ingenics DIGITAL

ENERGY-EFFICIENT AI AND SIGNAL PROCESSING IN EMBEDDED SYSTEMS

Dipl.-Ing. Ludwig Kürzinger
Ingenics Digital GmbH

SOFTWARE DEVELOPMENT & EMBEDDED SYSTEMS

Introduction

Introduction

Myself: Ludwig Kürzinger

Electrical Engineering at TUM

Background in Embedded Systems and Security

(Soon) PhD in Machine Learning

Ingenics Digital GmbH

Software Development and Embedded Systems

Founded in 1988

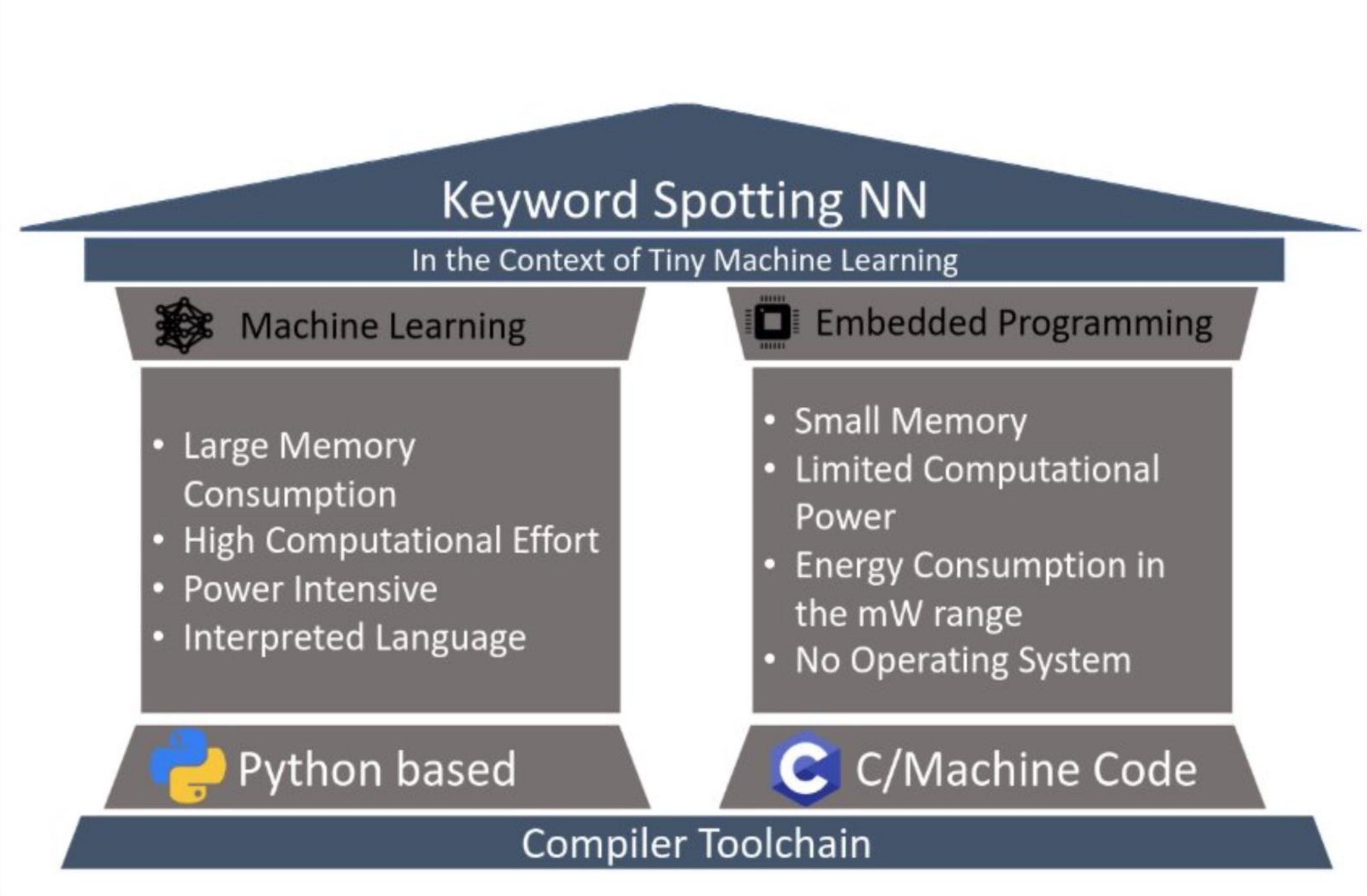
200+ Employees

Outline

1. Introduction
2. Modern Signal Processing in Embedded Systems
3. Practical Examples
4. Conclusion

Modern Signal Processing in Embedded Systems

Model Development vs. Model Deployment



Keyword Spotting from Model implementation to deployment
Image source: N. Günther. "Implementation and Evaluation of Keyword Spotting Neural Networks on Microcontrollers".

Toolchain Used in our Talk

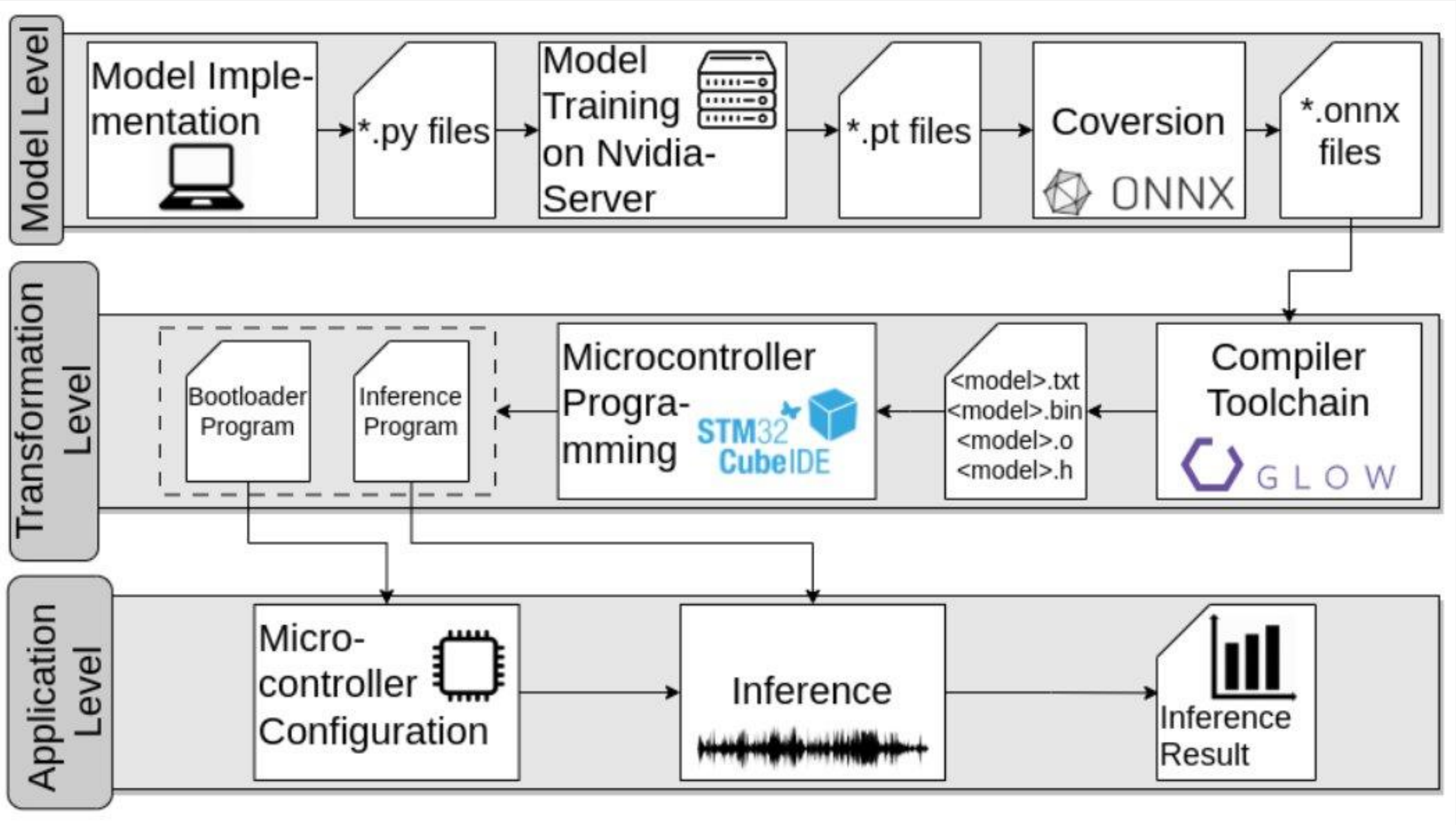
Tools:

- Pytorch
- ONNX file format
- Glow compiler

But these are ML Libraries!
– what about Signal Processing?

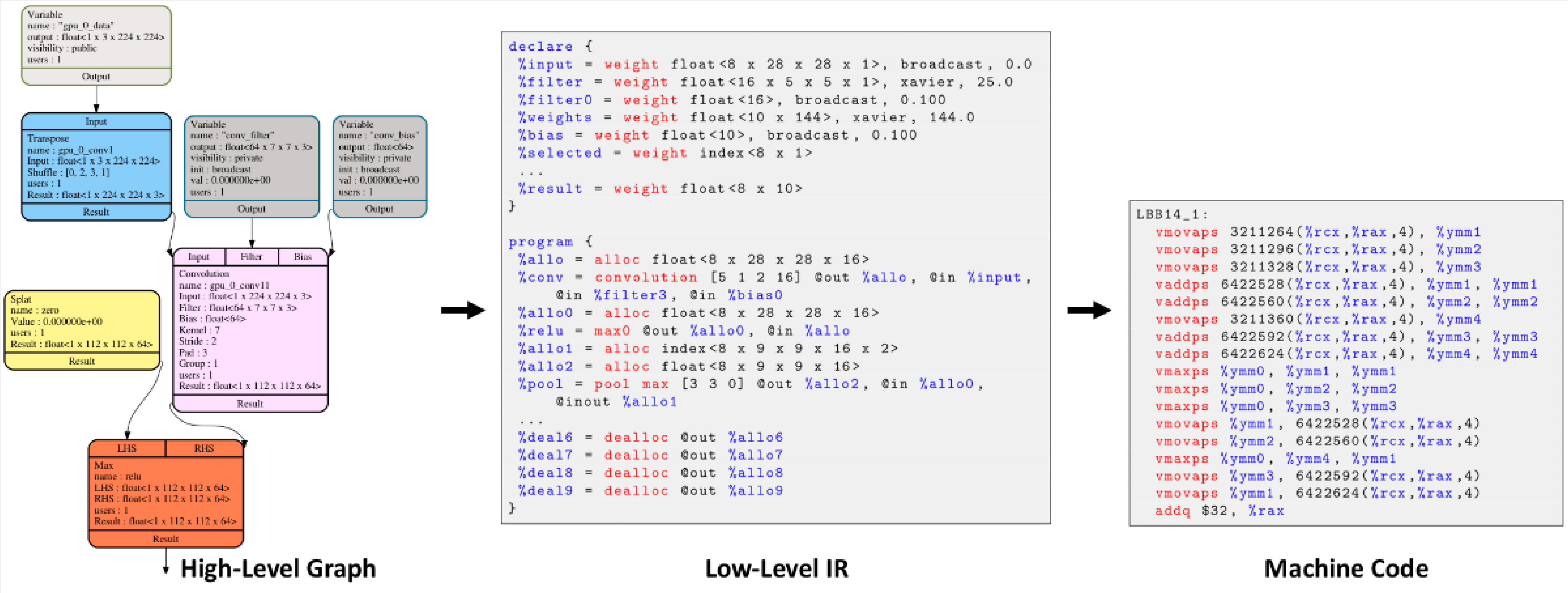
IP Protection?
– can be combined with secure boot!

Generic Model Transfer Strategy



Source: N. Günther. "Implementation and Evaluation of Keyword Spotting Neural Networks on Microcontrollers".

Graph Lowering



Workflow of the Pytorch Glow compiler toolchain

Energy Consumption of Operations

Integer	(@ 40nm)
Add	
8 bit	0.03pJ
32 bit	0.1pJ
Mult	
8 bit	0.2pJ
32 bit	3pJ

FP	
FAdd	
16 bit	0.4pJ
32 bit	0.9pJ
FMult	
16 bit	1pJ
32 bit	**4pJ**

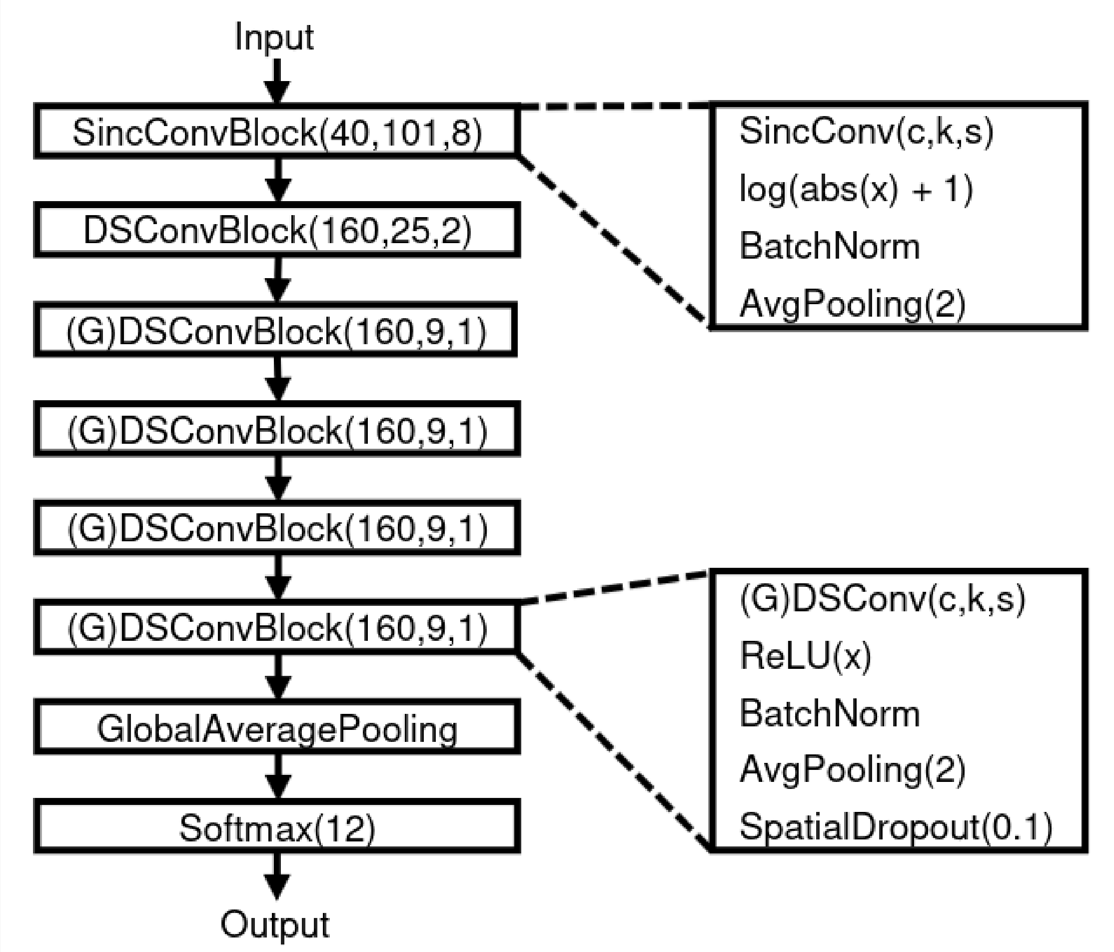
Memory	
Cache	(64 bit)
8KB	10pJ
32KB	20pJ
1MB	**100pJ**
DRAM	1.3-2.6nJ

Computation is relatively cheap, while memory accesses are expensive.

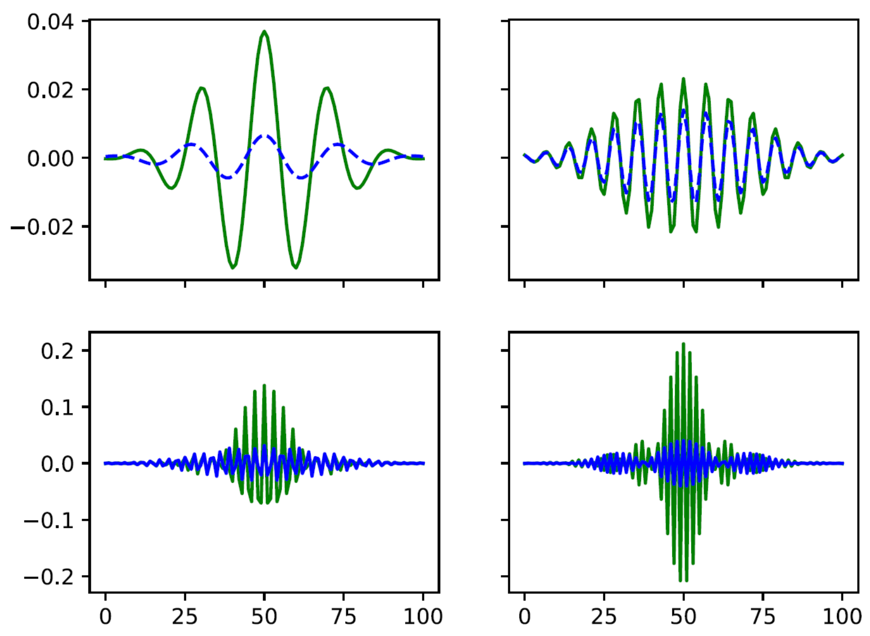
Source: M. Horowitz, “1.1 computing’s energy problem (and what we can do about it)". in 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), Feb 2014, pp. 10–14.

Practical Examples

Our Example Model



Always-On keyword spotting on raw audio



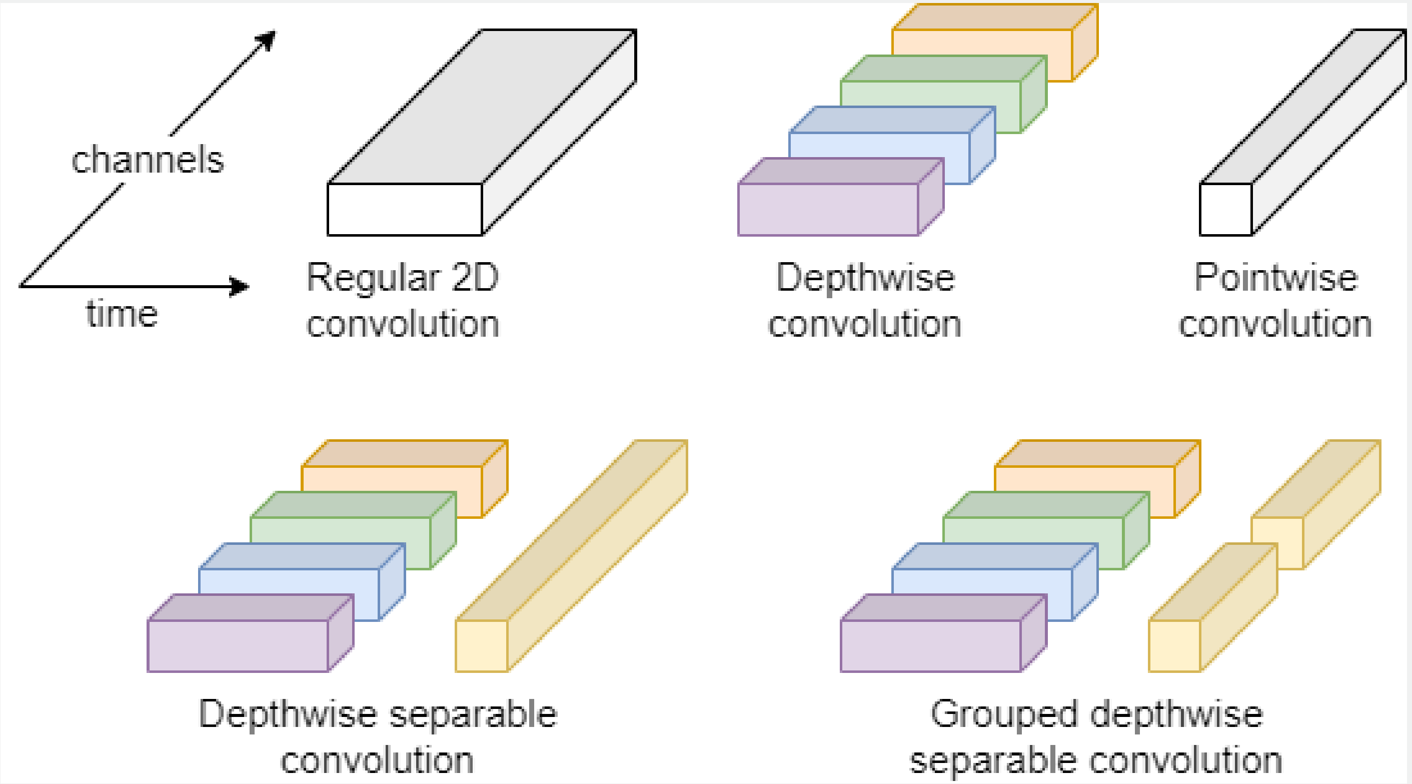
Learned filters at the input

Yes	No	Up	Left	On	Off	…

Keyword probabilities as output

Image Sources: Mittermaier et al., Small-Footprint Keyword Spotting on Raw Audio Data with Sinc-Convolutions
Kürzinger et al., Lightweight End-to-End Speech Recognition from Raw Audio Data Using Sinc-Convolutions

Depthwise Separable Convolutions



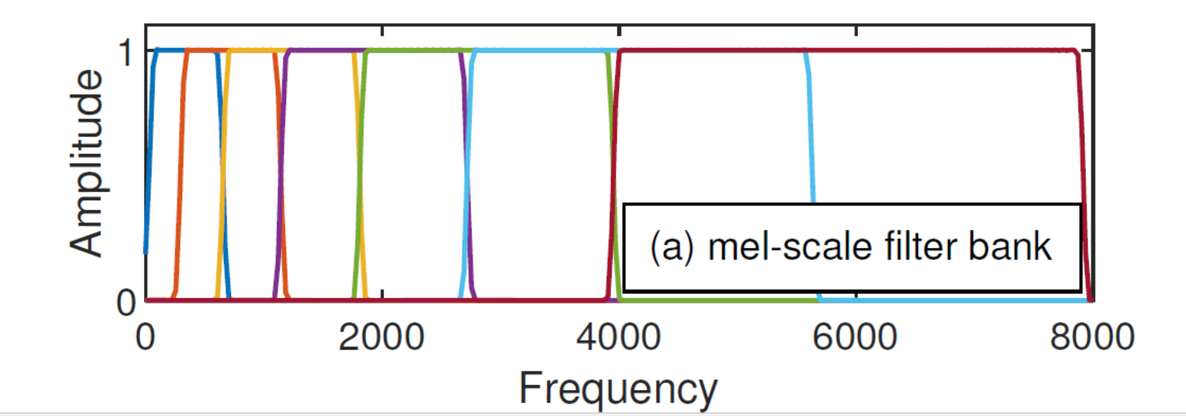
Depthwise separable convolutions:

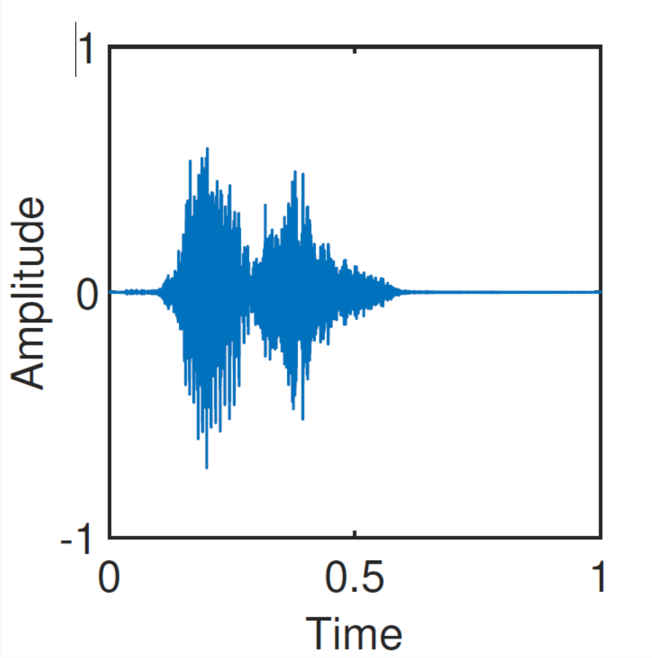
1x convolution along time
1x convolution across channels

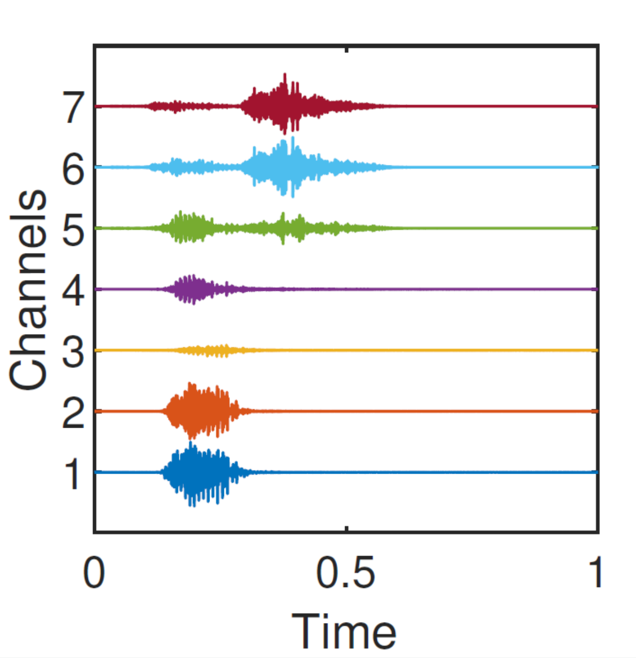
-> Less Parameters than regular convolutions

Image Source: Kürzinger et al., Lightweight End-to-End Speech Recognition from Raw Audio Data Using Sinc-Convolutions

Time-domain Bandpass using Sinc Convolutions







Sinc Convolutions:

$$\mathrm{sinc}(x) = \frac{\sin(x)}{x}$$

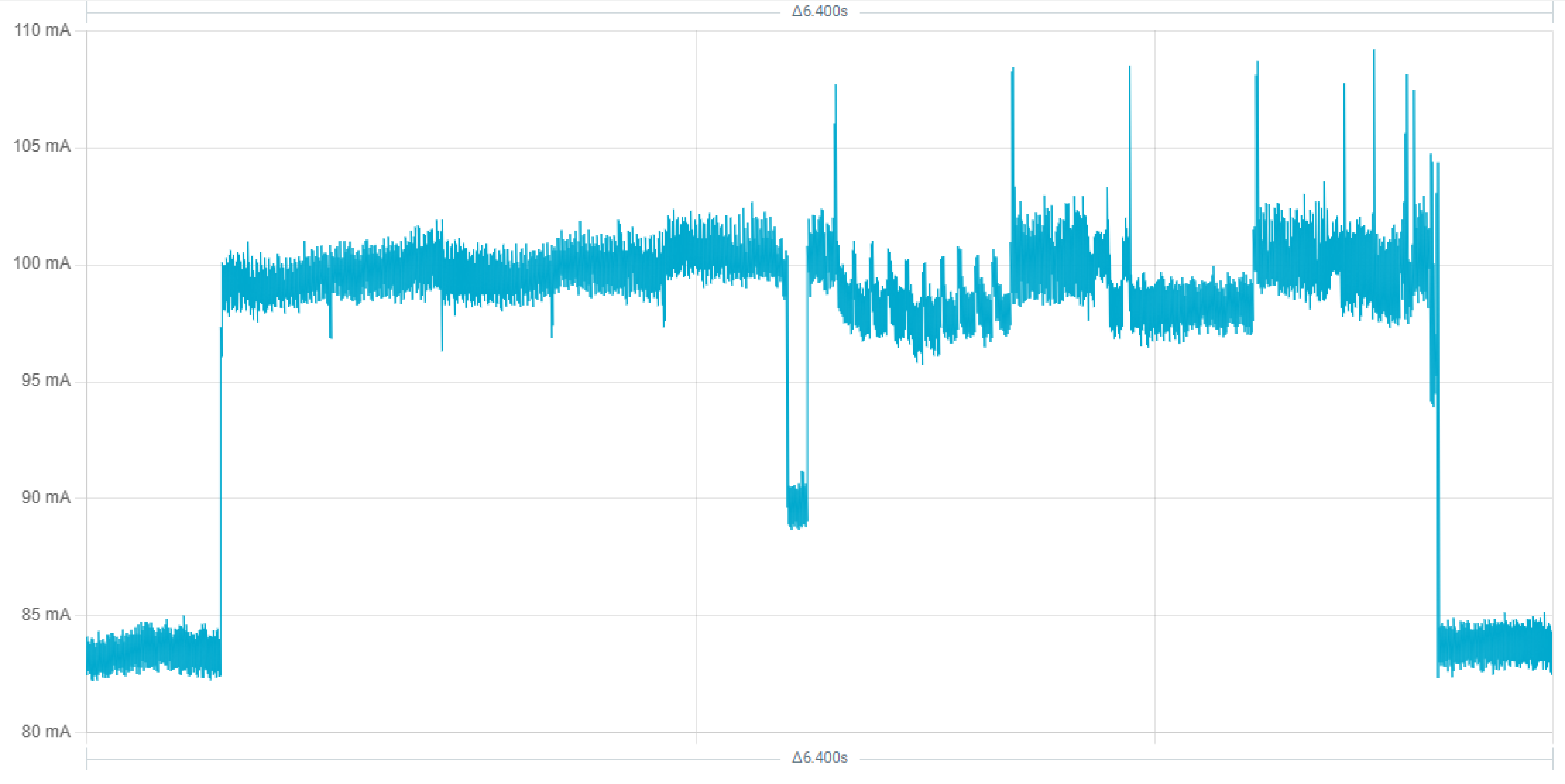
=> jump function in spectral domain

Only two parameters are needed to describe a bandpass:

Start and stop frequency

Image Source: Mittermaier et al., Small-Footprint Keyword Spotting on Raw Audio Data with Sinc-Convolutions

Power Draw Measurements



Example FP32 inference running at 5V on a Cortex-M33 Board

Conclusion

ingenics
DIGITAL

This talk discussed how to

- export your model for signal processing,
- deploy it on a microcontroller,
- and make it energy-efficient.

ingenics
DIGITAL

Questions?