Quality Assurance of Test Specifications for Reactive Systems

Benjamin Zeiß

Outline

- Motivation and Foundations
- A Quality Model for Test Specifications
- Model-Based Analysis of Test Specifications
- Case Study
- Contributions and Outlook

Outline

- **Motivation and Foundations**
- A Quality Model for Test Specifications
- Model-Based Analysis of Test Specifications
- Case Study
- Contributions and Outlook

Motivation

- Test specification sizes grow:
    - ETSI SIP test suite: ~62.000 LOC
    - ETSI LTE test suite in development: more than 200.000 LOC

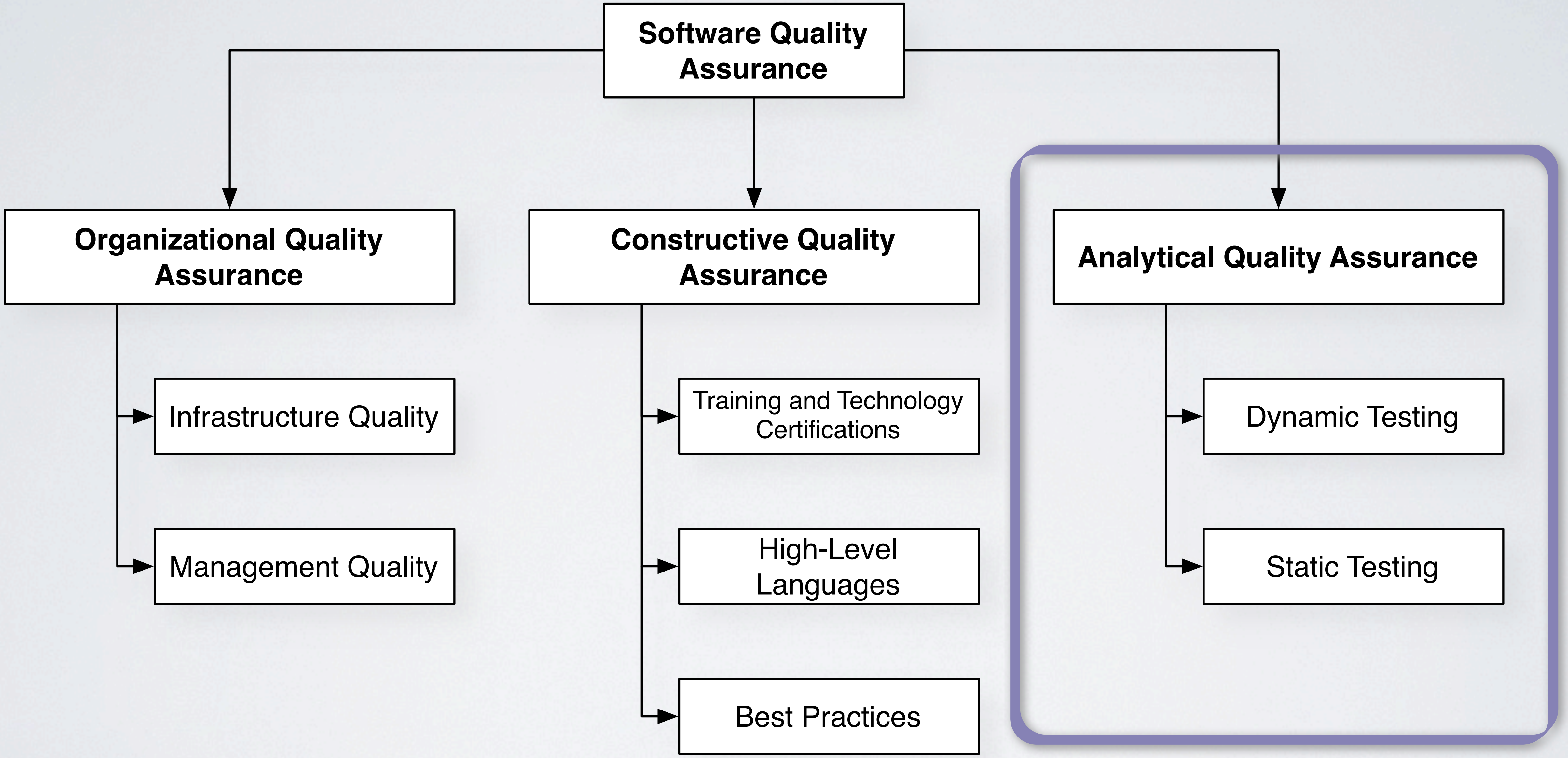
- More complex and error-prone.

  →Implications on the System Under Test (SUT)!

- Quality Assurance (QA) is necessary!

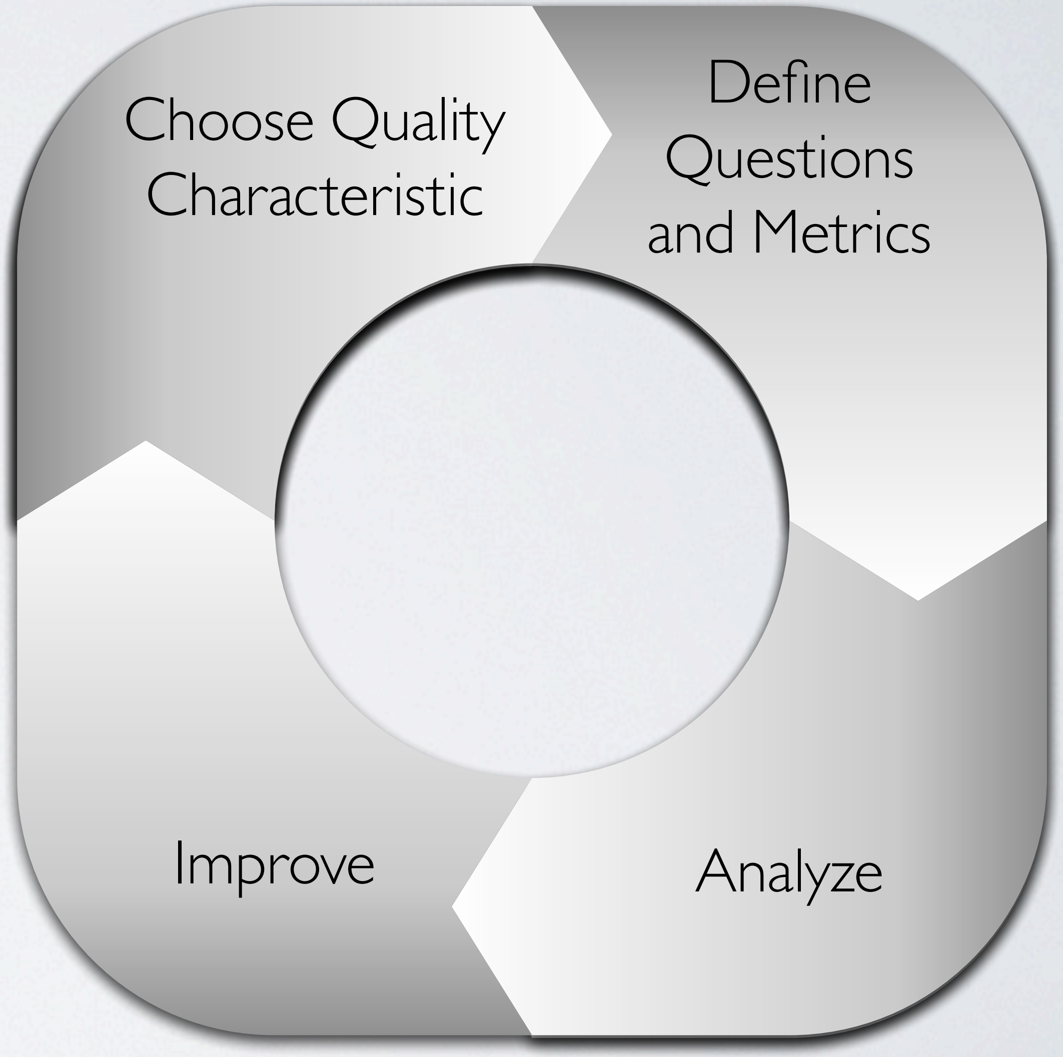
- **How can we apply QA to test specification development?**

Software Quality


Software Quality Assurance
Organizational Quality Assurance
Constructive Quality Assurance
Analytical Quality Assurance
Infrastructure Quality
Management Quality
Training and Technology Certifications
High-Level Languages
Best Practices
Dynamic Testing
Static Testing

Roadmap

- Define quality for test specifications.
- Develop quality assessment methodology for test specifications.
- Develop techniques for the analysis of test specifications.



Test Specification Properties

- Evaluation of specific aspects of a System Under Test (SUT).
- Conclusion with a test verdict.
- Repeatability (failure reproduction).
- Execution often unsupervised (test automation).
- No test specifications for test specifications!

Related Work (1/2)

- Quality of software tests:
  - G. Meszaros: **xUnit Test Patterns: Refactoring Test Code**. Addison-Wesley, 2007.
  - D. Vega and I. Schieferdecker. **Towards Quality of TTCN-3 Tests**. In Proceedings of SAM'06: Fifth Workshop on System Analysis and Modelling, volume 4320 of Lecture Notes in Computer Science (LNCS). Springer, 2006.
  - D. Vega, G. Din, S. Taranu, and I. Schieferdecker. **Application of Clustering Methods for Analysing of TTCN-3 Test Data Quality**. In Proceedings of the 2008 The Third International Conference on Software Engineering Advances (ICSEA 2008). IEEE, 2008.
  - D. Vega, I. Schieferdecker, and G. Din. **Test Data Variance as a Test Quality Measure: Exemplified for TTCN-3**. In Testing of Software and Communicating Systems, volume 4581 of Lecture Notes in Computer Science. Springer, 2007.

Related Work (2/2)

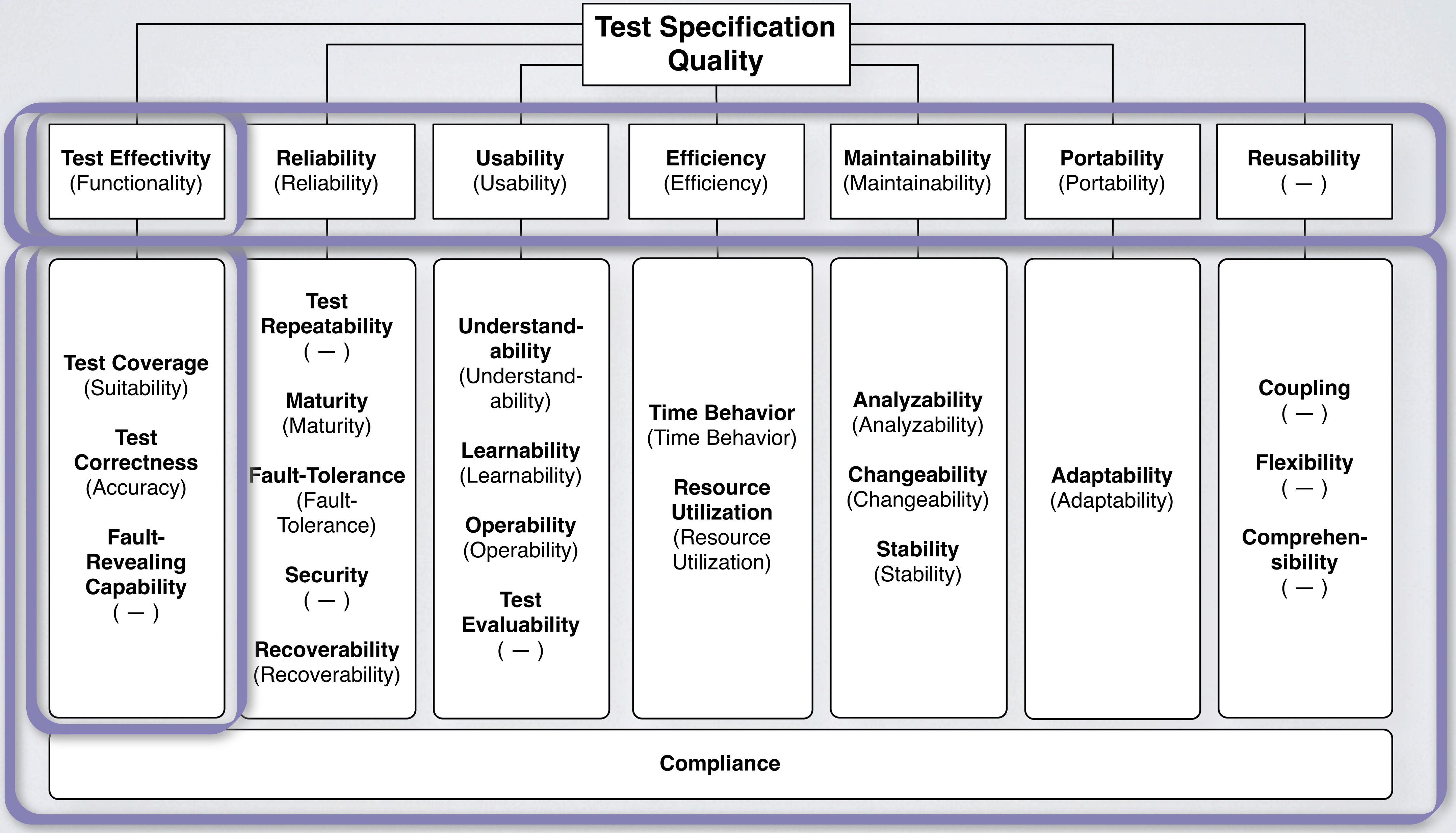
- Model-based analysis, test and system validation:
  - S. Boroday, A. Petrenko, and A. Ulrich. **Test Suite Consistency Verification**. In Proceedings of the 6th IEEE East-West Design & Test Symposium (EWDTS 2008), Ukraine, 2008.
  - H. Hallal, S. Boroday, A. Petrenko, and A. Ulrich. **A Formal Approach to Property Testing in Causally Consistent Distributed Traces**. Formal Aspects of Computing,18(1), 2006.
  - C. Flanagan and S. N. Freund. **Dynamic Architecture Extraction**. In Formal Approaches to Software Testing and Runtime Verification, volume 4262 of Lecture Notes in Computer Science (LNCS). Springer, 2006.

Outline

- Motivation and Foundations
- **A Quality Model for Test Specifications**
- Model-Based Analysis of Test Specifications
- Case Study
- Contributions and Outlook

Quality Model for Test Specifications

(ISO 9126 Adaptation)



Test Specification Quality

Test Effectivity (Functionality)
- **Test Coverage** (Suitability)
- **Test Correctness** (Accuracy)
- **Fault-Revealing Capability** (—)

Reliability (Reliability)
- **Test Repeatability** (—)
- **Maturity** (Maturity)
- **Fault-Tolerance** (Fault-Tolerance)
- **Security** (—)
- **Recoverability** (Recoverability)

Usability (Usability)
- **Understandability** (Understandability)
- **Learnability** (Learnability)
- **Operability** (Operability)
- **Test Evaluability** (—)

Efficiency (Efficiency)
- **Time Behavior** (Time Behavior)
- **Resource Utilization** (Resource Utilization)

Maintainability (Maintainability)
- **Analyzability** (Analyzability)
- **Changeability** (Changeability)
- **Stability** (Stability)

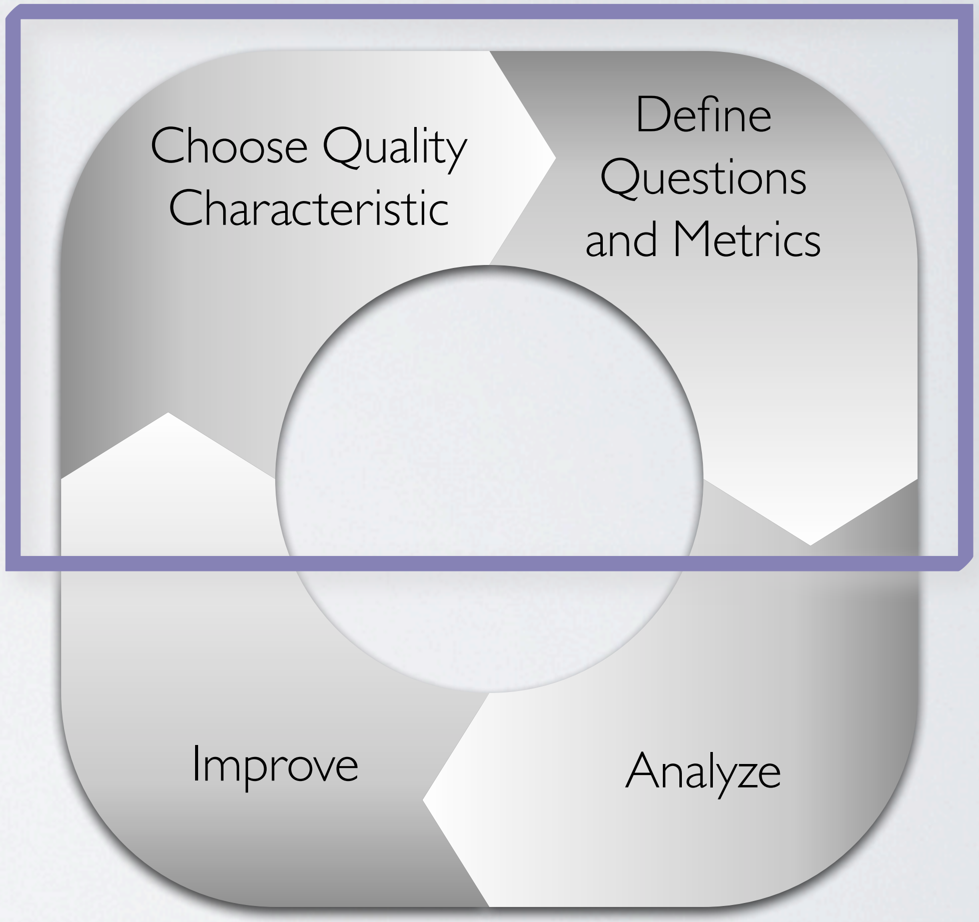
Portability (Portability)
- **Adaptability** (Adaptability)

Reusability (—)
- **Coupling** (—)
- **Flexibility** (—)
- **Comprehensibility** (—)

Compliance

Quality Model Instantiation

- Goal, Question, Metric (GQM) approach from Basili and Weiss (1984)
    - State the goal to be achieved.
    - Define questions that break the goal into its major components.
    - Select metrics that answer the questions.



Quality Model Instantiation: Test Correctness

- **Questions**:
  - Does the test case deliver consistent test verdicts?
- **Metrics**:
  - Test verdict completeness:

$$\text{tvc} := \begin{cases} 0 & \text{if no. of paths in test case setting no test verdict} > 1 \\ 1 & \text{otherwise} \end{cases}$$

  - Early test verdict:

$$\text{etv} := \begin{cases} 0 & \text{if no. of paths in test case setting a test verdict} \\ & \text{before any communicating behavior} > 1 \\ 1 & \text{otherwise} \end{cases}$$

Quality Model Instantiation Issues

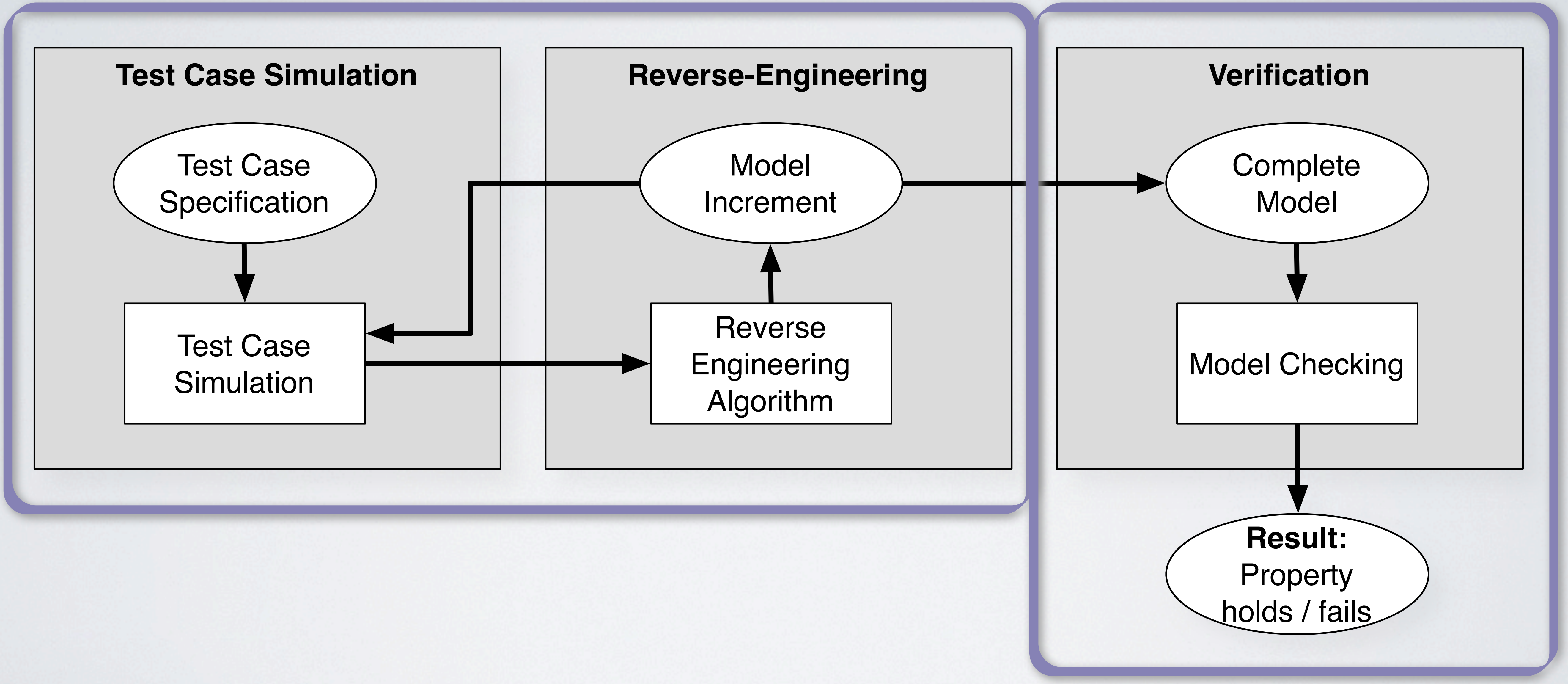
- State of the art for test specification QA is static analysis.

- Both metrics cannot be detected using static analysis:
    - Analysis of all paths in the test behavior.

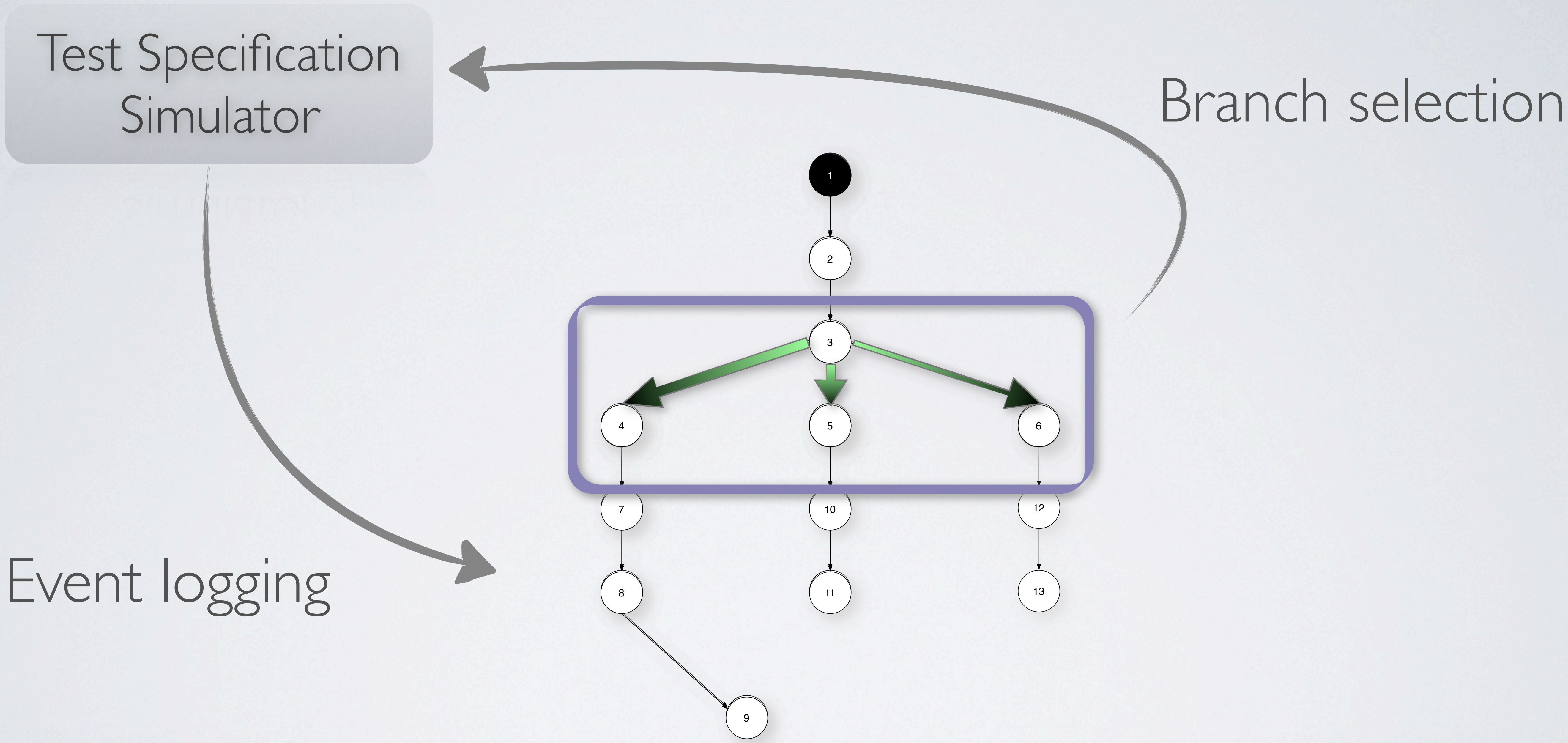
Outline

- Motivation and Foundations
- A Quality Model for Test Specifications
- **Model-Based Analysis of Test Specifications**
- Case Study
- Contributions and Outlook

Dynamic Analysis Methodology



Model Reverse-Engineering



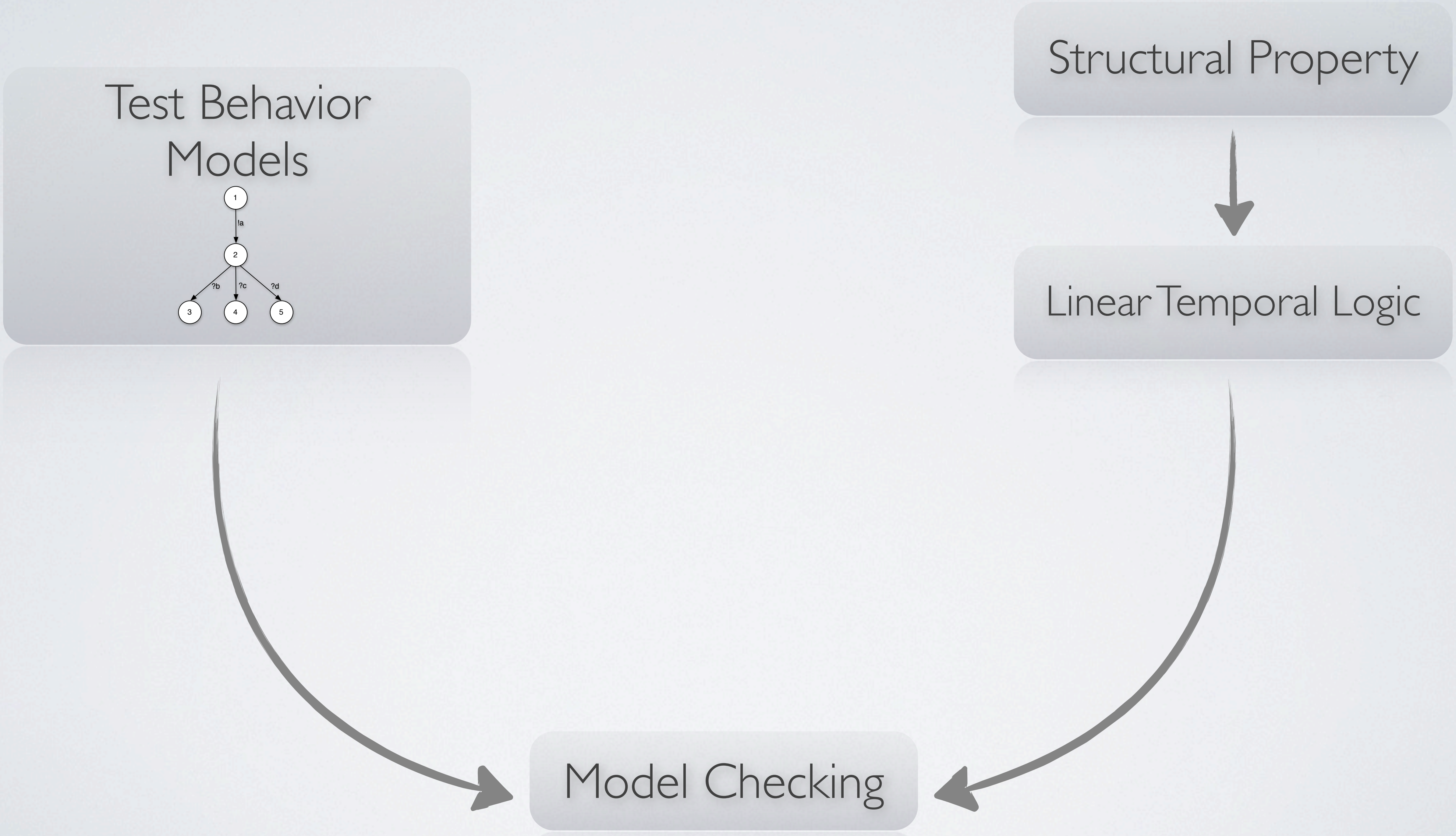
Test Specification Simulator

Branch selection

Event logging

Abstract partial test behavior models

Test Specification Model Verification



Test Behavior Models

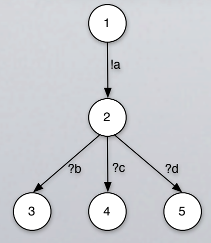
Structural Property

Linear Temporal Logic

Model Checking

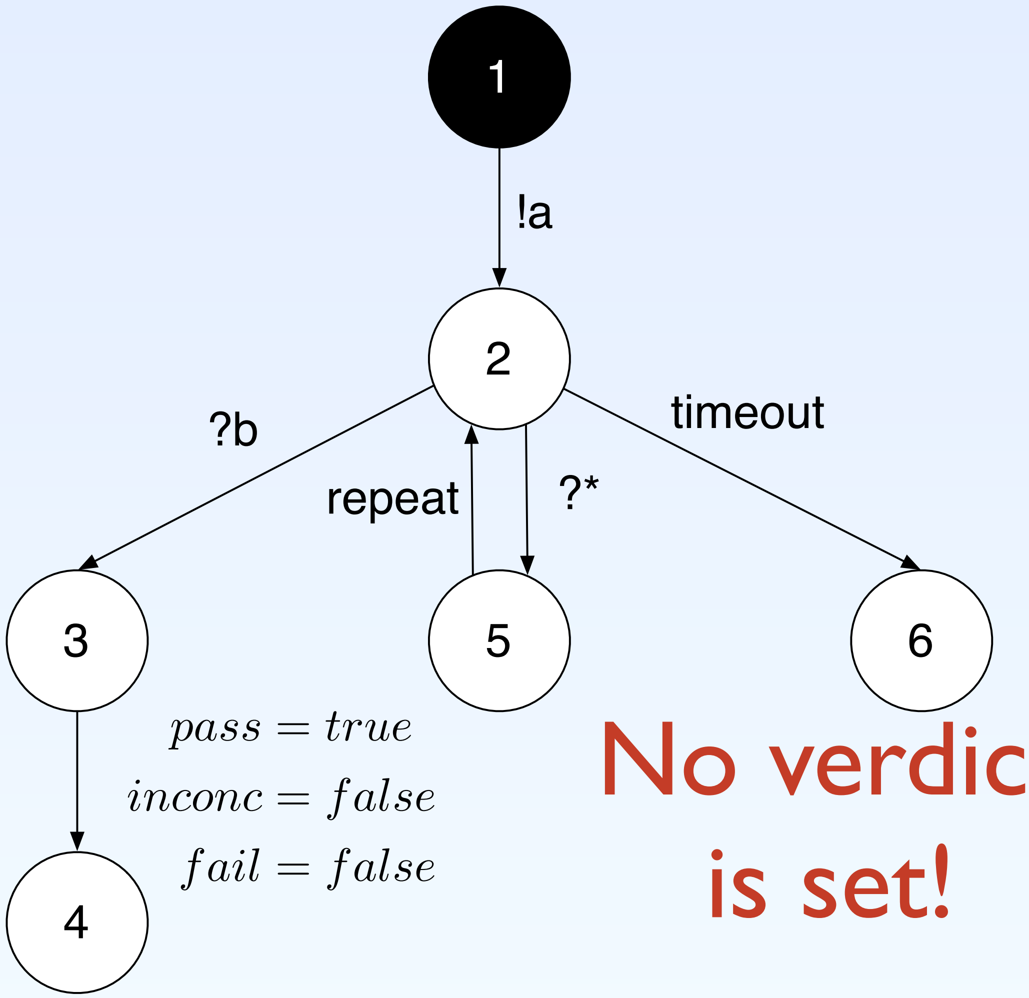
Test Specification Model Verification

Test Behavior Models



Structural Property

$V = \{pass = false, inconc = false, fail = false\}$



$pass = true$
$inconc = false$
$fail = false$

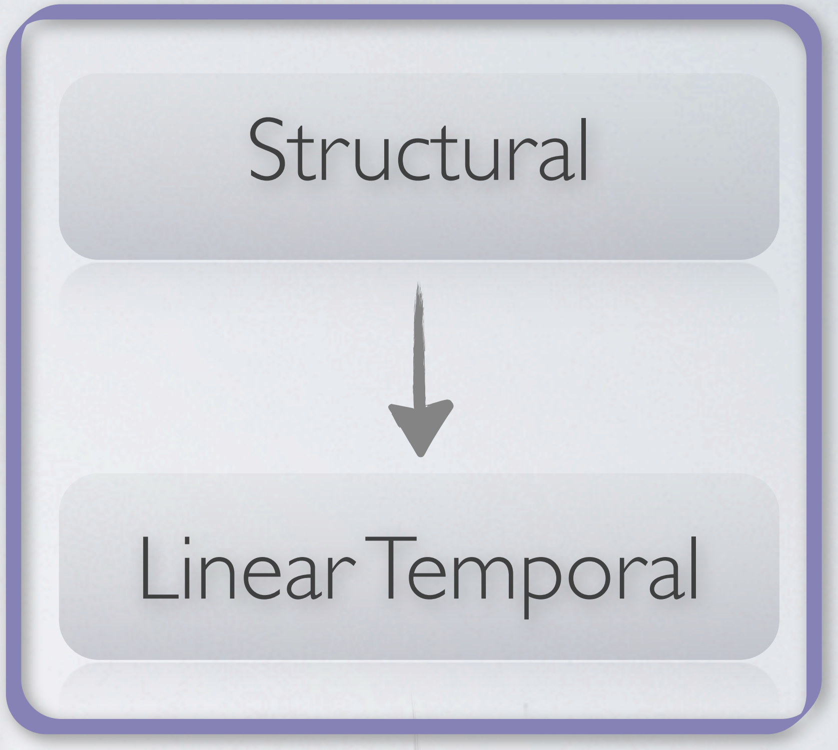
No verdict is set!

Model Checking

Test Specification Model Verification

- Possible LTL formula:
  - For each path, no verdict is set until a verdict becomes either pass, inconclusive, or fail.

$$\phi := (\neg pass \wedge \neg inconc \wedge \neg fail)\,\mathcal{U}\,(pass \vee inconc \vee fail)$$

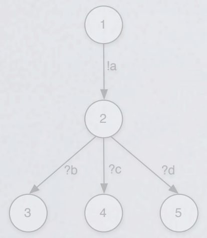

Structural

Linear Temporal

Model Checking

Test Specification Model Verification

Test Component Models



Structural

- Result:
  - The structural property does **not** hold in **all possible paths**.
  - The failing traces.

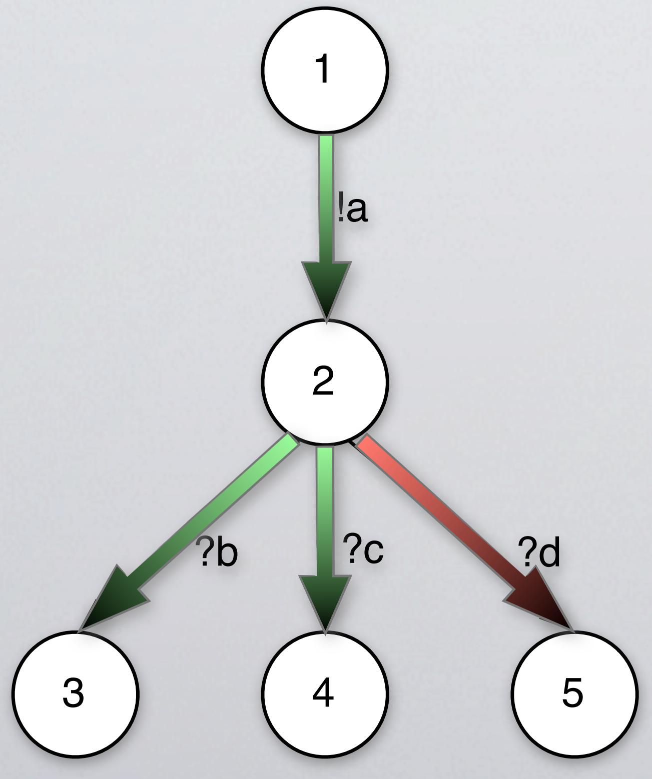
Model

Further Test Specification Analysis

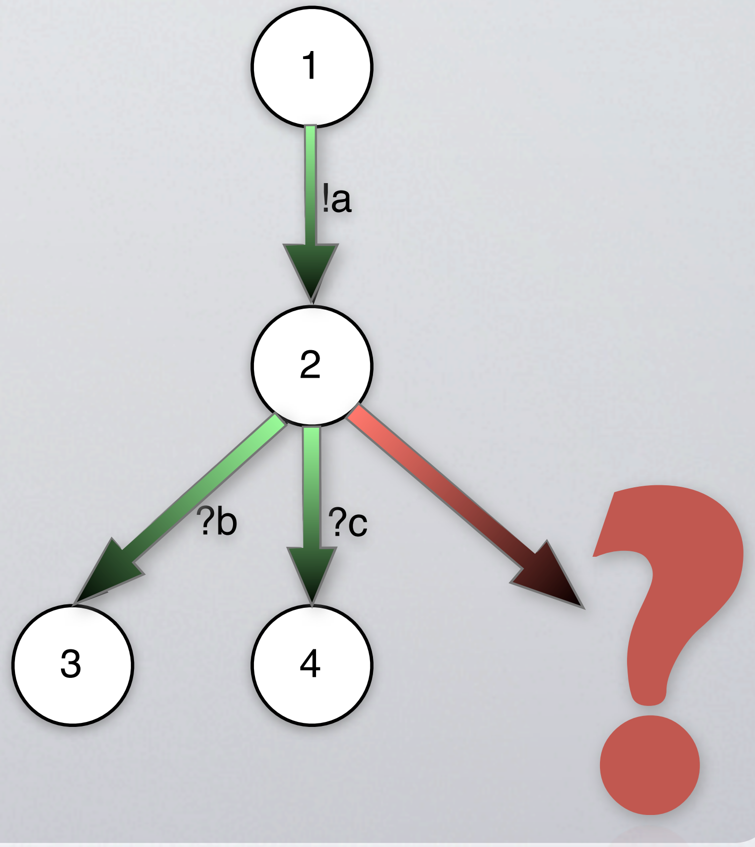
- Model-based analysis so far:
  - Test case analysis.
  - Catalog with 11 test case anomalies.
- Anomalies between test cases:
  - Are two test cases in a test suite similar?
  - Anomalies between two similar test cases?

Response Inconsistencies

Test Case 1



Test Case 2



Outline

- Motivation and Foundations
- A Quality Model for Test Specifications
- Model-Based Analysis of Test Specifications
- **Case Study**
- Contributions and Outlook

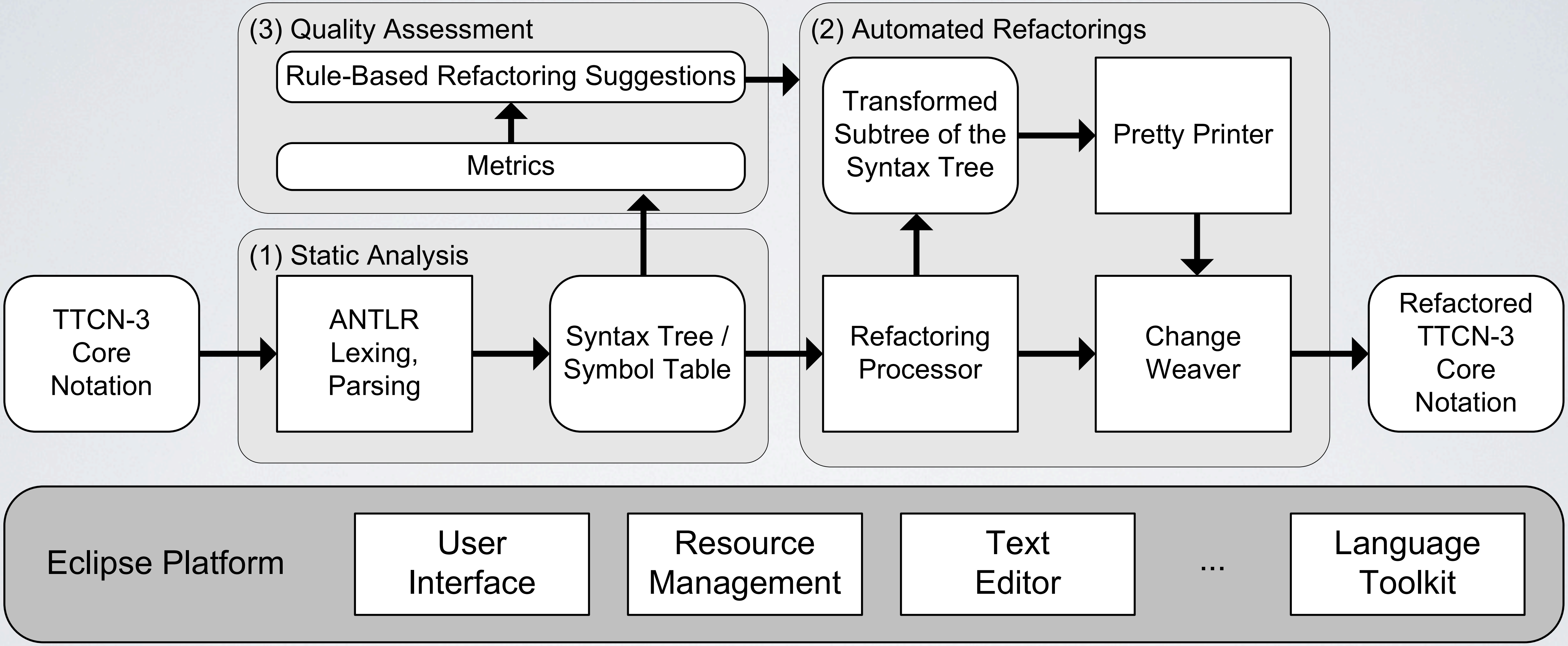
Experiments

- Two experiments:

  1. Static analysis of test specifications and their improvement.

  2. Dynamic model-based analysis of test cases in a test

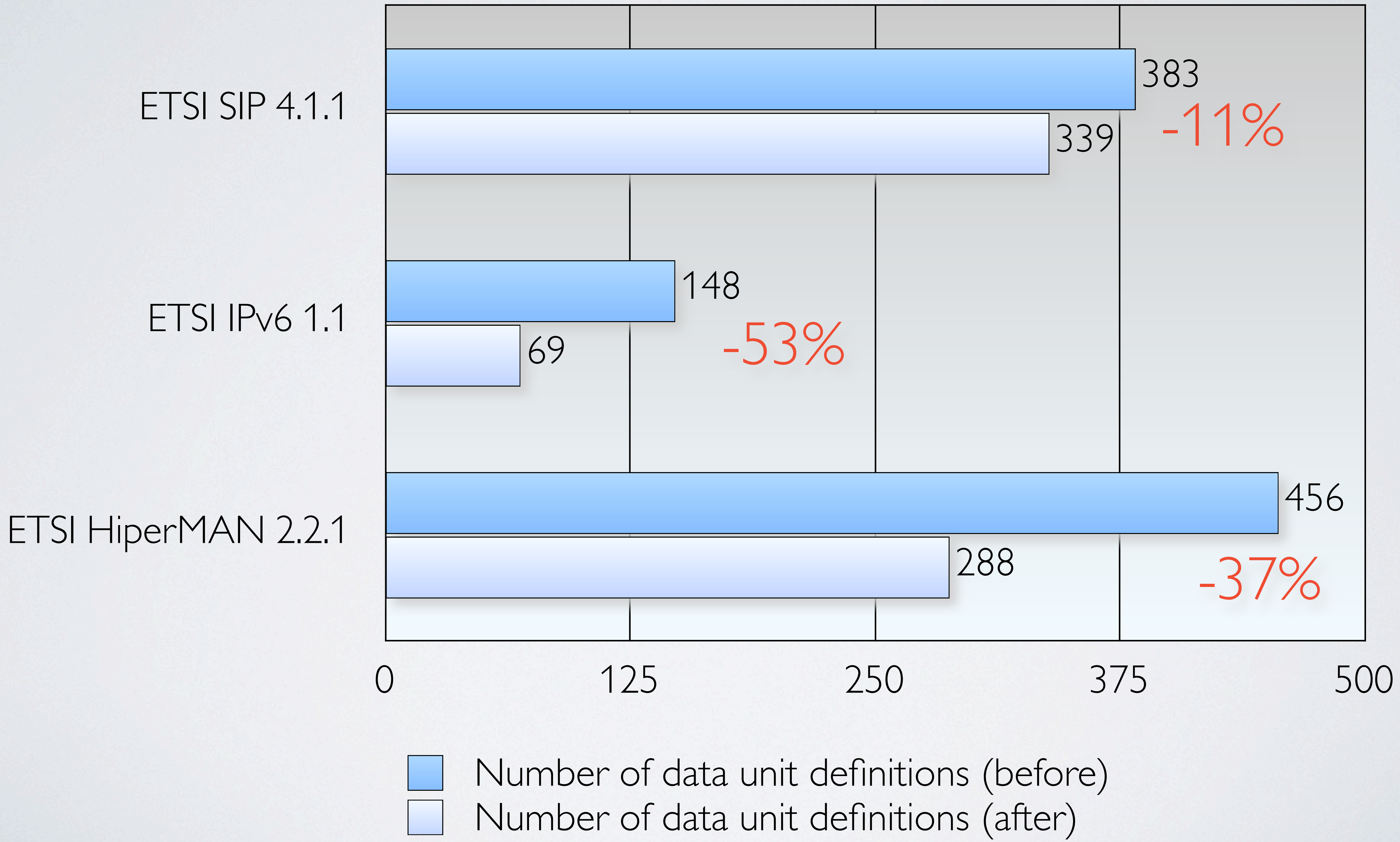
Static Analysis: Prototype Tool



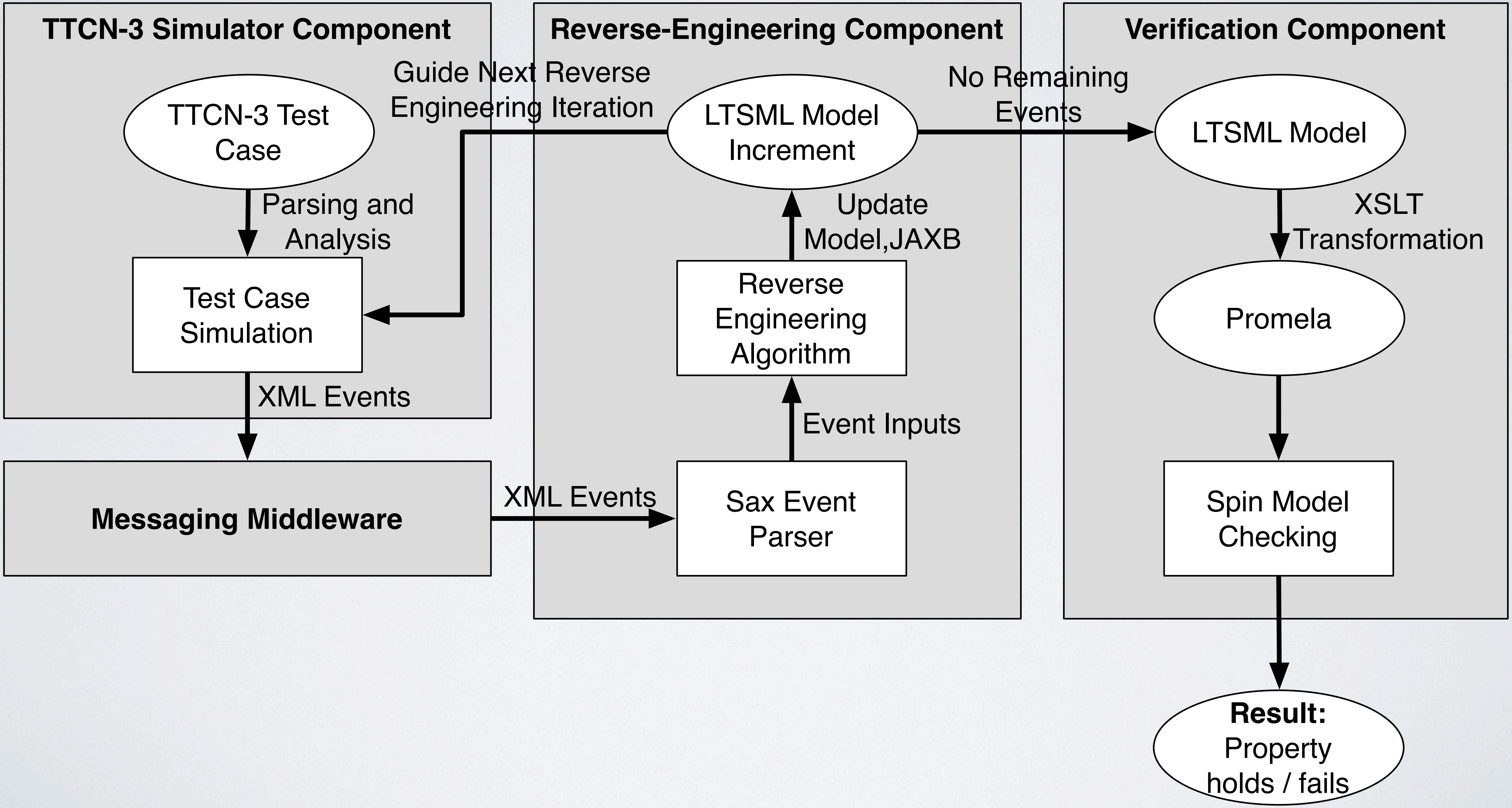
Static Analysis: Case Study (1/2)

- **Case study subjects**:
  - ETSI SIP v4.1.1, ETSI IPv6 v1.1, ETSI HiperMAN v2.2.1 (TTCN-3)
- **Objective:** improve the *maintainability* quality characteristic.
- **Quality subcharacteristic:** *changeability* of test data
  - Are there removable data unit definitions?
  - Are there similar data unit definitions?
- **Improvement rules**:
  - No. of references to a data unit definition = 0, then **remove data unit definition**.
  - No. of references to a data unit definition = 1, then **inline data unit definition**.
  - No. of different fields in similar data unit definitions account for at most 30% of all fields, then **parameterize data unit definition** at these fields.

Static Analysis: Case Study (2/2)



Dynamic Analysis: Prototype Tool



Dynamic Analysis: Case Study (1/3)

- **Case study subject**: subset of ETSI SIP v4.1.1 and v4.2.5 test suites (22 test cases).
    - Models have between 100-3000 states, 60-250 actions, 100-3200 transitions.
- **Objective**: practical feasibility, precision, find possible anomalies in the SIP test suite.
- **Quality subcharacteristic:** *reliability*, are there timeout inconsistencies?
- **Quality characteristic:** *compliance*, what is the degree of verdict / timer inconsistency?
- **Quality subcharacteristic:** *test completeness,* are there any missing or early test verdicts?

Dynamic Analysis: Case Study (2/3)

- **Reliability** (timeout inconsistencies):
    - All analyzed test cases exhibited the anomaly.
- **Compliance** (verdict/timer inconsistencies):
    - One test case among the analyzed exhibited the anomaly (SIP v4.2.5)
- **Test correctness** (missing or early test verdicts):
    - None of the SIP test cases exhibited the anomaly.
    - Mutation variant: test cases were identified correctly.

Dynamic Analysis: Case Study (3/3)

- **Interpretation**:
  - Coding guidelines may influence the presence of anomalies.
  - Anomaly selection is project-specific.
  - Catch human mistakes.
- **Recall and precision of the analysis**:
  - All test cases with anomalies were correctly identified.
  - No false positives among the reported test cases.
  - The behavior abstractions did not have any negative impact.

Outline

- Motivation and Foundations
- A Quality Model for Test Specifications
- Model-Based Analysis of Test Specifications
- Case Study
- **Contributions and Outlook**

Contributions

- **Four main contributions**:

  C1 A quality model for test specifications.

  C2 An instantiation of the quality model for test specifications.

  C3 A model-based analysis method for test specifications.

  C4 A method for the detection of inconsistent responses in a test suite.

Impact (1/2)

- **7 Conference papers (selection):**
  - SAM 2006: **Refactoring and Metrics for TTCN-3 Test Suites**. B. Zeiss, H. Neukirchen, J. Grabowski, D. Evans, P. Baker. LNCS 4320. C2
  - SE 2007: **Applying the ISO 9126 Quality Model to Test Specifications – Exemplified for TTCN-3 Test Specifications**. B. Zeiss, D. Vega, I. Schieferdecker, H. Neukirchen, J. Grabowski. LNI 105. C1 C2
  - TESTCOM/FATES 2008: **Reverse-Engineering Test Behavior Models for the Analysis of Structural Anomalies** (Short Paper). B. Zeiss, J. Grabowski. C3
  - TESTCOM/FATES 2009: **Analyzing Response Inconsistencies in Test Suites**. B. Zeiss, J. Grabowski. LNCS 5826. C4

Impact (2/2)

- 2 Journal articles:
  - STTT Vol. 10(4): **An Approach to Quality Engineering of TTCN-3 Test Specifications**. H. Neukirchen, B. Zeiss, J. Grabowski. 2008. C1 C2
  - STVR Vol. 18(2): **Quality assurance for TTCN-3 test specifications**. H. Neukirchen, B. Zeiss, J. Grabowski, P. Baker, D. Evans. 2008. C2
- Overview articles:
  - OBJEKTspektrum Online Themenspezial Testing: **Systematische Qualitätssicherung für Testartefakte**. J. Grabowski, P. Makedonski, T. Rings, B. Zeiss. 2009. C1 C2 C3 C4

Outlook

- Extension of the test case anomalies catalogue.
- Further test suite analyses.
- Refinement of the model definition.
- Domain-specific language for the tracing data.

Questions?