



VLSI Implementation for DNA Sequencing

Krishna Teja Malladi
Karthika Periyathambi
Piyush Keshri

Outline

- Motivation
- Overview
 - Smith Waterman Algorithm
- Hardware Architecture
 - Processing Grid
 - Path Tracker
- Validation
- Synthesis
- Hardware vs Software
- Place and Route
- Future Work

Smith Waterman Algorithm

- Local Alignment DNA Match/Mismatch
- Determine Best possible Match from genome database
- Locate exact position of insertion/deletion/mutation
- Most accurate Bioinformatics algorithm since 1981, but slowest too...

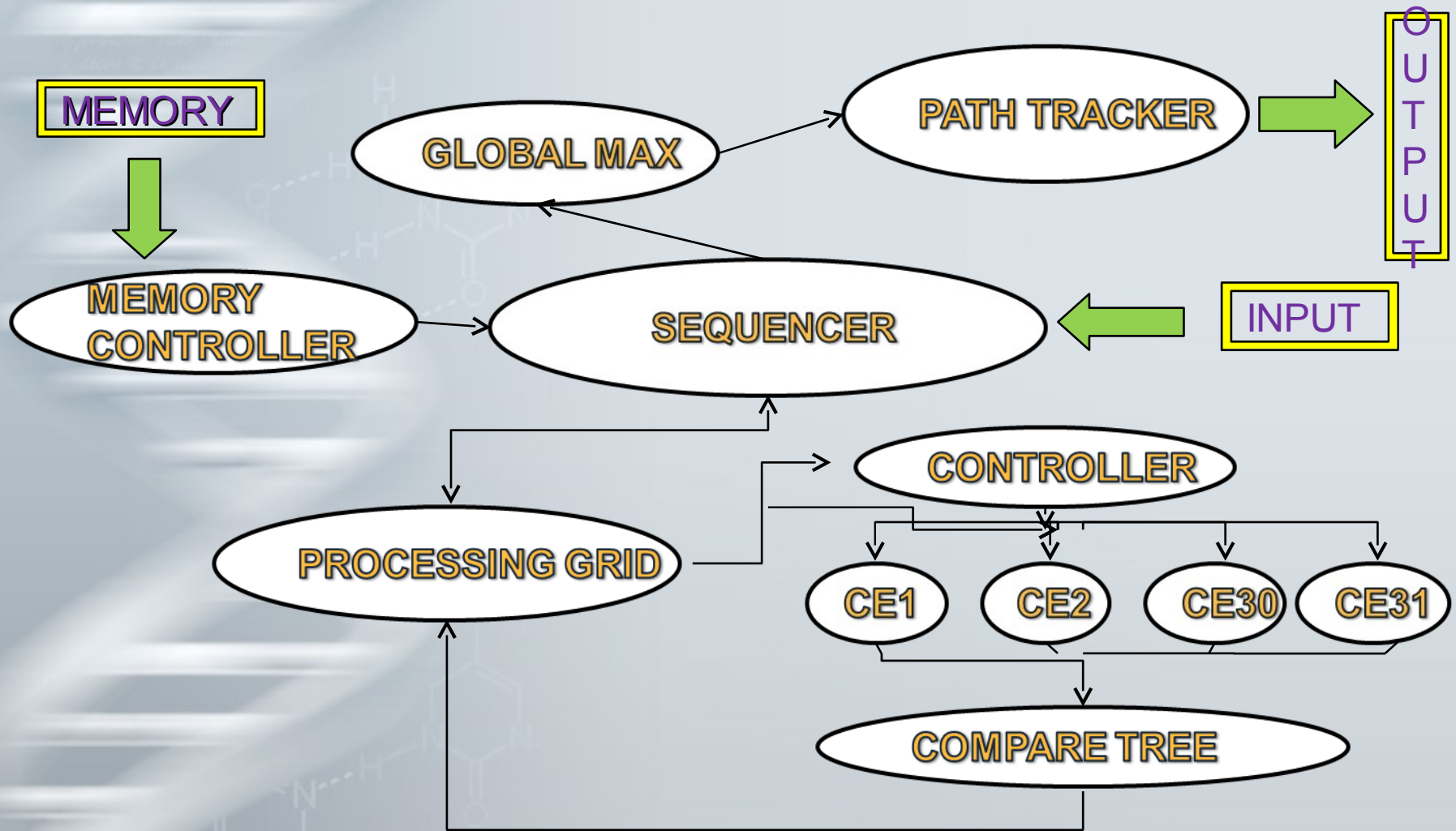
Demo

Math behind...

$$H(i,0) = 0, 0 \leq i \leq m \quad 0$$
$$H(i,j) = \max \left\{ \begin{array}{ll} H(i-1, j-1) + w(a_i, b_j) & \text{Match/Mismatch} \\ H(i-1, j) + w(a_i, -) & \text{Deletion} \\ H(i, j-1) + w(-, b_j) & \text{Insertion} \end{array} \right\}, 1 \leq i \leq m, 1 \leq j \leq n$$

- a, b = Input Sequences (Query + Database)
- m = length(a)
- n = length(b) (m==n here)
- $H(i,j)$ - is the maximum Similarity-Score between a suffix of a[1...i] and a suffix of b[1...j]
- '-' is the gap-scoring scheme
- A, T, C, G sequence elements represented by 00,01,10,11

Hardware Control Flow



Verilog Code

Memory

- Register files
- Latency with 8 bit communication

Processing Grid

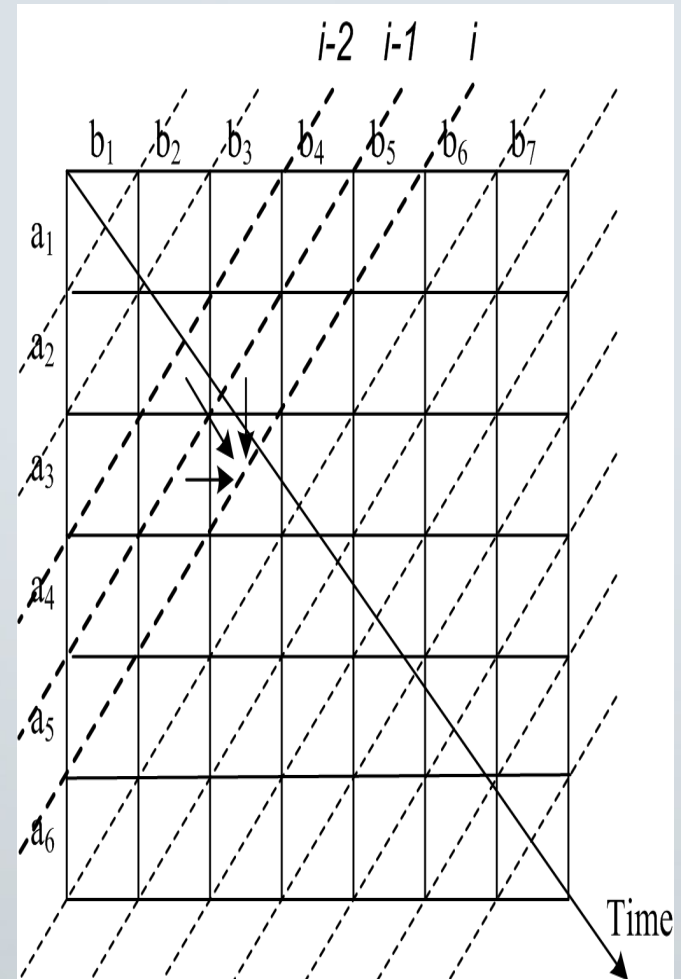
- Diagonal parallelism
- Emphasis on Hardware reuse
- Parallelization and Streaming Techniques
- Optimization of matrix non-storage: mimic FIFO

Path Tracker

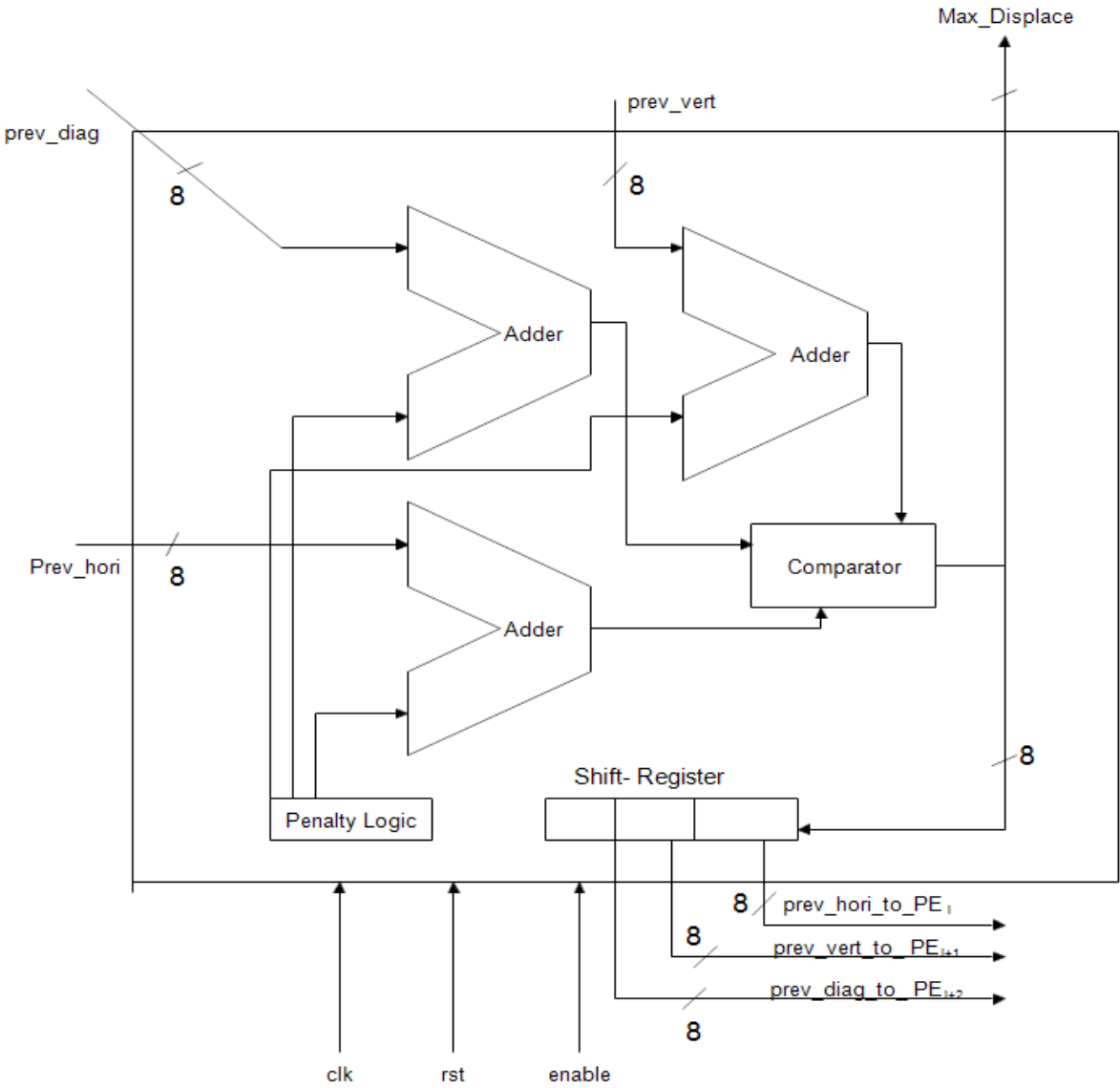
- Computes Matrix for global Max
- Stores H-matrix & Direction of Traversal
- Tracks back path

Output

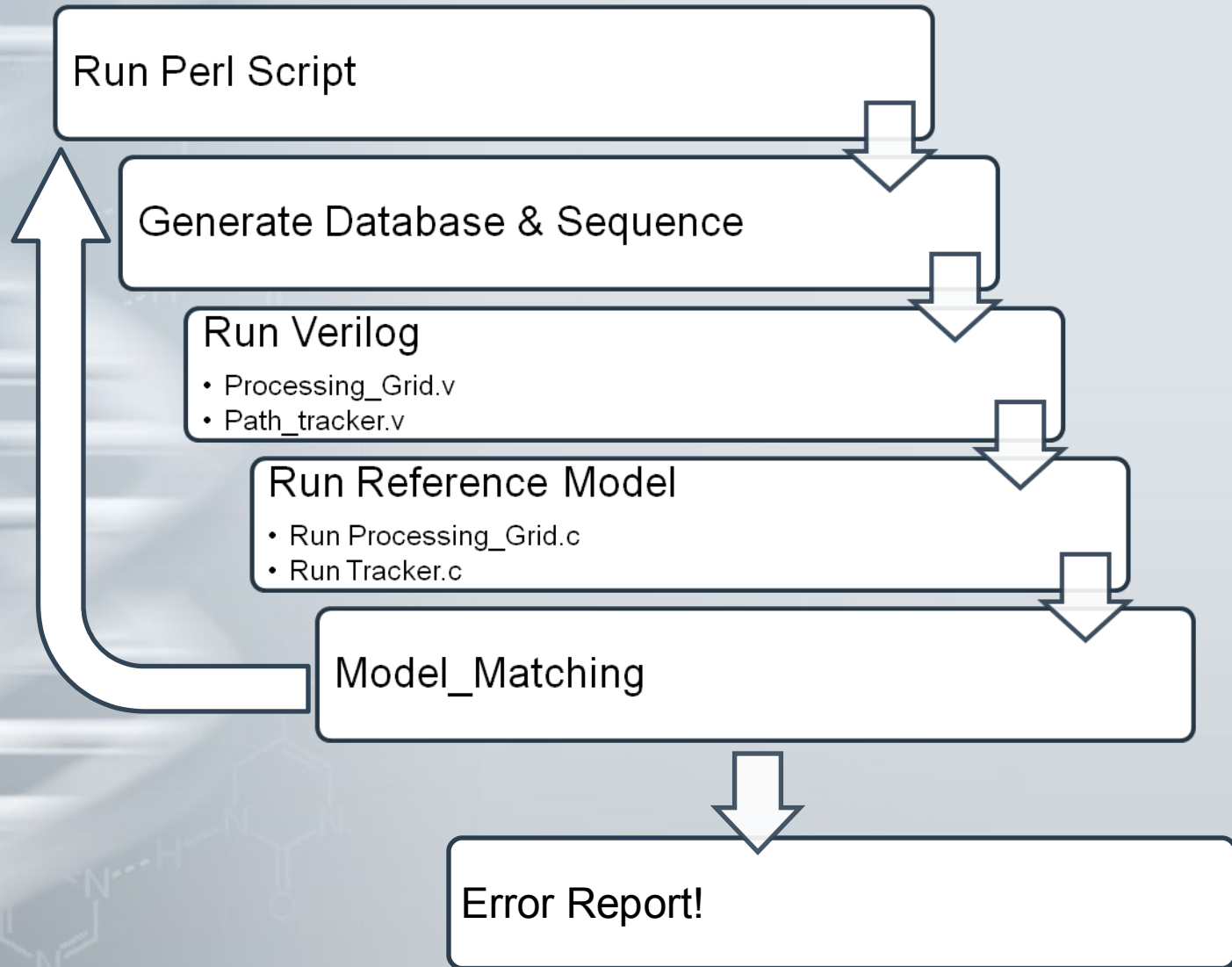
- A 192bit + 192bit mutations.
- Serialized to 8 bit output
- A done_assert signal.



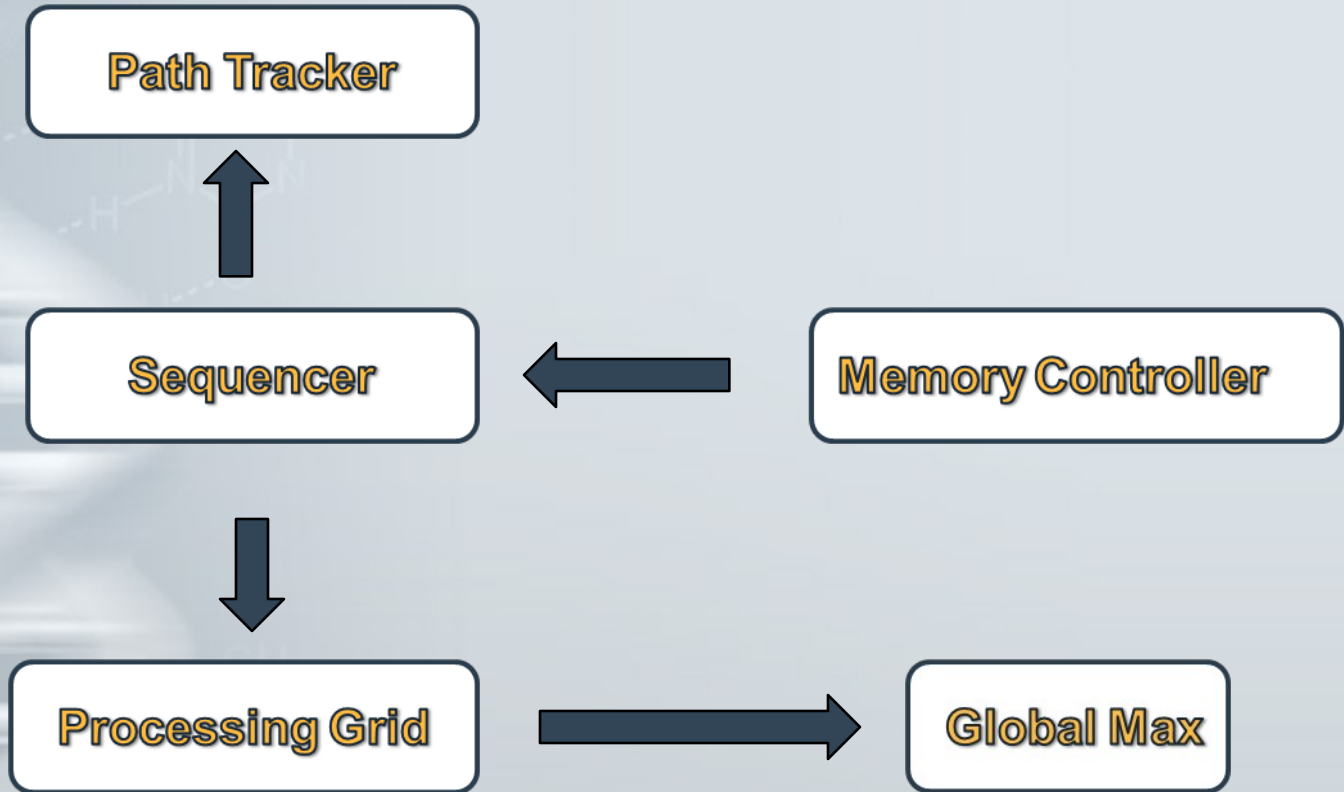
Compute Element



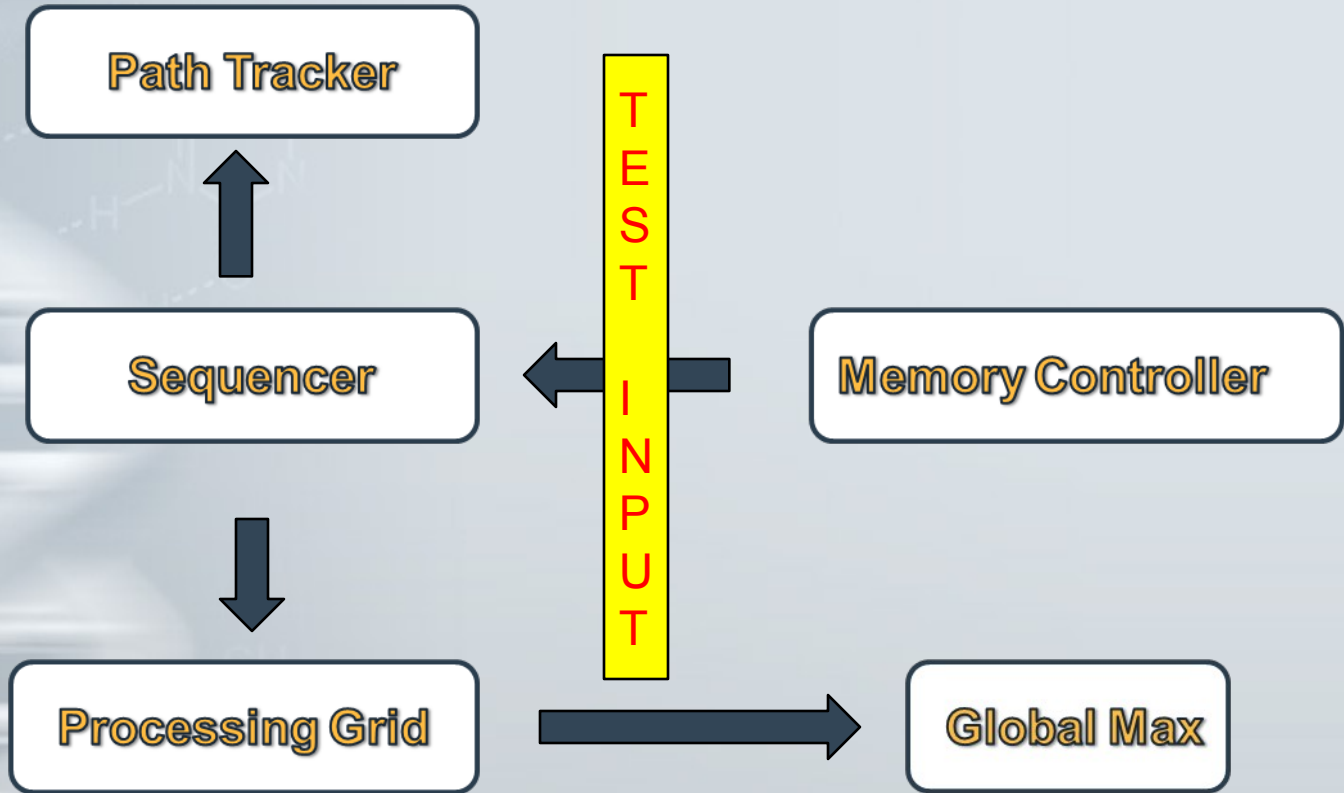
Validation of Design



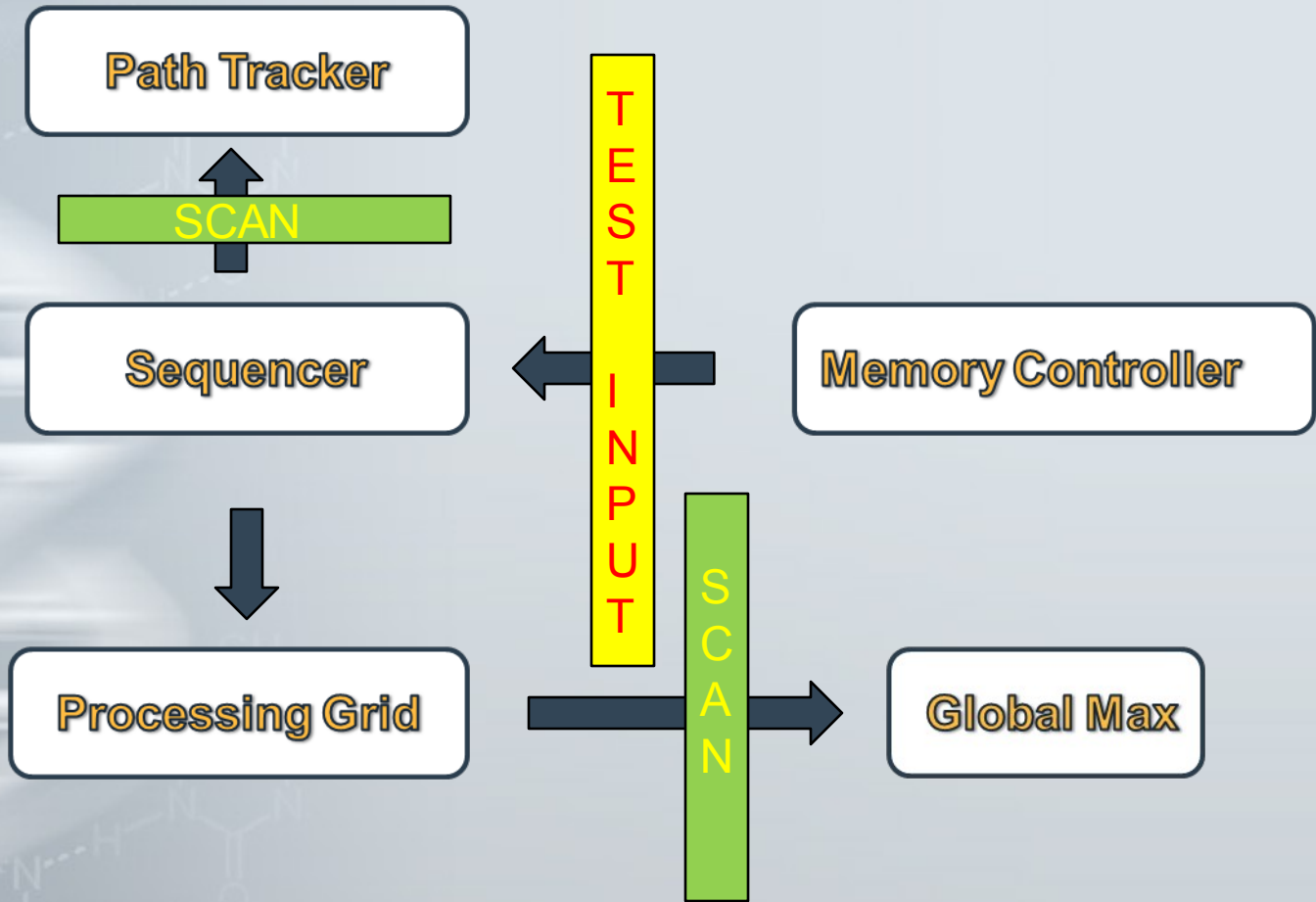
Validation



Validation



Validation



Test cases

- Random Cases Generator
- Corner cases
 - Perfect match
 - Perfect mismatch
 - zig-zag patterns
 - All zeroes
- “Grand catch”

ambiguous case-> shift by one

AAAAAAAAA ⇔ GAAAAAAAA

Results: 1) Insert: -AAAAAAAAA ⇔ GAAAAAAAA

2) Delete: AAAAAAAAA ⇔ AAAAAA-

3) Mismatch: -AAAAAAAA ⇔ -AAAAAAAA

Synthesis

- Synthesized phy_top :
 - Wraps Pads, template
 - Instantiating sequencer: Top module
- Included Pad Libraries
- Synthesized for 333 MHz.
- Set_dont_touch on Vdd, Vss instances (for non-removal of nets)
- Minor Internal errors' debug.
- Less Clock gating due to high activity factor.

Synthesis Summary

Clock Frequency	333 MHz
Negative Slack	None
Throughput	~9.8 Million QPS
Voltage	0.9 V
Dynamic Power	~373 mW
Leakage Power	~ 0.8 mW
I/O Pins	26

Hardware vs. Software

Sequential version: Throughput = 0.14 Million QPS

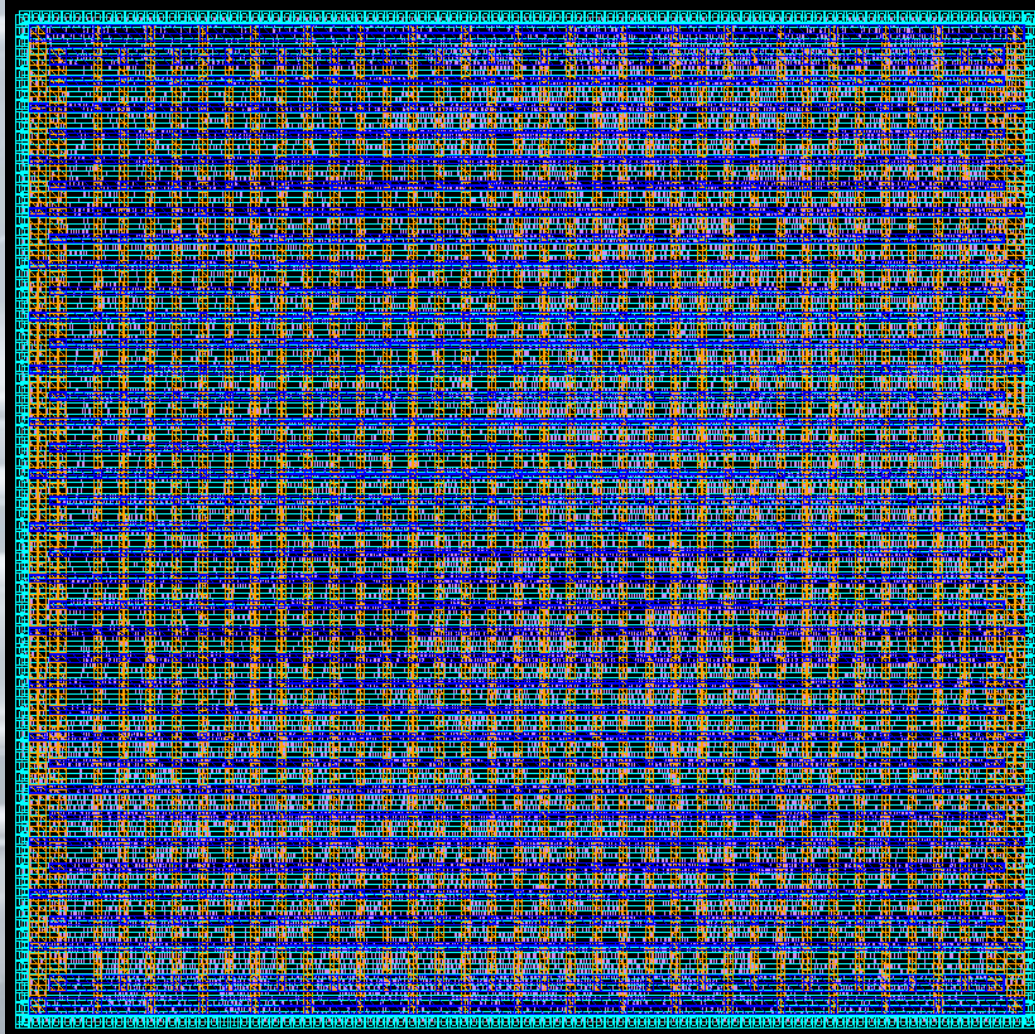
Pthread version : Throughput = 1.8 Million QPS

Hence, **Hardware at least 5X faster than parallelized code.**

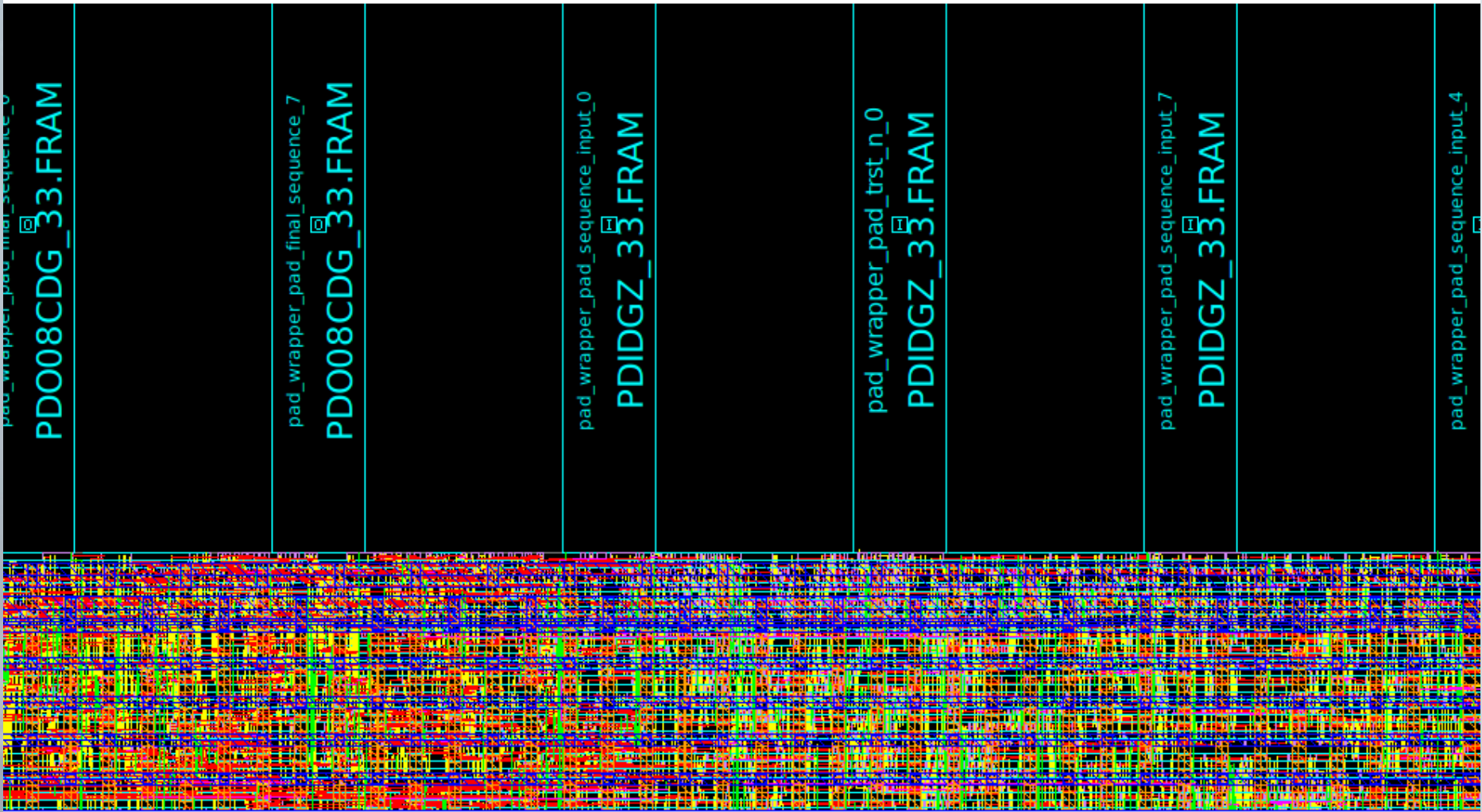
Place and Route

- Phy_top netlist placed and routed.
- Tie Cells included to join Vdd and Vss nets.
- Preroute VDD, VSS to deliver power to pads.
- Set_pad_physical constraints and place pads
- Reduced core_utilization to 0.5.
- Sign off
 - Filler cells
 - Antenna fixing

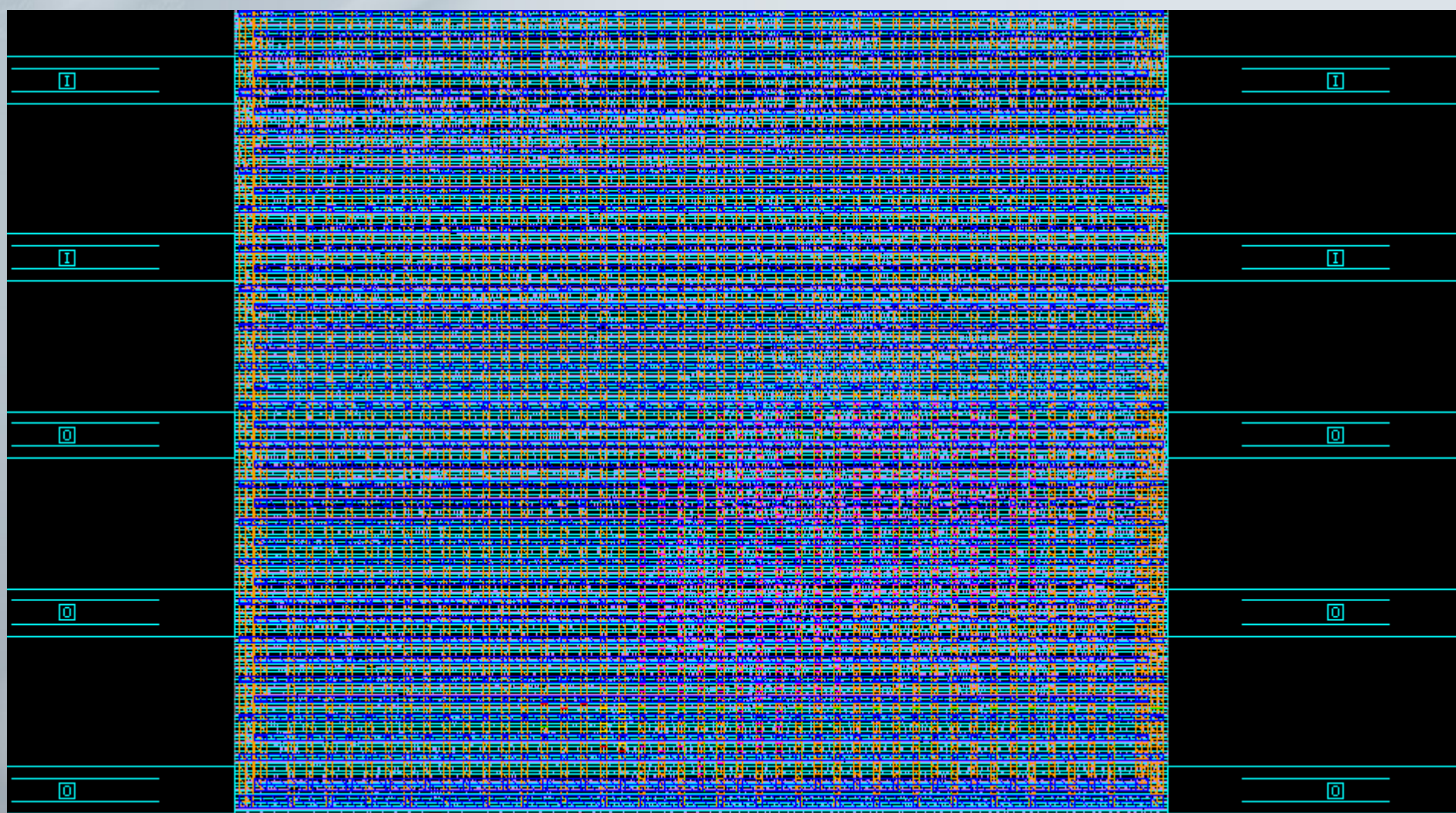
P&R of Base Design (without Memory and Pads)



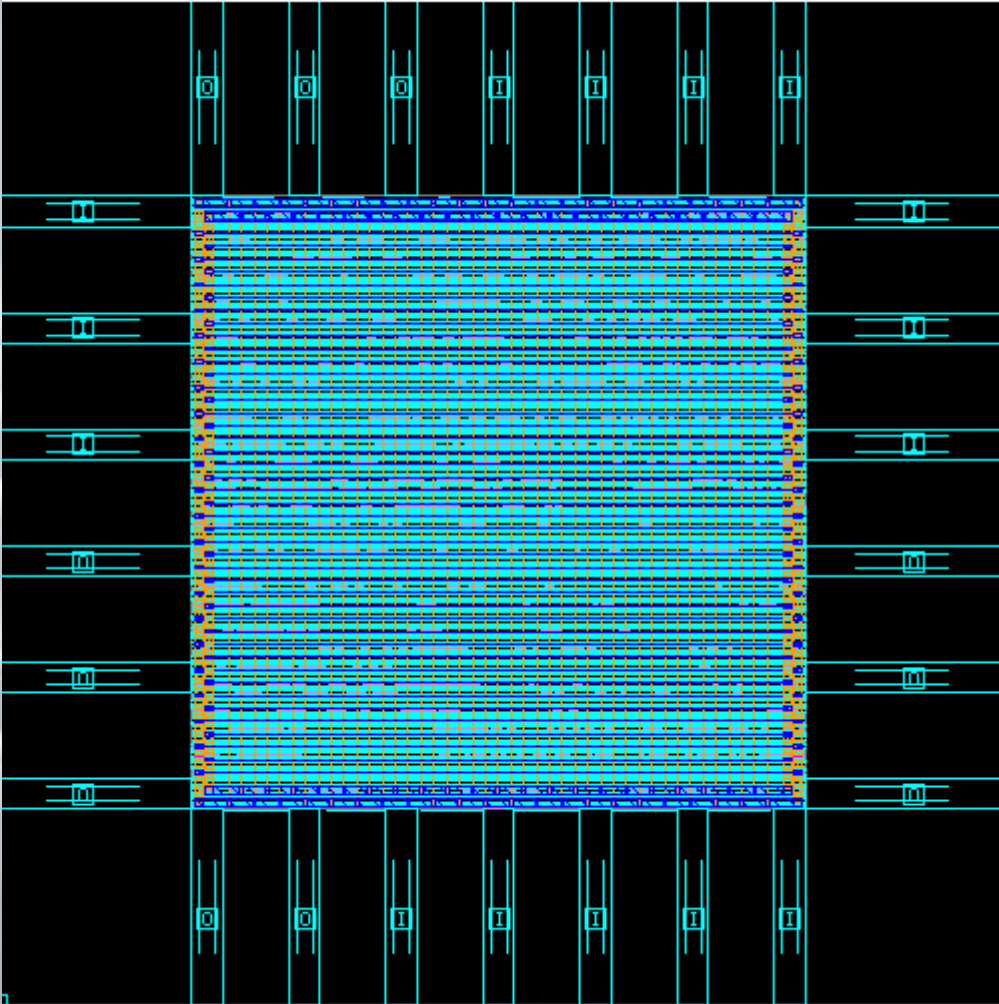
Pads on the Inputs



Overall Design



Final Layout



P&R Statistics

Clock Frequency	333 MHz
Timing Violations	None
Area	0.32 mm ²
Cell Count	~23,000
Standard Cell Utilization	91.4%
Core Utilization	0.5

Smith and Waterman



Thank You!