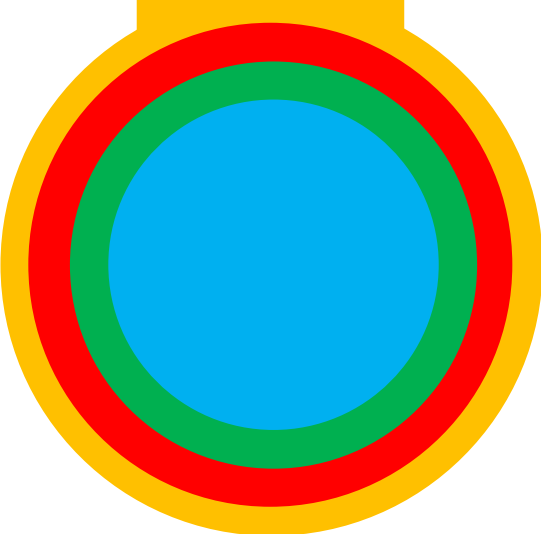


Functional TPU: A Proof-of-Concept AI Accelerator

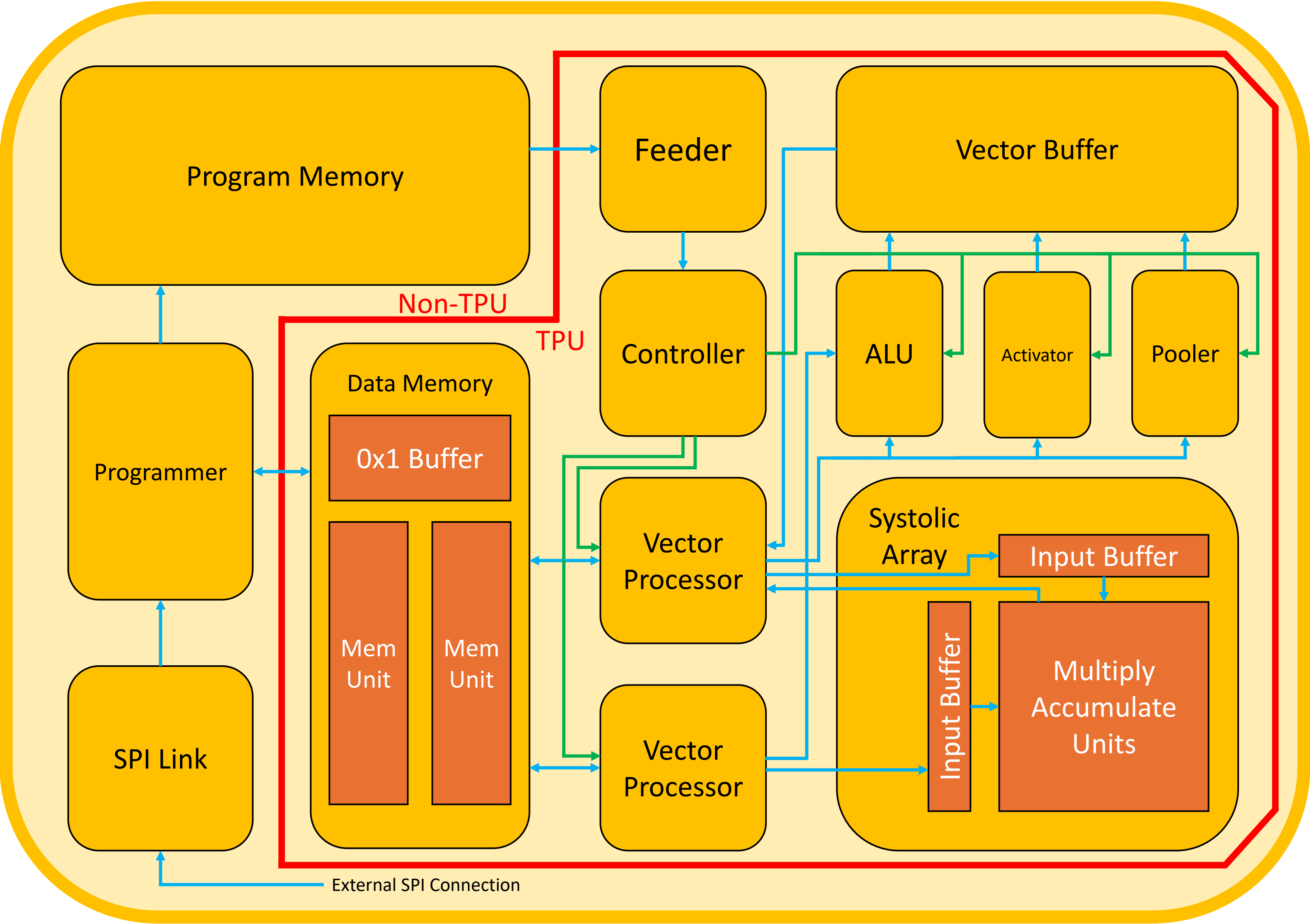
Gabe Litteken,
Peter Jamieson
Department of Electrical
and Computer Engineering,
Miami University



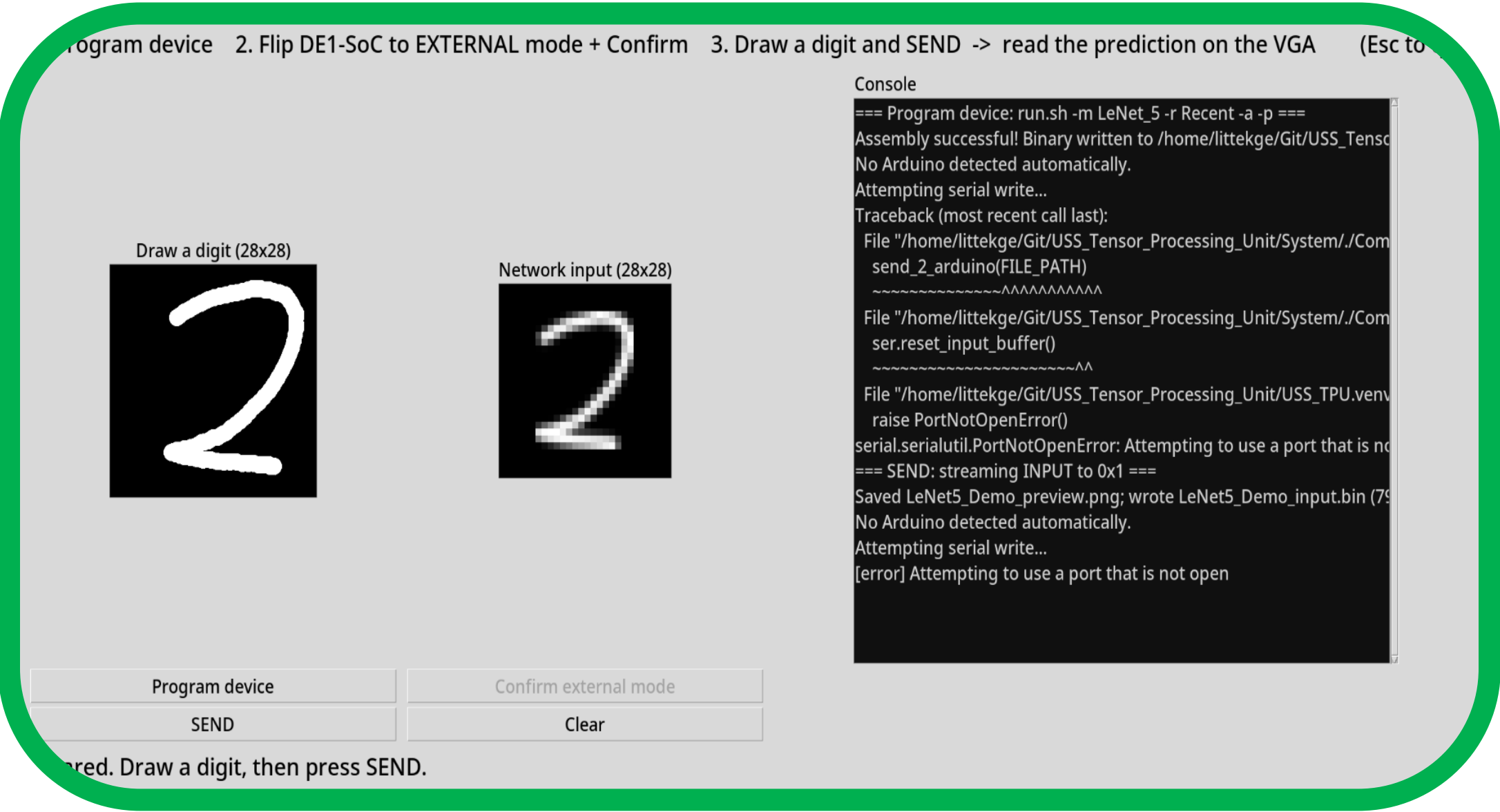
Functional TPU (Tensor Processing Unit) is a processor (in Verilog) that runs small neural networks using hardware acceleration to speed up computation. Our *Functional TPU* illustrates a possible architecture that students and researchers can freely explore and reuse to create their own TPUs.

The Heart of a TPU: The Systolic Array

The primary method of hardware acceleration in the *Functional TPU* is the *systolic array*. A systolic array is composed of many *multiply-accumulate units* linked together to maximize reuse of data during matrix multiplications.

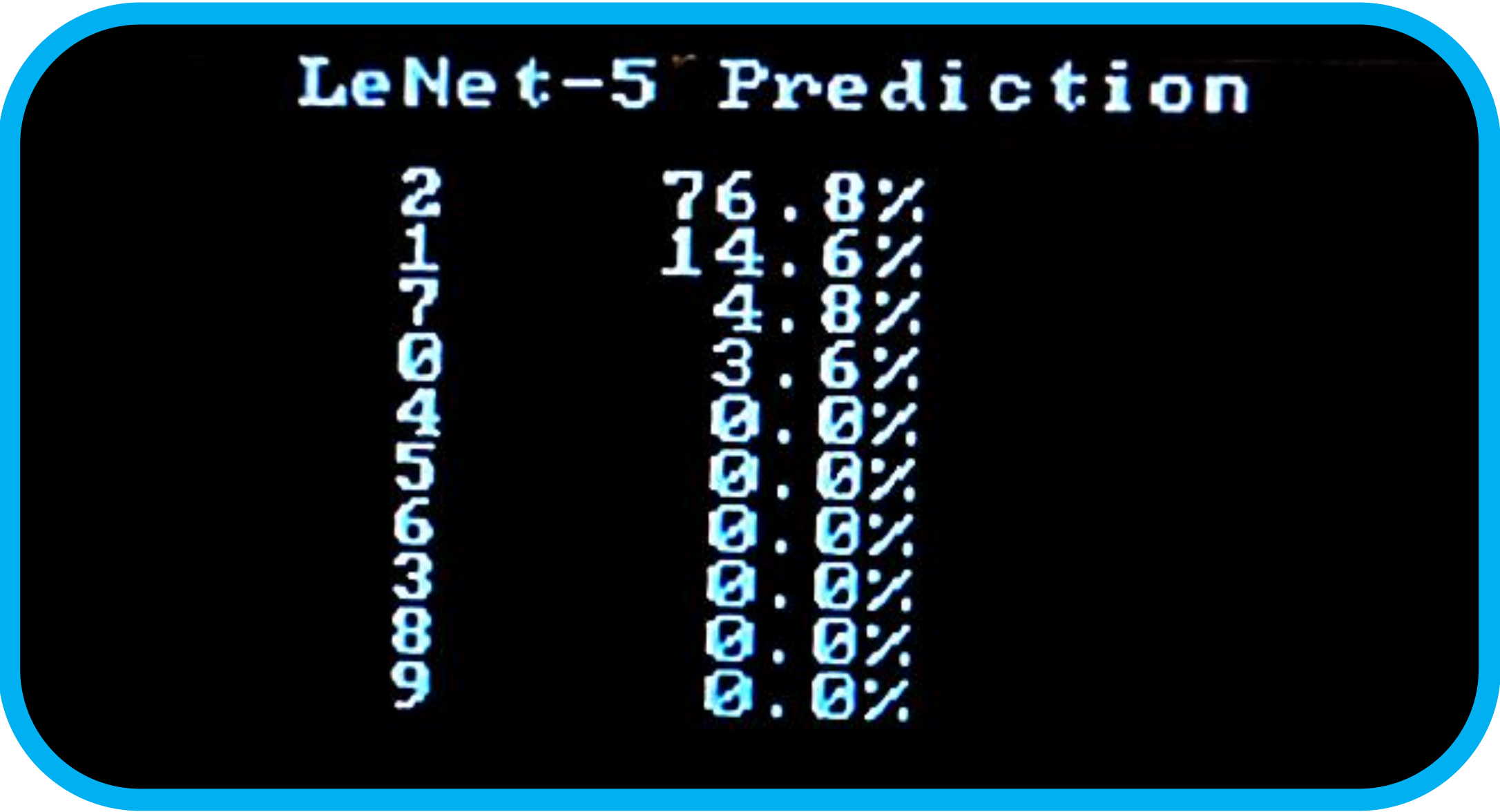


Digit Recognition!



TPU Diagram

Control Signal Data Signal



The Dark Factory Approach

The *Dark Factory development approach* (using AI to write code) made a functional TPU possible. The TPU's layout and functions were planned and documented in multiple spec documents, then converted to Verilog HDL using Claude Code. The design was implemented on an FPGA.

Module Descriptions

Listed below are descriptions of each module following the general path of data.

Note: Control signals for each module are described as follows:

- Each module defines control signals relevant to their own functionality (e.g. a module that reads an *idle* signal from another module would not define that signal as an input).
- Many combinational signals (usually data signals) are left up to user implementation unless explicitly defined.
- A user may define other control signals for each module as long as the explicit signals are implemented as described.

Tensor_Processing_Unit

Description: Top level module for the entire project. This module only exists as a wrapper for the TPU module so that the TPU module can interface with clock signals and IO from the target hardware. It also allows for easy modification of *clk* and *rst* signals for debugging.

TPU

Description: Top level module for the TPU. In accordance with the module hierarchy, this module instantiates the majority of the other modules. This module also defines connection routing between its various submodules and serves as a passthrough for relevant signals (*clk*, *rst*, *SPI* Signals, etc.).

Synchronous Control Signals:

Inputs:

Access Design At:

https://github.com/littekge/USS_Tensor_Processing_Unit