Xilinx Vision & Strategy for the Adaptable World

> Victor Peng, 社長兼最高経営責任者

XILINX

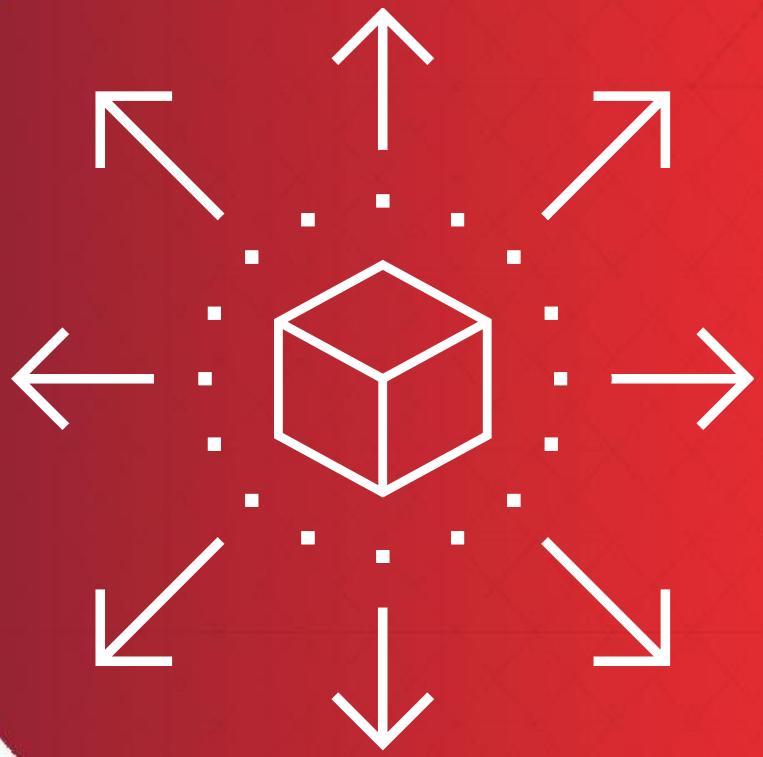
01

データの爆発的増加

> 90% が非構造化データ

> ビデオ/画像コンテンツ

> より高いスループットとリアルタイム
コンピューティングが必要



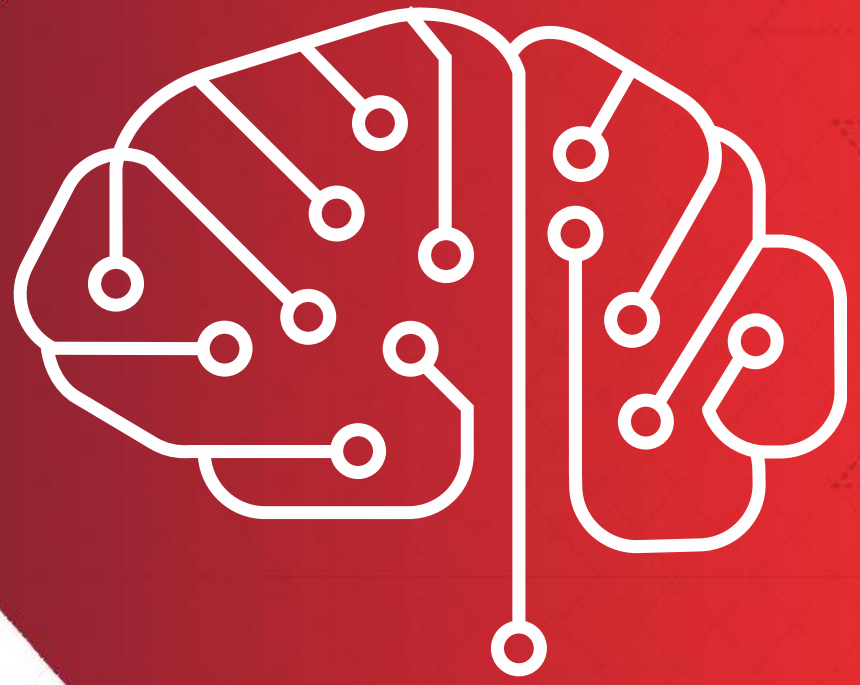
02

AI 時代の到来

> あらゆる業界で導入

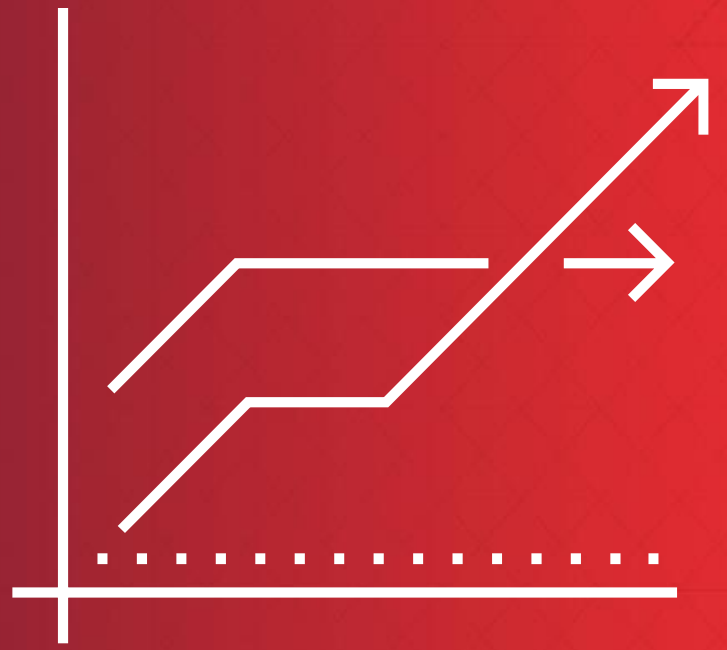
> 新しい知能をアプリケーションに注入

> エンドポイントからエッジ、クラウドへ



03

「ムーアの法則」後のコンピューティング



> アクセラレータを使用したヘテロジニアスコンピューティング

> アプリケーションの幅が広く、何種類ものアーキテクチャが必要

> イノベーションの速度がシリコン デザインサイクルを凌駕

適応性を備えたインテリジェンスの必要性

インテリジェントなコネクテッド ワールドでは
適応性のあるアクセラレーテッド コンピューティングが必要

すべてがインテリジェント/
コネクテッドに

グローバル規模での
展開

動的なニーズ、
急速なイノベーション

適応型世界の
実現に向けた戦略

XILINX®

適応型世界の実現に向けた戦略

データ センター ファースト

XILINX®

ソフトウェア アプリケーション開発者へのリーチ

アクセラレーテッド
オープン フレームワーク

TensorFlow™

gatk

FFMPEG

アクセラレーテッド
ライブラリ

機械学習

H.265
HEVC
High Efficiency Video Coding

データベース
解析

OpenCV

開発環境

SDAccel™
Environment

SDSoC™
Environment

開発 ボード

ソフトウェア
アプリケーション
開発者

開発スタック

システム開発者

成長を続けるデータ センター コンピューティングのエコシステム

アプリケーション、
ツール、
コミュニティ

bitfusion
DEEPHI 深鑑科技
edico genome
NGCODEC NEXT GENERATION VIDEO COMPRESSION
RYFT

クラウド開発
および展開
(FPGA-as-a-Service)

aws
Alibaba Cloud
HUAWEI
Baidu 百度
NIMBIX
Tencent Cloud

テクノロジ
およびシステム

CCIX
AMD
arm
HUAWEI
IBM
Mellanox TECHNOLOGIES
QUALCOMM

XILINX

コンピューティング アクセラレーション

コンピューティング
アクセラレーション

機械学習 **40x***

ビデオ/画像処理 **10x***

データ解析 **90x***

ゲノミクス **100x***

XILINX

ゲノミクスのユース ケース: 個別化医療

全遺伝情報の診断による
危篤新生児の治療

解析時間が 1 日から
20 分に短縮

患者のゲノムに応じて
治療を動的最適化

GUINNESS WORLD RECORDS
RECORD HOLDER

Rady Children's Hospital San Diego

edico genome

XILINX

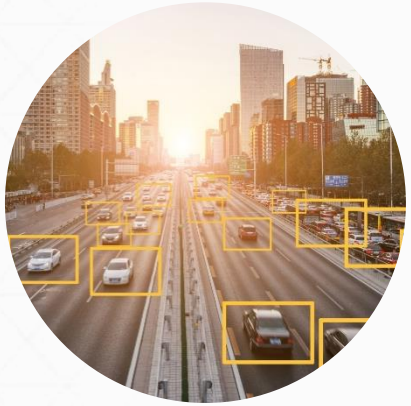
コンピューティング ストレージとネットワーク アクセラレーション

コンピューティング
ストレージ

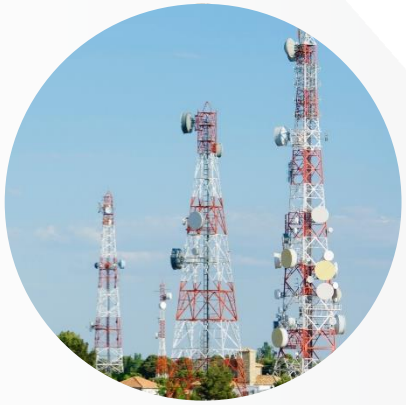
SmartNIC および
ネットワーク アクセラレーション

重要市場の成長を加速

重要市場の成長を加速



オートモーティブ



無線インフラストラクチャ

有線通信

オーディオ、ビデオ、放送

航空宇宙、防衛

産業、科学、医療

テスト、計測、エミュレーション

民生

XILINX®

適応型コンピューティングを推進

FPGA から ACAP (Adaptive Compute Acceleration Platform) へ

ワークロードの種類に適応した
アクセラレーションを可能にする
新しいデバイス カテゴリ

> HW/SW プログラマブル エンジン

> IP サブシステムとネットワーク オン チップ

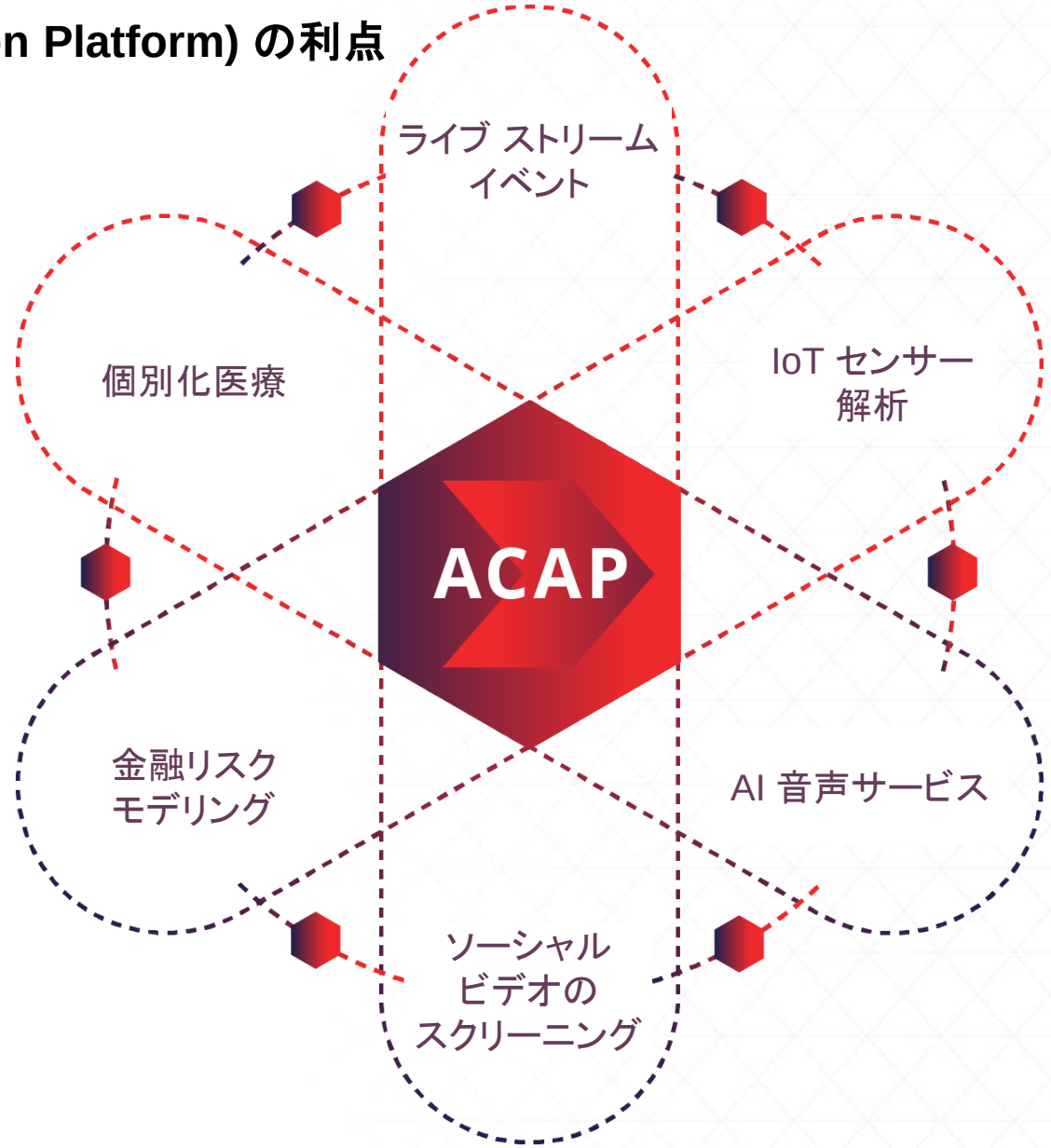
> 高度に統合されたプログラマブル I/O

ACAP (Adaptive Compute Acceleration Platform) の利点

ワークロードに
動的に適応

アクセラレーション性能が
指数関数的に向上

ソフトウェアで
プログラム可能



あらゆるワークロードにミリ秒単位で適応する
カスタム アクセラレーション

XILINX

Project Everest

初の 7nm ACAP
製品ファミリ

XILINX

Project Everest

4	1,500	50B	>$1B
年	エンジニア人数	トランジスタ数	研究開発費

Everest のブレークスルーと現行世代の比較

革新的な適応性
ワークロードに合わせた動的最適化

ソフトウェア/ハードウェア ユーザー
イノベーションから展開までの期間を短縮

20x**
AI 演算性能

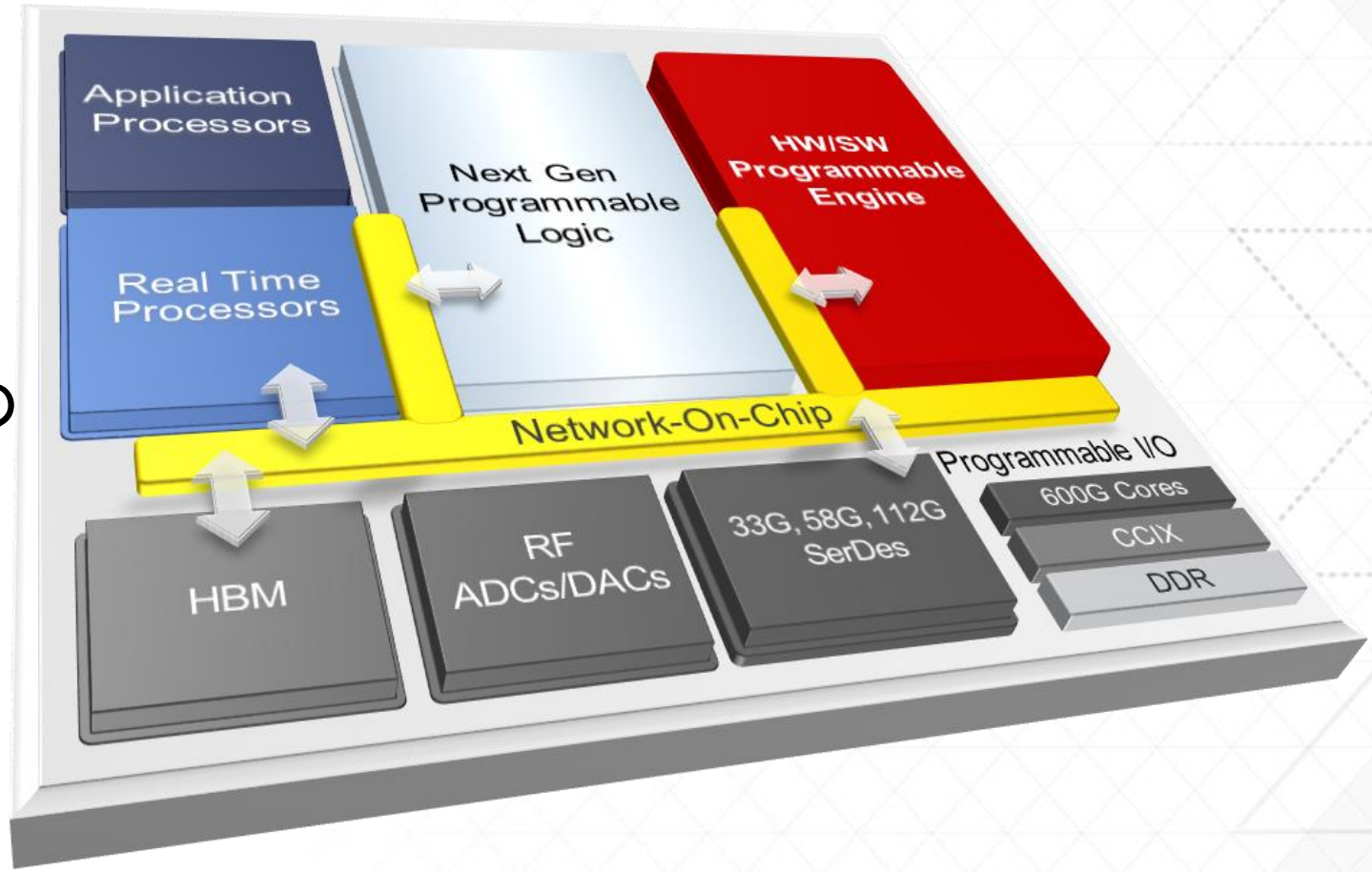
4x**
5G 通信帯域幅

今後のスケジュール

> 一部顧客にソフトウェア ツールのアーリー アクセス版を提供

> 年内にシリコン テープアウト

> 来年に出荷開始



ビジョン - 適応性のあるインテリジェントな世界

戦略

データ センター ファースト
重要市場の成長を加速
適応型コンピューティングを推進

ACAP – 新しい製品カテゴリ

動的な適応性を備えたシリコンにより、幅広いワークロードのアクセラレーションをサポート
注目のワークロードでは CPU に比べ 10 ～ 100 倍の性能で、
GPU や ASIC に比べ幅広いユース ケース

Project Everest

初の 7nm ACAP インプリメンテーション、2018 年内にテープアウト
SW および HW プログラマブル
パフォーマンスと単位ワットあたりの性能比が 10 倍以上に

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.