Error Analysis

Based on the VerilogEval benchmark (comprising 156 programming problems), we collected 83, 89, and 65 erroneous programming samples from the Qwen2.5-Coder-32B-Instruct, GPT-3.5-turbo, and GPT-4-turbo (various parameter sizes, including both code-specific and general-purpose versions) large language models, respectively. Using a total of 237 error samples (with temperature = 0.0 to ensure output stability), we studied the causes and types of errors in LLM-generated RTL.

The following analysis examines the error types corresponding to different erroneous code samples (identified by problem number).

(Note: A single erroneous code sample may contain multiple different types of errors, and we simultaneously categorize it into different classes accordingly.)

1. Qwen-Coder-32B-Instruction Model

Type I: Insufficient knowledge of specialized RTL programming

- **Wire in always block**：068，079，088，100，119，138，140，144，156
- **Numerical Processing Logic Error**：009，141
- **Vector Bit Selection Error**： 148
- **Inversion of Vector Slice Selection**： 092

Type II: Misinterpretation of design specification

1. **Insufficient Understanding of Circuit Concepts**
   - Time Related Concepts: 035, 037, 038, 040, 041, 046, 060, 067, 068, 073, 088, 096, 107, 109, 111, 120, 136, 137, 138, 139, 148 (Synchronous/Asynchronous logic), 066 (Recording pre-emptive signals), 080 (Erroneous sample output lags golden output by one cycle; Difference between `always` blocks and `assign` statements)
   - Special Design: 070 (SOP and POS), 078 (DualEdge), 082 (LFSR32), 086 (LFSR5)
   - Numerical and Vector Processing： 063, 084 (Understanding `most-significant-bit`), 071, 146 (`least significant`), 112 (Case condition vector setup), 115 (Sign bit setup)

- State Machine Design for Complex Scenarios： 089, 096, 128, 129, 133, 137, 139, 140, 142 143, 146, 149, 151, 152, 154, 155, 156 (Missing states or incorrect transition logic)
  Erroneous designs caused by lack of knowledge: 45

2. **Ambiguous Design Descriptions**
   - **Unclear Overall Module Functionality**： 045，062，074，104
   - **Ambiguous Module Input/Output Description**： 079，143，150
   - **Missing Module Initialization**： 028，034，053
     Erroneous designs caused by ambiguous descriptions: 10
3. **Misinterpretation of Multimodal Data**
   - **Tables:** 069, 121, 124, 134
   - **K-Maps (Karnaugh Maps):** 050, 057, 093, 113, 116, 122, 125
   - **Waveforms:** 098, 101, 102, 103, 117, 145, 147, 154
   - **State Transition Diagrams:** 091, 099, 109, 135, 136, 138, 143, 150
     Erroneous designs caused by multi-format data: 27
4. **Missing Details of Long Descriptions**
   093，139，140，142，150，151，152，153，155，156
   Erroneous designs caused by missing details of long descriptions: 10

Summary: 6 samples contain only Type I errors, 70 samples contain Type II errors, and 7 samples contain both Type I and Type II errors.

2. GPT-3.5 Turbo Model

Type I: Insufficient knowledge of specialized RTL programming

- **Wire in always block**: 079, 100, 107, 111, 139, 140, 143, 146
- **Numeric Processing Logic Errors:** 071
- **Vector Bit Selection Errors:** 068, 089, 109
- **Inversion of Vector Slice Selection**： 105
- **Incomplete Code:** 016 (missing submodule), 018, 021 (missing bit-selection logic), 048 (incomplete functionality)
- **Variable Redefinition:** 041, 047, 073, 121
- **Use of Undefined Variable:** 093
- **Incorrect Generate Statement Usage:** 092
- **Mixed Blocking/Non-Blocking Assignments:** 108
- **Incorrect Usage of `integer` Variables:** 144

Type II: Misinterpretation of design specification

1. **Insufficient Understanding of Circuit Concepts**
   - **Time Related Concepts:** 037, 060, 066, 073, 088, 111, 120, 121, 128, 133, 136, 137, 138, 139, 140, 146, 148, 149, 151, 154, 156 (sync/async logic)
     054, 066, 139 (prev_in recording), 049 (posedge/negedge logic)
   - **Special Design**: 070 (SOP and POS), 078 (DualEdge), 082 (LFSR32), 086 (LFSR5)
   - **Numerical and Vector Processing：** 033 (signed addition overflow), 071, 134 (case condition setup), 105 (right-shift error), 112 (LSB interpretation), 115 (sign bit handling), 141 (timing processing), 146 (numeric recording)
   - **State Machine Design for Complex Scenarios：** 089，091，095，096，127，128，129，133，137，139，140，142，146，148，149，151，152，154，155，156 (Missing states or incorrect transition logic)
     Erroneous designs caused by lack of knowledge: 43
2. **Ambiguous Design Descriptions**
   - **Unclear Overall Module Functionality：** 045，062，074，104
   - **Ambiguous Module Input/Output Description：** 028, 079，143，150
   - **Missing Module Initialization：** 028，034，053
     _Erroneous designs caused by ambiguous descriptions: 10
3. **Misinterpretation of Multimodal Data**
   - **Tables:** 069，124，134
   - **K-Maps (Karnaugh Maps):** 050，057，093，113，116，122，125
   - **Waveforms:** 098，101，102，103，117，126，130，131，145，147，150，154
   - **State Transition Diagrams:** 088，091，099，135，138，143，150
     Erroneous designs caused by multi-format data: 29
4. **Missing Details of Long Descriptions**
   092，093，094，142，150，151，152，153，154，155，156
   Erroneous designs caused by missing details of long descriptions: 11

Summary: 12 samples contain only Type I errors, 64 samples contain Type II errors, and 13 samples contain both Type I and Type II errors.

3. GPT-4 Turbo Model

Type I: Insufficient knowledge of specialized RTL programming

- **Incomplete Code:** 016 (missing submodule), 037, 048 (missing end)
- **Wire in always block**: 068
- **Use of Undefined Variable:** 060
- **Incorrect Generate Statement Usage:** 092
- **Incorrect Usage of `integer` Variables:** 144, 153

Type II: Misinterpretation of design specification

1. **Insufficient Understanding of Circuit Concepts**
   - **Time Related Concepts:** 120, 133 (sync/async logic), 066 (prev_in recording)
   - **Special Design**: 070 (SOP and POS), 078 (DualEdge), 082 (LFSR32), 086 (LFSR5)
   - **Numerical and Vector Processing：** 033 (addition overflow check), 112 (case condition), 141 (timing handling)
   - **State Machine Design for Complex Scenarios：** 089，095，128，133，137，139，140，142，146，149，151，154，155，156 (Missing states or incorrect transition logic)

   Erroneous designs caused by lack of knowledge: 23
2. **Ambiguous Design Descriptions**
   - **Unclear Overall Module Functionality：** 045，062，074
   - **Ambiguous Module Input/Output Description：** 079，150 (state signal)
   - **Missing Module Initialization：** 028，034，053

   _Erroneous designs caused by ambiguous descriptions: 8
3. **Misinterpretation of Multimodal Data**
   - **Tables:** 134
   - **K-Maps (Karnaugh Maps):** 050，057，093，113，116，122，125
   - **Waveforms:** 098，101，102，103，117，131，145，147，154
   - **State Transition Diagrams:** 088，091，099，135，143，150

   Erroneous designs caused by multi-format data: 23
4. **Missing Details of Long Descriptions**

   093，094，124，142，150，151，153，154，155，156

   Erroneous designs caused by missing details of long descriptions: 10

Summary: 7 samples contain only Type I errors, 57 samples contain Type II errors, and 1 samples contain both Type I and Type II errors.

We hope this classification effort will provide insights—including for training and usage—to support LLMs in generating more robust and higher-quality RTL code in the future.

Note: As the classification is based on manual evaluation and limited to Qwen and GPT model families, subjective judgment may introduce inaccuracies or omissions. We welcome your feedback for further refinement.