



2019年中国嵌入式技术大会
EMBEDDED TECHNOLOGY
Conference China 2019

嵌入式AI处理器

林金龙

北京大学软件与微电子学院



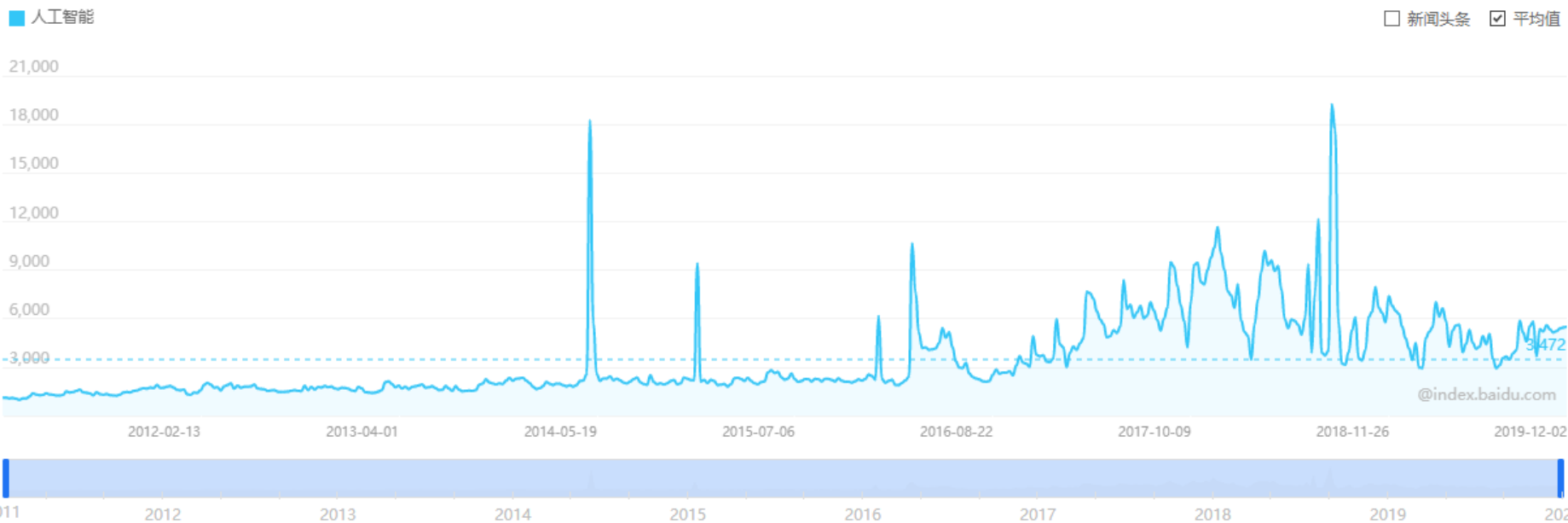
北京大学



“人工智能”成为热点

搜索指数 ?

2011-01-01 ~ 2019-12-05 | 自定义 | PC+移动 | 全国 |



百度搜索指数



清华大学



“AI 芯片” 获得关注



人工智能

网页 资讯 视频 图片 知道

百度为您找到相关结果约43,000,000个



人工智能芯片

网页 资讯 视频 图片 知道

百度为您找到相关结果约17,300,000个



人工智能处理器

网页 资讯 视频 图片 知道

百度为您找到相关结果约10,700,000个



Searched for acmdlTitle:(AI processor) [new search] [edit/save query]

Searched The ACM Full-Text Collection: 579,224 records [Expand your search to The A

4,703 results found



[35 videos found](#)

Result 1 - 20 of 4,703



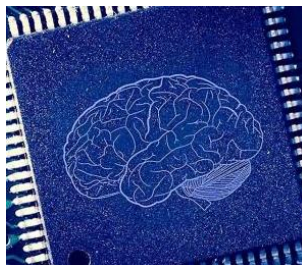
清华大学



AI chip



Embedded AI Processor

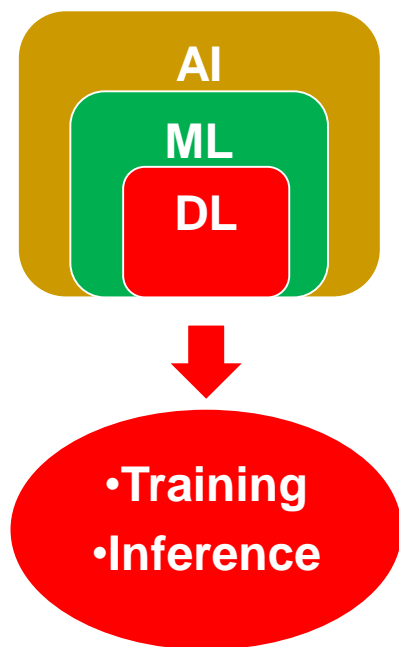


SNNP

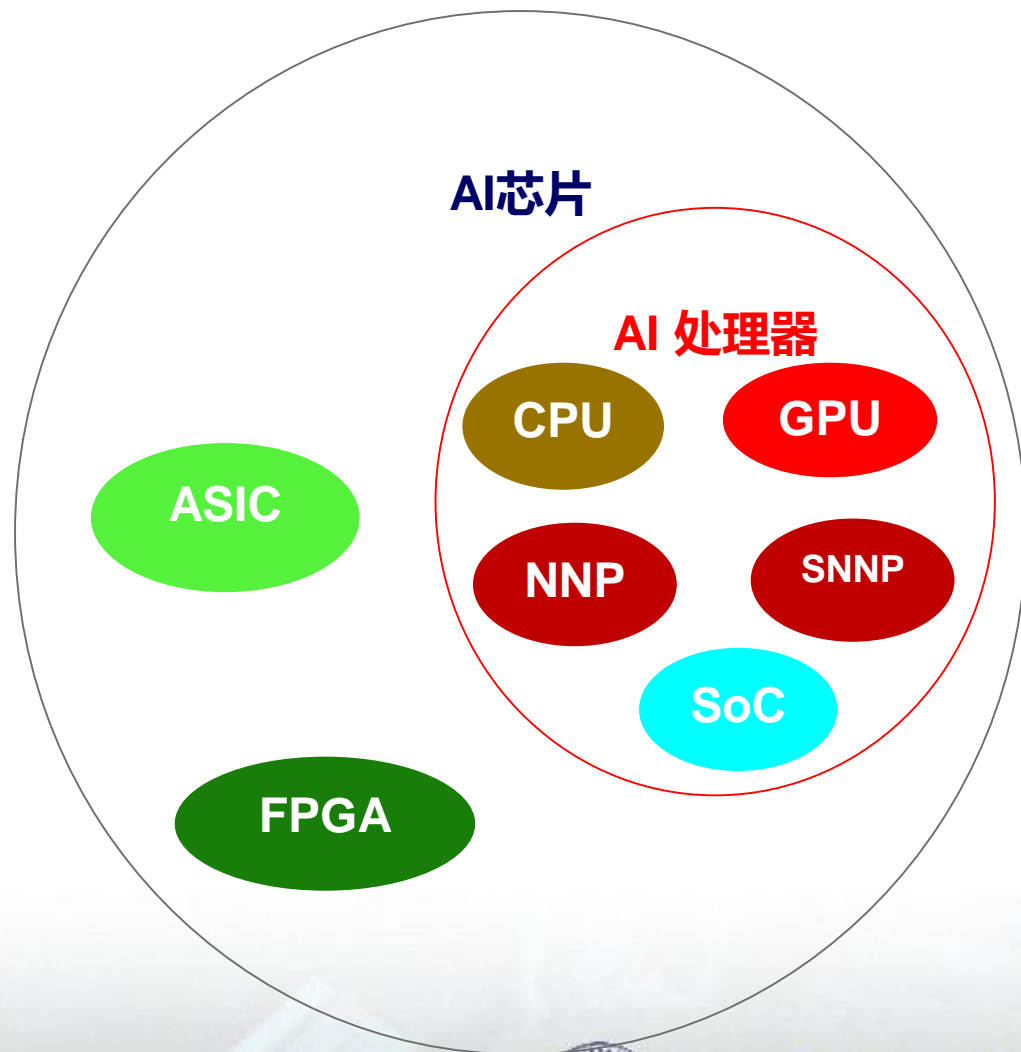


The future

“AI 芯片”？

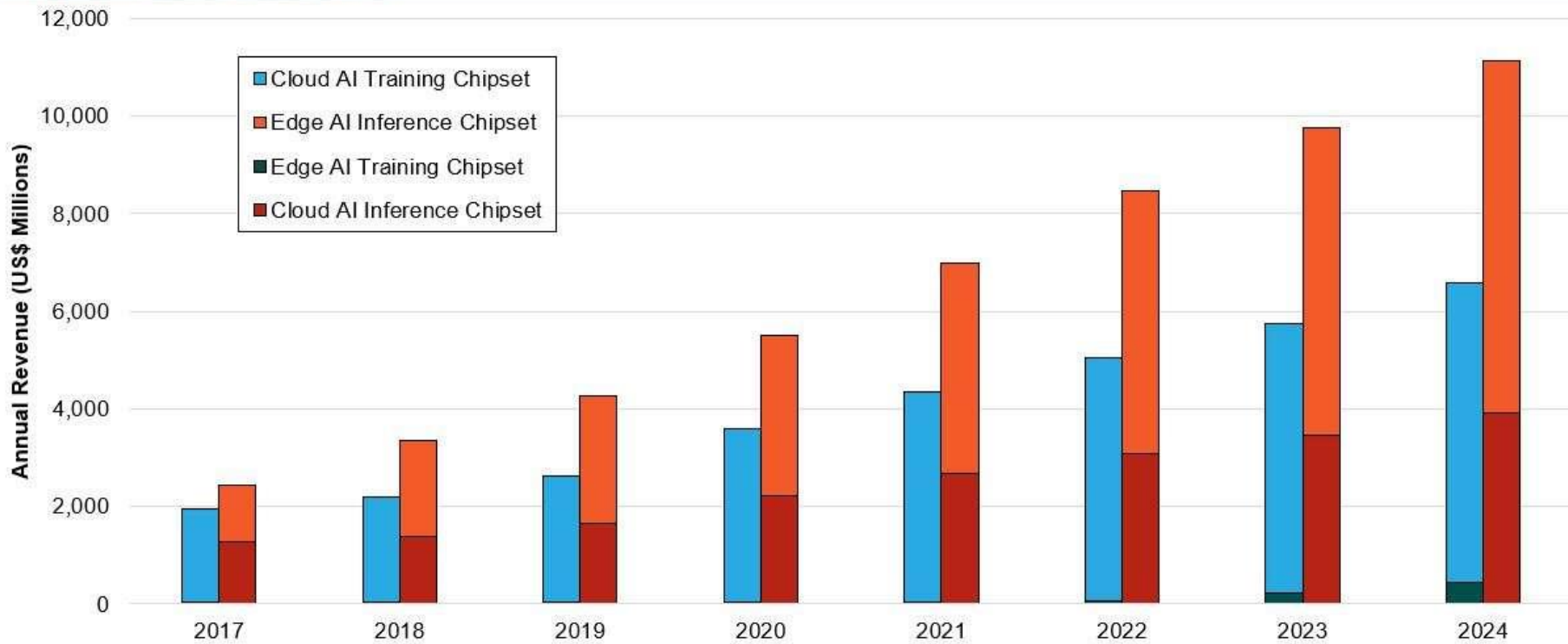


➔
加速





Total Annual Revenue from AI Chipset Sales by AI Inference and Training
World Markets, Forecast: 2017 to 2024



Source: ABI Research

<https://www.eet-china.com/news/201909171033.html>



清华大学



The CompassIntel A-List Index: AI Chipset Index -2018

Vender	Index	Vender	Index
Nvidia	85.3	Synopsys	61
Intel	82.9	MediaTek (联发科)	59.5
IBM	80.2	Imagination	59
Google	78	Marvell	58.5
Apple Inc.	75.3	Xilinx	58
AMD	74.7	CEVA	54
ARM	73	Cadence	51.5
Qualcomm	73	Rockchip (瑞芯微)	48
SAMSUNG Electronics	72.1	Vensilcon (芯原)	47
NXP	70.3	General Vision	46
Broadcom	68.2	Cambricon (寒武纪)	44.5
Huawei (华为)	64.5	Horizon Robotics (地平线)	38.5



Training

Cloud

Device

?!

ALTERA FPGA

HUAWEI 升腾910

NVIDIA GPU Tesla DGX

XILINX FPGA

Cambricon MUL100
寒武纪科技

Google TPU3.0

BITMAIN BM168X

Inference

NVIDIA GPU Tesla DGX

IBM TrueNorth

intel VPU Myriad X

Cambricon MUL100
寒武纪科技

XILINX FPGA

ALTERA FPGA

HUAWEI 升腾910

Google TPU3.0

BITMAIN BM168X

Cambricon 寒武纪科技 Cambricon-A Google Edge TPU

NVIDIA Jetson

Apple A12 Bionic

HUAWEI 麒麟990

Horizon Robotics Journey Sunrise

VeriSilicon VIP8000

Horizon Robotics

IBM TrueNorth

Rockchip 瑞芯微电子 RK3399Pro

QUALCOMM



NXP

ARM

BITMAIN BM1880



北京大学



不同神经处理器一帧图像算力需求

神经网络	图像尺寸	算力需求	i7-数量 (相当) (30fps)≈
AlexNet (8)	224X224	1.4GFLOPS	56
CaffeNet (8)	224X224	1.4GFLOPS	56
Vgg-VD-16	224X224	16GFLOPS	640
Vgg-VD-19	224X224	20GFLOPS	800
ResNet 34	224X224	8GFLOPS	320
SeNet	224X224	21GFLOPS	840
GoogleNet	224X224	8GFLOPS	320

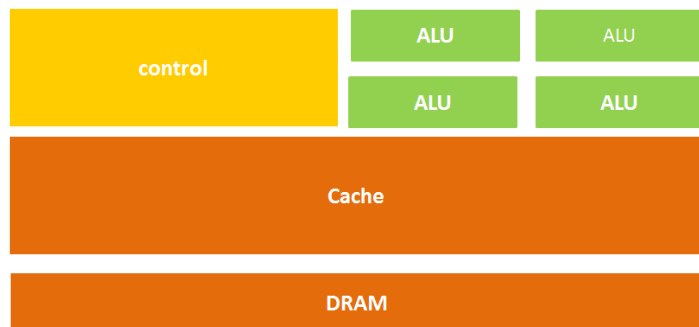
<https://www.jianshu.com/p/5f6a74490286?from=groupmessage>
https://blog.csdn.net/weixin_36474809/article/details/85675504



北京大學



CPU



GPU



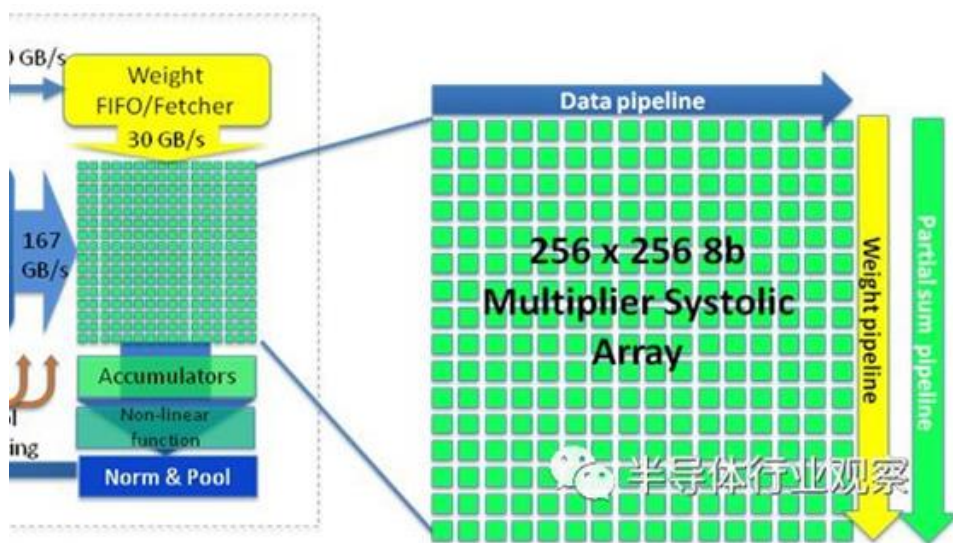
https://www.sohu.com/a/200698604_160923



北京大学

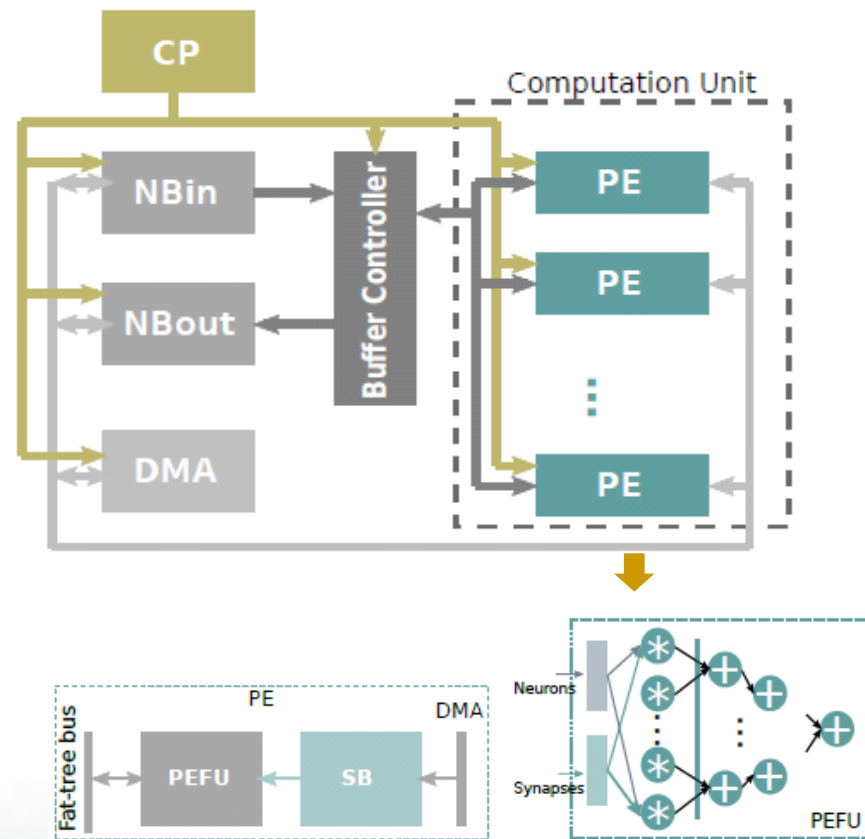
加速实现?

TPU



https://www.sohu.com/a/245428319_132567

Cambricon-1A



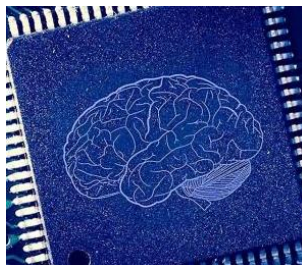
<https://blog.csdn.net/evolone/article/details/82594264>



AI chip



Embedded AI Processor



SNNP



The future



应用在哪儿？



<https://image.baidu.com/>



北京大學

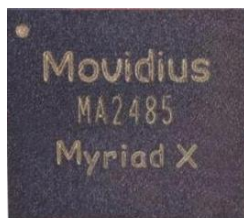
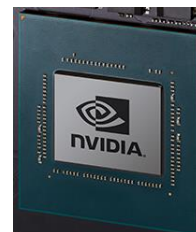
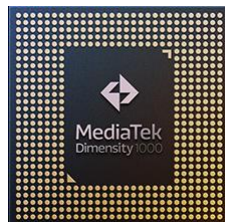
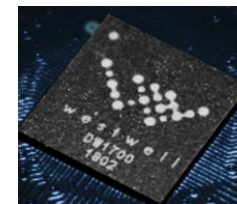
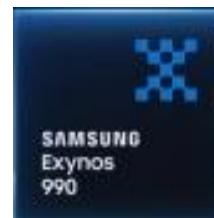


2019年中国嵌入式技术大会

EMBEDDED TECHNOLOGY

Conference China 2019

处理器



..... **SOC!**

<https://image.baidu.com/>



北京大學



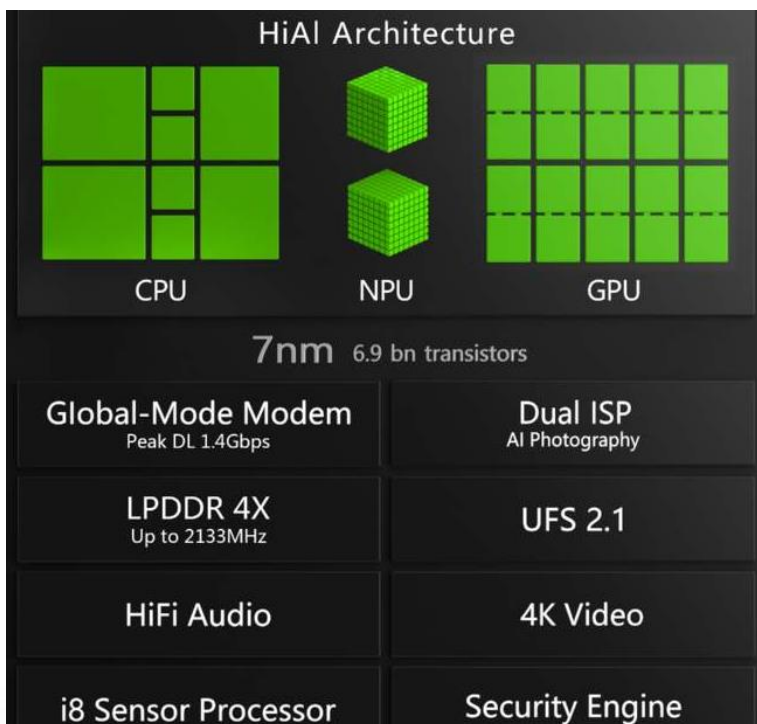
性能/功耗

处理器	Myriad 2	Xavier	A12	BM1880	Journey	RK3399Pro	sunrise
性能 (Tops)	1.0	32	5	2	4	3	4
功耗 (W)	0.5	10	3	2.5	2	2	2

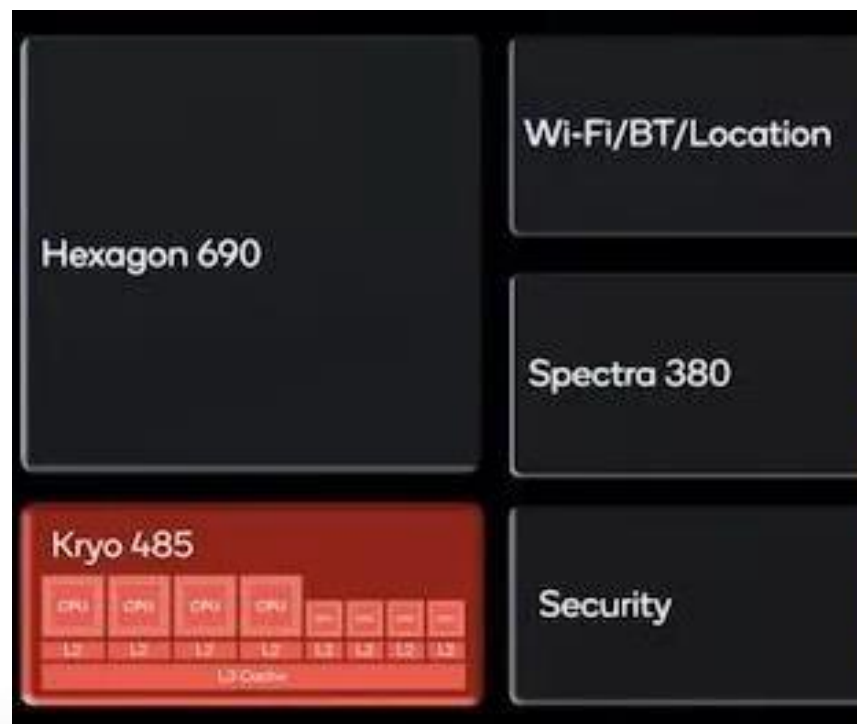




Kirin 980



Snapdragon 855



http://www.360doc.com/content/19/0109/02/29478554_807587672.shtml

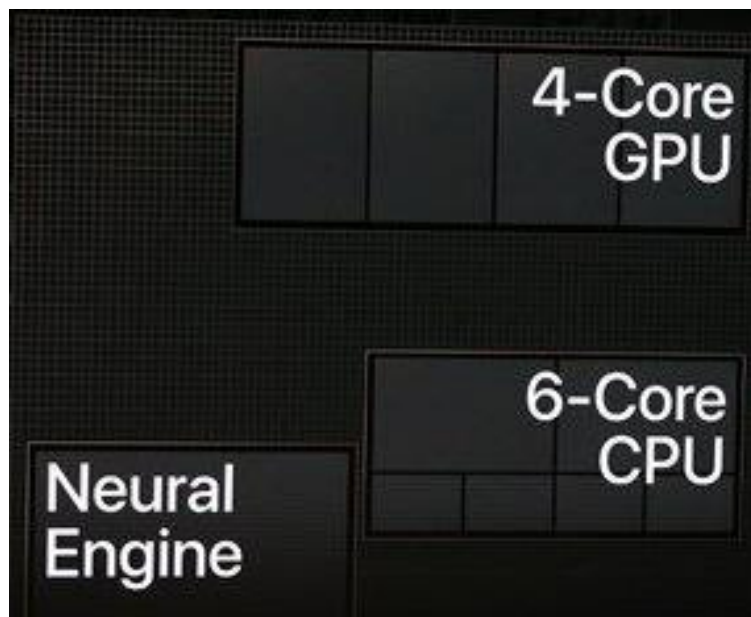
<http://www.cctime.com/html/2018-10-19/1413137.htm>



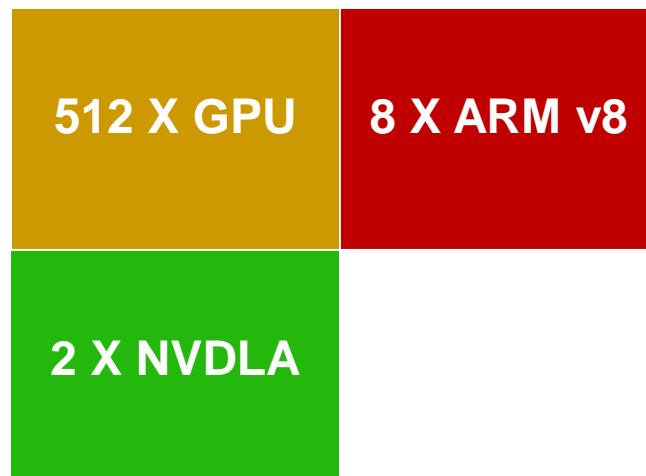
北京大學



A12



Jetson AGX Xavier

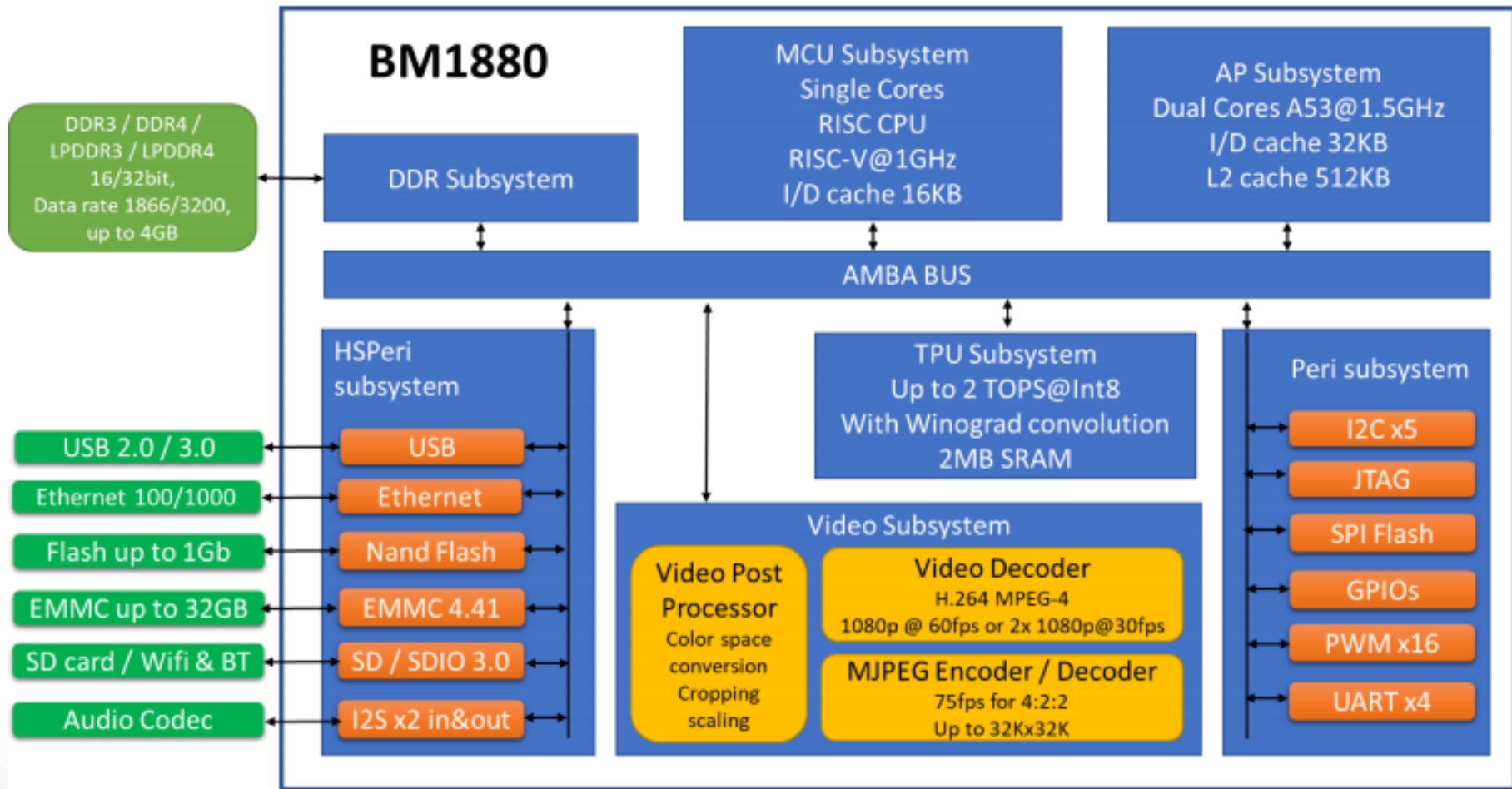


<http://www.cctime.com/html/2018-10-19/1413137.htm>

<http://www.Nvidia.com>



北京大學



<https://www.bitmain.com/>



北京大學



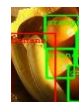
2019年中国嵌入式技术大会

EMBEDDED TECHNOLOGY

Conference China 2019

应用

Kirin
A1X
Snapdragon
.....



AI



Postmates



JD.COM



Jetson

IFM

FOXCONN

Journey



Face++ 旷视

Canon



上汽集团
SAIC MOTOR

HIKVISION

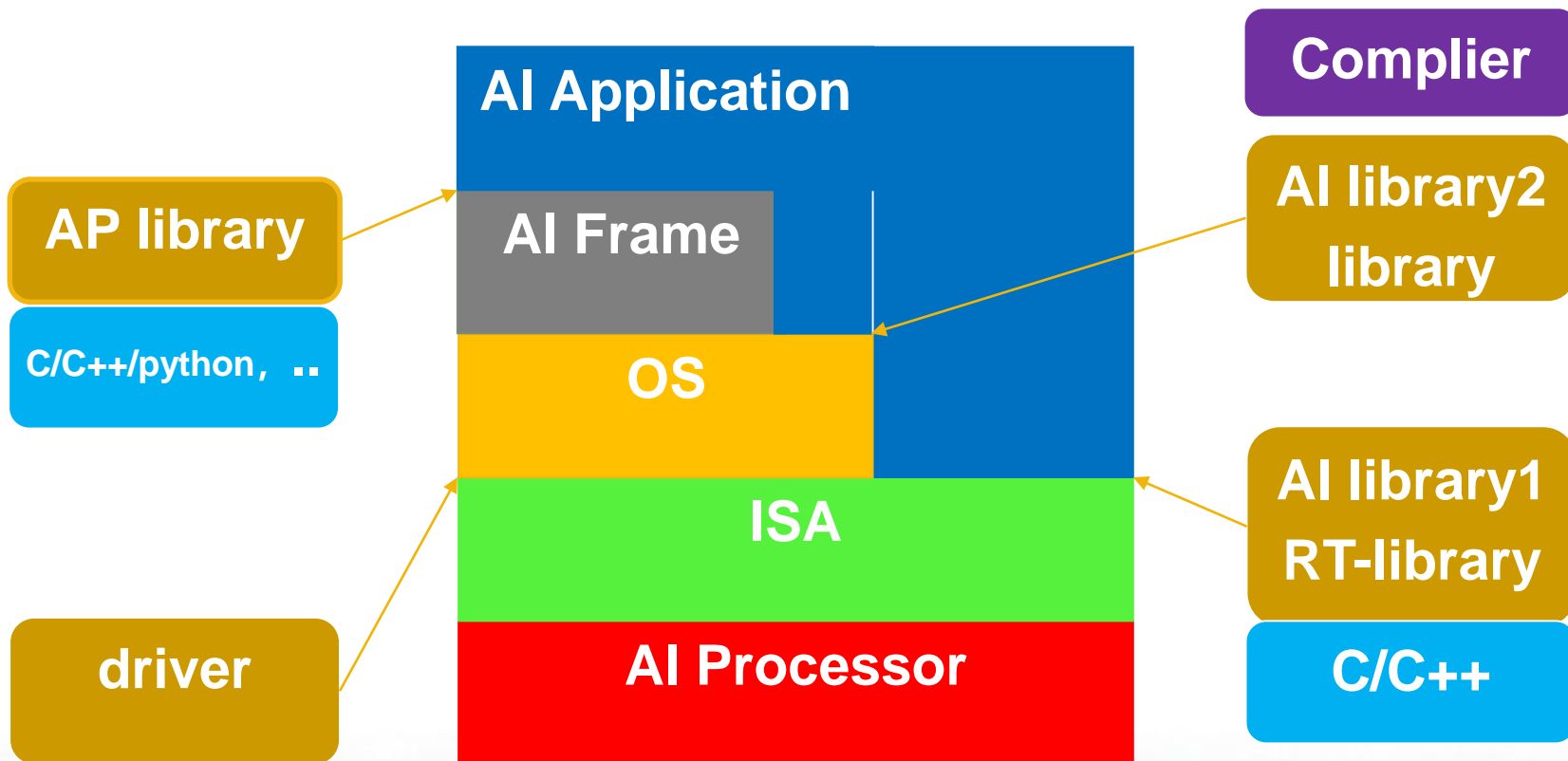


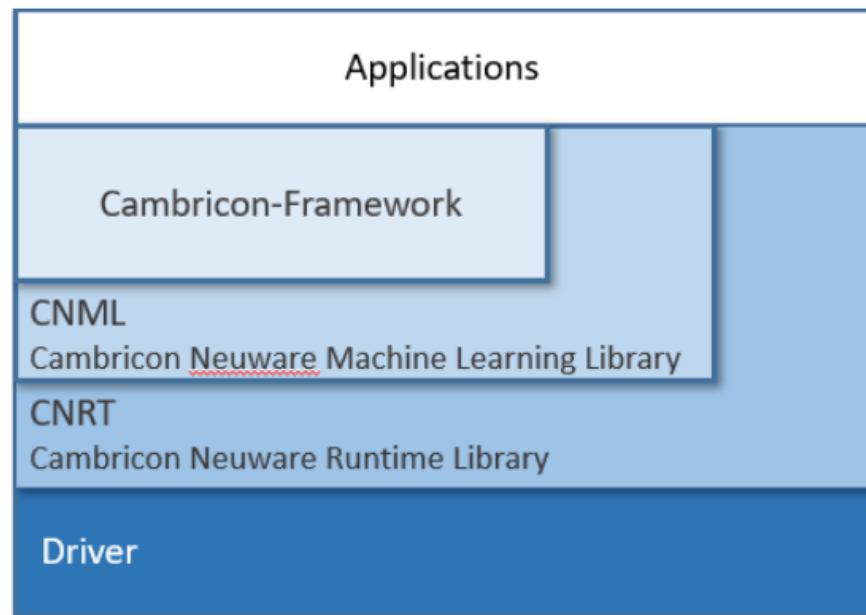
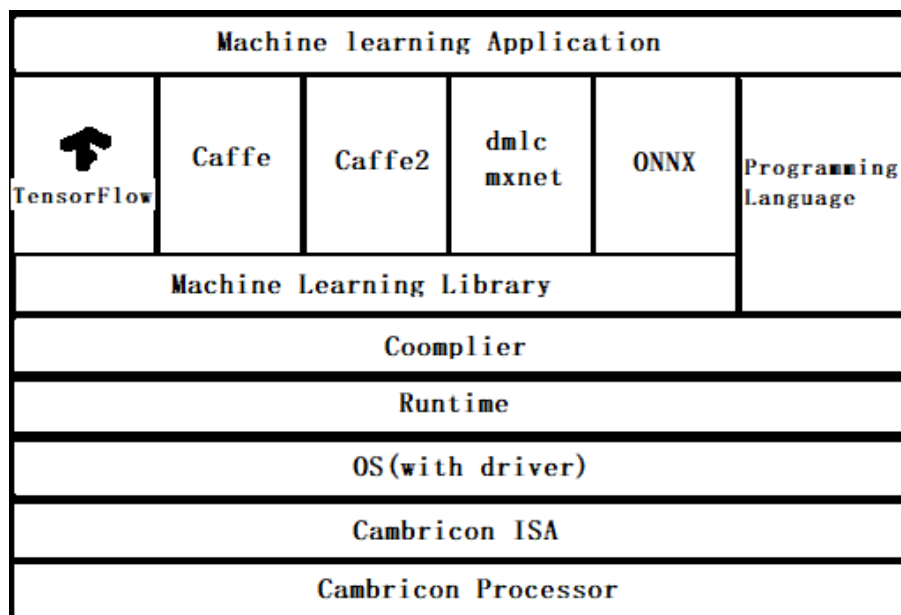
清华大学



运行环境

开发工具







2019年中国嵌入式技术大会

EMBEDDED TECHNOLOGY

Conference China 2019

Nvidia Jetson

DEEPCSTREAM SDK

ISAAC SDK

SOFTWARE MODULES

Depth Estimation

Object Detection

Pose Estimation

Gesture Recognition

Path Planning

Autonomous Navigation

Ecosystem Modules

INSIGHT DEVELOPER TOOLS

JETPACK SDK

Deep Learning

TensorRT
cuDNN

Computer Vision

VisionWorks
OpenCV

Accelerated Computing

cuBLAS
cuFFT

Graphics

Vulkan
OpenGL

Multimedia

libargus
Video API

Sensors

Drivers
Ecosystem

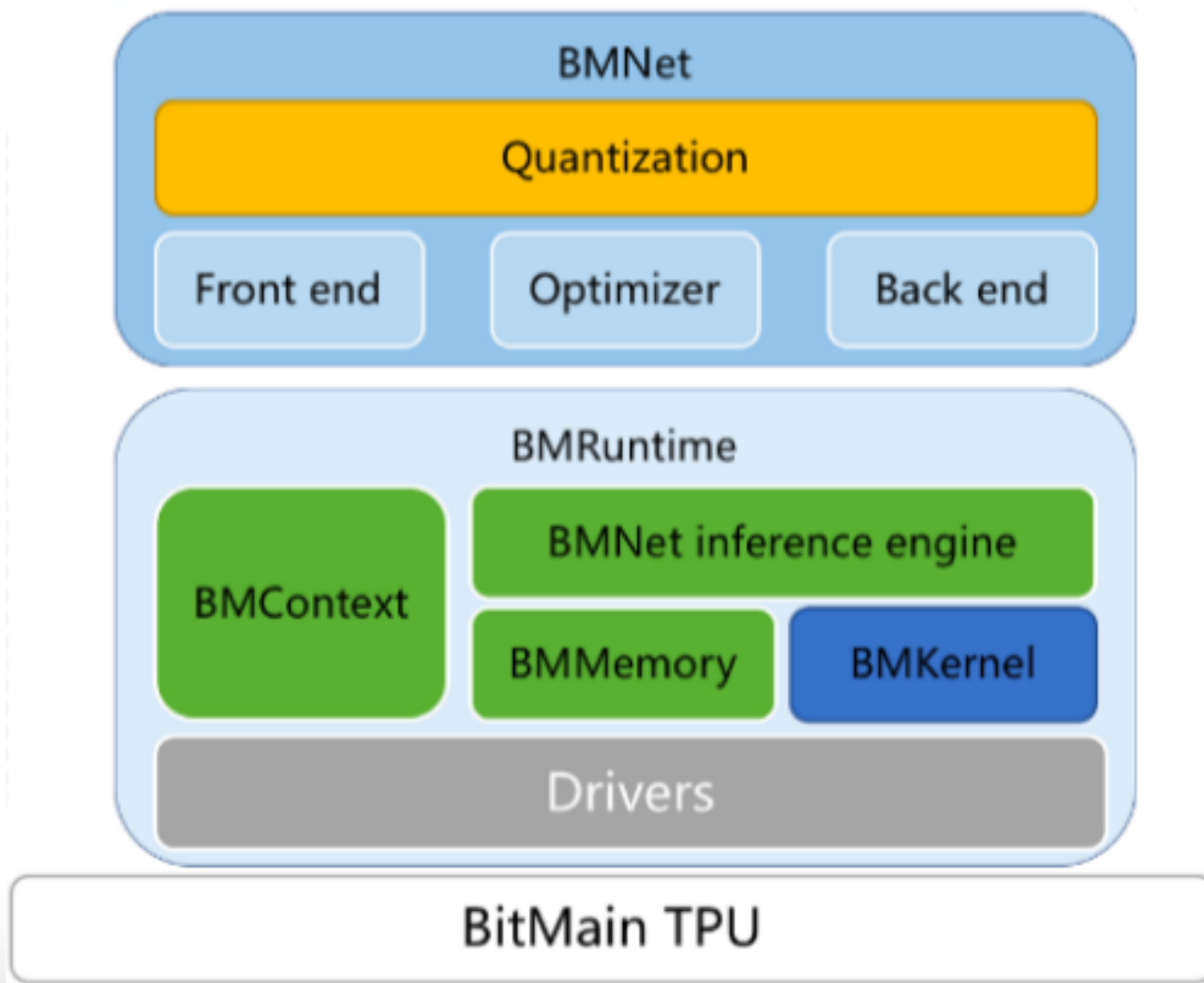
CUDA | CUDA-X AI | LINUX FOR TEGRA | RTOS

JETSON COMPUTER

<https://developer.nvidia.com/embedded/develop/software>



清华大学

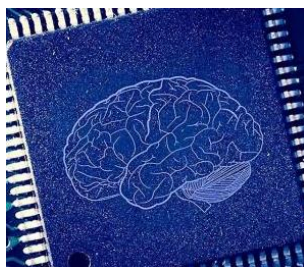




AI chip



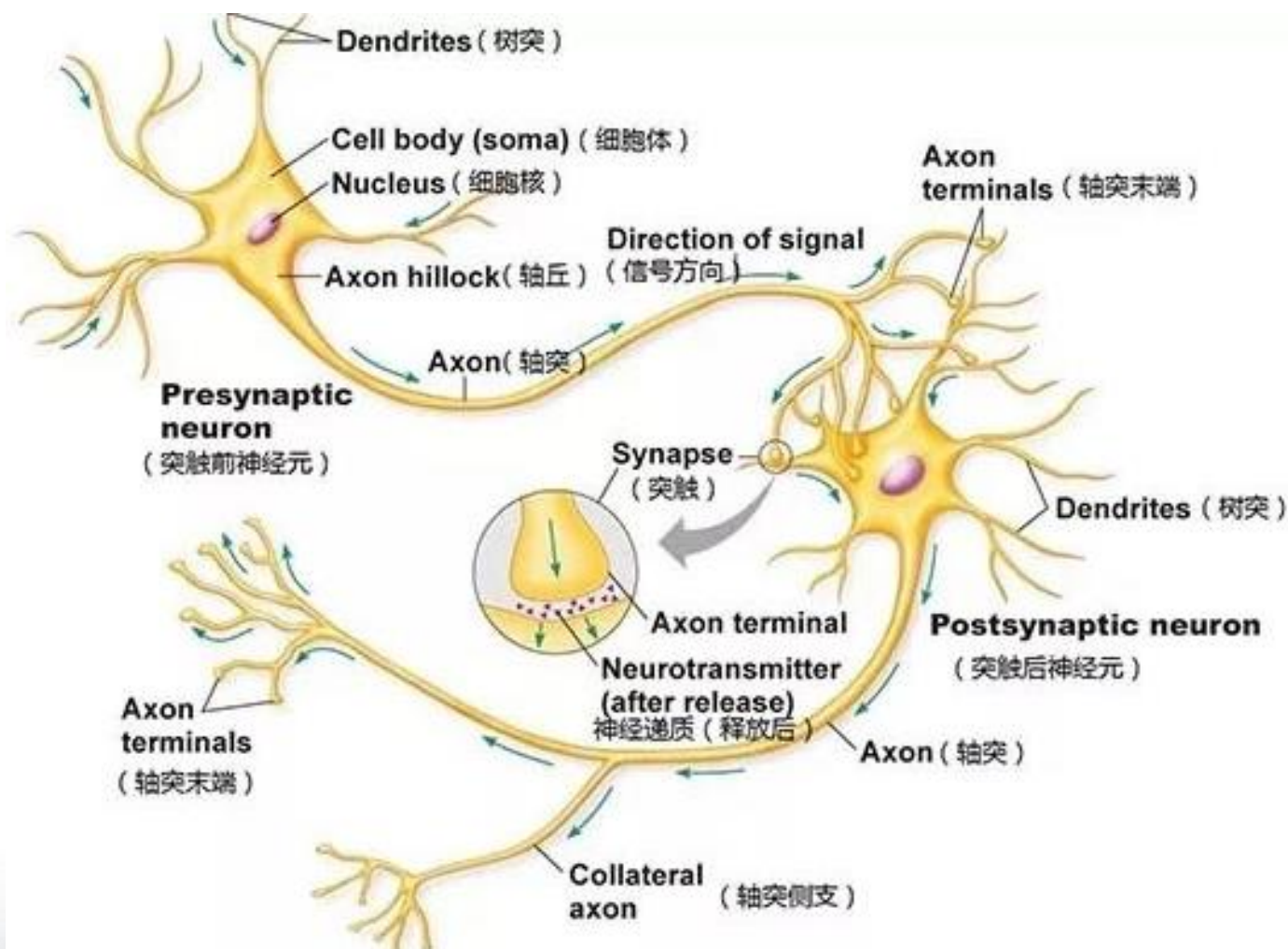
Embedded AI Processor

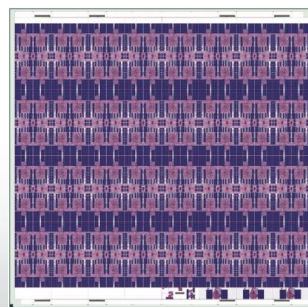
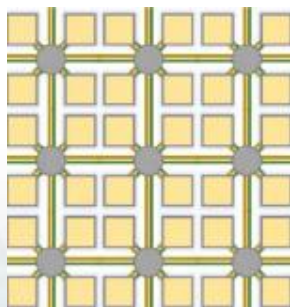
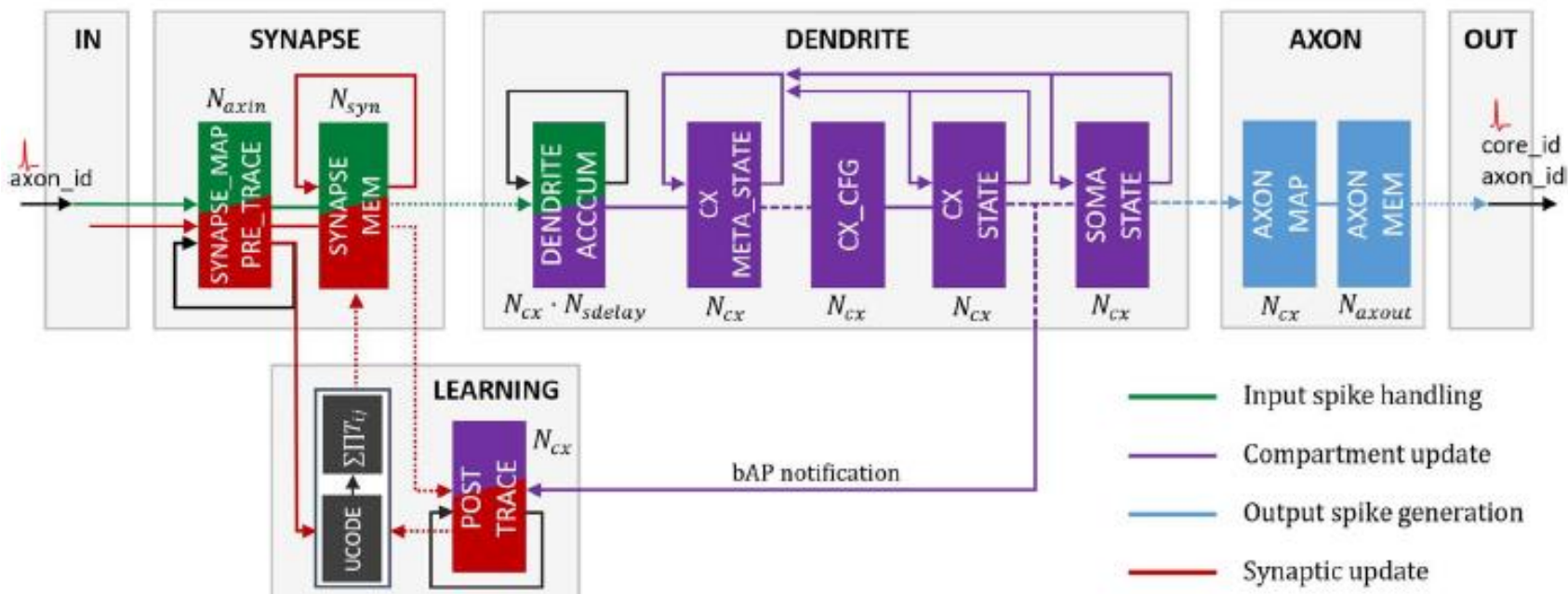


SNNP



The future





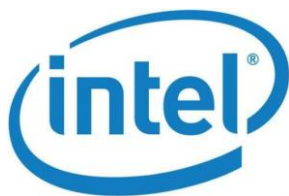


2019年中国嵌入式技术大会

EMBEDDED TECHNOLOGY

Conference China 2019

厂商/团队



北京大學



IBM
TrueNorth
1M neurons
250 million synapses
0.046Tops
70mW



Inel
Loihi
0.13 M neurons
130 million synapses



浙江大学
DARWIN
0.15 M neurons
10 million synapses



清华大学
Tianjic
Hybrid
950mw

<http://www.elecfans.com/d/641626.html>

http://st.zjol.com.cn/kj/sb/201908/t20190826_10885846.shtml

- Jing Pei et al, Towards artificial general intelligence with hybrid
- Tianjic chip architecture, Nature, 1st Augst, 2019



北京大學

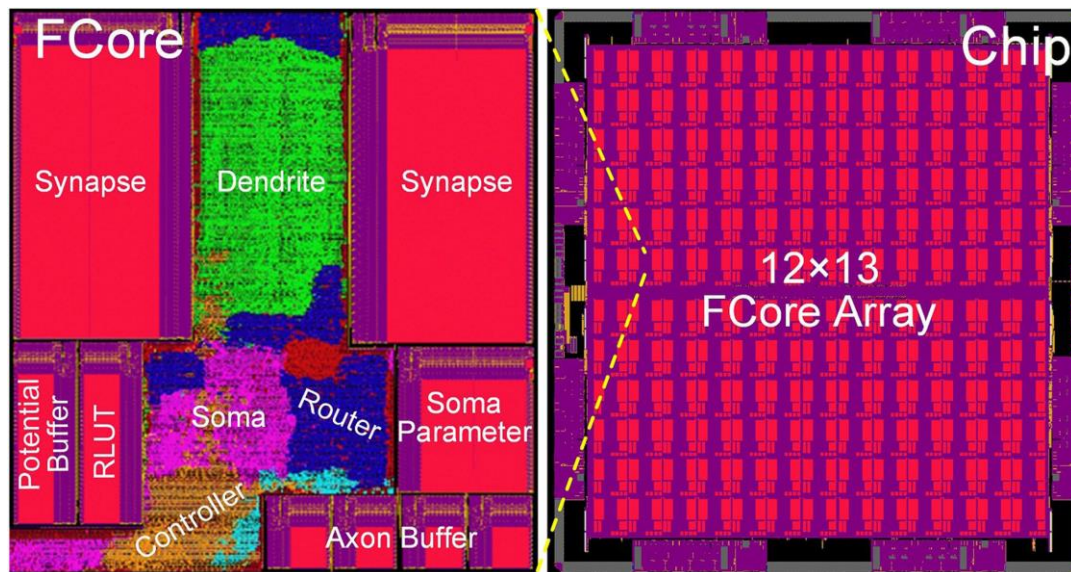


Loihi

Tianjic

128x
Neuromorphic
Core

3X
X86

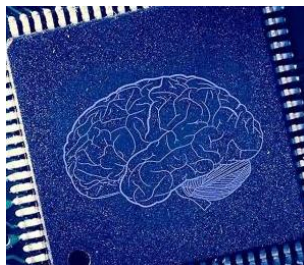




AI chip



Embedded AI Processor



NNP



The future



- 面向应用的SoC 将是嵌入式AI处理器的主流;
- 深度学习单元将动态可配置;
- 软件开发环境和模型转换工具将更通用;
- 类脑处理器将进入实际应用。





2019年中国嵌入式技术大会
EMBEDDED TECHNOLOGY
Conference China 2019

Thank you !

