

深度学习电路与系统的展望

汪玉

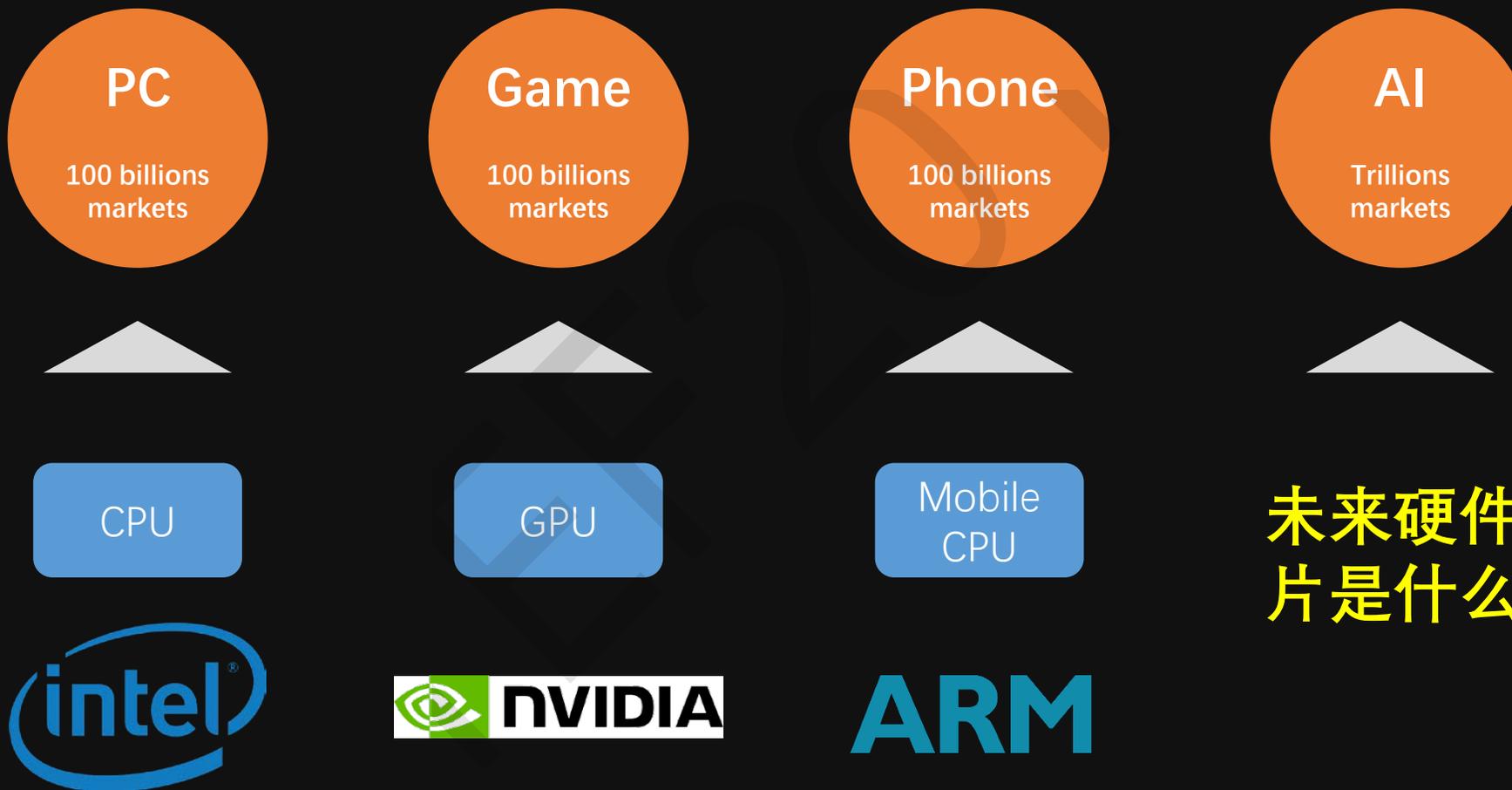
清华大学电子系

yu-wang@tsinghua.edu.cn

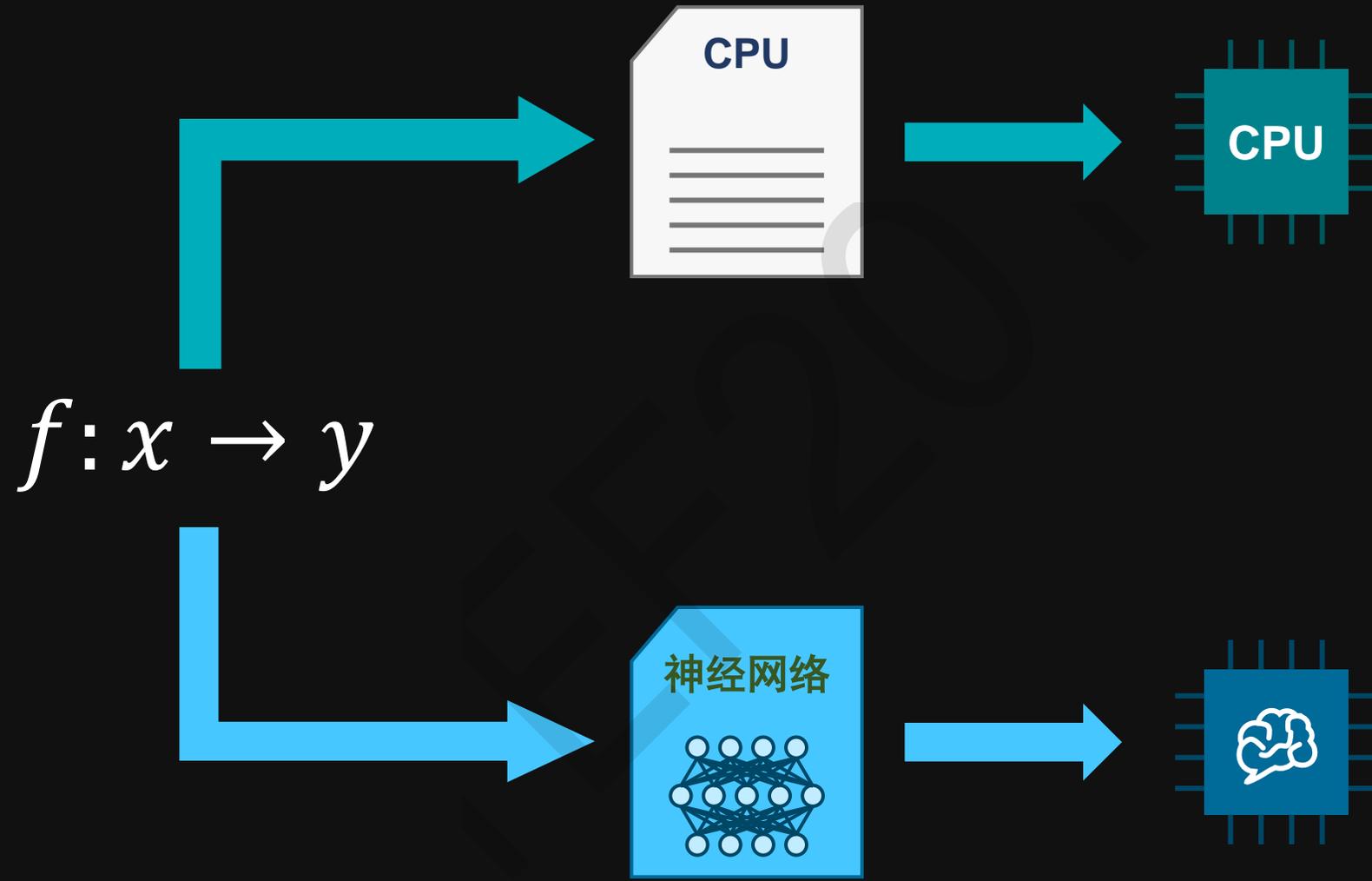
Part

Discover the Philosophy behind Deep Learning Computing

为什么要加速深度学习

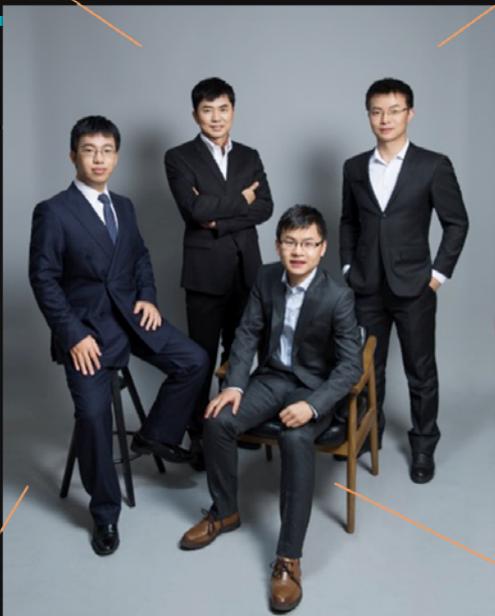


未来硬件芯片是什么？



汪玉 联合创始人

11年体系结构研究经验
 清华大学电子系长聘副教授
 清华大学电子系党委副书记
 国家自然科学基金“优青”获得者
 ACM FPGA技术委员会亚太唯一成员
 IEEE/ACM Senior Member



韩松 联合创始人

全球最前沿深度学习压缩与体系结构研究
 MIT EECS助理教授
 斯坦福大学博士，师从Bill Dally教授
 Deep Compression技术提出者
 ICLR 2016 最佳论文

姚颂 联合创始人&CEO

清华大学电子系毕业
 斯坦福大学电子系访问学者
 前清华大学电子系科协主席
 MIT Tech Review 35岁以下35名创新者
 福布斯亚洲30 Under 30

单羿 联合创始人&CTO

清华大学电子系博士
 伦敦帝国理工联合培养
 IBM PhD Fellowship
 前百度IDL异构计算方向创始成员
 前地平线机器人FPGA技术负责人

以场景切入

深鉴核心平台



注：DNNDK即 Deep Neural Network Development Kit, 深度学习SDK, DPU即 Deep-learning Processing Unit, 深度学习专用处理器

开始智能加速器研发

大量研究者进入该领域

2016.3

公司成立

2016.4, AlphaGo

2016.8, 火热投资

2017, 出现大量相关公司

深鉴产品大量应用

2018.7, 赛灵思完成收购

2012

2013

2014

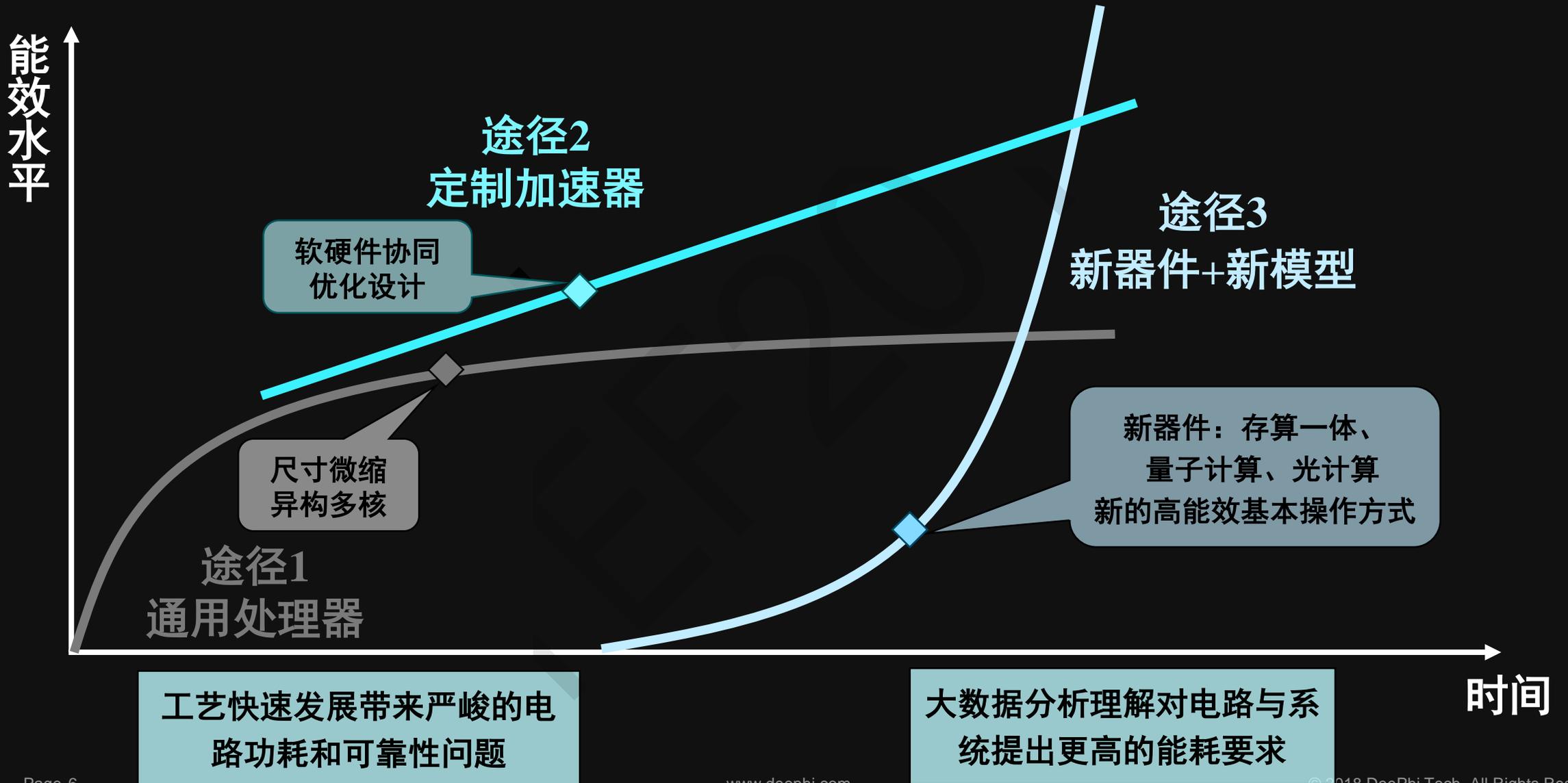
2015

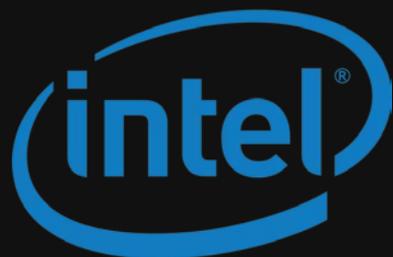
2016

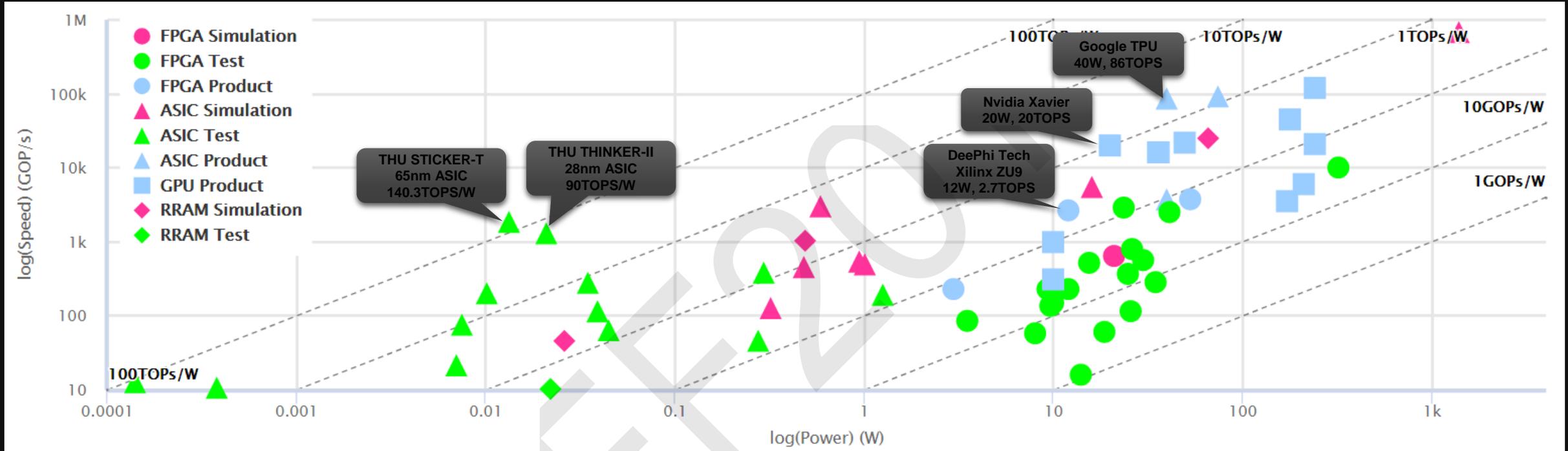
2017

2018

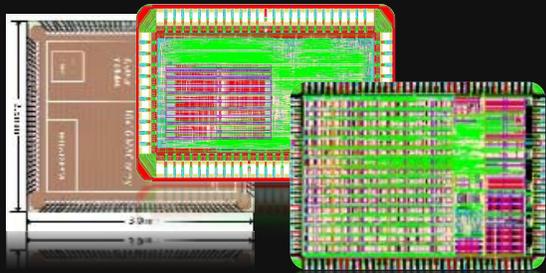
2019



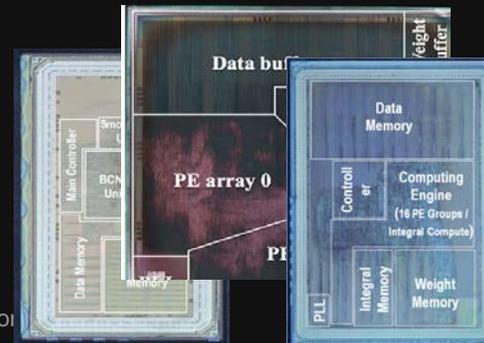




Source: <https://nicsefc.ee.tsinghua.edu.cn/projects/neural-network-accelerator/>



STICKER—serial
边缘计算加速器
清华电子 刘勇攀
ISSCC19 TCAS117 VLSI18



人工智能加速芯片
清华微电 尹首一
ISLPED17 设计竞赛奖
VLSI2018 VLSI2018 VLSI2019

存算一体 Processing - In- Memory

国内:



国外:

国内外起步在
同一起跑线上



www.crossbar-inc.com

MYTHIC

SYNTIANT

量子计算 Quantum Computing

国内:

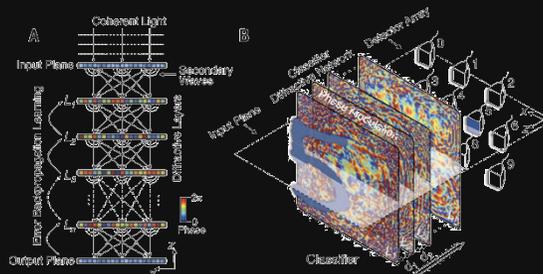


国外:



光计算 Optical Computing

学术界:



ref:Lin, Xing, et al. "All-optical machine learning using diffractive deep neural networks." Science 361.6406 (2018): 1004-1008.

国外:



© 2018 DeePhi Tech. All Rights Reserved.

Part 02

Discover the Philosophy behind Deep Learning Computing

应用定义智能芯片

AI Chip Landscape

Source: 唐杉, "AI Chip List", <https://basicmi.github.io/AI-Chip/>

科技巨头/系统厂商



芯片厂商



IP/设计服务



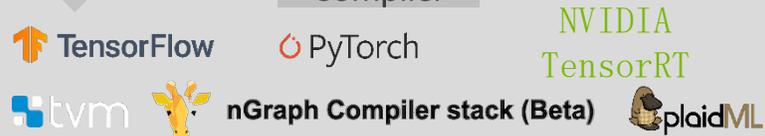
中国初创企业



世界范围初创企业



Compiler



Benchmarks



一般用途

领域定义

应用定义

机器学习

机器学习
(OpenCV+DL)

深度学习
(仅前向推理)

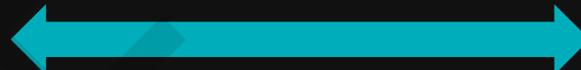
典型的神经网络

数据中心业务

边缘上的视觉应用

特定应用解决方案

通用性



效率

每个人都在做自己的SOC方案：
制造+IP+工具+集成技术

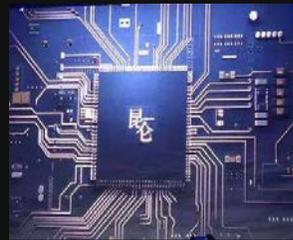
针对特定应用

寒武纪 MLU100 2018年5月



<https://www.1caifu.com/school/news/detail/84/38667>

百度 昆仑 2018年7月



http://m.sohu.com/a/239609885_505862

华为 昇腾 2018年10月



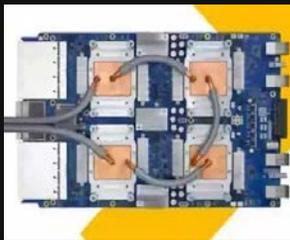
<http://www.eetop.cn/view-6904631.html>

比特大陆 算丰 2018年10月



<http://www.elecfans.com/d/847284.html>

Google TPU 3.0 2018年5月



http://www.sohu.com/a/243238471_99970711

Amazon Inferentia 2018年11月



<https://news.mydrivers.com/1/605/605344.htm>

Intel SpringCrest 2018年5月



<https://www.ithome.com/html/digi/361281.htm>

Xilinx ACAP Versal 2018年10月



<http://www.fmups.com/html/1498025114.html>



云端
计算+存储+分析

边缘
计算+通信

终端
展示+交互

需要提升边缘处理能力

基站, 车辆, 路由器里是什么智能芯片?

Part 03

Discover the Philosophy behind Deep Learning Computing

全球化芯片+系统软件



原料&设备



制造



设计



系统软件



应用软件

Intel创立 8080 286 386 奔腾 奔腾II 酷睿II 酷睿i5 酷睿i9
 1968 1974 1982 1985 1993 1997 2005 2009 2017

CPU: 差距较大

2002 2007 2014
 龙芯一号 龙芯二号 麒麟910T

2006年
 Movidius 创立

AI芯片: 差距较小

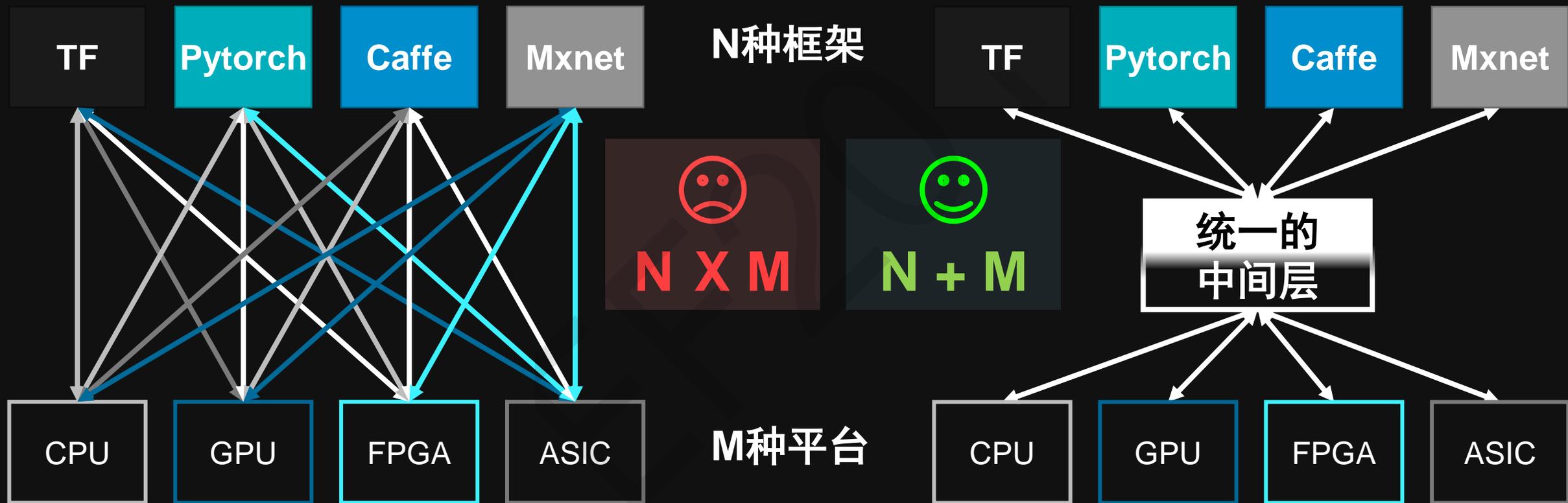
Nvidia TX1 2016年3月 TPU by Google 2016年5月

2015年7月 地平线创立 2016年3月 深鉴科技创立 2016年3月 寒武纪科技创立

新器件: 同一起跑线

MYTHIC创立 2016年 SYNTIANT创立 2017年

2017年
 知存科技

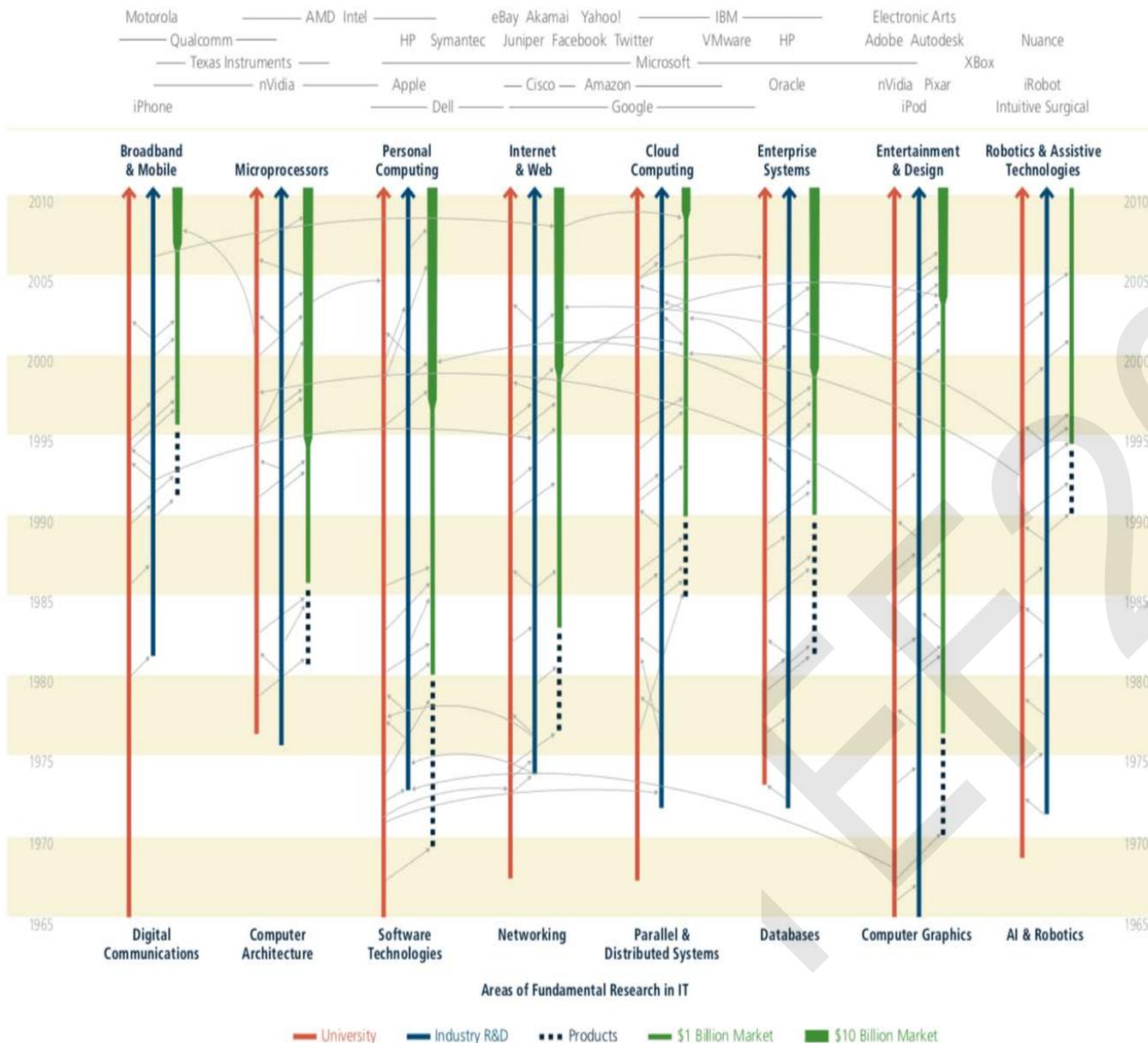


Part 04

Discover the Philosophy behind Deep Learning Computing

中国迫切需要产学研合作

IT Sectors With Large Economic Impact

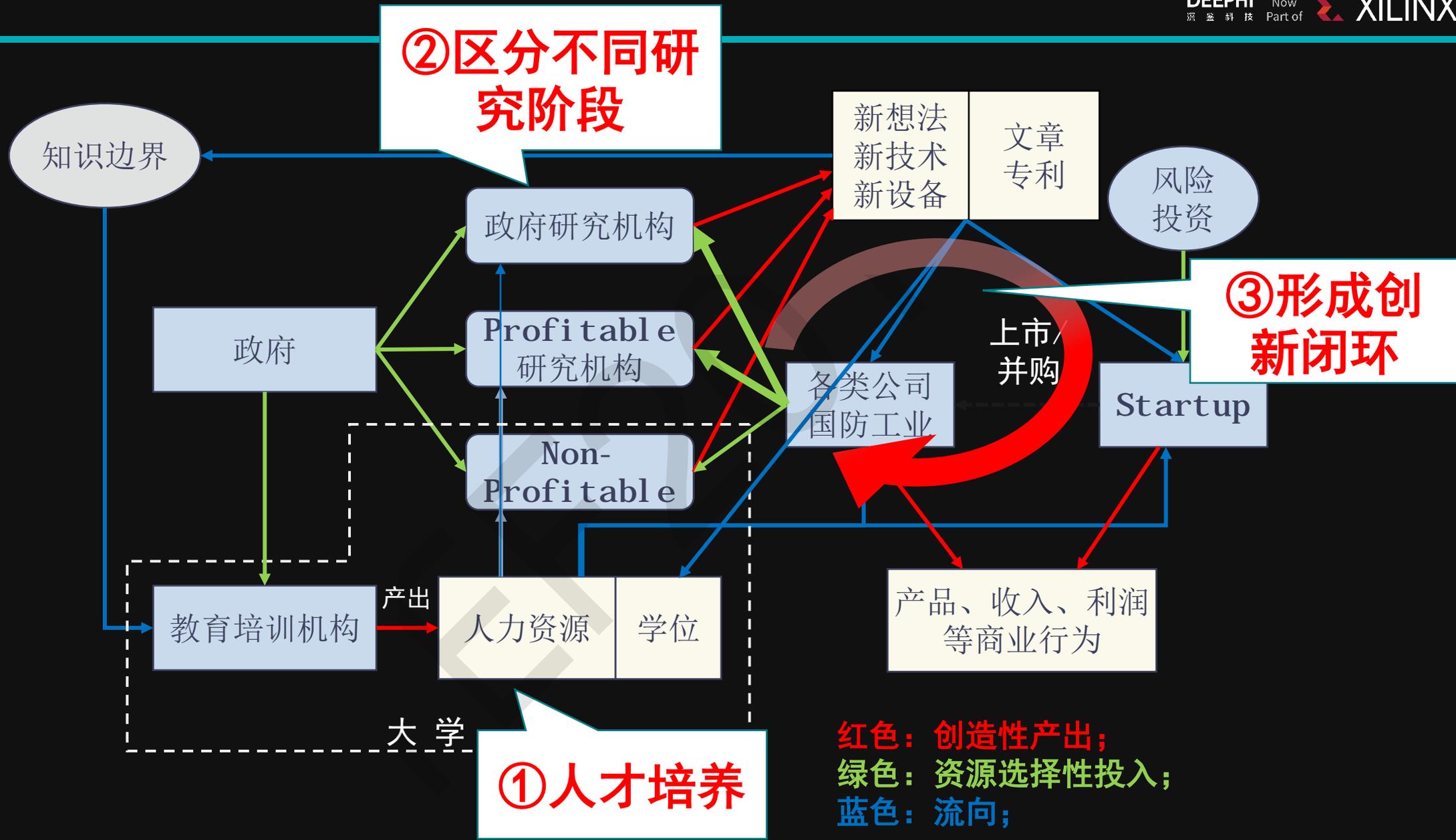


信息产业发展的特点：

- 1、**时间长**：一个颠覆性方向依旧需要10年才能从想法到10亿美金的市场规模
- 2、**依赖于同时期其他诸多先进技术**
- 3、**迭代快**：同一方向的研究迭代快，往往在**工业和学术界之间迭代**
- 4、**系统一般比较大**（复杂），不可能完全由一个公司（机构）完成，对人的需求高

Continue Innovation in Information Technology, NAP, 2012

FIGURE 1 Examples of the contributions of federally supported fundamental research to the creation of IT sectors, firms, and products with economic impact. Tracks added since the 2003 update of the figure are described in Appendix B. See also Box 1 and Appendix C.



- 加强教育，区分阶段，形成创新闭环
- 对于学校
 - 加强**基础理论**的研究投入，以更开放的态度对待知识产权
 - 加强和企业的**共同育人**工作（智能算法基础理论、应用、芯片）
- 产学研合作：
 - 基础研究和应用研究时间维度不同，基础研究可面向潜在的应用开展
 - 建立**信用体系**，以技术为基本出发点，形成**创新闭环**



官方微信



www.deepi.com

THANK YOU!

Thursday, August 1, 2019

Discover the Philosophy behind Deep Learning Computing