SRI HARI A S
sriharias2204@gmail.com

Predictive Runtime Hybrid Floating-Point Unit for Energy-Efficient AI and Signal Processing

---

Abstract

This project presents a predictive hybrid floating-point unit (FPU) that dynamically selects between exact FP32 arithmetic, approximate FP32 arithmetic, and BF16 trans-precision based on operand characteristics. The design uses exponent- and mantissa-based look-ahead logic to determine the safest and most energy-efficient computation mode before execution. To ensure practical ASIC feasibility, operand isolation and pipelined prediction are used so that only the selected datapath toggles during operation. The architecture is verified through RTL simulation, synthesized using Cadence Genus on a 90 nm standard-cell library, validated through logical equivalence checking and done Physical Flow using Cadence Innovus.

---

1. Introduction

Modern machine-learning and signal-processing workloads do not always require full IEEE-754 FP32 precision. In many operations, reduced precision or approximate arithmetic can achieve nearly identical application-level accuracy while consuming significantly less energy. However, fixed-precision or fixed-approximation designs risk numerical instability when operands exhibit cancellation or extreme exponent values.

This project proposes a **runtime-predictive hybrid FPU** that automatically selects the most appropriate arithmetic mode (exact FP32, approximate FP32, or BF16) on a per-operation basis using operand-aware logic.

---

2. Proposed Architecture

The FPU consists of three parallel arithmetic datapaths:

- Exact FP32 adder and multiplier
- Approximate FP32 adder and multiplier
- BF16 adder and multiplier

A lightweight **predictor** examines the input operands before execution by analysing:

- Exponent difference (for dominance and cancellation)
- Exponent sum (for multiplication stability)
- Mantissa similarity
- Mantissa activity and exponent range (for BF16 suitability)

Based on this analysis, a control FSM selects one of the three modes. **Operand isolation** ensures that only the selected datapath receives active inputs, preventing unnecessary power consumption in unused units.

---

3. ASIC Implementation

For the ASIC flow, only the **exact FP32** and **approximate FP32** datapaths were enabled and synthesized. The BF16 path was retained in the RTL for architectural completeness but was excluded from the Genus synthesis to ensure a fair and focused evaluation.

The design was synthesized using:

- Cadence Genus
- 90 nm CMOS standard-cell library

Power, area, and timing were analysed under typical operating conditions.

Power Report :

```
      : Info=6, Warn=2, Error=0, Fatal=0
Instance: /fpu_mode
Power Unit: W
PDB Frames: /stim#0/frame#0
  ------------------------------------------------------------------------------
    Category         Leakage     Internal    Switching        Total    Row%
  ------------------------------------------------------------------------------
      memory     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
    register     4.34057e-08  4.56083e-04  7.00905e-06  4.63135e-04  80.64%
       latch     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
       logic     1.70381e-08  5.21883e-05  3.33037e-05  8.55090e-05  14.89%
        bbox     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
       clock     0.00000e+00  0.00000e+00  2.56520e-05  2.56520e-05   4.47%
         pad     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
          pm     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
  ------------------------------------------------------------------------------
    Subtotal     6.04438e-08  5.08271e-04  6.59647e-05  5.74296e-04 100.00%
  Percentage           0.01%       88.50%       11.49%      100.00% 100.00%
  ------------------------------------------------------------------------------
@genus:root: 6>
```

Area Report :

```
      : Info=6, Warn=2, Error=0, Fatal=0
Instance: /fpu_mode
Power Unit: W
PDB Frames: /stim#0/frame#0
  ------------------------------------------------------------------------------
    Category         Leakage     Internal    Switching        Total    Row%
  ------------------------------------------------------------------------------
      memory     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
    register     4.34057e-08  4.56083e-04  7.00905e-06  4.63135e-04  80.64%
       latch     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
       logic     1.70381e-08  5.21883e-05  3.33037e-05  8.55090e-05  14.89%
        bbox     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
       clock     0.00000e+00  0.00000e+00  2.56520e-05  2.56520e-05   4.47%
         pad     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
          pm     0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00   0.00%
  ------------------------------------------------------------------------------
    Subtotal     6.04438e-08  5.08271e-04  6.59647e-05  5.74296e-04 100.00%
  Percentage           0.01%       88.50%       11.49%      100.00% 100.00%
  ------------------------------------------------------------------------------
@genus:root: 6> report_area
============================================================
  Generated by:           Genus(TM) Synthesis Solution 21.14-s082_1
  Generated on:           Sep 04 2025  02:54:10 pm
  Module:                 fpu_mode
  Technology libraries:   fast_vdd1v0 1.0
                          fast_vdd1v0 1.0
  Operating conditions:   PVT_1P1V_0C (balanced_tree)
  Wireload mode:          enclosed
  Area mode:              timing library
============================================================

Instance Module  Cell Count  Cell Area  Net Area   Total Area   Wireload
-----------------------------------------------------------------------------
fpu_mode               242   1100.898     0.000     1100.898 <none> (D)
  (D) = wireload is default in technology library
@genus:root: 7>
```

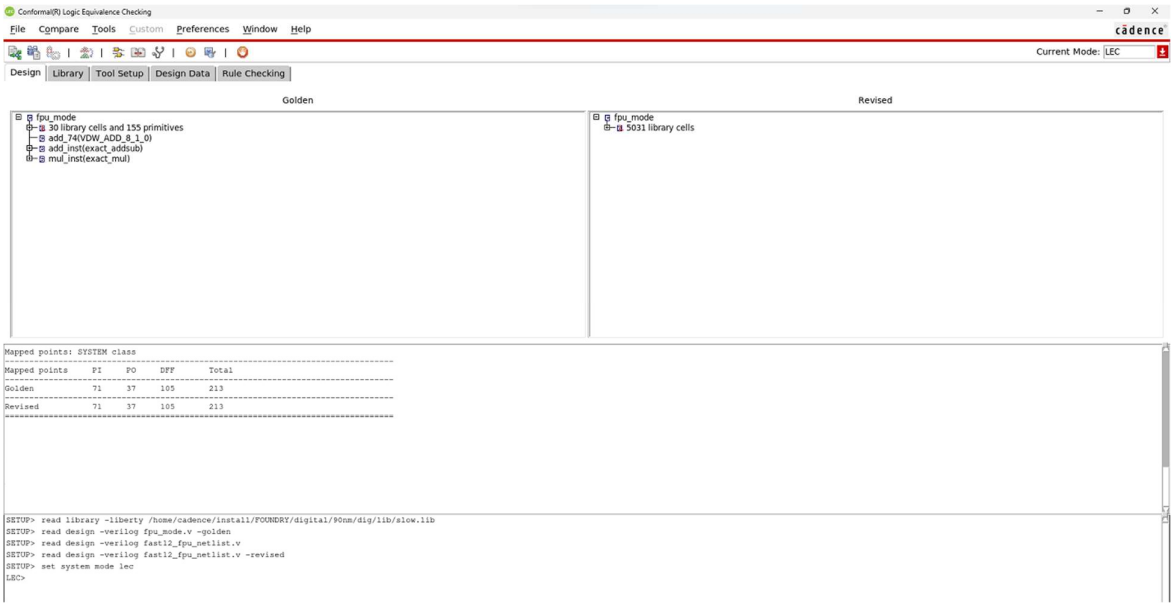
Timing Report :

```
nstance Module  Cell Count  Cell Area  Net Area   Total Area   Wireload
------------------------------------------------------------------------
pu_mode                242   1100.898     0.000     1100.898 <none> (D)
 (D) = wireload is default in technology library
genus:root: 7> report_timing
============================================================
 Generated by:           Genus(TM) Synthesis Solution 21.14-s082_1
 Generated on:           Sep 04 2025  02:54:57 pm
 Module:                 fpu_mode
 Operating conditions:   PVT_1P1V_0C (balanced_tree)
 Wireload mode:          enclosed
 Area mode:              timing library
============================================================


ath 1: MET (484 ps) Setup Check with Pin B_buf_reg[30]/CK->SE
        Group: clk
   Startpoint: (R) state_reg[0]/CK
        Clock: (R) clk
     Endpoint: (F) B_buf_reg[30]/SE
        Clock: (R) clk

                  Capture       Launch
      Clock Edge:+    1000            0
     Src Latency:+       0            0
     Net Latency:+       0 (I)        0 (I)
         Arrival:=    1000            0

           Setup:-      56
     Uncertainty:-      10
   Required Time:=     934
    Launch Clock:-       0
       Data Path:-     450
           Slack:=     484

-------------------------------------------------------------------------------------------------
  Timing Point    Flags   Arc   Edge    Cell       Fanout Load Trans Delay Arrival Instance
                                                      (fF)  (ps)  (ps)   (ps)  Location
-------------------------------------------------------------------------------------------------
 state_reg[0]/CK  -       -     R     (arrival)      106    -   100     0       0    (-,-)
 state_reg[0]/Q   -       CK->Q F     DFFRHQX1HVT      3  1.0    14   102     102    (-,-)
 g6991__6260/Y    -       B->Y  R     NOR3BX1HVT       3  1.2    53    45     147    (-,-)
 g6965__8428/Y    -       B->Y  F     NAND2X2HVT      52 20.8   225   151     298    (-,-)
 drc_bufs7024/Y   -       A->Y  R     INVX1HVT         1  0.4    38    97     395    (-,-)
 drc_bufs7023/Y   -       A->Y  F     INVX1HVT        16  6.4    70    55     450    (-,-)
 B_buf_reg[30]/SE <<<     -     F     SDFFQX1HVT      16    -     -     0     450    (-,-)
-------------------------------------------------------------------------------------------------
```

---

4. LEC

Logical equivalence between RTL and the synthesized netlist was confirmed using Cadence Conformal LEC, ensuring that synthesis optimizations did not alter the design behaviour.

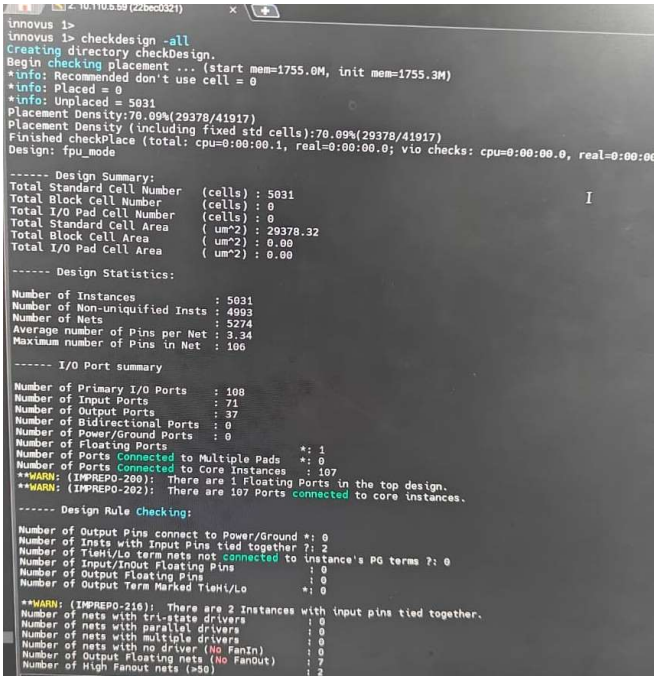


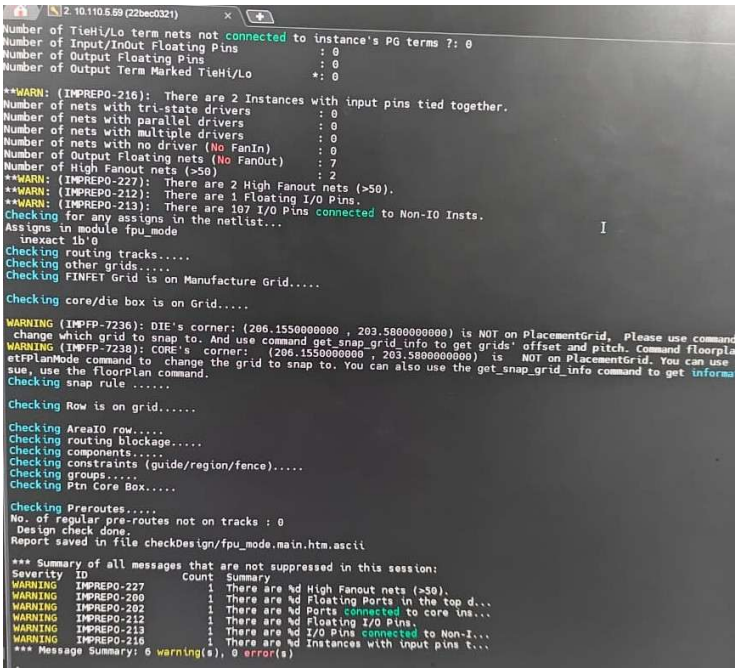
5.ASIC Physical Design Flow

After successful RTL simulation, synthesis, and logical equivalence checking, the proposed **Predictive Hybrid FPU** was taken through a complete **ASIC physical design flow** using **Cadence Innovus Implementation System**.

The **gate-level netlist** generated from Cadence Genus was imported into Innovus along with the **90 nm standard-cell technology library**. The following physical design steps were carried out:

Design Check :





Floorplanning

A suitable die and core area were defined to accommodate the hybrid FPU datapath and control logic. I/O pins were placed to minimize routing congestion for critical datapaths such as the multiplier and adder buses.

Power Planning

Power rings and standard-cell power rails (VDD and VSS) were created to ensure stable power delivery across the design. Core-level power grids and horizontal/vertical metal straps were inserted to guarantee low IR-drop and robust current delivery, especially for the arithmetic units which have high switching activity.

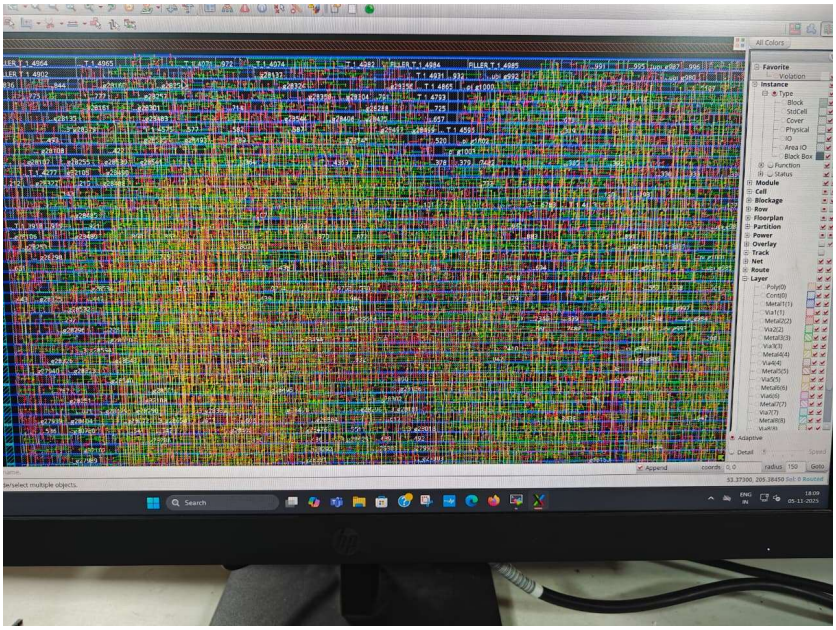
Standard Cell Placement

The synthesized standard cells were placed using timing-driven placement. Critical blocks such as the FP32 multiplier units were placed close to reduce interconnect delay. Filler cells were inserted to maintain well continuity and to satisfy DRC requirements.

Clock Tree Synthesis (CTS)

Clock tree synthesis was performed to distribute the clock signal with minimal skew and balanced insertion delay. Clock buffers and inverters were inserted to meet setup and hold timing requirements across all registers in the FPU.

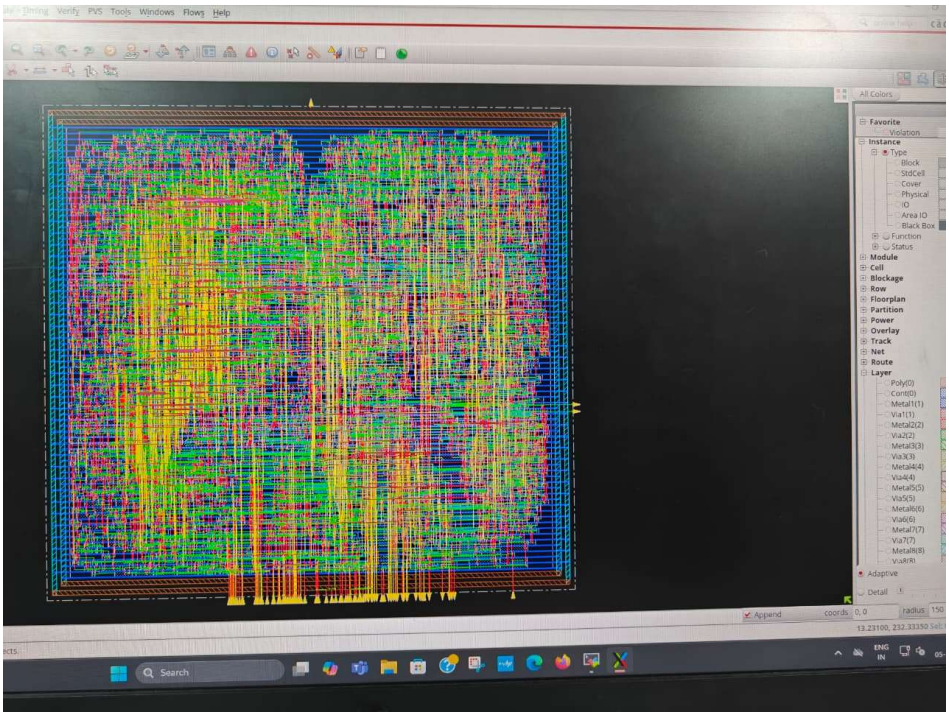
Post Placement and CTS :



Routing

Global and detailed routing were executed to connect all standard cells and macro pins while meeting design rule constraints. Special attention was given to high-fanout control signals and wide datapaths in the arithmetic units.

Post-Routing :



Timing and DRC Signoff

Post-route static timing analysis confirmed that the design met timing constraints at the target clock frequency. Design Rule Check (DRC) and connectivity checks were performed to ensure the layout was manufacturable.

GDS-II Generation

Finally, the completed physical layout was exported as a **GDS-II file**, which represents the final mask-level description of the Predictive Hybrid FPU. This file is suitable for fabrication and serves as proof that the design is physically realizable.

---

6. Results

The hybrid FPU achieved:

- Lower power consumption compared to always-exact FP32
- Reduced area overhead relative to multiple independent FPUs
- Correct IEEE-754 behaviour when operating in exact mode
- Safe and stable approximation when the approximate mode is selected

---

7. Conclusion

This project demonstrates that **predictive, operand-aware mixed-precision arithmetic** can be implemented in hardware without sacrificing correctness or ASIC feasibility. By enabling exact and approximate FP32 computation within a single FPU, significant energy savings can be achieved for AI and signal-processing workloads.