ESE532:
System-on-a-Chip Architecture

Day 20: November 7, 2018
Verification 1

Today

- Motivation
- Challenge and Coverage
- Golden Model / Reference Specification
- Automation and Regression

Message

- If you don't get it, it doesn't work.

- Verification is important and challenging
- Demands careful thought
  - Tractable and adequate coverage
- Value to a simple functional reference
- **Must** be automated and rerun with changes
  - Often throughout lifecycle of design

Goal

- Assure design works correctly
  - Not fail and lose consumer confidence.
  - Not kill anyone
    - Ethical issue
    - …or lose them money, privacy, service availability….
  - Not lose points on your grade ☺

Challenge

- Designs are complex
  - Lots of ways things can go wrong
  - Lots of subtle ways things can go wrong
  - Lots of tricky interactions
- Designs are often poorly specified
  - Complex to completely specify

Verification

- Often dominant cost in product
  - Requires most manpower (cost)
  - Takes up most of schedule

Correct?

- How do we define correctness for a design?
- How do we know the design is correct?
- How do we know the design remains correct when?
  – Add a some feature
  – Perform an optimization
  – Fix a bug

Life Cycle

- Design
  – specify what means to be correct
- Development
  – Implement and refine
  – Fix bugs
  – Optimize
- Operation and Maintenance
  – Discover bugs, new uses and interaction
  – Fix and provide updates
- Upgrade/revision

Testing and Coverage

Strawman Testing

Validate the design by testing it:
- Create a set of test inputs
- Apply test inputs
- Collect response outputs
- Check if outputs match expectations

Strawman: Inputs and Outputs

Validate the design by testing it:
- Create a set of test inputs
  – How do we generate an adequate set of inputs? (know if a set is adequate?)
- Apply test inputs
- Collect response outputs
- Check if outputs match expectations
  – How do we know if outputs are correct?

Try 1: Inputs and Outputs

- Create a set of test inputs
  – How do we generate an adequate set of inputs? (know if a set is adequate?)
    • All possible inputs
- Check if outputs match expectations
  – How do we know if outputs are correct?
    • Manually identify correct output

How many input cases?

Combinational:
- 10-input AND gate?
- Any N-input combinational function?
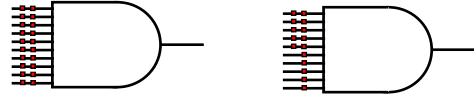
Add Pipelining

- The output doesn't correspond to the input on a single cycle
- Need to think about inputs sequences to output sequences
- How many input cases?

Add Feedback State

- When have state
  – Different inputs can produce different outputs
- Behavior depends on state
- Need to reason about all states the design can be in

How many input cases?

- Process 1000 Byte packet
  – No state kept between packets
- Process 1000 Byte packets
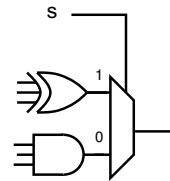 – Keep 32b of state between packets

Observation

- Cannot afford
  – Exhaustively generate input cases
  – Manual write output expectations

- Will need to be smarter about test case selection

Structural Simplifications

- How many cases if treat as 7-input function?
- How many useful cases
  – If hold s at 0?
  – If hold s at 1?
  – Together total cases?

3

Useful Test Cases

```
int fun(int s,a,b,c,d) {
  if (s>20)
    if (s>100)
      return(a+b); else return(b+c);
  else
    if (s<0)
      return(c+d); else return(a+d);
}
```

What values of s will be interesting?

When s=10, what values of a, b, c, d interesting?

Finite State Machine

- What input cases should we try to exercise for an FSM? (goal for test cases)

```
int state;
while (true) {
  switch (state) {
    case (ST1): out=1; state=ST2; break;
    case (ST2):  if (in>0) {out=2; state=ST3;}
                 else {out=0; state=ST2;} break;
    case (ST3): ...
```

Coverage

- Do our tests execute every line of code?
  - What percentage of the code is exercised?
- Gate-level designs
  - Can we toggle every gate output?
- Necessary but not sufficient
  - Not exercised or not toggled, definitely not testing some functionality
    - Remember: If you don't test it, it doesn't work.
- Measurable

So far…

- Identifying test stimulus important and tricky
  - Cannot generally afford exhaustive
  - Need understand/exploit structure
- Coverage metrics a start
  - Not complete answer

Reference Specification (Golden Model)

Problem

- Manually writing down results for all input cases
  - Tedious
  - Error prone
  - …simple not viable for large number cases need to cover
    - Definitely not viable exhaustive
    - …and still not viable when select intelligently

4

Specification Model

- Ideally, have a function that can
  - compute the correct output
  - for any input sequence
- ``Gold Standard" – an oracle
  - Whatever the function says is truth
- Could be another program
  - Written in a different language? Same language?

Testing with Reference Specification

Validate the design by testing it:
- Create a set of test inputs
- Apply test inputs
  - To implementation under test
  - To reference specification
- Collect response outputs
- Check if outputs match

Test against Specification

- Relieved ourselves of writing outputs
- Still have to select input cases
  - Can freely use larger set since not responsible for manually generating output match

Random Inputs

- Can use random inputs
  - Since can generate expected output for any case
- Use coverage metric to see how well random inputs are exercising the code
- Can be particularly good to identify interactions and corner cases didn't think of manually
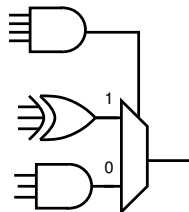- Still unlikely to generate very obscure cases

Random Inputs

- Expected number of tests to exercise both cases?

- What would we like to happen to reduce?
  - What would we select manually?

Random Testing

- Completely random may be just as bad as exhaustive
  - Expected time to exercise interesting piece of code
  - Expected time to produce a legal input
    - E.g. – random packets will almost always have erroneous checksums
  - E.g. random bytes won't generate duplicate chunks, or much opportunity for LZW compression

Biased Random

- Non-uniform random generation of inputs
  - Compute checksums correctly most of the time
    - Control rate and distribution of checksum errors
- Randomize properties of input, E.g.
  - Lengths of repeated sequences
  - Distance between repeated sequences
  - Edit sequence applied to differentiate files

Testing with Reference Specification

Validate the design by testing it:

- Create a set of test inputs
- Apply test inputs
  - To implementation under test
  - To reference specification
- Collect response outputs
- Check if outputs match

Specification

- Where would we get a reference specification?
  - and why should we trust it?

  - Isn't this just another design that can be *equally* buggy?

Standard

- Many standards includes a reference implementation.

Existing Product

- Many times there's an existing product or open-source implementation…

Develop Specification

- Maybe develop a simple, functional implementation as part of early design

Specification Correct?

- How would we know the specification is correct? -- why should we trust it?
  - Simpler/smaller
    - Less opportunity for bugs
    - Written for function/clarity not performance
  - Different
    - Ok as long as reference and implementation don't have same bugs
      - Debug and test them against each other

Simpler Functional

- Examples of functional specification being simpler than implementation?

Simpler Functional

- Sequential vs. parallel
- Unpipelined vs. pipelined
- Simple algorithm
  - Brute force?
- No data movement optimizations
- Use robust, mature (well-tested) building blocks

Common Bugs

- Combinational (for simplicity)
- 10 input function
- Assume two specifications have 1% error rate (1% of input cases wrong)
- Assume independent
  - (key assumption – weaker to extent wrong)
- Probability of both giving same wrong result?
  - For a particular input case?
  - Across all input cases?

Testing with Reference Specification

Validate the design by testing it:

- Create a set of test inputs
- Apply test inputs
  - To implementation under test
  - To reference specification
- Collect response outputs
- Check if outputs match

Coverage

- Of specification or implementation?
  - Almost certainly both
- Specification may have a case split that implementation doesn't have
  - E.g. handle exceptional case
- Implementation typically have many more cases to handle in general

Automation and Regression

Automated

- Testing suite **must** be automated
  - Single script or make build to run
  - Just start the script
  - Runs through all testing and comparison without manual interaction
  - Including scoring and reporting a single pass/fail result
    - Maybe a count of failing cases

Regression Test

- Regression Test -- Suite of tests to run and validate functionality

Regression Tests

- One big test or many small tests?
- Benefit of many small tests?
- Benefit of big test(s)?

Automation Mandatory

- Will run regression suite repeatedly during Life Cycle
  - Every change
  - As optimize
  - Every bug fix

Life Cycle

- Design
  - specify what means to be correct
- Development
  - Implement and refine
  - Fix bugs
  - optimize
- Operation and Maintenance
  - Discover bugs, new uses and interaction
  - Fix and provide updates
- Upgrade/revision

Automation Value

- Engineer time is bottleneck
  - Expensive, limited resource
  - Esp. the engineer(s) that understand what the design should do
- Cannot spend that time evaluating/running tests
- Reserve it for debug, design, creating tests
- Capture knowledge in tools and tests

When find a bug

- If regression suite didn't originally find it
  - Add a test (expand regression suite) so will have a test to cover
- Make sure won't miss it again
- Test suite monotonically improving

When add a feature

- Add a test to validate that feature
  - And interaction with existing functionality

- Maybe add the test first…
  - See test identifies lack of feature before add functionality
  - …then see (correctly added) feature satisfies test

Continuous Integration

- When commit code to shared repo (git, svn)
  - Build and run regression suite
  - Perhaps before allow commit
  - Guarantee not break good version
    - Or, at least, know how function/broken the current version is
- Alternately, nightly regression
 - Automation to check out, build, run tests

Regression Test Size

- Want to be comprehensive
  - More tests better….
- Want to run in tractable time
  - Few minutes once make change or when checkin
  - Cannot run for weeks or months
  - Might want to at least run overnight
- Sometimes forced to subset
  - Small, focused subset for immediate test
  - Comprehensive test for full validation

Unit Tests

- Regression for individual components
- Good to validate independently
- Lower complexity
  - Fewer tests
  - Complete quickly
- Make sure component(s) working before run top-level design tests
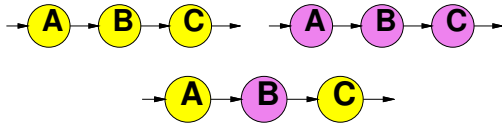 - One strategy for long top-level regression
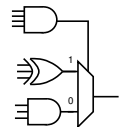
Functional Scaffolding

- If functional decomposed into components like implementation
- Replace individual components with implementation
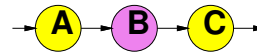 - Use reference/functional spec for rest

55

Functional Scaffolding

- If functional decomposed into components like implementation
- Replace individual components with implementation
  - Use reference/functional spec for rest
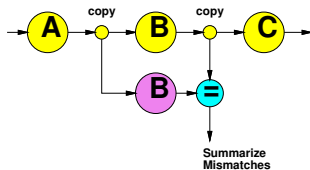- Independent test of integration for that module

56

Functional Scaffolding

- If functional decomposed into components like implementation
- Run reference component and implementation together and check outputs

copy   copy

Summarize
Mismatches

57

Decompose Specification

- Should specification decompose like implementation?
  - ultimate golden reference
    - Only if that decomposition is simplest
- But, worth refining
  - Golden reference simplest
  - Intermediate functional decomposed
    - Validate it versus golden
    - Still simpler than final implementation
    - Then use with implementation

58

Big Ideas

- Testing
  - Designs are complicated, need extensive validation – If you don't test it, it doesn't work.
  - Exhaustive testing not tractable
  - Demands care
  - Coverage one tool for helping identify
- Reference specification as "gold" standard
  - Simple, functional
- Must automate regression
  - Use regularly throughout life cycle

59

Admin

- P2 due Friday
- P3 ... Later today?
  - Not out, yet … but coming soon

60

10