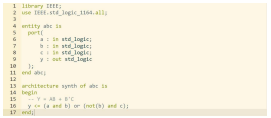
Analyzing Code Trajectories to Understand VHDL Learning

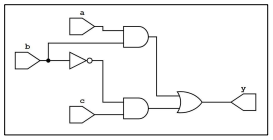
Tufts University School of Engineering
Brendan Amorin, Shuchin Aeron, Steven Bell, David Hammer, Mark Hempstead, Eric Miller

What is VHDL and why is it Difficult?

- VHDL is a hardware description language that can model the behavior of digital systems.

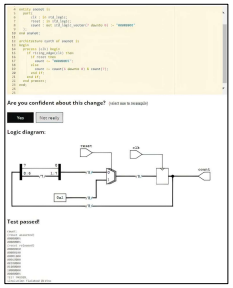


Simple VHDL Example

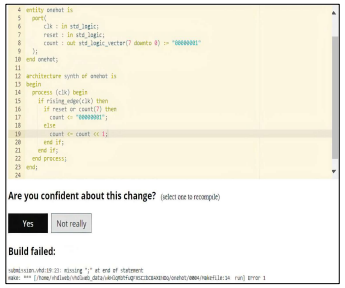


Corresponding Hardware

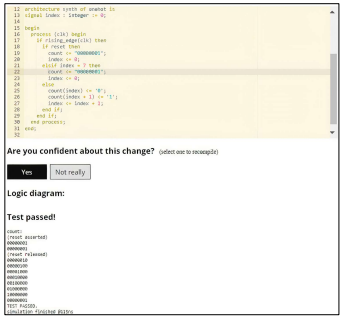
- It is concurrent: statements generally occur simultaneously, as opposed to sequentially.
- Requires a basic understanding of hardware being designed.
- Typically, novice VHDL programmers first learned a sequential programming language.
- Results in a variety of methods, some erroneous, for completing problems.



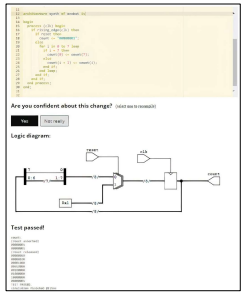
Concatenate-bits Method



C-code Method

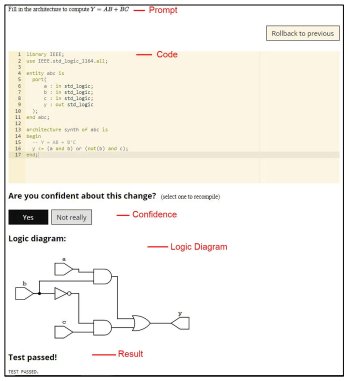


Index-variable Method



For-loop Method

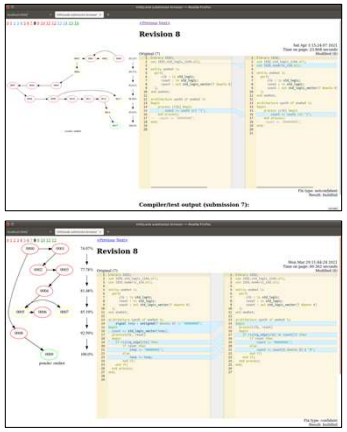
What Data We Have



Also, have date and time when code was submitted.

Drawing Solution Trajectories

- Goal of connecting a student's submissions based on their different approaches to solving a problem.
- Use text-based comparison of code.
- Find first passing submission and compare all previous submissions to it to indicate progress.
- Submissions equally like the final submission have the same rank, and submissions that are identical share the same node.
- Edge between current submission and most similar previous submission.

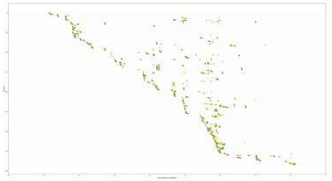


Sample Student Solution Trajectories along with Diffviewer Tool

Solution Trajectory Metrics

- Use Dijkstra's algorithm to determine the shortest path from initial submission to final submission, which suggests the percentage of submissions that contributed to the student's final solution.
- Timestamps can reveal trends about when students typically submit their assignments.

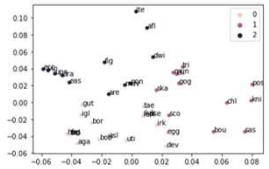
Percentages for a Selection of Problems



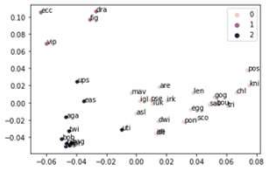
Timeline of when Students Submitted Their Code Relative to the Deadline

Grouping Students by Code:

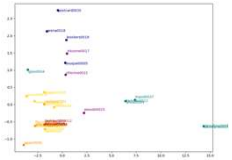
- Goal of grouping students based on similarities in coding process and solution.
- Affinity matrix calculated with two methods: using Wasserstein metric from optimal transport between different students' code submissions and computing similarity between students' solutions to a problem.
- Spectral Embedding performed on affinity matrix and then k-means clustering used.
- Alternate approach: counting occurrences of keywords in solutions and use these totals as features in k-means clustering.



Optimal-Transport Method



Similarity Method



Bag-of-Keywords Method