Xilinx Vision & Strategy for the Adaptable World

> Victor Peng, President & CEO

XILINX

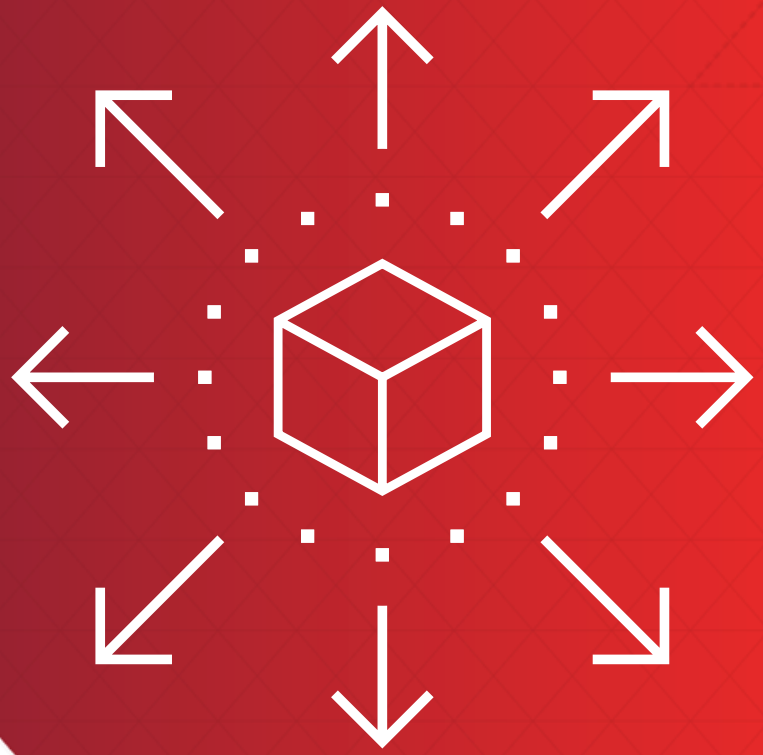
01

Explosion of Data

> 90% unstructured

> Video & image content

> Needs higher throughput & real-time computing



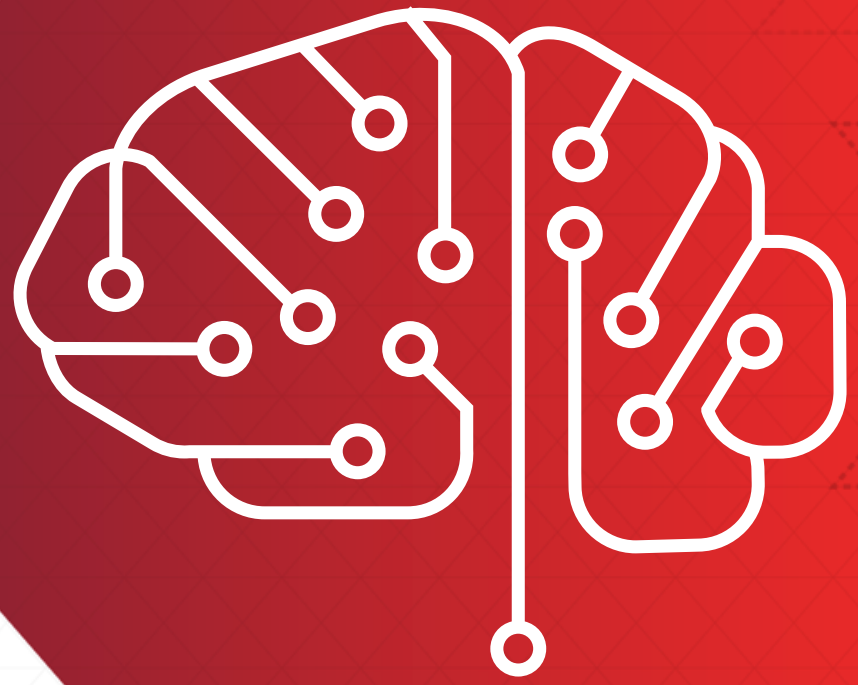
02

Dawn of AI

> Adoption across all industries

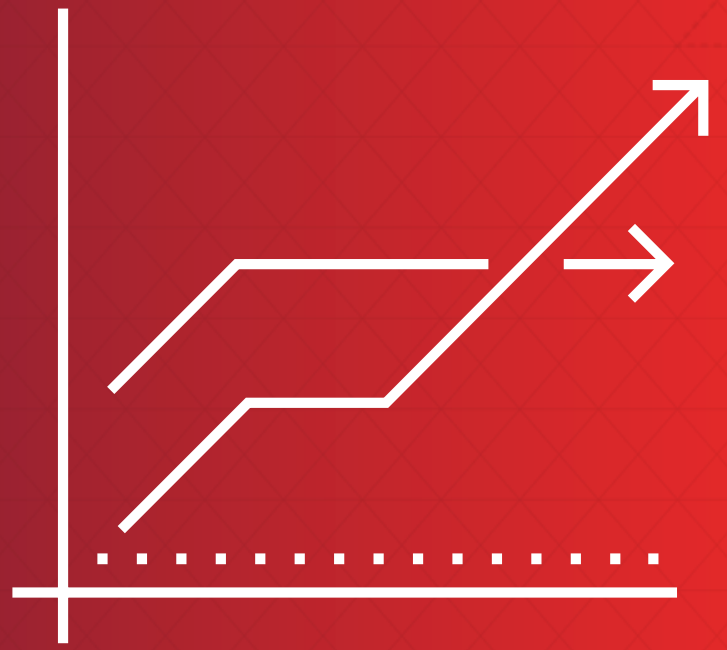
> Injecting new intelligence into apps

> From endpoints to edge to cloud



03

Computing After Moore's Law



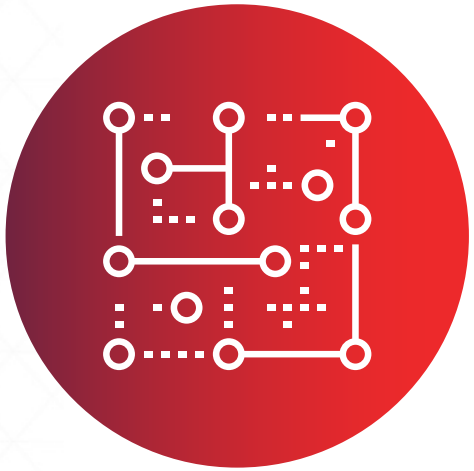
> Heterogeneous computing with accelerators

> Breadth of apps require different architectures

> Speed of innovation outpacing silicon design cycles

The Need for Adaptable Intelligence

The intelligent connected world needs adaptable accelerated computing.



Everything Intelligent & Connected

Deployed at Global Scale

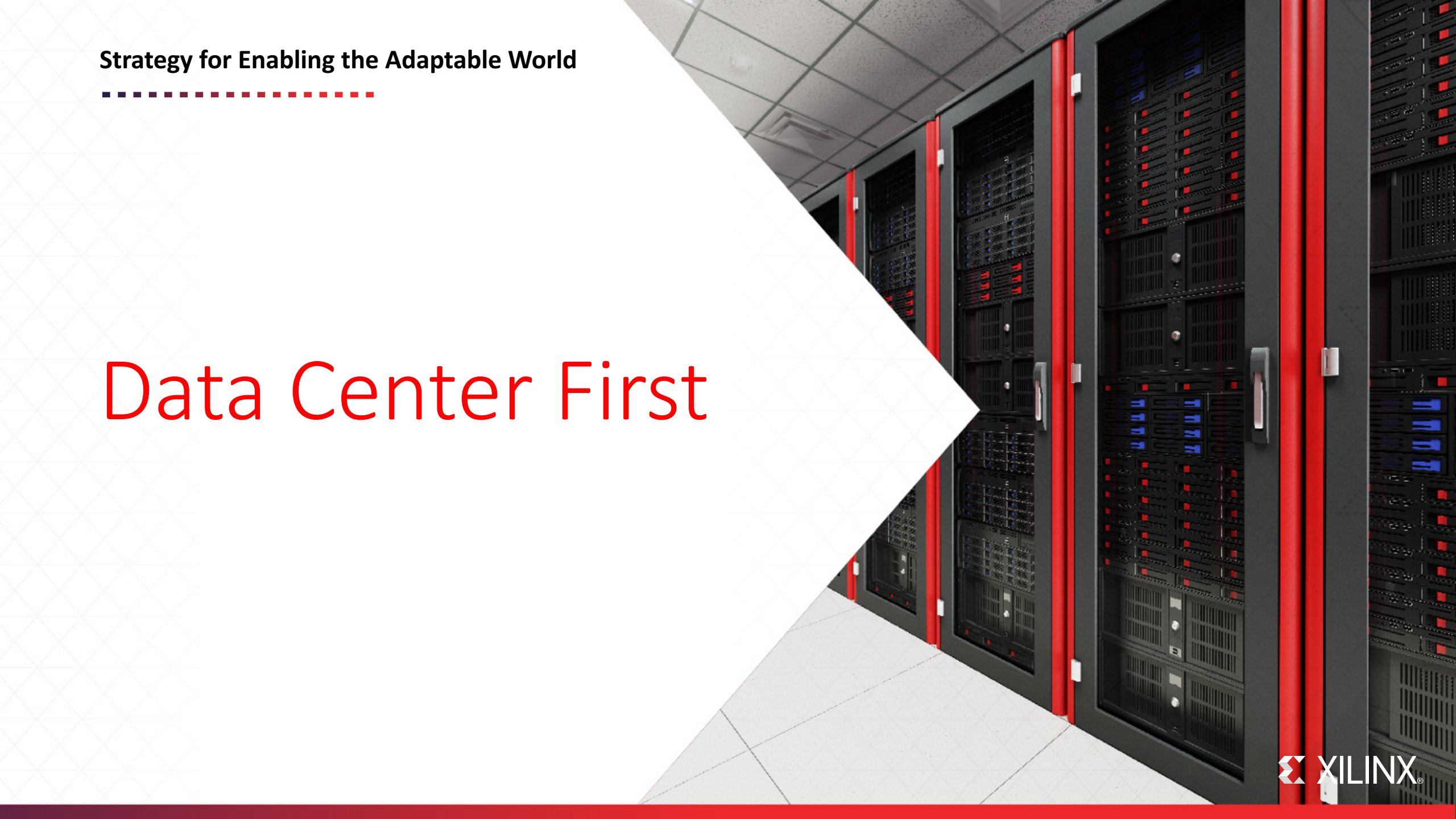
Dynamic Needs & Rapid Innovation

Strategy for Enabling the Adaptable World

XILINX®

Strategy for Enabling the Adaptable World

Data Center First



XILINX

Reaching Software Application Developers

Accelerated Open Frameworks

TensorFlow™ gatk FFMPEG

Accelerated Libraries

Machine learning

H.265 HEVC High Efficiency Video Coding

Database analytics

OpenCV

Development Environment

SDAccel™ Environment

SDSoC™ Environment

Development Boards

Software Application Developers

Development Stack

System Developers

XILINX®

Growing Data Center Compute Ecosystem

Applications, Tools & Communities

bitfusion
DEEPHI 深鉴科技
edico genome
NGCODEC NEXT GENERATION VIDEO COMPRESSION
RYFT

Cloud Development & Deployment (FPGA as a Service)

aws
Alibaba Cloud
HUAWEI
Baidu 百度
NIMBIX
Tencent Cloud

Technology & Systems

CCIX
AMD
arm
HUAWEI
IBM
Mellanox TECHNOLOGIES
QUALCOMM

XILINX

Compute Acceleration

Compute Acceleration

- Machine Learning **40x***
- Video & Image Processing **10x***
- Data Analytics **90x***
- Genomics **100x***

XILINX

Genomics Use Case: Personalized Medicine

Whole genome diagnosis to treat critically ill newborns

Analysis reduced from 1 day to 20 minutes

Patient-specific genomics dynamically optimized

GUINNESS WORLD RECORDS™ RECORD HOLDER

Rady Children's Hospital San Diego

edico genome

XILINX®

Computational Storage & Network Acceleration

Computational Storage

SmartNICs & Network Acceleration

Strategy for Enabling the Adaptable World

Accelerate Growth in Core Markets

XILINX

Accelerate Growth in Core Markets

Automotive

Wireless Infrastructure

Wired Communications

Audio, Video, & Broadcast

Aerospace & Defense

Industrial, Scientific & Medical

Test, Measure, & Emulation

Consumer

XILINX®

Strategy for Enabling the Adaptable World

Drive Adaptive Computing

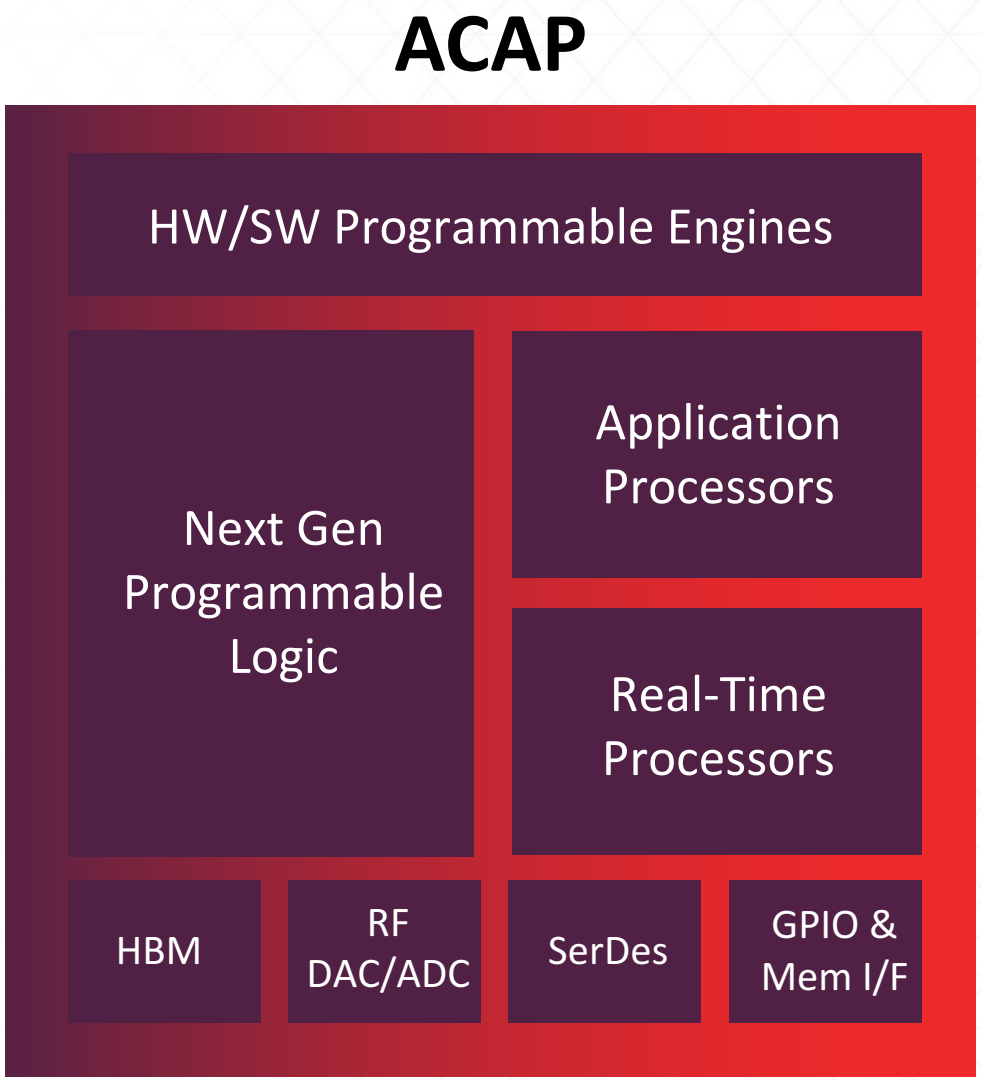
XILINX

New Device Category for Adaptive Workload-Specific Acceleration

> HW/SW programmable engines

> IP subsystems and a network-on-chip

> Highly integrated programmable I/O

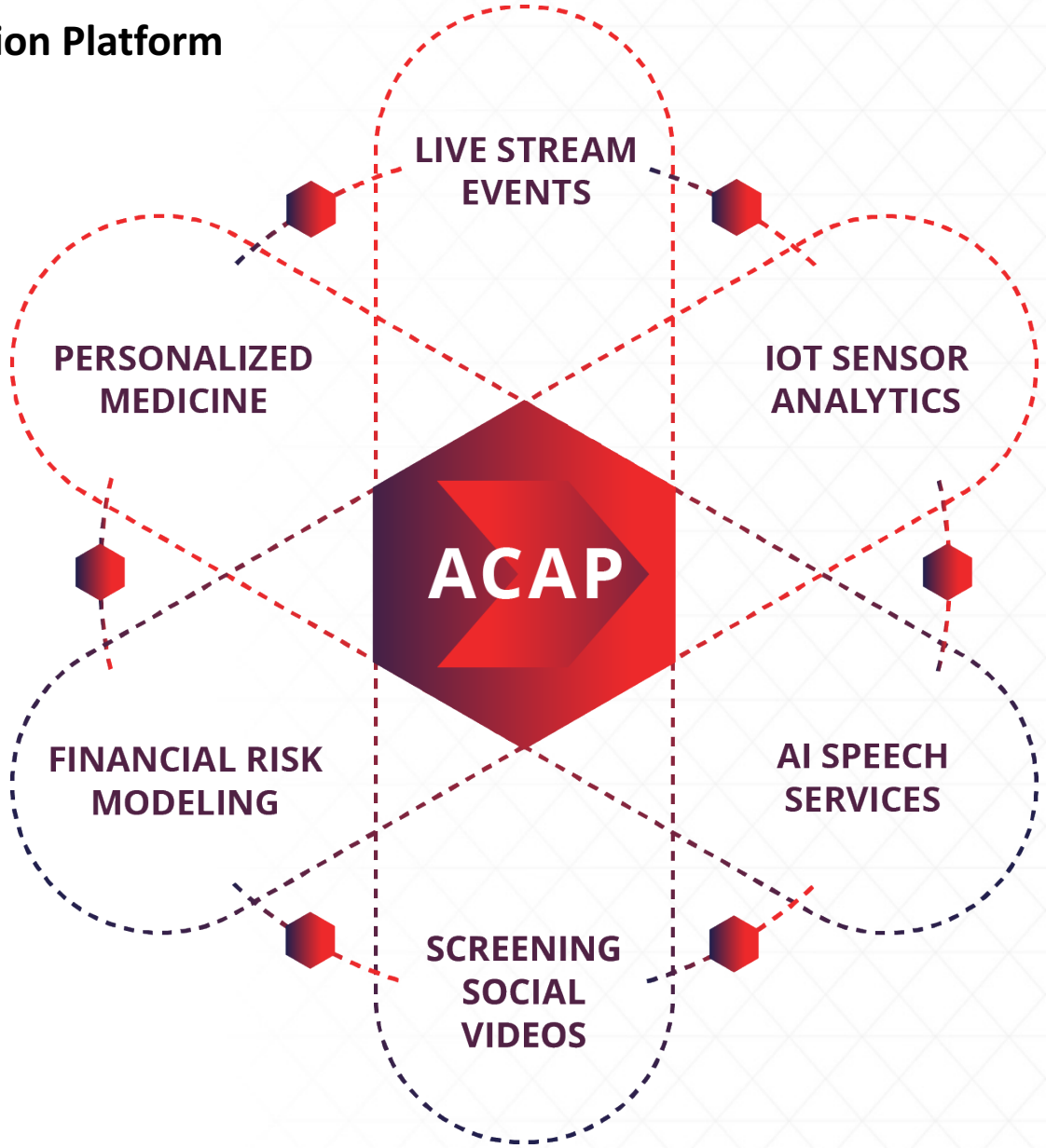


Benefits of an Adaptive Compute Acceleration Platform

Dynamically Adaptable to Workloads

Exponential Increase in Acceleration

Software Programmable



Custom acceleration for any workload – in milliseconds

XILINX®

Project “Everest”

The First 7nm ACAP
Product Family

XILINX®

Project “Everest”

4	1,500	50B	>$1B
Years	Engineers	Transistors	R&D

Everest Breakthroughs vs Current Generation

Revolutionary Adaptability
Dynamic Optimization for Workloads

Software & Hardware Users
Rapid Innovation & Deployment

20x**
AI Compute Capability

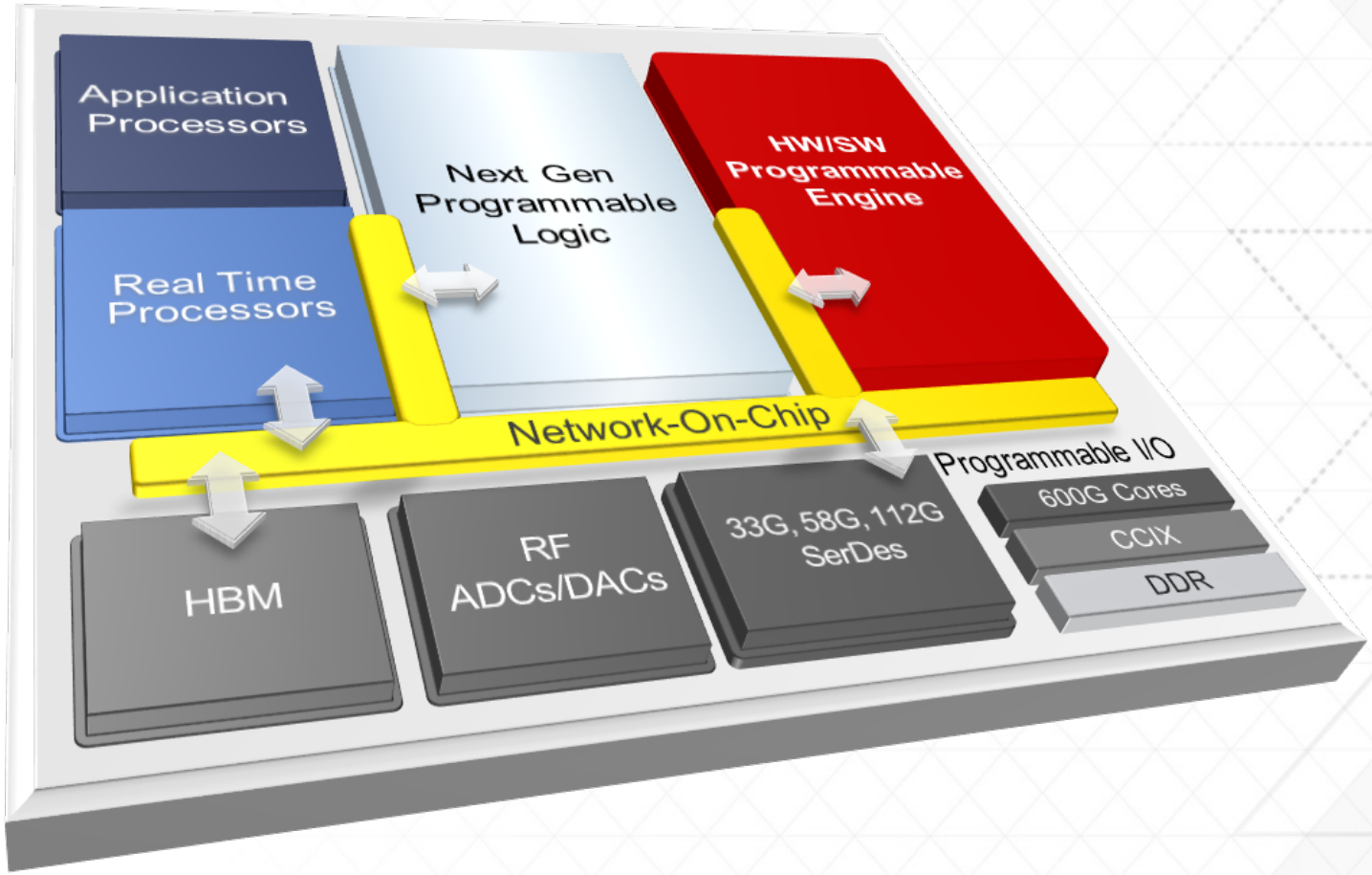
4x**
5G Communications Bandwidth

Timeline

> Early software tools available to strategic customers

> Silicon tape out this year

> First shipment next year



Vision - Adaptable Intelligent World

Strategy

Data center first

Accelerate growth of core vertical markets

Drive adaptive computing

ACAP – a new product category

Accelerating broad range of workloads with dynamically adaptable silicon

10-100x faster than CPUs for new workloads, more use cases than GPUs or ASICs

Project “Everest”

1st ACAP implemented in 7nm, tape out 2018

SW and HW programmable

>10X in performance, performance/watt

Thank you

XILINX®

Footnotes

Machine Learning: 40x Deephi. LSTM inference. KU060 vs Xeon Core i7 5930k. Xilinx delivers 43x perf and 40x perf/watt versus Xeon. Perf = Latency reduction. Benchmark: TIMIT, an acoustic-phonetic continuous speech dataset.This effort won best paper at FPGA2017. **Video & Image Processing: 10x** NGCodec. Transcoding on HEVC. Comparison on AWS F1 vs C4.8xlarge instances. (VU9P vs Xeon E5-2666v3) H.265 v2.3, 1080p50 @ 2Mbps. High quality profile. One F1 instance transcodes at same throughput as 10 C4 instances. **Data Analytics: 90x** Ryft. ElasticSearch on 1 TB logfile (unstructured data) Comparison on AWS F1 vs C4.8xlarge instances. F1 ElasticSearch analysis takes 41 minutes (.68 hours) C4 ElasticSearch analysis takes 62.5 hours **Genomics: 100x** Edico Genome. Next Gen Sequencing (NGS) analytics on whole human genome. Guinness World record set by Stephen Kingsmore, M.D., D.Sc., president and CEO of Rady Children's Institute for Genomic Medicine at Rady Children's Hospital CPU Server: 33 hours to complete NGS analytics using Xeon server. Edico Server: 20 minutes. Multiple Virtex 7 FPGAs in appliance Not related to 100x claim, but a second Guinness World record set by Children's Hospital Philadelphia for fastest simultaneous NGS for 1000 whole human genomes. Ran on 1000 AWS F1 instances http://www.bio-itworld.com/2017/10/23/childrens-hospital-of-philadelphia-edico-set-world-record-for-secondary-analysis-speed.aspx

****20X** AI Compute" is based on a NIC / Data Center comparison with Virtex UltraScale+ for Machine Learning inference for image recognition. Everest compute is equivalent to 20x VU9Ps (Data Center's most widely deployed FinFET FPGA) running all their DPS resources (7,000 DSP slices) at max performance. **4X** 5G Bandwidth" is based on a Massive-MIMO 16x16 radio implementation comparison, comparing Everest to our latest RFSoC devices at 16nm. Everest leverages advanced 5G accelerators for 4X signal processing performance to implement 16x16 **800MHz** digital radio (which can be deployed as part of larger Massive MIMO antenna arrays). Latest RFSoC devices have signal processing bandwidth (4,000 DSP slices) for **200MHz** of 5G spectrum.