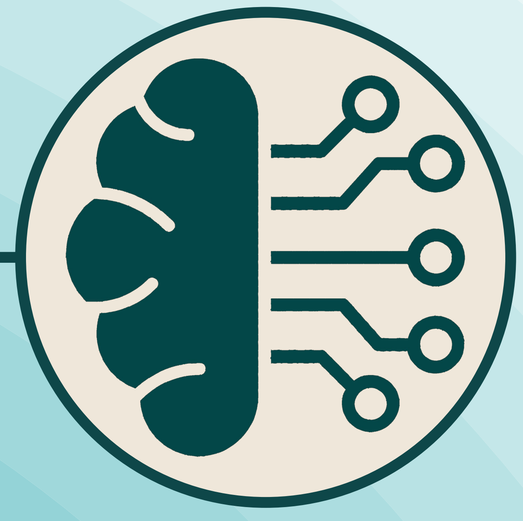
# High-Level Synthesis of Neural Networks for FPGAs

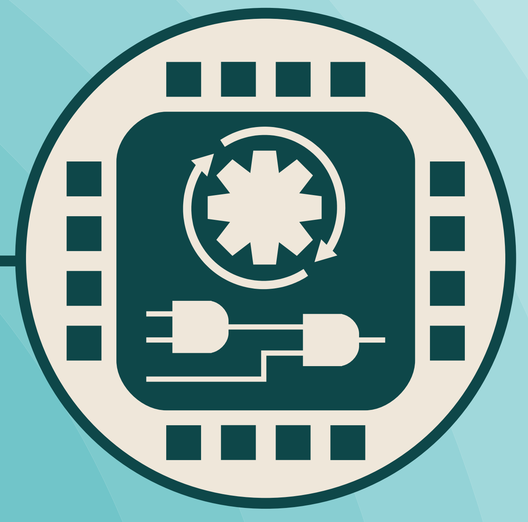
**Christof Schlaak** – Christophe Dubach

## All you need is Neural Networks



- For classification, prediction or clustering tasks
- *CHALLENGES*
  1) High computing needs & inherently parallel ➡ How can we **accelerate** NNs?
  2) Many different NN models
  3) **Implementations evolve** often and regularly

## All you need is FPGAs



- Highly **efficient** (better than CPU/GPU)!
- **Reconfigurable** (better than ASIC)➡ exploit different NN types & adapt to NN changes!
- *CHALLENGES*
  1) Designed in low-level HDL➡ error-prone and time-consuming
  2) Require HW design expertise
  3) Portability?

## All you need is ... LIFT ... is all you need

- LIFT is a **high-level functional, data-parallel language**
- LIFT offers **performance-portability**
- LIFT **generates optimised hardware implementations**

Dot product in LIFT

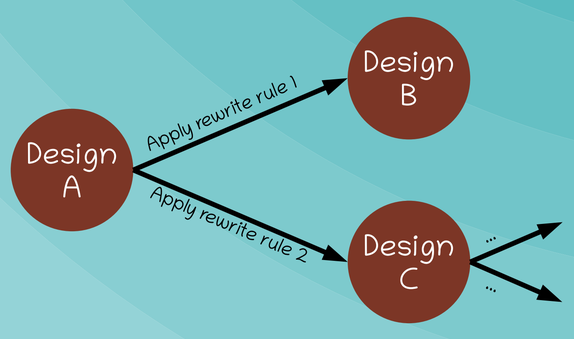
```
reduce(+,
    map(*,
        zip(s1,s2)
    )
)
```

### DSE

- **Design space exploration** using **rewrite rules** on algorithmic & hardware-specific level
- ML can be used to make the best rule choices

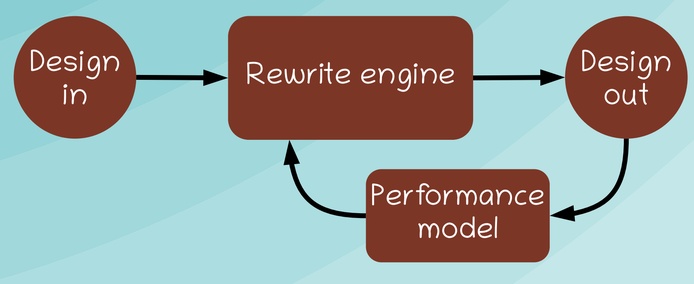
Map fission rule

```
map(f) =>
  split(k) >>
  map(map(f)) >>
  join
```



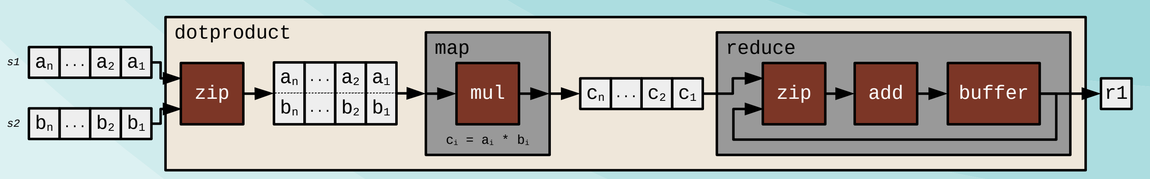

### Perf. Model

- **Performance model** to estimate design quality
- Feedback results into new design generation
- Generated design is precisely predictable
  ➡ very accurate estimations!




### HDL out

- Each functional expression implies a specific hardware design ➡ straightforward HDL generation
- No need for intermediate software-like representation




### NN types

- Support arbitrary NN architectures and tools (e.g. TF) through ONNX
- Support RNNs in future by adding further primitives (e.g. fold, mapAccum)

Microsoft Research

LIFT
lift-project.org

THE UNIVERSITY of EDINBURGH informatics