
# Squeezing the Last Milliwatt and Cubic Millimeter from Smart Cameras Using the Latest FPGAs and DRAMs

**Hussein Osman**
Segment Marketing Director
Lattice Semiconductor

**Richard Crisp**
VP & CTO, Etron Technology America

LATTICE SEMICONDUCTOR Etron

# Why FPGA for Edge AI

Accelerating Innovation in Low Power Applications

**ULTRA-LOW POWER**
1 milliwatt – 1 watt


**SCALABLE PERFORMANCE**
Multiple use cases in parallel or serial

**SECURE**
Secure device configuration

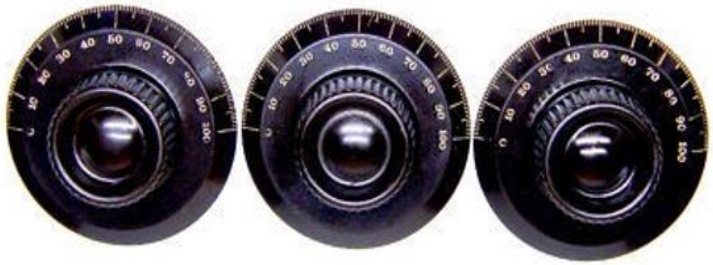
**FLEXIBLE COMPUTATION RESOURCES**
Pre and post processing
ISP, FFT and filtering

**HARDWARE PROGRAMMABLE**
Adapts to fast changing machine learning algorithms

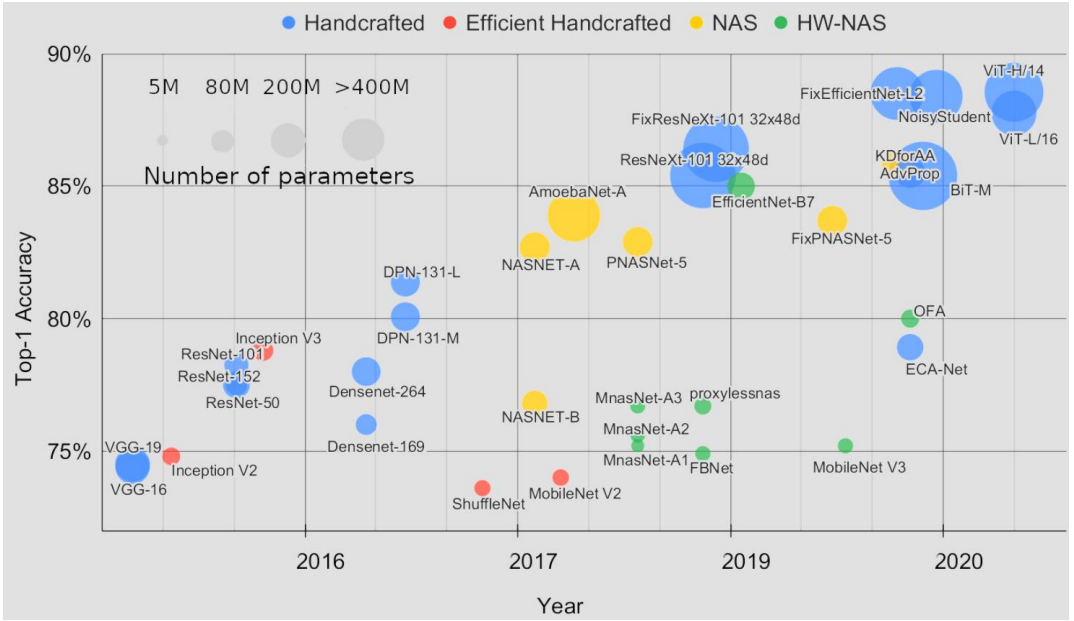
# FPGAs Speed / Power Optimized AI Innovation

Accelerating Innovation in Low Power Applications

**Hardware Optimization** & **Algorithm Optimization** > **Continuous Improvement New Use Cases**



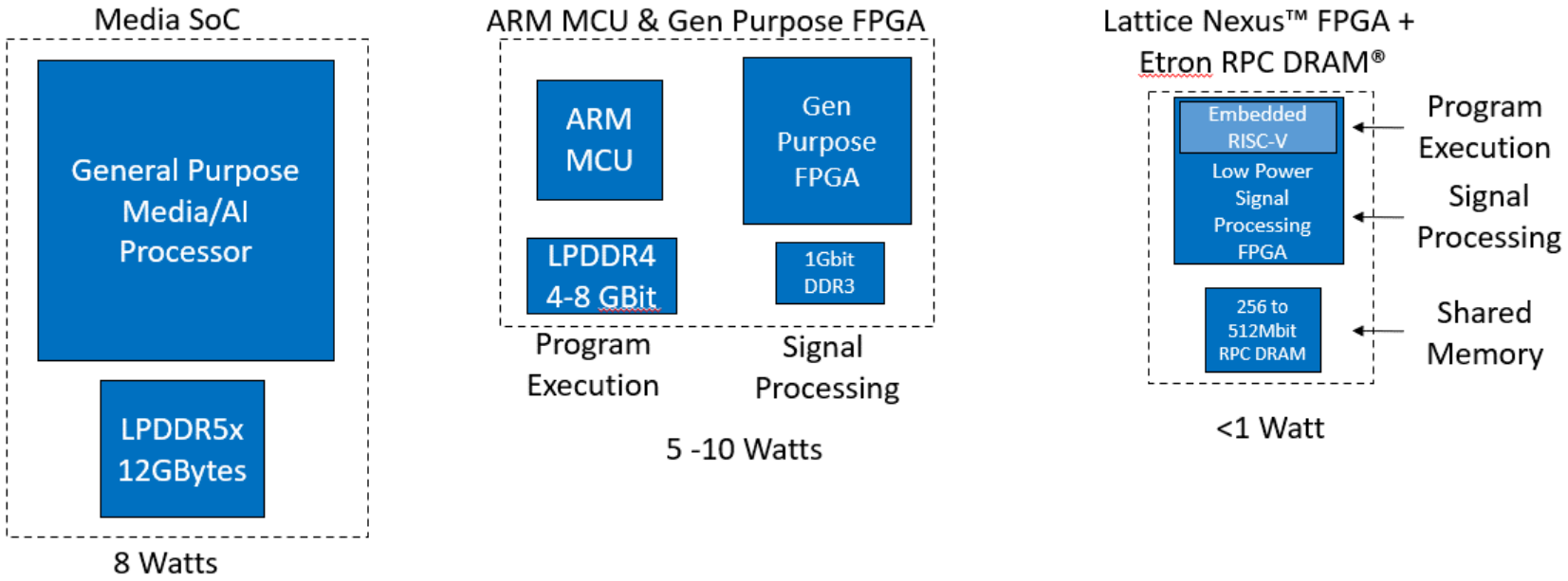
## AI models are rapidly evolving




Source: BENMEZIANE et al.: A COMPREHENSIVE SURVEY ON HARDWARE-AWARE NEURAL ARCHITECTURE SEARCH

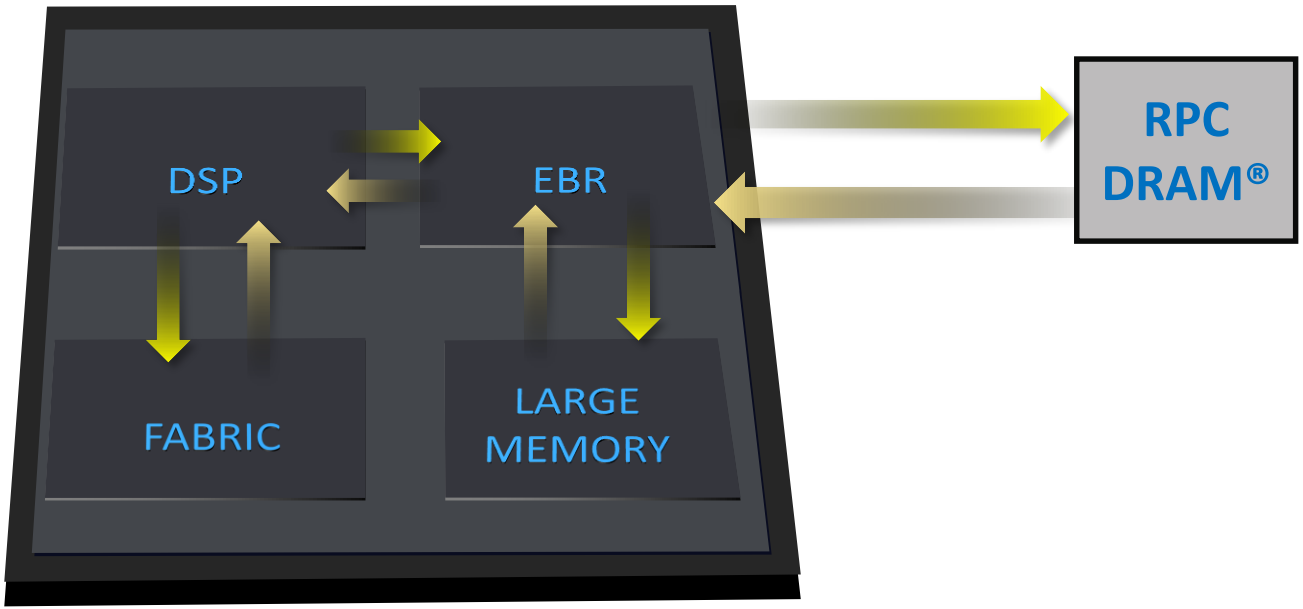
# Edge AI Camera Architectural Options

Accelerating Innovation in Low Power Applications



# Power Efficient FPGA Inferencing Resources

Accelerating Innovation in Low Power Applications



# Scalable Efficient CNN Acceleration Engine

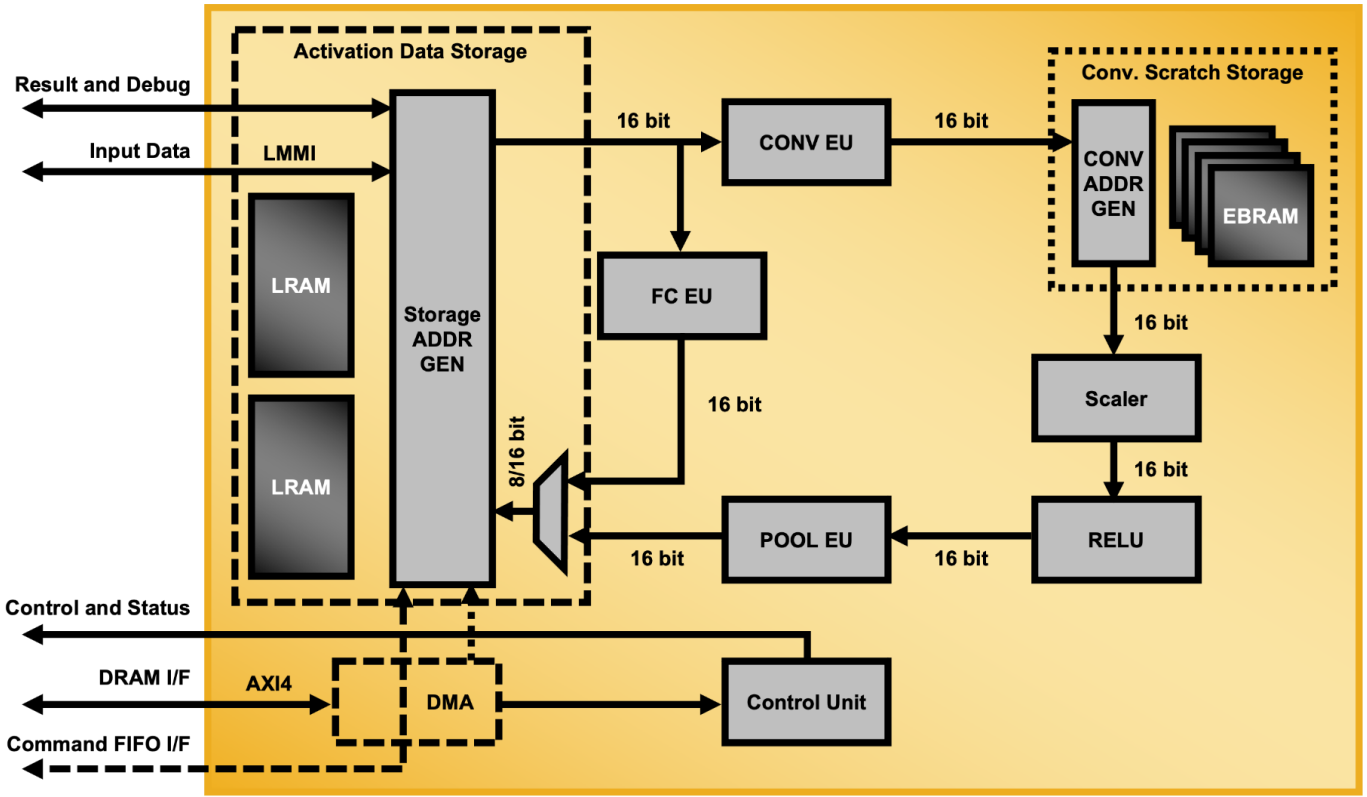
Accelerating Innovation in Low Power Applications

Compact, Optimized or Extended CNN

AXI4 or FIFO interface

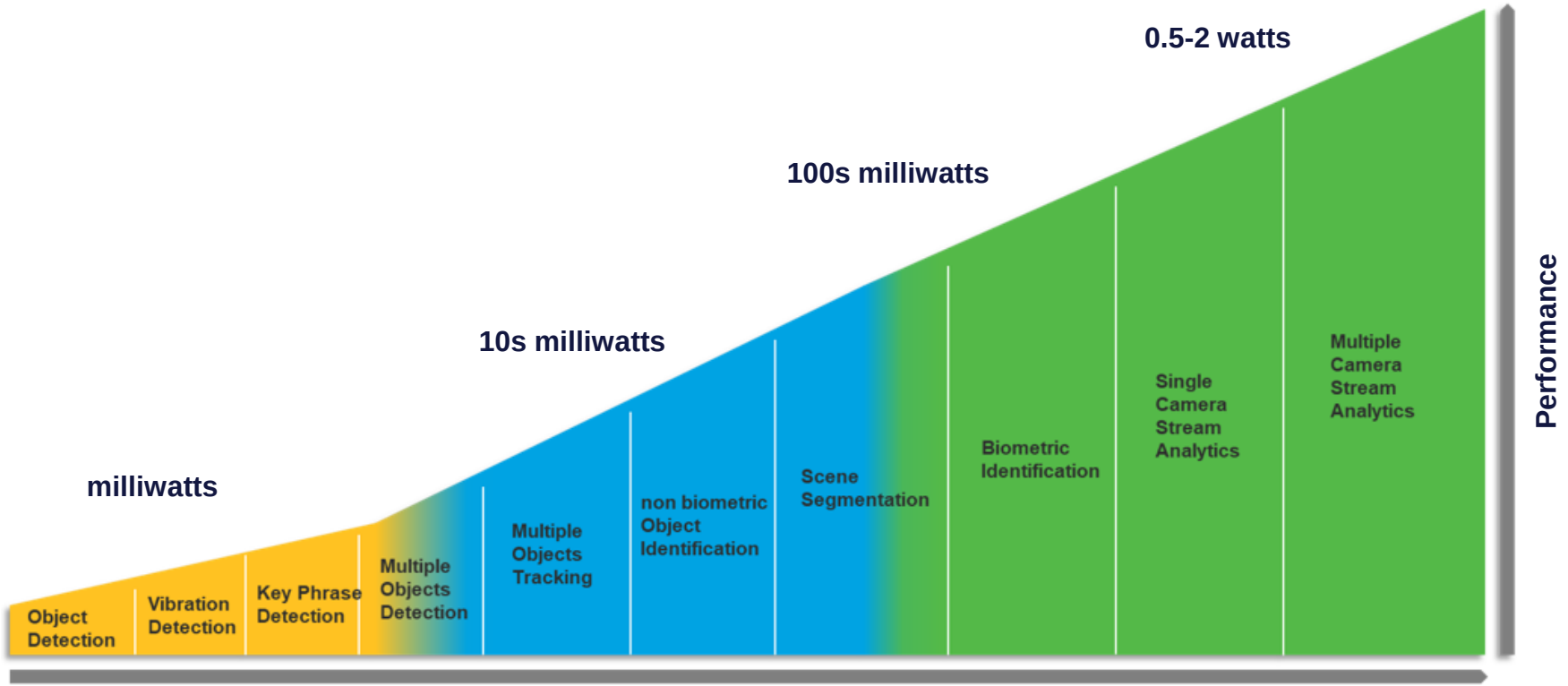
Layer Support

- convolution
- max pooling
- global average pooling layer
- batch normalization fully connected



# Scalable CNN Acceleration Engine

Accelerating Innovation in Low Power Applications

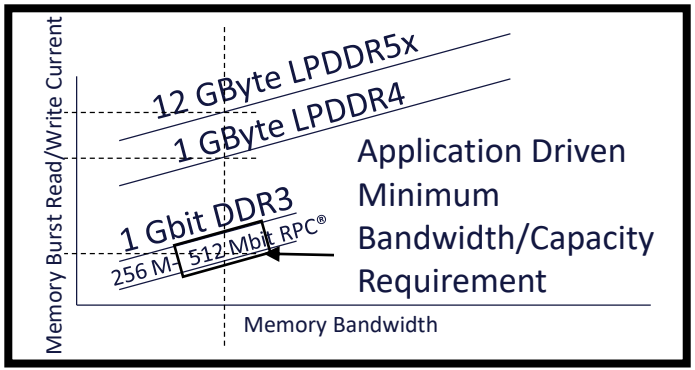


# Object Detection and Counting

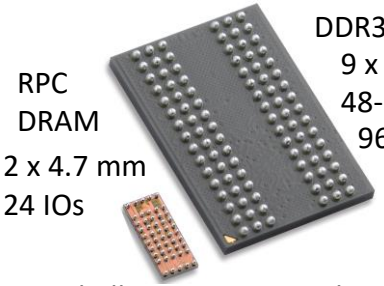
- Accelerated, low-power human presence detection and counting using neural network model
- VGG, MobileNetv1, MobileNetv2, ResNet, and SSD type structures are supported
- TF Lite based implementation for ease of use
- Reference designs are provided to enable design replication and transfer learning
- Total power consumption of less than 200 mW
- Processing at up to 60 FPS and VGA resolution

# Avoid Overprovisioning Memory:

Look for Opportunity for Size, Weight, Power & Cost Savings




Extra Bits and Excess Memory Bandwidth increase cost, power dissipation, and WLCSP memory PCB footprint




"RPC DRAM: Less than half the I/Os with < 1/10 the footprint"

"With same # of FPGA I/Os, RPC DRAM can provide twice the bandwidth of DDR3 at same clock frequency or the same bandwidth but at half the clock frequency"


Conventional DDR & FPGA in BGA (Overkill)

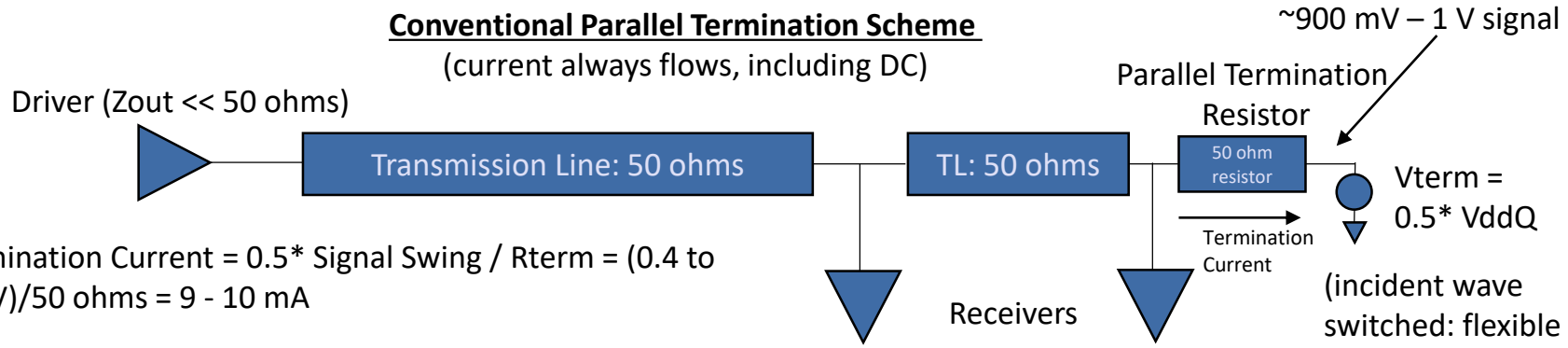

Right Sized Memory/FPGA In CSPs

# Power Savings from Series Termination For Memory Bus:

Applying Basic E.E. Transmission Line Principles

**Conventional Parallel Termination Scheme**

(current always flows, including DC)

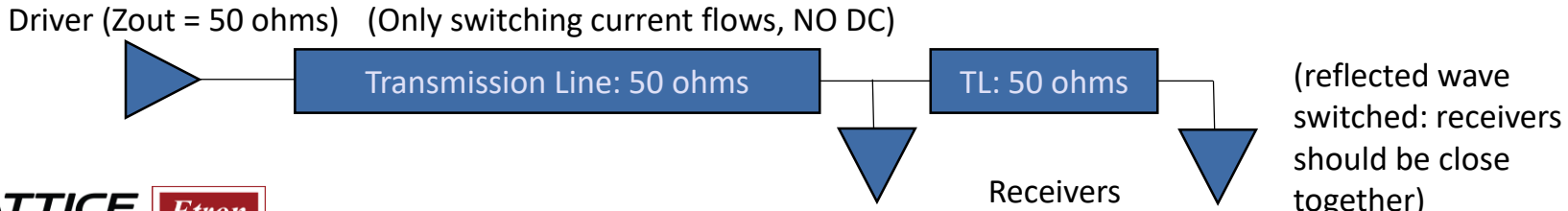



Termination Current = 0.5* Signal Swing / Rterm = (0.4 to 0.5 V)/50 ohms = 9 - 10 mA

For a DDR3 two rank system there are 54 terminated signals = 486 - 540 mA termination current!

(incident wave switched: flexible receiver placement)

**Series Termination Scheme**

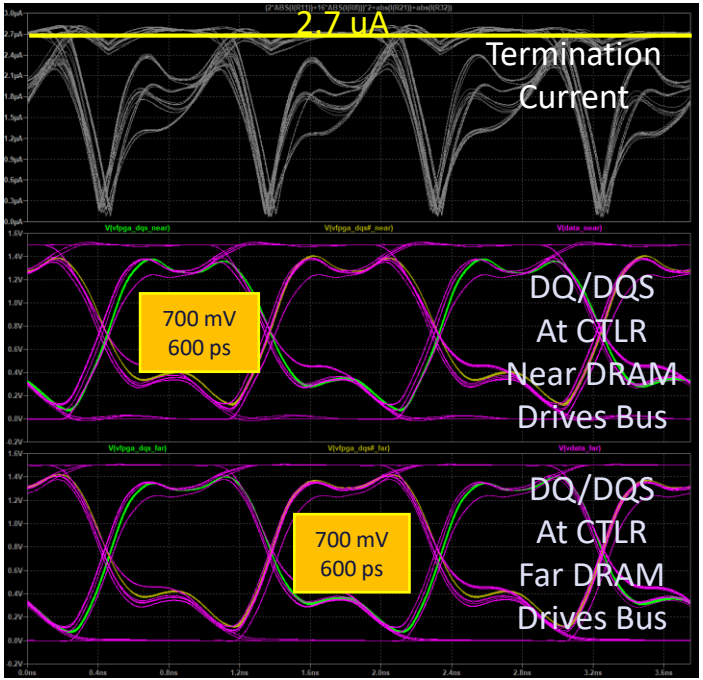



(reflected wave switched: receivers should be close together)

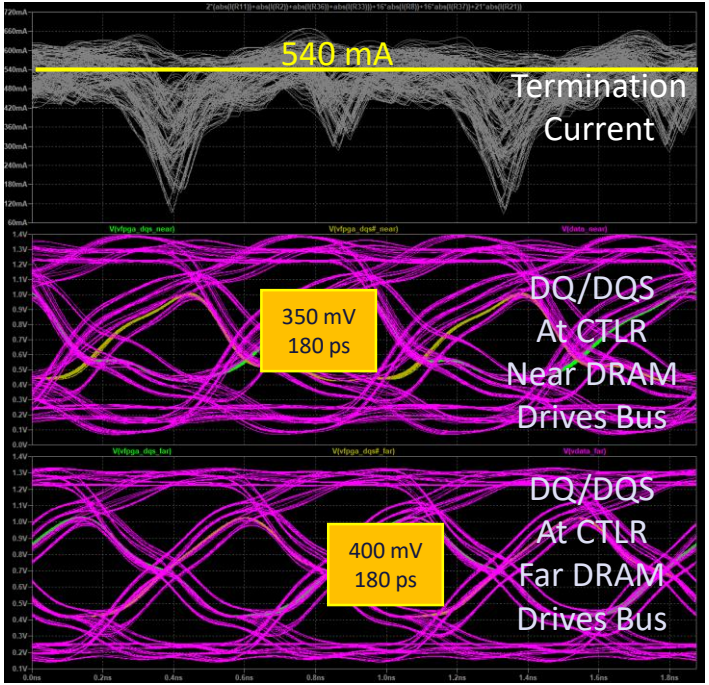
# Two Rank Configuration SI Comparison:

Series Term @ 533 MHz vs Parallel Termination @ 1066 MHz

@ Lower frequencies (like soft FPGA I/F): series termination works fine & saves significant power




Series terminated: 2 Rank RPC w/Soft FPGA I/F @ DDR1066




Parallel terminated: 2 Rank DDR3 w/Hard ASIC I/F @ DDR2133

# Component Availability and System Reliability

28 nm FD-SOI → low power; 100x lower SER

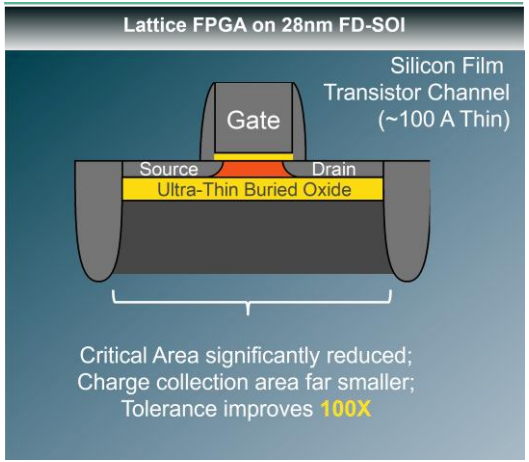
FuSa certified FPGA design tools

RISC-V with Green Hills µ-velOSity

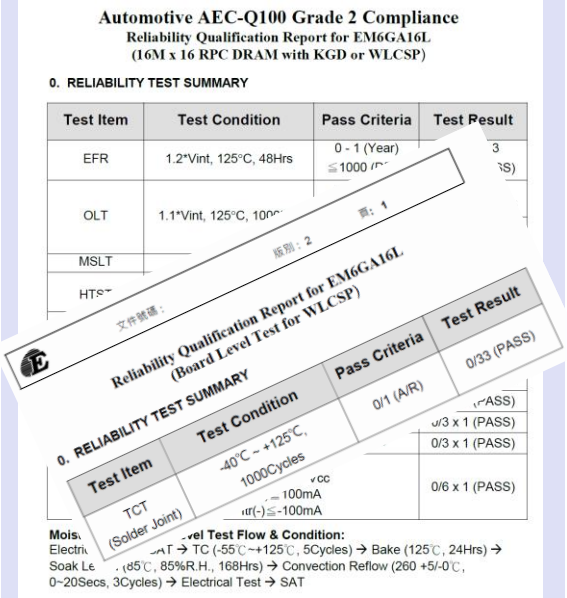
ECC protected memory

AEC Q100 Level 2 qualified FPGAs and memories

Extensive tier 1 automotive environmental testing




DRAM Component Reliability Reports




**RPC DRAM and Lattice CertusPro™-NX FPGA components are available NOW in volume**

**(including from DigiKey)**

# Key Takeaways

## For minimum Size, Weight, Power and Cost:

- FPGAs offer parallel processing and adaptability suitable for rapidly evolving AI use cases
- Optimizing and tuning edge AI models and image signal processing to reduce complexity while still meeting application needs is key for reducing power and cost
- Avoid Overprovisioning Compute Horsepower and Memory: cut it back to minimum requirements with a small margin: no extra credit for unused excess capability vs application requirements
- Using CHIP SCALE packages can enable significant power savings by using Series Bus Termination due to faster bus settling times vs BGA.
    - Parallel termination – power 3.6 W
    - Series termination – power 840 mW -> 2.76 W savings!

# Resources: Find Out More, Where to Buy

**CrossLink™-NX**

CrossLink™-NX


**Embedded Vision Processing**

For Etron RPC Design Info

https://etron.com/innovative-dram-pl/rpc-dram//

**CertusPro™-NX**

CertusPro™-NX


**Advanced General Purpose Processing**

Buy RPC DRAM from DIGIKEY

https://www.digikey.com/en/products/detail/etron-technology-inc/EM6GA16LCAEA-12H/13169828

# Thank You