

Earl J. Lum
 +1-650-430-2221
elum@ejlwireless.com

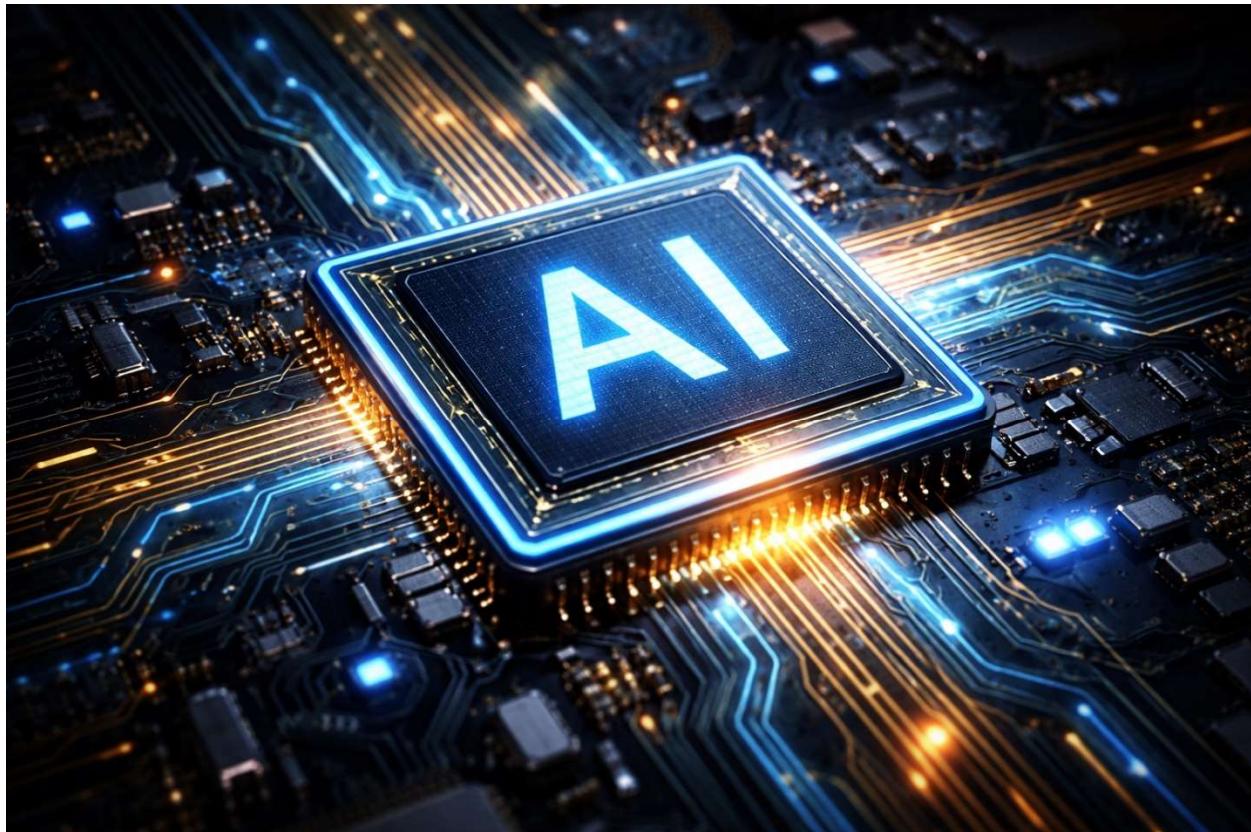


AWS Annapurna Labs

Global ASIC NPU, CPU, DPU, and QPU Deployment Analysis

“Under the Radar” AIxPU Research Series

February 2026



Entire contents © 2026 EJL Wireless Research LLC. All Rights Reserved. Reproduction of this publication in any form without prior written permission is strictly forbidden and will be prosecuted to the full extent of US and International laws. The transfer of this publication in either paper or electronic form to unlicensed third parties is strictly forbidden. The information contained herein has been obtained from sources EJL Wireless Research LLC deems reliable. EJL Wireless Research disclaims all warranties as to the accuracy, completeness, or adequacy of such information. EJL Wireless Research LLC shall bear no liability for errors, omissions, or inadequacies in the information contained herein or for the interpretation thereof. The reader assumes sole responsibility for the selection of these materials to achieve their intended results. The opinions expressed herein are subject to change without notice. All product and company names are trademarks™ or registered® trademarks of their respective holders. Use of them does not imply any affiliation with or endorsement by them.

TABLE OF CONTENTS

RESEARCH BRIEF	3
Methodology	3
Background	3
EXECUTIVE SUMMARY	4
Conclusion.....	7
TRAINIUM AI NEURAL PROCESSING UNIT (NPU) ANALYSIS	8
Successfully replacing 3 rd party AI accelerator silicon with internally developed Trainium family.....	8
INFERENTIA AI NEURAL PROCESSING UNIT (NPU) ANALYSIS	14
GRAVITON COMPUTE PROCESSING UNIT (CPU).....	18
NITRO DATA PROCESSING UNIT (DPU)	25
OCELOT QUANTUM PROCESSING UNIT (QPU)	33

TABLES

Table 1: AWS Annapurna Labs Trainium NPU Specifications by Generation	8
Table 2: Global AWS Annapurna Labs Trainium NPUs Installed by Geographic Region	10
Table 3: Global AWS Annapurna Labs Trainium NPUs Installed by Generation and Data Center/Country ..	11
Table 4: AWS Annapurna Labs Inferentia NPU Specifications by Generation	14
Table 5: Global AWS Annapurna Labs Inferentia NPUs Installed by Geographic Region	15
Table 6: Global AWS Annapurna Labs Inferentia NPUs Installed by Generation and Data Center/Country	16
Table 7: AWS Annapurna Labs Graviton CPU Specifications by Generation	18
Table 8: Global AWS Annapurna Labs Graviton CPUs Installed by Geographic Region	20
Table 9: Global AWS Annapurna Labs Graviton CPUs Installed by Generation and Data Center/Country ..	20
Table 10: AWS Annapurna Labs Nitro DPUs Specifications by Generation	26
Table 11: Global AWS Annapurna Labs Nitro DPUs Installed by Geographic Region	28
Table 12: Global AWS Annapurna Labs Nitro DPUs Installed by Generation and Data Center/Country ..	28
Table 13: Global AWS Annapurna Labs Ocelot QPUs Installed by Geographic Region	34
Table 14: Global AWS Annapurna Labs Ocelot QPUs Installed by Data Center/Country.....	35

EXHIBITS

Exhibit 1: Global AWS Annapurna Labs Trainium NPUs Installed by Generation	4
Exhibit 2: Global AWS Annapurna Labs Inferentia NPUs Installed by Generation	5
Exhibit 3: Global AWS Annapurna Labs Graviton CPUs Installed by Generation	5
Exhibit 4: Global AWS Annapurna Labs Nitro DPUs Installed by Generation	6
Exhibit 5: Global AWS Annapurna Labs Ocelot QPUs Installed by Geographic Region	6
Exhibit 6: AWS Annapurna Labs Trainium NPU Generations.....	8
Exhibit 7: Global AWS Annapurna Labs Trainium NPUs Installed by Generation	9
Exhibit 8: Global AWS Annapurna Labs Trainium NPUs Installed by Geographic Region	12
Exhibit 9: Global AWS Annapurna Labs Trainium1 NPUs Installed by Geographic Region	12
Exhibit 10: Global AWS Annapurna Labs Trainium2 NPUs Installed by Geographic Region	13
Exhibit 11: Global AWS Annapurna Labs Trainium3 NPUs Installed by Geographic Region	13
Exhibit 12: AWS Annapurna Labs Inferentia NPU Generations.....	14
Exhibit 13: Global AWS Annapurna Labs Inferentia NPUs Installed by Generation.....	15
Exhibit 14: Global AWS Annapurna Labs Inferentia1 NPUs Installed by Geographic Region	17
Exhibit 15: Global AWS Annapurna Labs Inferentia2 NPUs Installed by Geographic Region	17
Exhibit 16: AWS Annapurna Labs Graviton CPU Generations.....	18
Exhibit 17: Global AWS Annapurna Labs Graviton CPUs Installed by Generation	19
Exhibit 18: Global AWS Annapurna Labs Graviton CPUs Installed by Geographic Region.....	21
Exhibit 19: Global AWS Annapurna Labs Graviton1 CPUs Installed by Geographic Region.....	21
Exhibit 20: Global AWS Annapurna Labs Graviton2 CPUs Installed by Geographic Region.....	22
Exhibit 21: Global AWS Annapurna Labs Graviton3 CPUs Installed by Geographic Region.....	22
Exhibit 22: Global AWS Annapurna Labs Graviton3E CPUs Installed by Geographic Region	23
Exhibit 23: Global AWS Annapurna Labs Graviton4 CPUs Installed by Geographic Region.....	23
Exhibit 24: Global AWS Annapurna Labs Graviton5 CPUs Installed by Geographic Region.....	24
Exhibit 25: AWS Nitro Cards.....	25
Exhibit 26: AWS Nitro System Server Example	25
Exhibit 27: AWS Annapurna Labs Nitro DPU Generations.....	26
Exhibit 28: Global AWS Annapurna Labs Nitro DPUs Installed by Generation	27
Exhibit 29: Global AWS Annapurna Labs Nitro DPUs Installed by Geographic Region.....	29
Exhibit 30: Global AWS Annapurna Labs Nitro v1 DPUs Installed by Geographic Region	29
Exhibit 31: Global AWS Annapurna Labs Nitro v2 DPUs Installed by Geographic Region	30
Exhibit 32: Global AWS Annapurna Labs Nitro v3 DPUs Installed by Geographic Region	30
Exhibit 33: Global AWS Annapurna Labs Nitro v4 DPUs Installed by Geographic Region	31
Exhibit 34: Global AWS Annapurna Labs Nitro v5 DPUs Installed by Geographic Region	31
Exhibit 35: Global AWS Annapurna Labs Nitro v6 DPUs Installed by Geographic Region	32
Exhibit 36: AWS Annapurna Labs Ocelot QPU Packaged	33
Exhibit 37: AWS Annapurna Labs Ocelot QPU Dual Chip Stacked Design	33
Exhibit 38: AWS Annapurna Labs Ocelot QPU in Packaged Configuration.....	34
Exhibit 39: AWS Annapurna Labs Ocelot QPU in Dilution Chamber System	34
Exhibit 40: Global AWS Annapurna Labs Ocelot QPUs Installed by Geographic Region	36