Exploring AI/ML Applications on the Ultra-Low Power SiWx917 Wi-Fi 6 Solution

2025 tech talks WEBINAR SERIES

Wi-Fi CERTIFIED

TNBTECH 박광철

Wi-Fi

Designed from the ground up for IoT - SiWx917 Wi-Fi 6 SoC

Bluetooth® matter

Wi Fi 6 CERTIFIED

DIFFERENTIATED FEATURES

Ultra-Low Power

- Increases Battery life and Recharging Interval

IoT-Optimized Wireless Performance

- 2.4GHz: Long-range, low-power, effective wall penetration, high-throughput
- Wi-Fi, Bluetooth LE, and Matter in single package

Edge Computing + System Integration

- **AI/ML accelerator enablement available in June in SiSDK 25Q2 release**
- Application MCU and Wireless Processor with Networking off loads
- Rich Peripherals, High GPIO count, and Large Memory

DEVICE SPECIFICATIONS

Wide Range of Memory configurations

- 672 kB on-die SRAM with configurable split between Cortex-M4 and Network Wireless processor
- In-package 8MB Flash/PSRAM, 16MB External Flash/PSRAM,
- Single-Chip Matter over Wi-Fi Solution

Multiprotocol Co-Existence

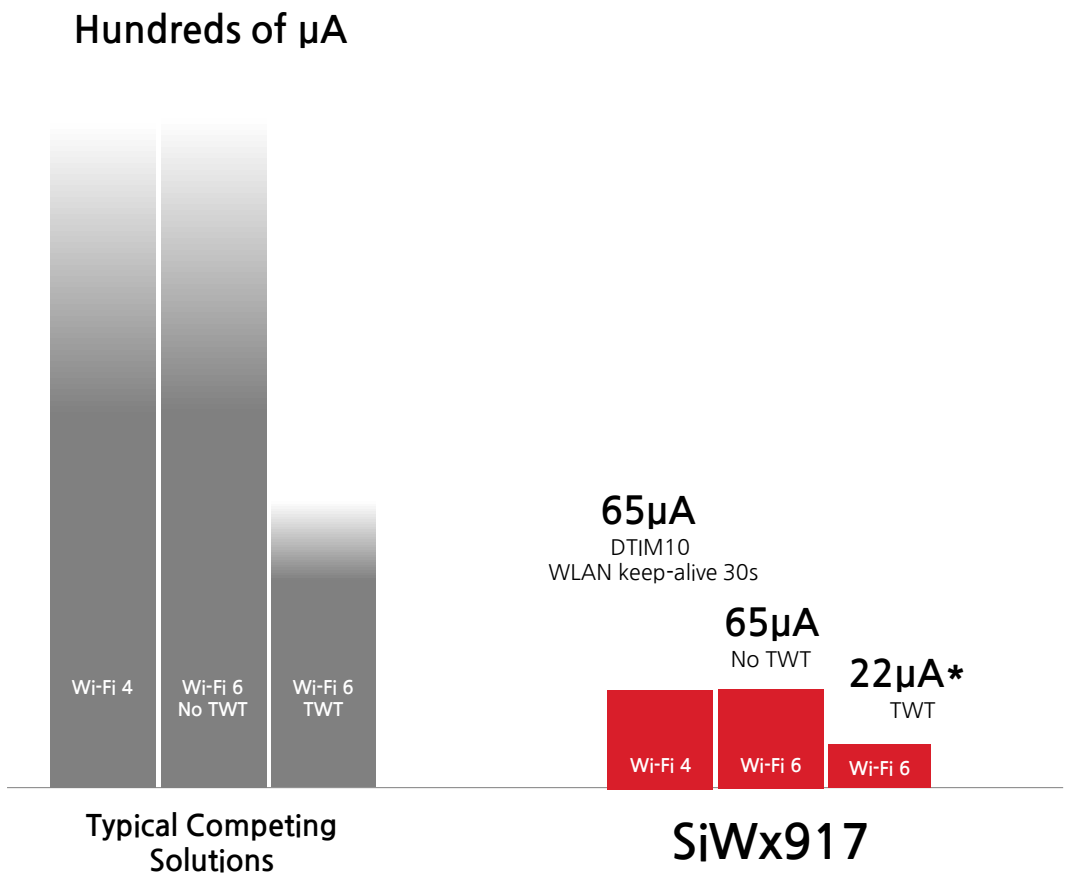
- High-performance Wi-Fi 6 and Bluetooth Low Energy 5.4

Robust Security

- A High Level of Security for the Device, Wi-Fi Protocol, and Networking

SiWx917: Lowest Wi-Fi Power - Longest IoT Battery Life

Wi-Fi Standby Current Consumption

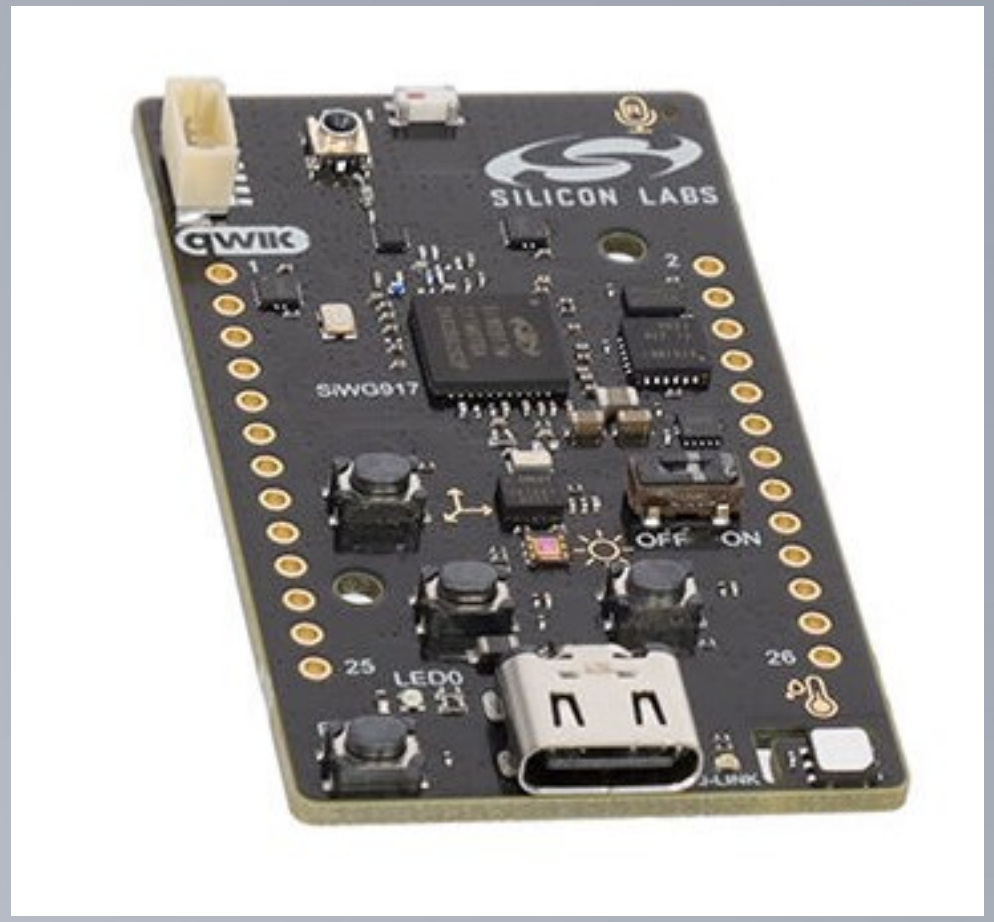


* Wi-Fi 6 TWT with auto-config feature enabled. TWT Rx latency 60s with 8ms wakeup duration. WLAN keep-alive every 60s. No TCP keepalive. 352kB SRAM retention. Does not include application MCU operation.

SiWG917 SoC Battery Life Estimation

How the SiWG917 SoC battery life of up to 2 years was estimated:

- Average current consumption for wireless and application **37µA at 3.3V**
- Associated standby low-power mode
- SiWG917 SoC as TCP client maintains socket connection
- 60 secs TCP keep-alive used.
- WLAN keep-alive 30 secs. 352kB NWP SRAM retention
- TWT Auto Config feature enabled. TWT Rx latency 60 secs w s wakeup duration
- Arm Cortex-M4 in sleep mode (PS4). 320kB SRAM retenti
- Measurements are taken in optimal conditions (RF chamber)
- Battery capacity 1000mAh (example AAA rechargeable battery)



- **On-board sensors**: Temperature Sensor (Si7021), Humidity Sensor (Si7021), 6-axis inertial sensor (ICM-20689), 2 x Digital Microphones (ICS-43434), Ambient Light Sensor (VEML6035)
- **Probe points** for power measurements
- **Qwiic Connector:** For compatibility with Sparkfun's expansion hardware (Sensors, Cameral, LCD, etc)
- USB Port: Board Power, Serial Communication
- **On-board Debugger**: SWD, VCOM

NEW:
Development Kit for AI/ML SiWG917 Wi-Fi, BLE and Sensors

- Arm Cortex-M4F 180Mhz
- 8MB flash, 8 MB on-board PSRAM and 320kB SRAM
- Matrix Vector Processor for AI/ML

  Co-processor for offloading matrix math operations

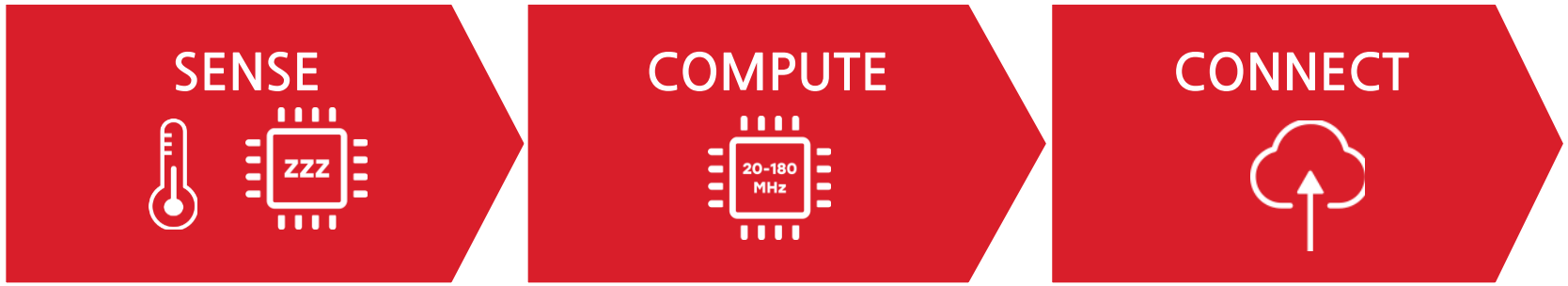
  Delivers faster ML inference with lower power consumption

  Performs Real and Complex Matrix and Vector operations, providing computing efficiency
- Wide variety of AI/ML applications

  Key-word detection demo

  Accelerometer based demos
- TensorFlow Lite support Available in June
- OPN: SiWx917-DK2605A

SiWx917 Ultra-Low-Power Sensor Processing



Receive & Store Sensor Data in the ULP Mode

- Receive through ULP peripherals
- Store in ULP SRAM
- Cortex-M4 sleeps
- PS1 power state is supported for ADC-based sensors, PS2 for other sensors.

Process Sensor Data in Cortex-M4 High efficiency Mode

- Cortex-M4
- High efficiency computing
  - PS2 at 20MHz (32uA/MHz)
- High-performance computing
  - PS4 at 180MHz (65uA/MHz)
- SW controlled Power manager
- Fast wake-up time

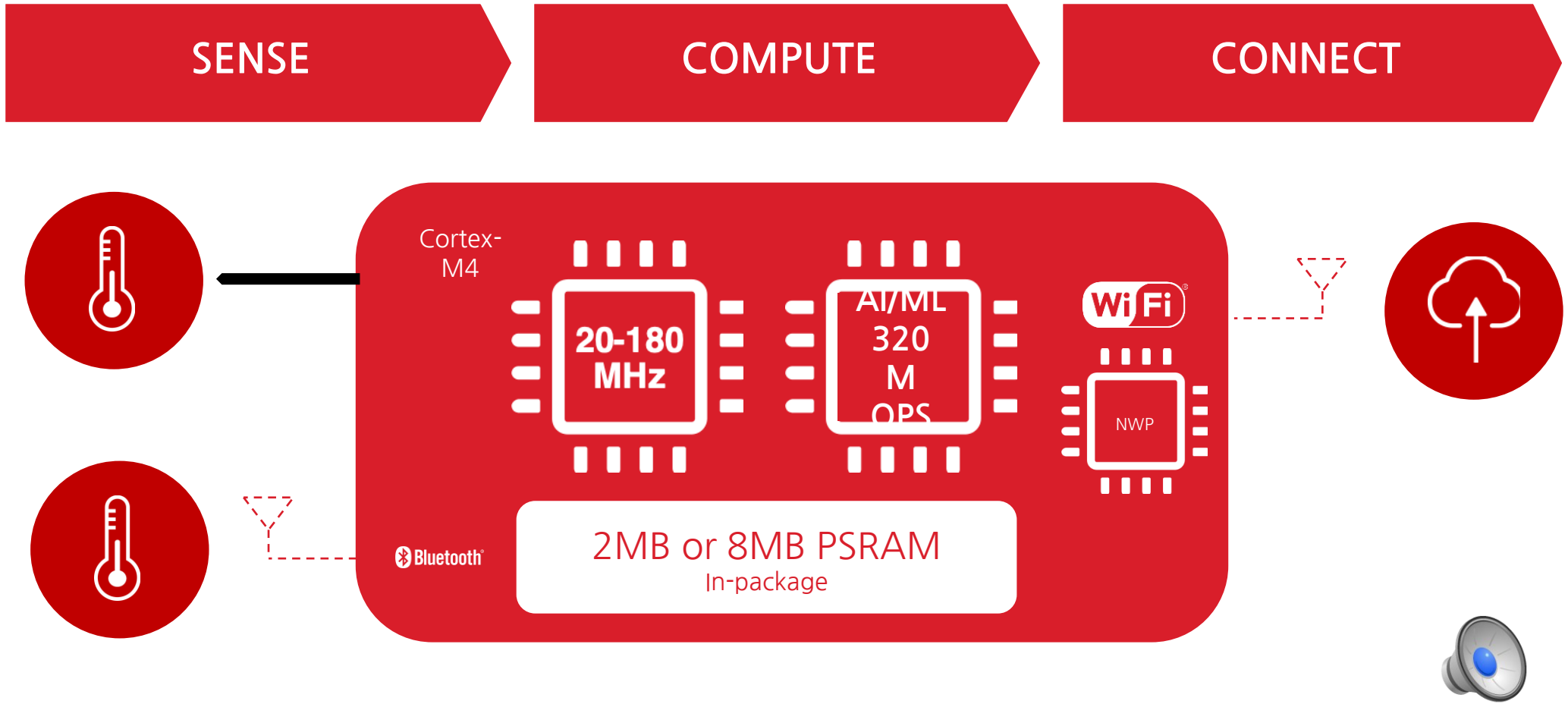
Enter High-performance Mode to Send Data to Cloud

- Maintain Wi-Fi Cloud connection
- Network Wireless Processor maintains TCP connection

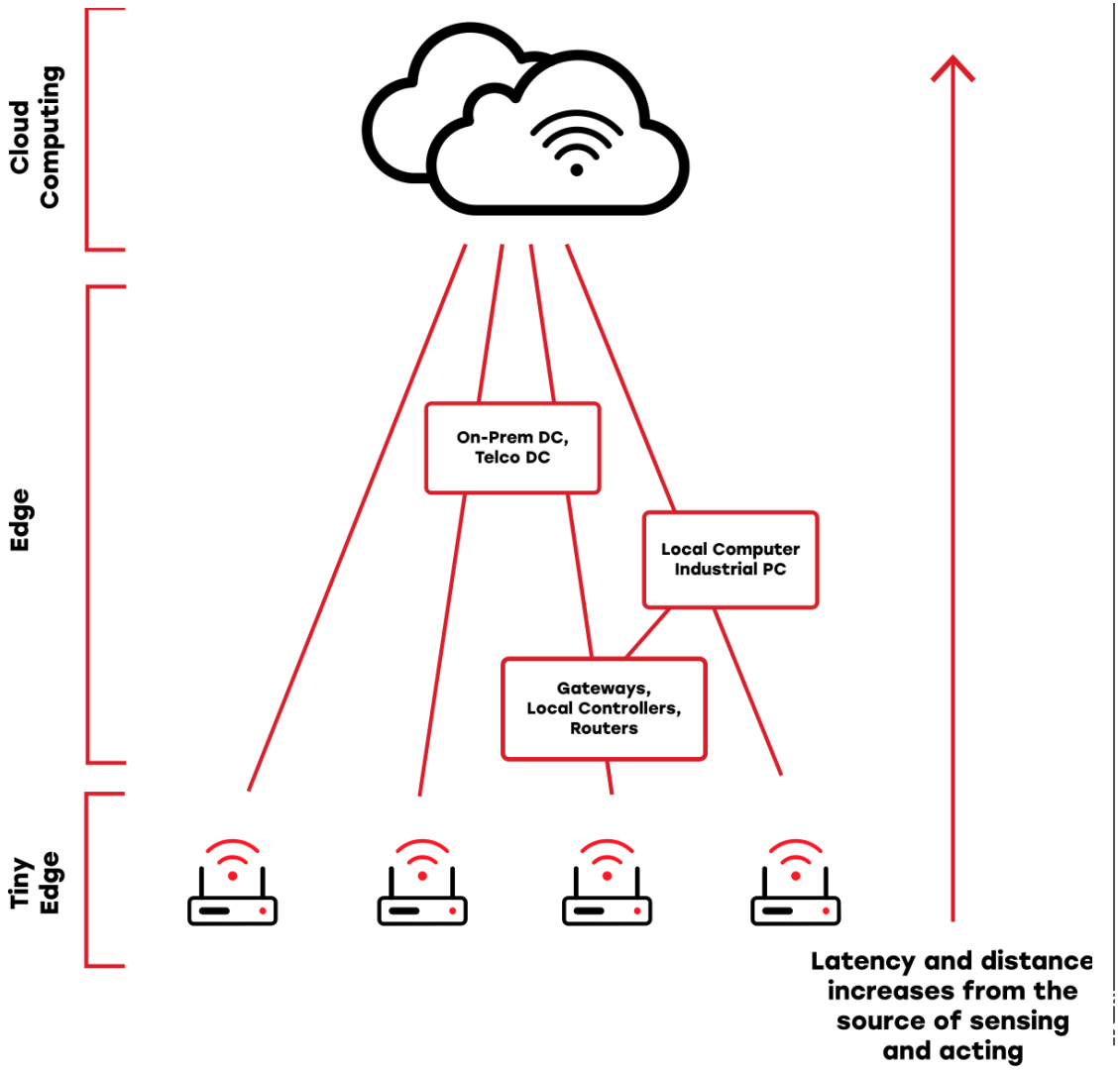
BENEFITS

- **Minimize Power and Extend Battery Life**
- **Offload MCU**
- **Compute Locally at Low-power**

SiWx917 Ultra-Low-Power Sensor Processing



Artificial Intelligence(AI) and Machine Learning(ML) at the Tiny Edge



Key Benefits

Low Latency

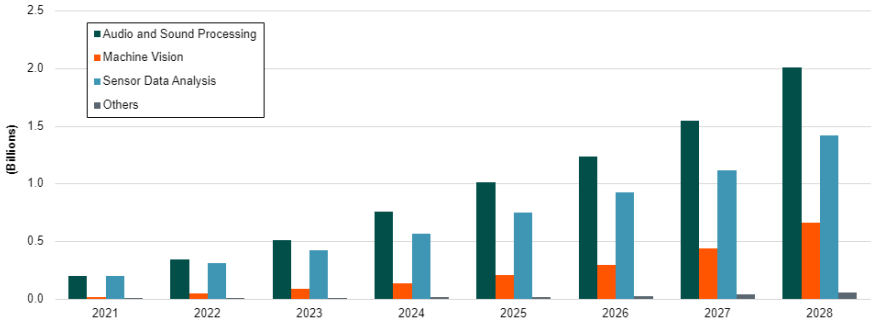
Privacy, IP Protection, Security

Bandwidth Constraints

Offline Mode Operation

Cost Reduction

>4B Devices deployed with TinyML in 2028

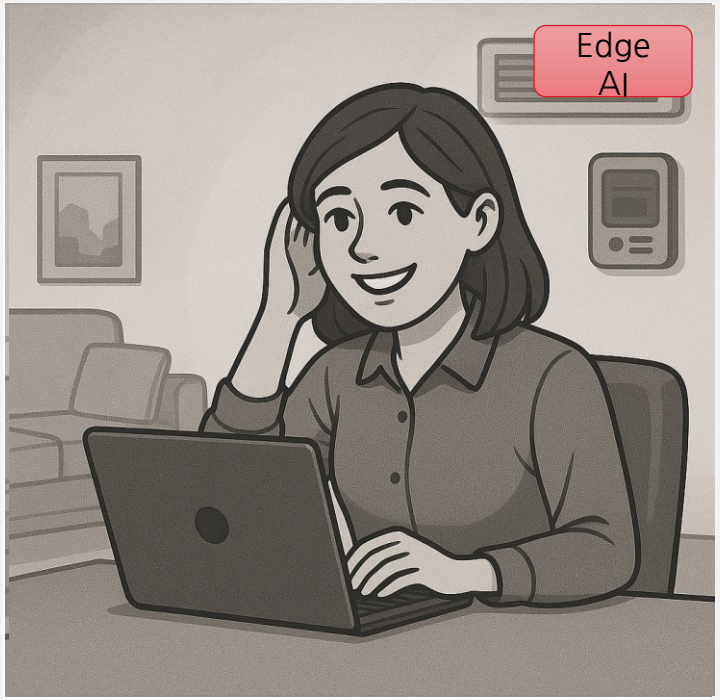


*Source: ABI Research, Artificial Intelligence and Machine Learning, 2 QTR 2024

Amy's Day - Transformed by Edge AI

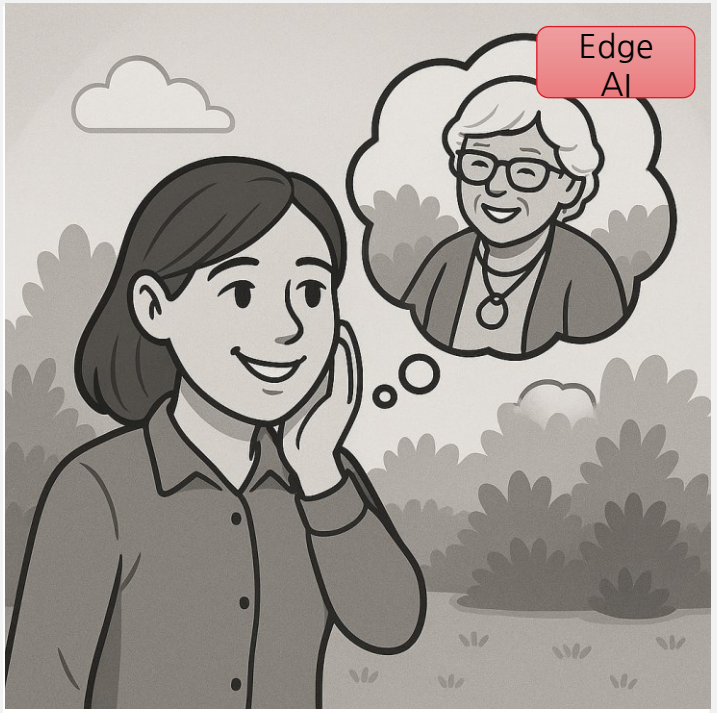
9:00 FACTORY

- ✓ Detect anomalies in real time
- ✓ Increased accuracy
- ✓ Works reliably



18:00 HOME

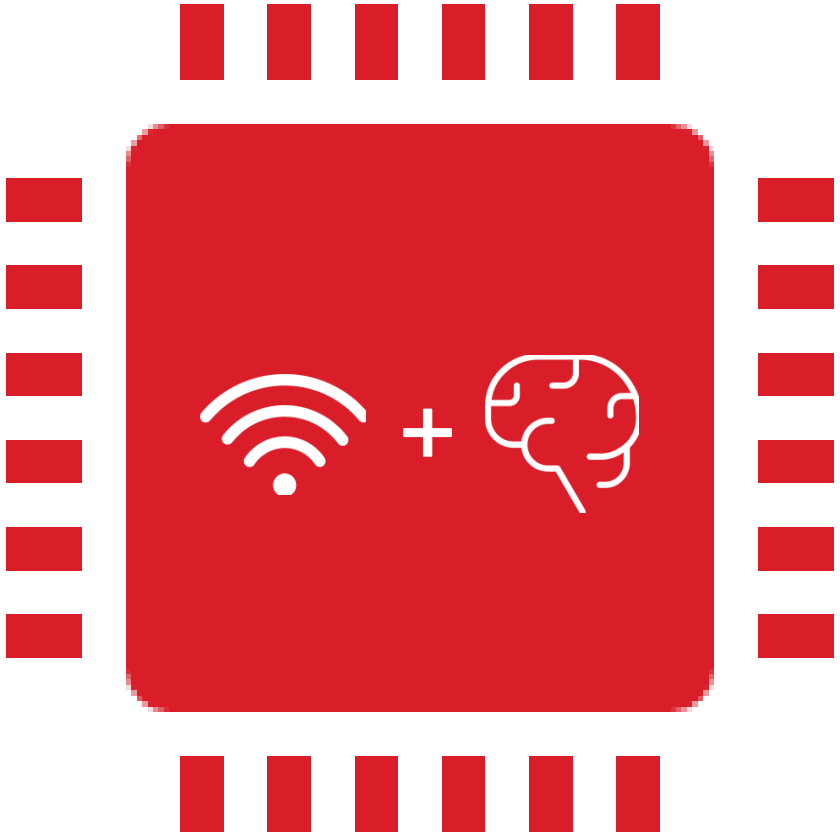
- ✓ Keep data private and secure
- ✓ Devices respond instantly
- ✓ Operate perfectly even offline



20:00 FAMILY CALL

- ✓ Instant, reliable alerts
- ✓ Reduces false positives
- ✓ Improves battery life

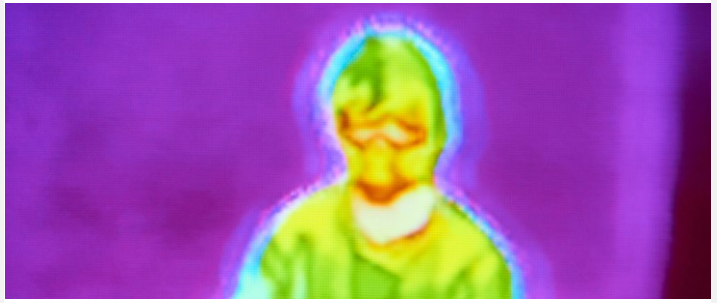
When Should you Use Edge AI?



- Why moving from rule-based logic to AI?
  - Not adaptable, fragile with noisy, variable or edge-case inputs.
  - Labor-intensive, hard to maintain and scale.
- What problems can you solve?
  - Spotting abnormal sensor patterns (e.g., vibration, motion).
  - Detect wake-words or acoustic events
  - Locally classifying images, gestures or activities.
- When does local AI make sense?
  - When decisions need to be fast, private or offline
  - When traditional logic can't handle variability
  - When you want to trigger cloud AI only selectively

- Data considerations:
  - Are sensors onboard? (accelerometers, microphones, etc.)
  - Are you capable of capturing meaningful and consistent data?
  - Will the model fit within compute and memory budgets

Machine Learning Applications Supported by Silicon Labs



VISION

Low-resolution

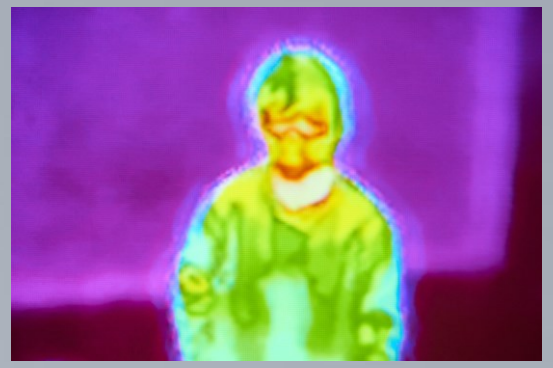
AUDIO

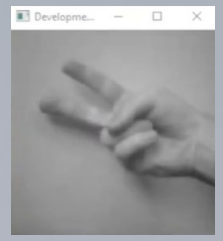
Voice commands
Audio pattern matching

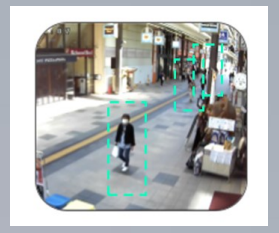
SENSOR

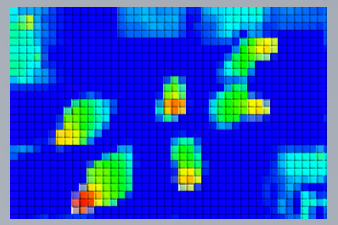
Vision

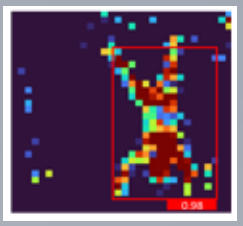
Low-resolution vision











Audio

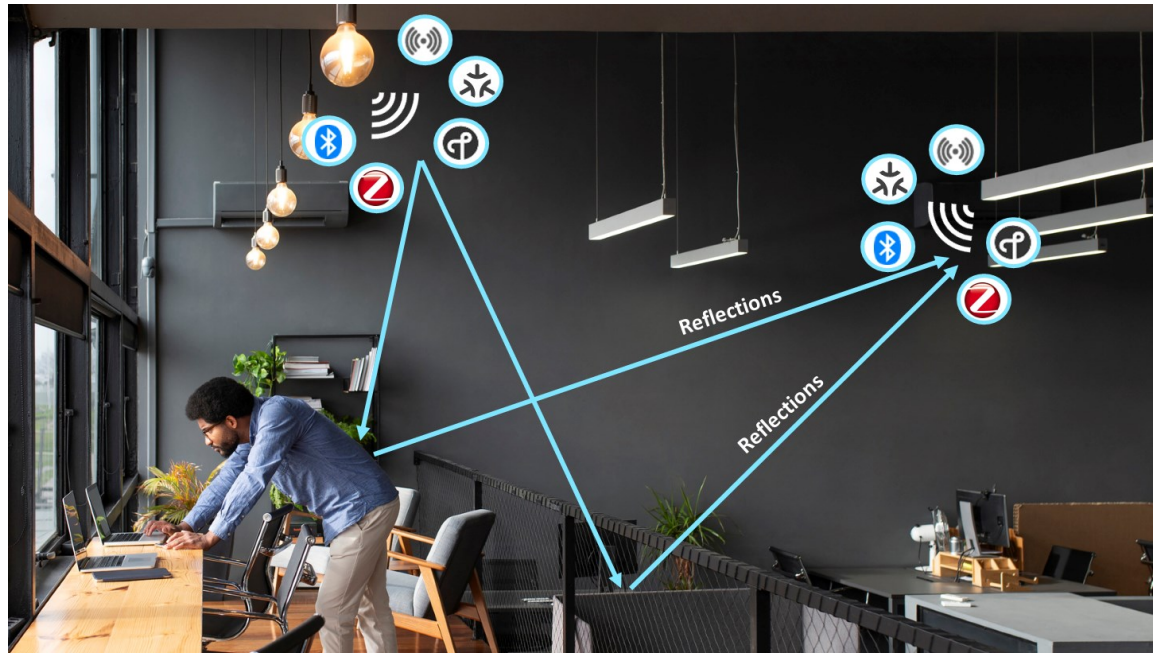
Voice commands
Audio pattern matching

Sensor

Sensing with Channel State Information

Sensing with Wi-Fi 802.11bf

- CSI based sensing is an optional feature in 802.11ax available but ONLY with Wi-Fi (802.11.bf)
- Superior to simple RSSI based solutions



Demo: Voice Control Introduction

OUR GOAL

Can a tiny chip understand human speech?

1. Add "Ears" to SiWX917 Chip
2. Key Commands: "ON" and "OFF"

음성 입력 "On" 또는 "Off" 명령어 수신

AI 분석 소형 칩(MCU)에서 실시간 추론

LED 제어 결과에 따른 하드웨어 제어

인터넷 없이 기기에서 직접 음성 인식!

1단계: Develop Model (모델 개발)

Create Dataset

학습용 데이터 수집 및 전처리

- 음성/이미지/센서 데이터 수집
- 데이터 정제 및 라벨링
- 데이터 분할 (80/10/10)

Train Model

신경망 모델 학습

- CNN/DNN 아키텍처 설계
- 하이퍼파라미터 튜닝
- 반복 학습으로 가중치 최적화

Test Model

모델 성능 검증

- 테스트 데이터로 평가
- 정확도/손실 측정 및 분석
- 과적합 여부 확인

출력: .h5 모델 파일

2단계: Prep Model (모델 준비)

Quantize & Convert

모델 양자화 및 TFLite 변환

- Float32 → INT8 양자화
- 모델 크기 75-90% 감소
- 추론 속도 향상

양자화 방식:

- 동적 범위: 가중치만 INT8
- 전체 INT8: 입출력 포함

Profile Model

타겟 하드웨어 리소스 분석

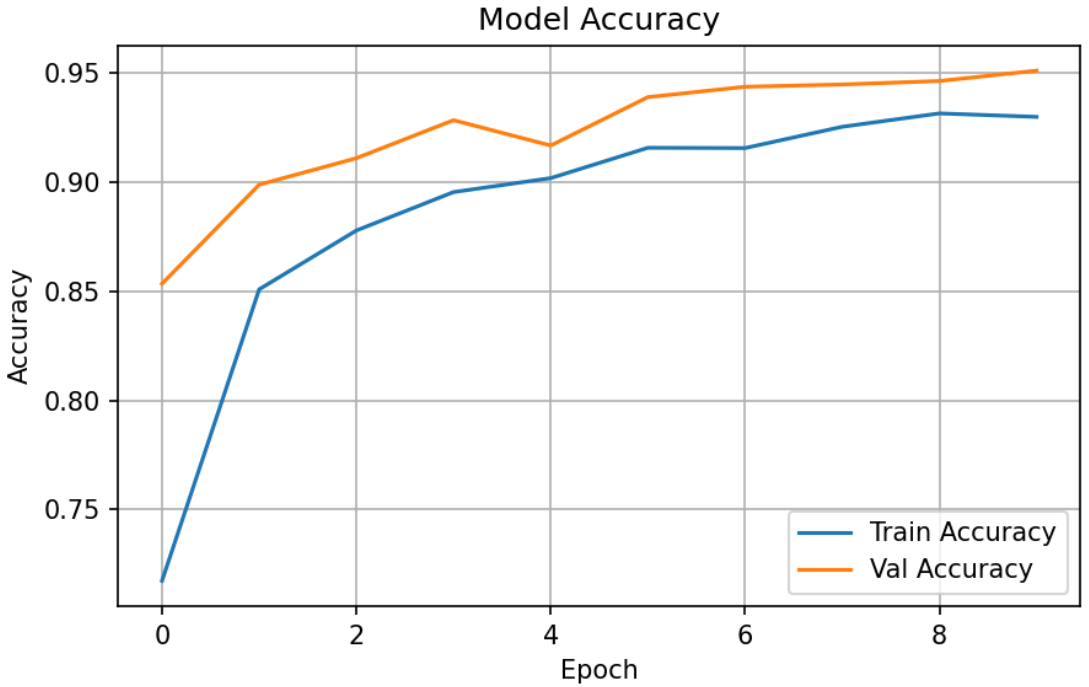
- Flash/RAM 사용량 측정
- 추론 시간 벤치마크
- Tensor Arena 크기 최적화
- 레이어별 성능 프로파일링

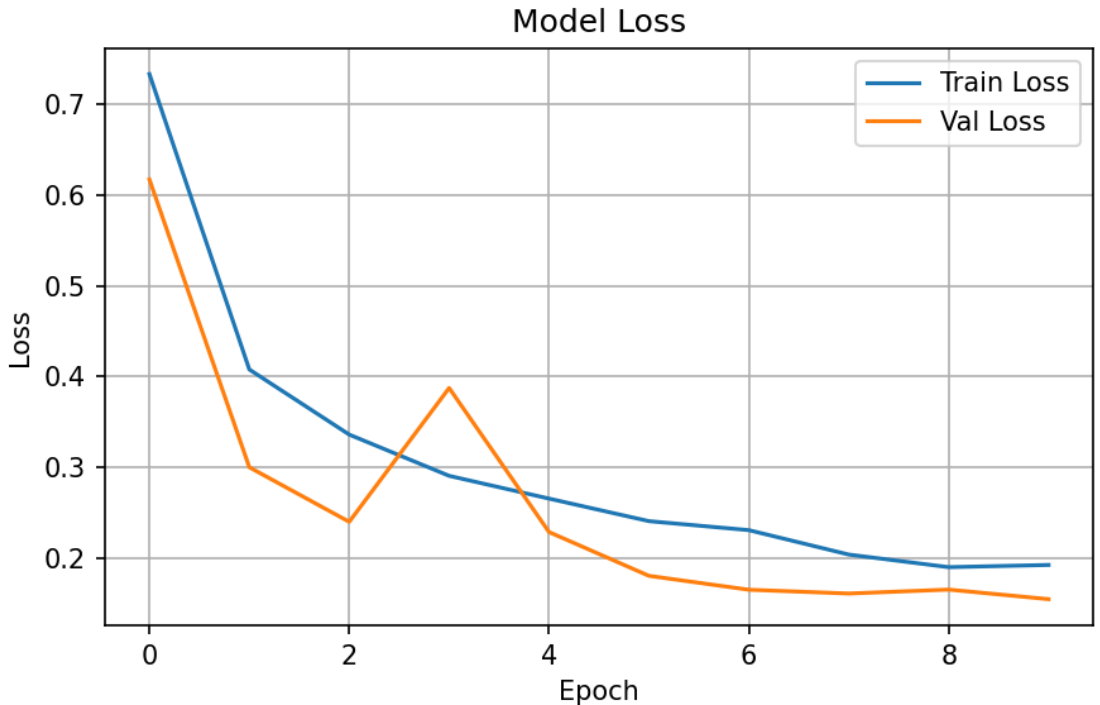
입력: .h5 (4.8MB) → 출력: .tflite (411KB) **91% 크기 감소**

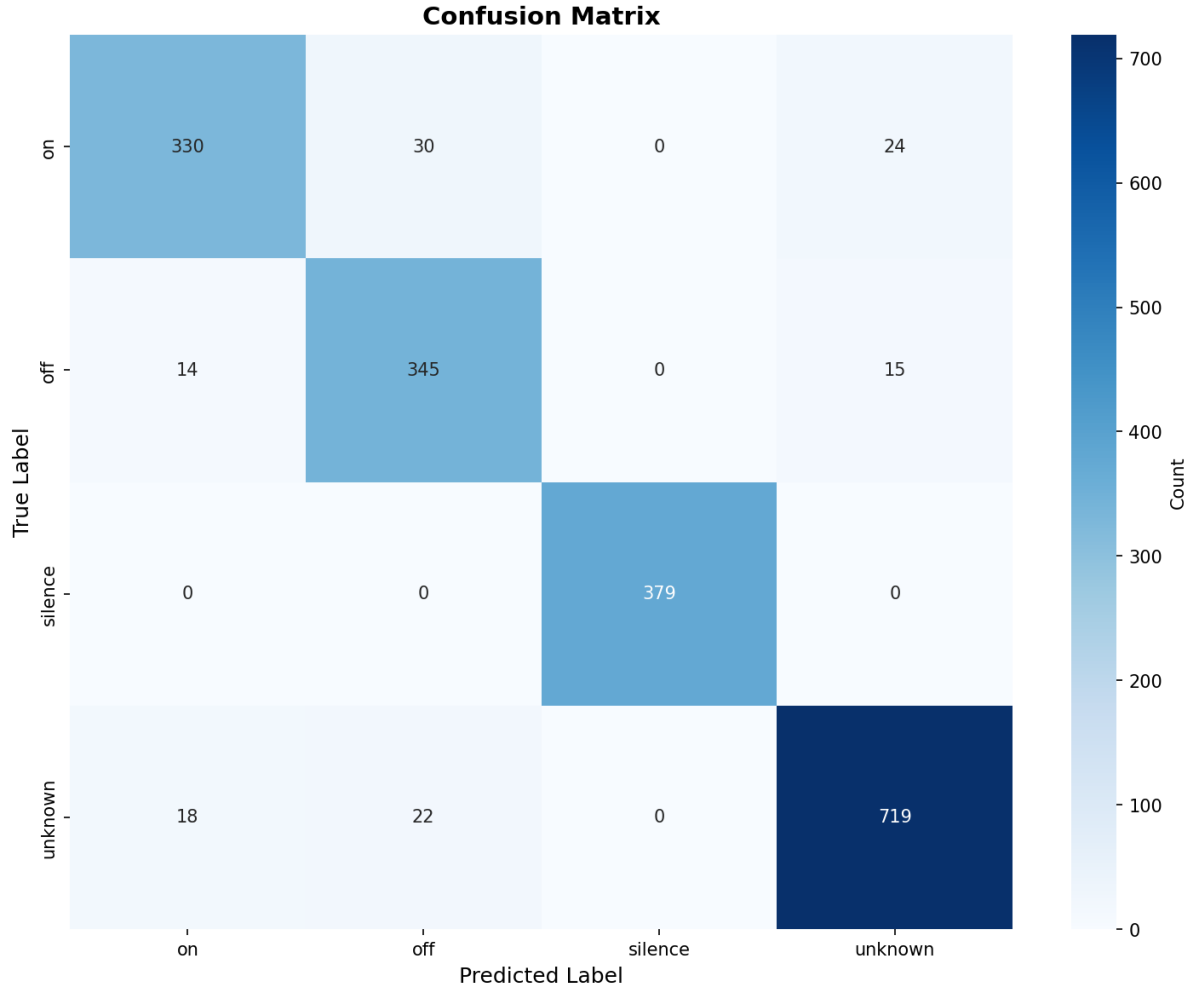
3단계: Deploy & Execute (배포 및 실행)

Layer 1	TFLite Flatbuffer	.tflite → C 배열 변환, 펌웨어에 포함하여 플래시 메모리에 저장	
Layer 2	TFLite Micro Interpreter	모델 로드 및 추론 실행 엔진, 메모리 할당 및 연산 스케줄링	
Layer 3	Backends	**CMSIS-NN kernel** ARM 최적화 커널, 2-5배 성능 향상	**SiLabs HW based kernel** MVP 가속기 활용, 전용 연산 오프로드
Layer 4	Hardware	**Cortex-M** CPU 연산 처리, 범용 계산	**MVP Accelerator** 행렬 연산 가속, 저전력 추론

```
(venv) →  keyword_spotting_project git:(main) ✗ bash run_all.sh
```


Model Accuracy
Accuracy
Epoch
Train Accuracy
Val Accuracy


Model Loss
Loss
Epoch
Train Loss
Val Loss


Confusion Matrix
True Label
Predicted Label
on
off
silence
unknown
330
30
0
24
14
345
0
15
0
0
379
0
18
22
0
719
Count

SILICON LABS

Q&A

SILICON LABS

Thank you

APAC Tech Talks - South Korea Series

Date (Wednesday)	Time (Seoul Time)	Topic
November 26	10:00-11:00	Exploring AI/ML Applications on the Ultra-Low-Power SiWx917 Wi-Fi 6 Solution
December 10	10:00-11:00	Exploring Multiprotocol Wireless Techniques
On-Demand		
June 25	What's new in Matter	
July 23	The Most Application-Optimized Bluetooth SoCs for Future-Ready Applications	
August 20	Bringing Bluetooth 6.0 Channel Sounding to Market: Precision Ranging for Secure & Smart Applications	
September 24	Harvesting Energy for Smart IoT with Silicon Labs xG22e	
October 22	Introducing MG26, PG26, and BG26: A Highly Flexible SoC Platform for All of Your IoT Needs	

SILICON LABS

Thank you