

# EdgeSense™



OEC Workshop, December 2019

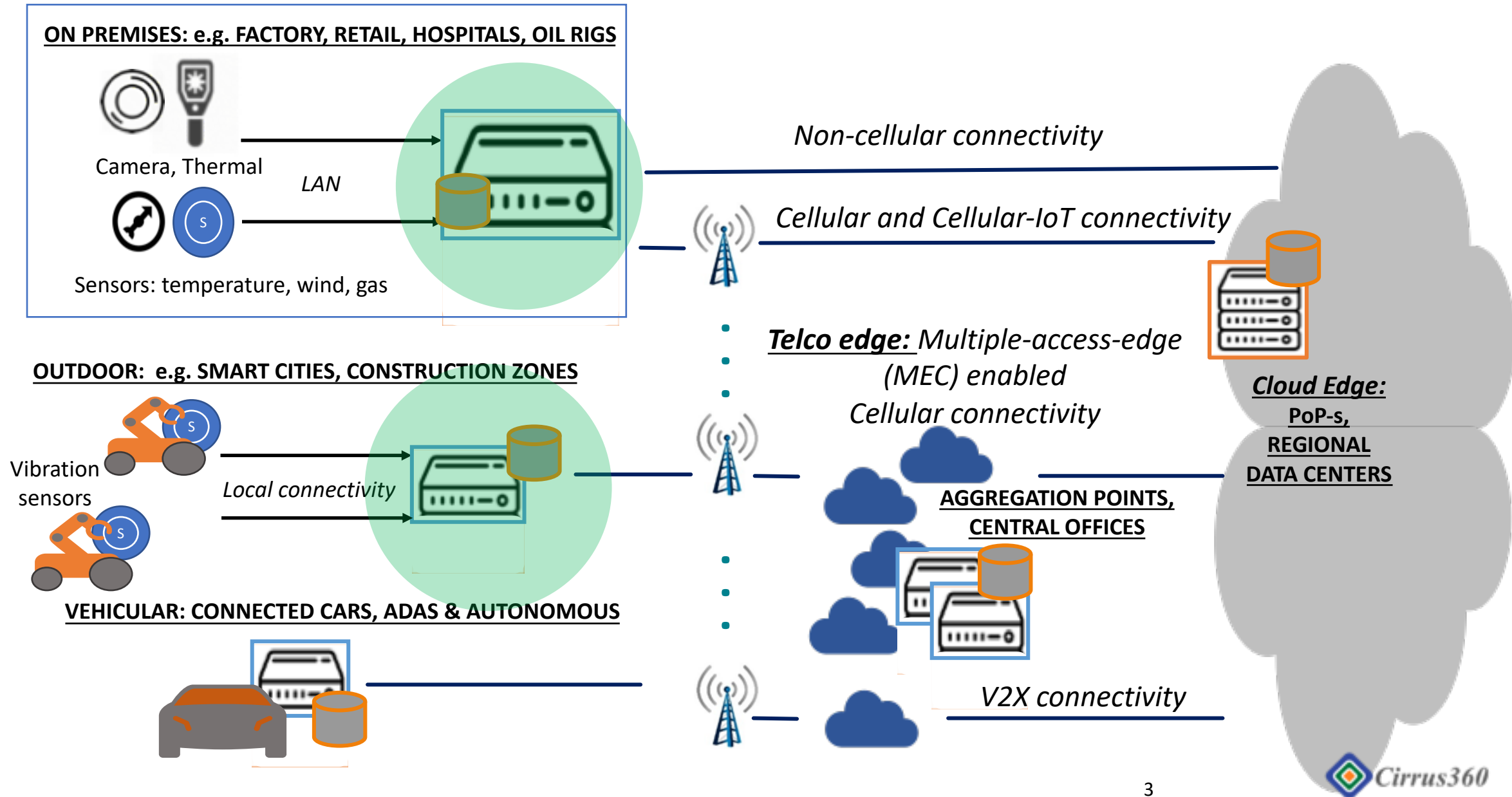
Sudipta Sen, [sudipta@cirrus3sixty.com](mailto:sudipta@cirrus3sixty.com) +1 972 822 0882

Chaitali Sengupta, [chaitali@cirrus3sixty.com](mailto:chaitali@cirrus3sixty.com) +1 214 733 2803

# Cirrus360

- Based in Richardson TX, Launched in 2018
- January 2019: US DOE Small Business Innovations & Research (SBIR) award:
  - EdgeSense™: Smart Sensors and FPGA-based Heterogenous Edge Computing Unit.
  - Motivated by use of edge computing by “beamline scientists” in government labs, within HPC workflow
- Founder: Chaitali Sengupta, PhD.:
  - Experience: VP Reliance Jio India; Sr. Director Qualcomm, Co-founder SNRLabs; Distinguished Member Tech. Staff, Texas Instruments.. Winner of MIT Tech Review’s top innovators under 35 award (TR35).
- Founder: Sudipta Sen:
  - Experience: VP Reliance Jio, India; Sr. Director Qualcomm; Co-founder and CEO at SNRLabs (acquired by SEVEN Networks), Texas Instruments, Compaq. Expertise in hardware and ODM based development.
- Rice University collaborator: Prof. Joseph R. Cavallaro and 4 member team
- Partnerships and support: Actian (data management), Intel (FPGA/hardware), Xilinx (FPGA)
- *Partner for smart factory: Aisleconnect, Taiwan led by Paul Lin (ex-Microsoft, ex-Oracle)*
- *Collaboration with customers (Manufacturing SMEs)*

# Edge: Many shapes, many sizes, ... many services!



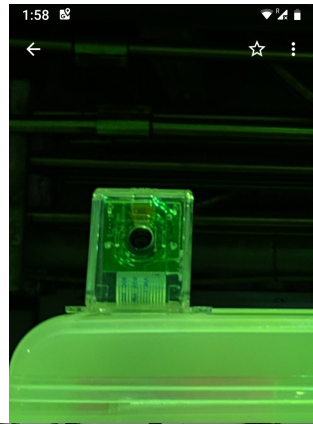
# Cirrus360 Edge/IIoT vision

- Target markets: Smart manufacturing (Industry 4.0) SMEs, Government, Smart “spaces”
- Current typical customer journey with IoT/Edge computing:
  1. Assessment of use cases: return on investments (ROI) vs. total cost of ownership (TCO) e.g. effective decision making across organization hierarchies, savings from predictive maintenance of equipment, increased revenue from improved customer visibility
  2. Solution identification: Often with help of (expensive!) consultants
  3. Experimentation: Prototyping, measurement of cost vs benefits
  4. Deployment options: (4a) Purchase vertically-integrated separate solutions “as-a-service” for each use case e.g. surveillance, asset tracking, predictive maintenance OR, (4b) Build custom solutions

Current IIoT/Edge customer journey => Time consuming, Expensive, In-efficient

## Cirrus360 vision:

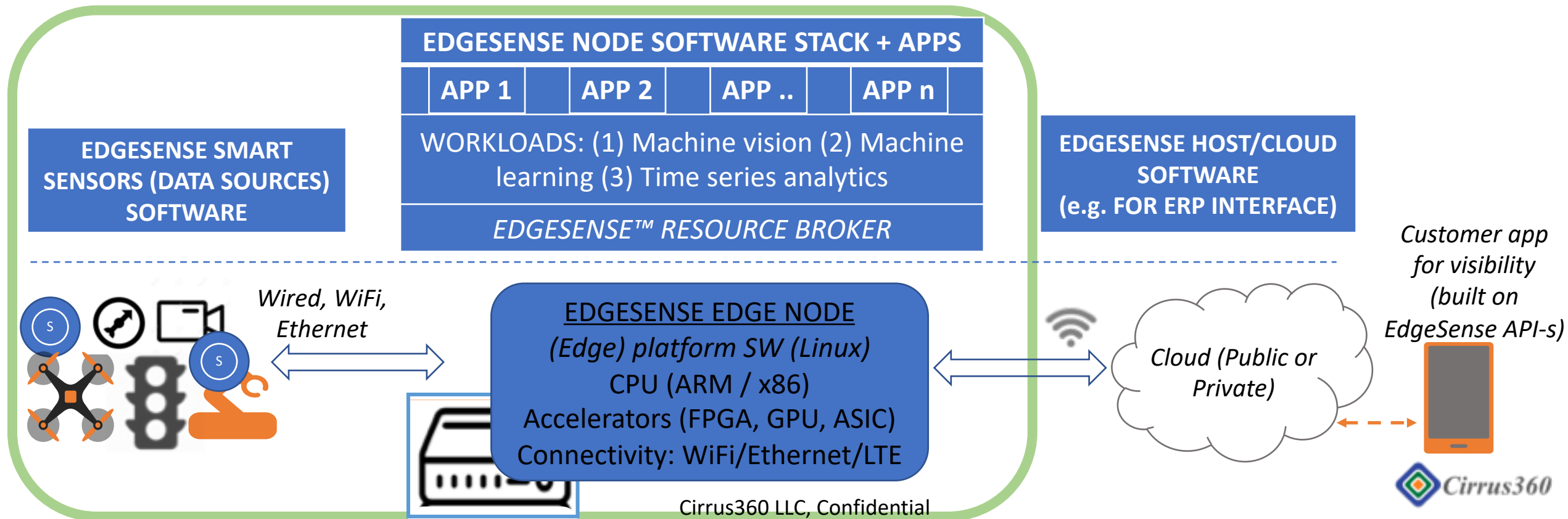
- (A) Enable our customers to *customize and deploy sensors-edge-cloud applications* on their factory floor, construction zones, workshops,
- (B) Lower TCO and increase ROI via a platform shared across use cases, and
- (C) Deploy in production based on tangible business value and ROI.



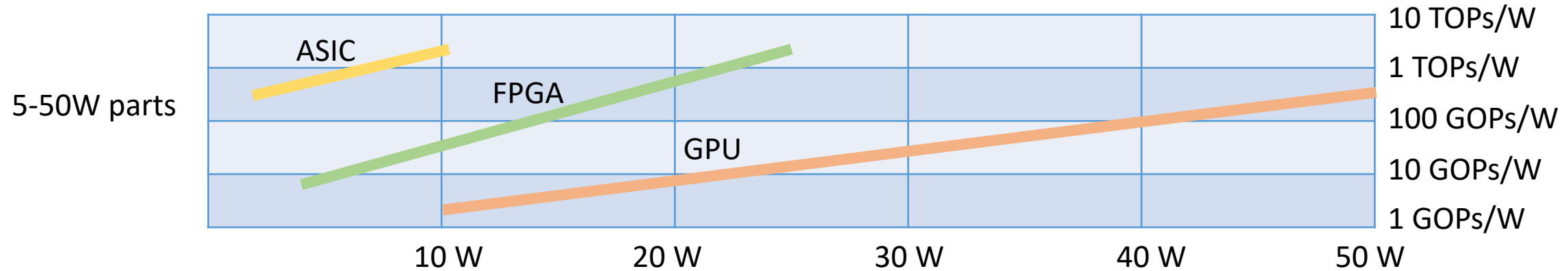
# Cirrus360 EdgeSense™

Cirrus360 designs  
in blue

- EdgeSense™ vision: **Application fabric for “smart spaces” edge-native application market place**
  - Support workflows (APP in picture) focused on data analytics and machine learning (AI/ML) inference, mapped to data sources (sensors in picture); running on edge node composed of off-the-shelf hardware
  - EdgeSense™ Resource Broker (ERB): Intelligent software to manage deployment and workflow including processing on accelerators such as FPGA-s, GPU-s



# GPU vs ASIC vs FPGA based ML (NN) acceleration



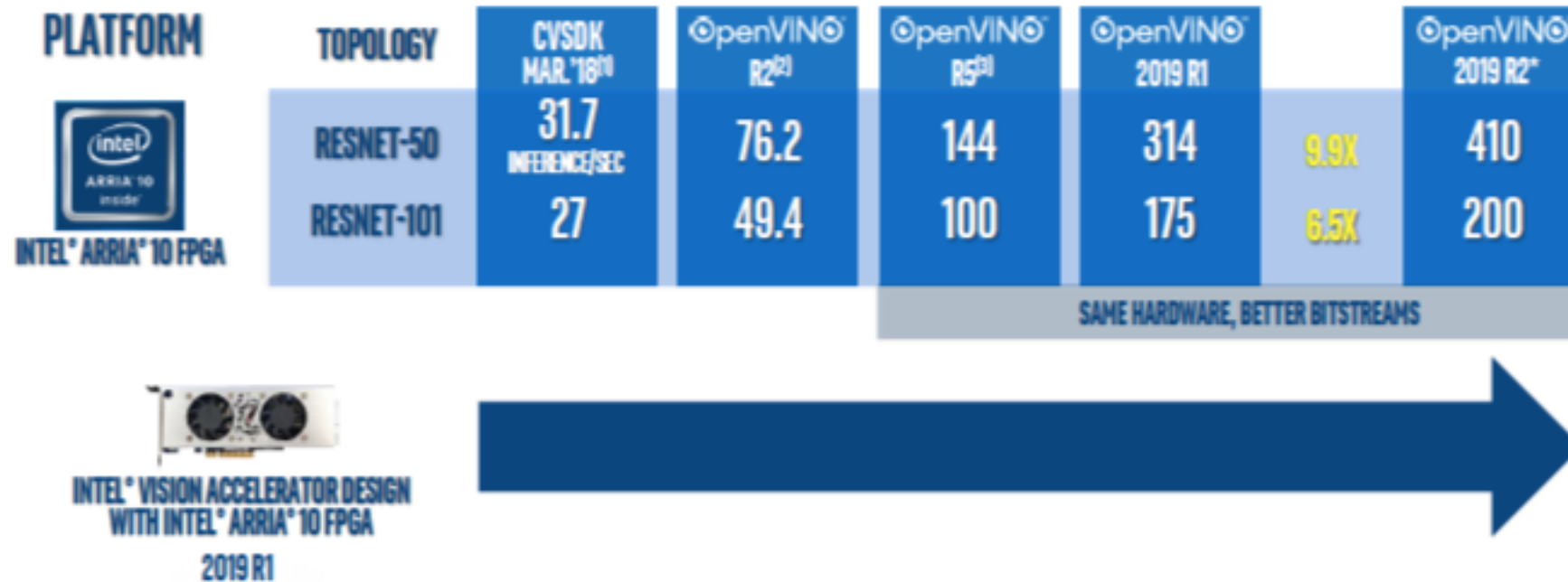
	Relative merits	Examples (Published numbers)
ASIC	Maximum efficiency TOPs/W Multi-year design cycle Programmable but limited flexibility	Google TPU v1: 2.1 TOPs/W (40W) Toshiba: 2.1 TOPs/W (10W) Hokkaido Univ.: 539 GOPs/W (3W)
GPU	Easy to program Relatively higher power	Jetson TX1: 31 GOPs/W (10W), Nvidia Tesla P4: 440 GOPs/W (50W), Nvidia
FPGA	Efficiency improves over time with same HW Flexible, with wide range of performance	Xilinx Zynq MPSoC Ultra96: 8.5 GOPs/W (5W) Syracuse U. Virtex-7 ADM-PCIE-7v3: 2.2 TOPs/W (25W)

*Cirrus360 ERB Accelerator-function-as-a-service enables use of accelerators within an edge software stack*



# From Intel AI Workshop – CERN – May, 8th 2019

## FPGA performance evolves over time



### Optimization Notice

Copyright © 2019, Intel Corporation. All rights reserved.  
\*Other names and brands may be claimed as the property of others.

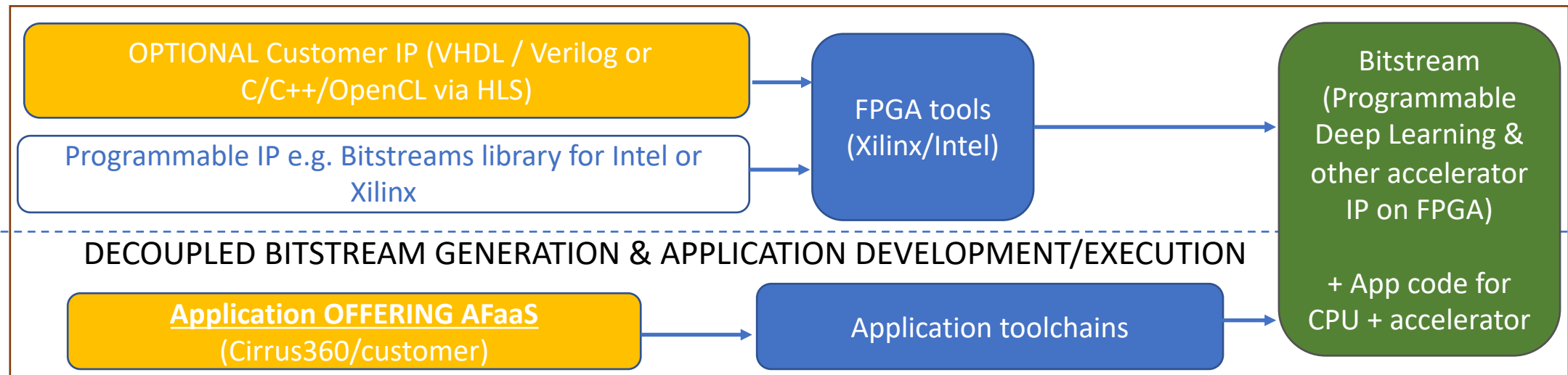


55



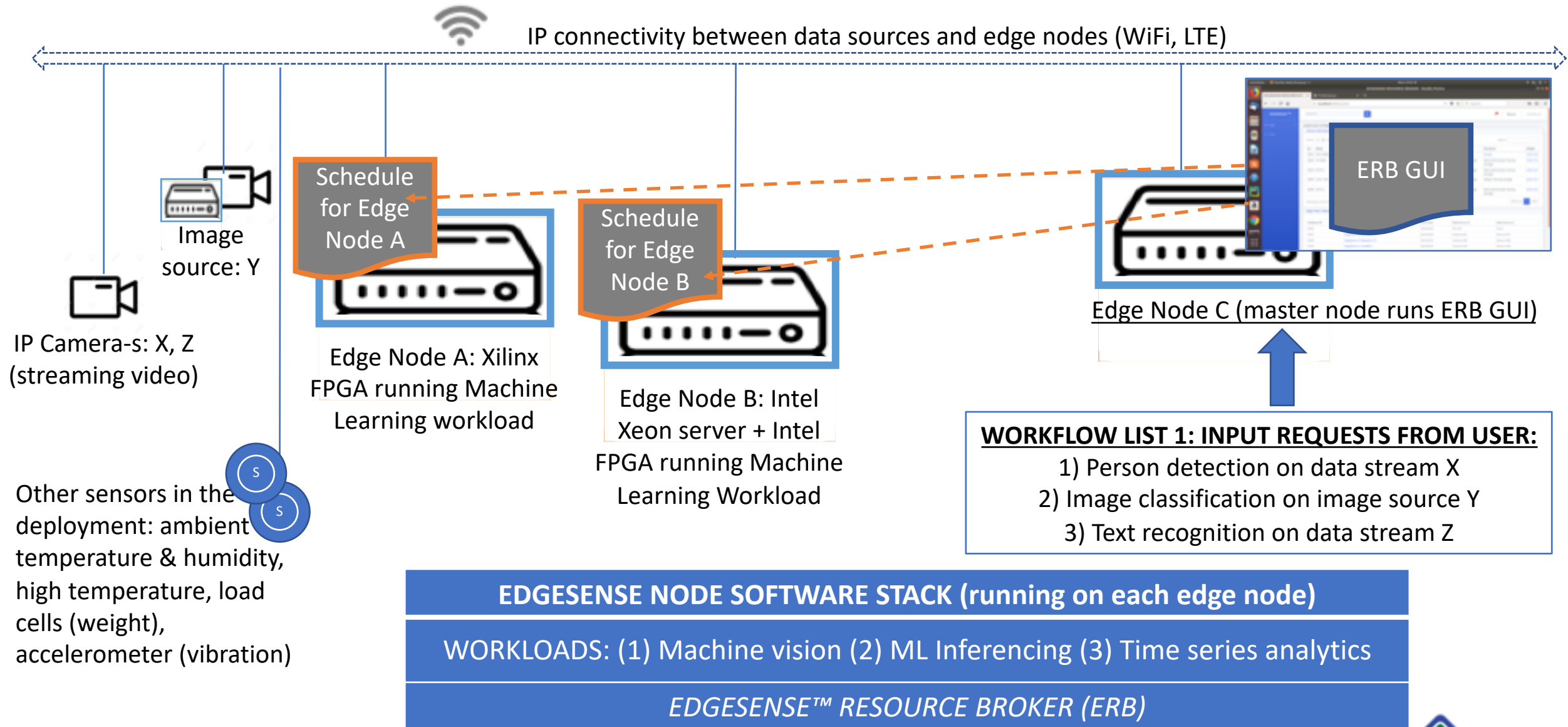
# AFaaS approach: motivation

- FPGA programming can take time and expertise, no matter which FPGA tool flow is used.
- Cirrus360 uses an Accelerator-Function-as-a-service (AFaaS) approach to utilize the FPGA accelerators for application task level work.
- FPGA scheduling considerations:
  - Resource constrained (by default) vs latency constrained (when application requires)
  - Latency constrained workloads: beware of overheads
  - More work executed for same FPGA bitstream and once input is loaded: the more efficient the workflow execution





# EdgeSense in action: multiple ML workflows on shared edge nodes



### WORKFLOW LIST 1: INPUT REQUESTS FROM USER:

- 1) Person detection on data stream X
- 2) Image classification on image source Y
- 3) Text recognition on data stream Z

### ERB OUTPUT: WORKFLOW LIST 1: SCHEDULE:

Tasks automatically scheduled by ERB on Edge Node A and  
Edge Node B  
(User does not have to worry about what to run on which edge node)

Person detection on Edge Node A and text recognition on Edge Node B  
on two video streams, both running real time

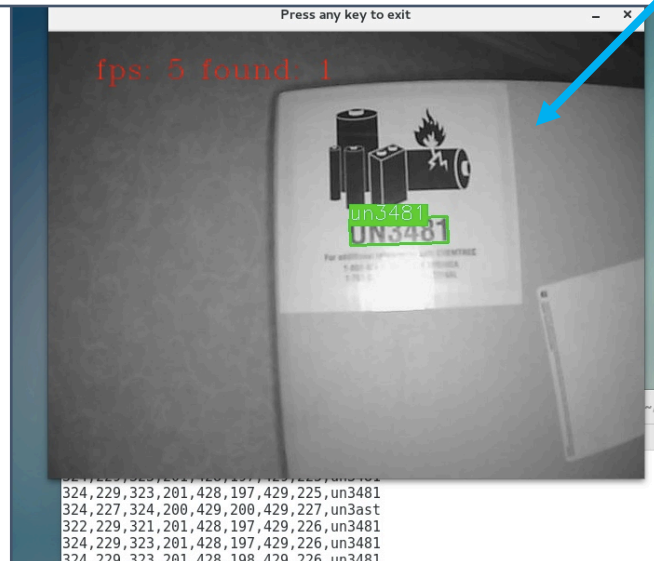
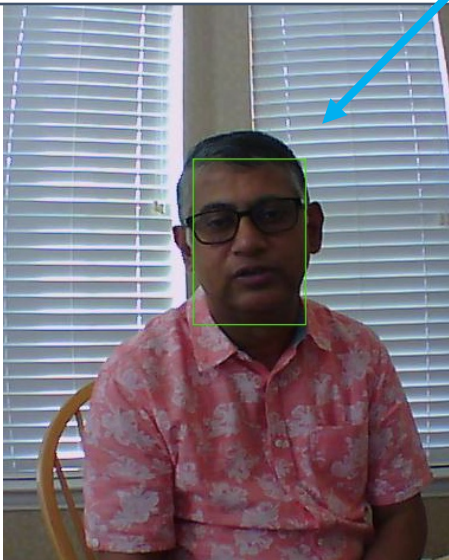


Image classification on Edge Node A:  
Running periodically

```
Load Image : PIC_190.jpg
Run DPU Task for GoogLeNet ...
DPU Task Execution time: 17171us
DPU Task Performance: 184.631GOPS
top[0] prob = 0.819910 name = leopard, Panthera pardus
top[1] prob = 0.142479 name = jaguar, panther, Panthera onca, Fells onca
top[2] prob = 0.024759 name = cheetah, chetah, Acinonyx jubatus
top[3] prob = 0.007094 name = snow leopard, ounce, Panthera uncia
top[4] prob = 0.002610 name = window screen

Load Image : PIC_499.jpg
Run DPU Task for GoogLeNet ...
DPU Task Execution time: 17774us
DPU Task Performance: 177.788GOPS
top[0] prob = 0.992892 name = Granny Smith
top[1] prob = 0.005210 name = banana
top[2] prob = 0.000705 name = grocery store, grocery, food market, market
top[3] prob = 0.000157 name = bell pepper
top[4] prob = 0.000074 name = cucumber, cuke
```

# Next steps in our journey

- EdgeSense roadmap:
  - Product that is of interest to US DOE but can also be leveraged in commercial sector
  - Customers pilots (in progress)
    - Market focus: Industry 4.0 for SMEs, Smart “spaces”, DOE labs
- *Collaboration and next steps:*
  - Smart sensors and AFaaS platform for smart factory
  - Predictive maintenance for remote heavy machinery

