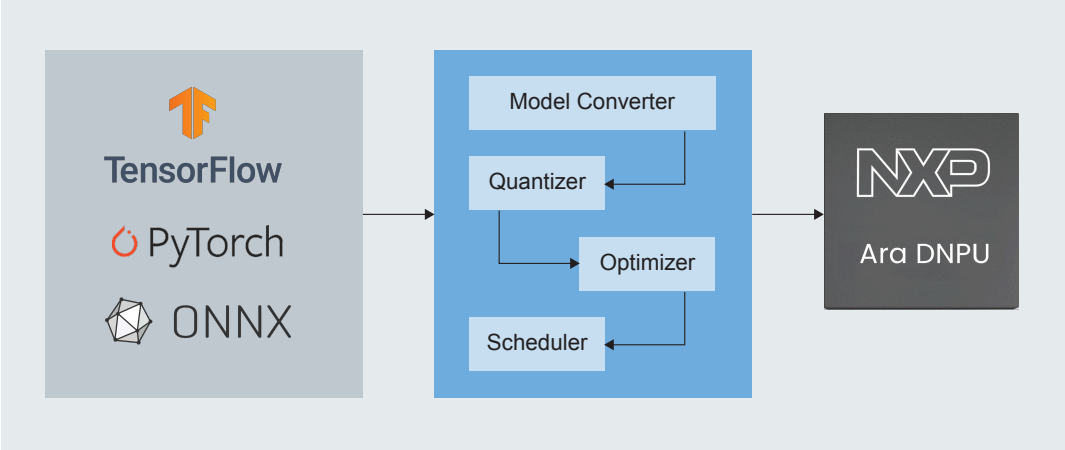
NXP Applications Processor and Ara DNPU Connection Block Diagram



Integration of Ara DNPU and Applications Processors Host with Drivers Block Diagram



Machine learning deployment flow Block Diagram



View additional information for Ara240 Discrete Neural Processing Unit (DNPU).

Note: The information on this document is subject to change without notice.