

ALVEO™ V70 AI INFERENCE ACCELERATOR CARD

OVERVIEW

With intelligent connected devices in our homes, cars, offices, factories, cities, and in the cloud, the cost for this proliferation of AI enabled applications is an exponential increase in the data-processing and energy efficiency requirements placed on the chips powering these devices. The challenge is not just how to deploy the AI model, but how to deploy the AI application most efficiently. The best implementation of an AI application doesn't need to be the fastest, it needs to be the most efficient, yet remain flexible. AMD XDNA architecture built on adaptive computing platforms provide the best of both worlds for AI inference workloads - from cloud, edge or endpoint.

The Alveo™ V70 accelerator card is the first AMD Alveo production card leveraging AMD XDNA™ architecture with AI Engines providing a tightly integrated heterogeneous compute platform for CNN, RNN, and NLP acceleration targeting cloud and edge applications. V70 is designed to be the most energy efficient AI Inference card in the AMD portfolio tuned for video analytics and natural language processing workloads and offers industry standard framework support, directly compiling models trained in TensorFlow and PyTorch. The card is a PCIe®-based half height, half length, single slot card that supports passive cooling for closed-loop thermal control in the server PCIe expansion slot. The card is equipped with a 7nm Versal® ACAP device which has an integrated AI Engine core to complement adaptable and scalar engines and 16 GB of DDR4 memory. Providing low power and a low-profile form factor, the V70 helps reduce cost per AI channel and provides high channel density for video applications.



TARGET APPLICATIONS

KEY VALUE PROPS

- Helps reduced cost per AI channel
- Entire pipeline execution on accelerator
- High channel density for video analytics applications

TARGET USERS

- Data Scientist
- AI Developers
- Software Developers
- Data Center Infrastructure Engineers

USE CASES

- Video Content Analysis
- Natural Language Processing
- Smart City/ Smart Retail
- Recommendation Ranking
- Medical Imaging

Алматы (7273)495-231
 Ангарск (3955)60-70-56
 Архангельск (8182)63-90-72
 Астрахань (8512)99-46-04
 Барнаул (3852)73-04-60
 Белгород (4722)40-23-64
 Благовещенск (4162)22-76-07
 Брянск (4832)59-03-52
 Владивосток (423)249-28-31
 Владикавказ (8672)28-90-48
 Владимир (4922)49-43-18
 Волгоград (844)278-03-48
 Вологда (8172)26-41-59
 Воронеж (473)204-51-73
 Екатеринбург (343)384-55-89

Иваново (4932)77-34-06
 Ижевск (3412)26-03-58
 Иркутск (395)279-98-46
 Казань (843)206-01-48
 Калининград (4012)72-03-81
 Калуга (4842)92-23-67
 Кемерово (3842)65-04-62
 Киров (8332)68-02-04
 Коломна (4966)23-41-49
 Кострома (4942)77-07-48
 Краснодар (861)203-40-90
 Красноярск (391)204-63-61
 Курск (4712)77-13-04
 Курган (3522)50-90-47
 Липецк (4742)52-20-81

Россия +7(495)268-04-70

Магнитогорск (3519)55-03-13
 Москва (495)268-04-70
 Мурманск (8152)59-64-93
 Набережные Челны (8552)20-53-41
 Нижний Новгород (831)429-08-12
 Новокузнецк (3843)20-46-81
 Ноябрьск (3496)41-32-12
 Новосибирск (383)227-86-73
 Омск (3812)21-46-40
 Орел (4862)44-53-42
 Оренбург (3532)37-68-04
 Пенза (8412)22-31-16
 Петрозаводск (8142)55-98-37
 Псков (8112)59-10-37
 Пермь (342)205-81-47

Казахстан +7(7172)727-132

Ростов-на-Дону (863)308-18-15
 Рязань (4912)46-61-64
 Самара (846)206-03-16
 Санкт-Петербург (812)309-46-40
 Саратов (845)249-38-78
 Севастополь (8692)22-31-93
 Саранск (8342)22-96-24
 Симферополь (3652)67-13-56
 Смоленск (4812)29-41-54
 Сочи (862)225-72-31
 Ставрополь (8652)20-65-13
 Сургут (3462)77-98-35
 Сыктывкар (8212)25-95-17
 Тамбов (4752)50-40-97
 Тверь (4822)63-31-35

Киргизия +996(312)96-26-47

Тольятти (8482)63-91-07
 Томск (3822)98-41-53
 Тула (4872)33-79-87
 Тюмень (3452)66-21-18
 Ульяновск (8422)24-23-59
 Улан-Удэ (3012)59-97-51
 Уфа (347)229-48-12
 Хабаровск (4212)92-98-04
 Чебоксары (8352)28-53-07
 Челябинск (351)202-03-61
 Череповец (8202)49-02-64
 Чита (3022)38-34-83
 Якутск (4112)23-90-97
 Ярославль (4852)69-52-93

SPECIFICATIONS

FEATURE	DETAILS
Application	• AI Inference
Architecture	• AMD XDNA – Versal AI Core
AI Engine	• 2nd-gen AIE-ML tiles
TOPS* (INT8)	• 404
TOPS* (BF16)	• 202
Memory Bandwidth	• 47600 GB/s for internal memory • 76.8 GB/s for external DDR
High Density Video Decoder**	• 96 channels of 1920x1080p • H.264/H.265
PCIe interface	• Gen 4/5 x 8
Form Factor	• Half Height, Half Length
Cooling	• Passive
Power (TDP)	• 75W

* Using 50% weight sparsity

** @10 fps, H.264/H.265

Алматы (7273)495-231
Ангарск (3955)60-70-56
Архангельск (8182)63-90-72
Астрахань (8512)99-46-04
Барнаул (3852)73-04-60
Белгород (4722)40-23-64
Благовещенск (4162)22-76-07
Брянск (4832)59-03-52
Владивосток (423)249-28-31
Владикавказ (8672)28-90-48
Владимир (4922)49-43-18
Волгоград (844)278-03-48
Вологда (8172)26-41-59
Воронеж (473)204-51-73
Екатеринбург (343)384-55-89

Иваново (4932)77-34-06
Ижевск (3412)26-03-58
Иркутск (395)279-98-46
Казань (843)206-01-48
Калининград (4012)72-03-81
Калуга (4842)92-23-67
Кемерово (3842)65-04-62
Киров (8332)68-02-04
Коломна (4966)23-41-49
Кострома (4942)77-07-48
Краснодар (861)203-40-90
Красноярск (391)204-63-61
Курск (4712)77-13-04
Курган (3522)50-90-47
Липецк (4742)52-20-81

Россия +7(495)268-04-70

Магнитогорск (3519)55-03-13
Москва (495)268-04-70
Мурманск (8152)59-64-93
Набережные Челны (8552)20-53-41
Нижегород (831)429-08-12
Новокузнецк (3843)20-46-81
Ноябрьск (3496)41-32-12
Новосибирск (383)227-86-73
Омск (3812)21-46-40
Орел (4862)44-53-42
Оренбург (3532)37-68-04
Пенза (8412)22-31-16
Петрозаводск (8142)55-98-37
Псков (8112)59-10-37
Пермь (342)205-81-47

Казахстан +7(7172)727-132

Ростов-на-Дону (863)308-18-15
Рязань (4912)46-61-64
Самара (846)206-03-16
Санкт-Петербург (812)309-46-40
Саратов (845)249-38-78
Севастополь (8692)22-31-93
Саранск (8342)22-96-24
Симферополь (3652)67-13-56
Смоленск (4812)29-41-54
Сочи (862)225-72-31
Ставрополь (8652)20-65-13
Сургут (3462)77-98-35
Сыктывкар (8212)25-95-17
Тамбов (4752)50-40-97
Тверь (4822)63-31-35

Киргизия +996(312)96-26-47

Тольятти (8482)63-91-07
Томск (3822)98-41-53
Тула (4872)33-79-87
Тюмень (3452)66-21-18
Ульяновск (8422)24-23-59
Улан-Удэ (3012)59-97-51
Уфа (347)229-48-12
Хабаровск (4212)92-98-04
Чебоксары (8352)28-53-07
Челябинск (351)202-03-61
Череповец (8202)49-02-64
Чита (3022)38-34-83
Якутск (4112)23-90-97
Ярославль (4852)69-52-93

<https://amd.nt-rt.ru/> || aid@nt-rt.ru