MONDRAGON UNIBERTSITATEA

GOI ESKOLA POLITEKNIKOA

ESCUELA POLITÉCNICA SUPERIOR

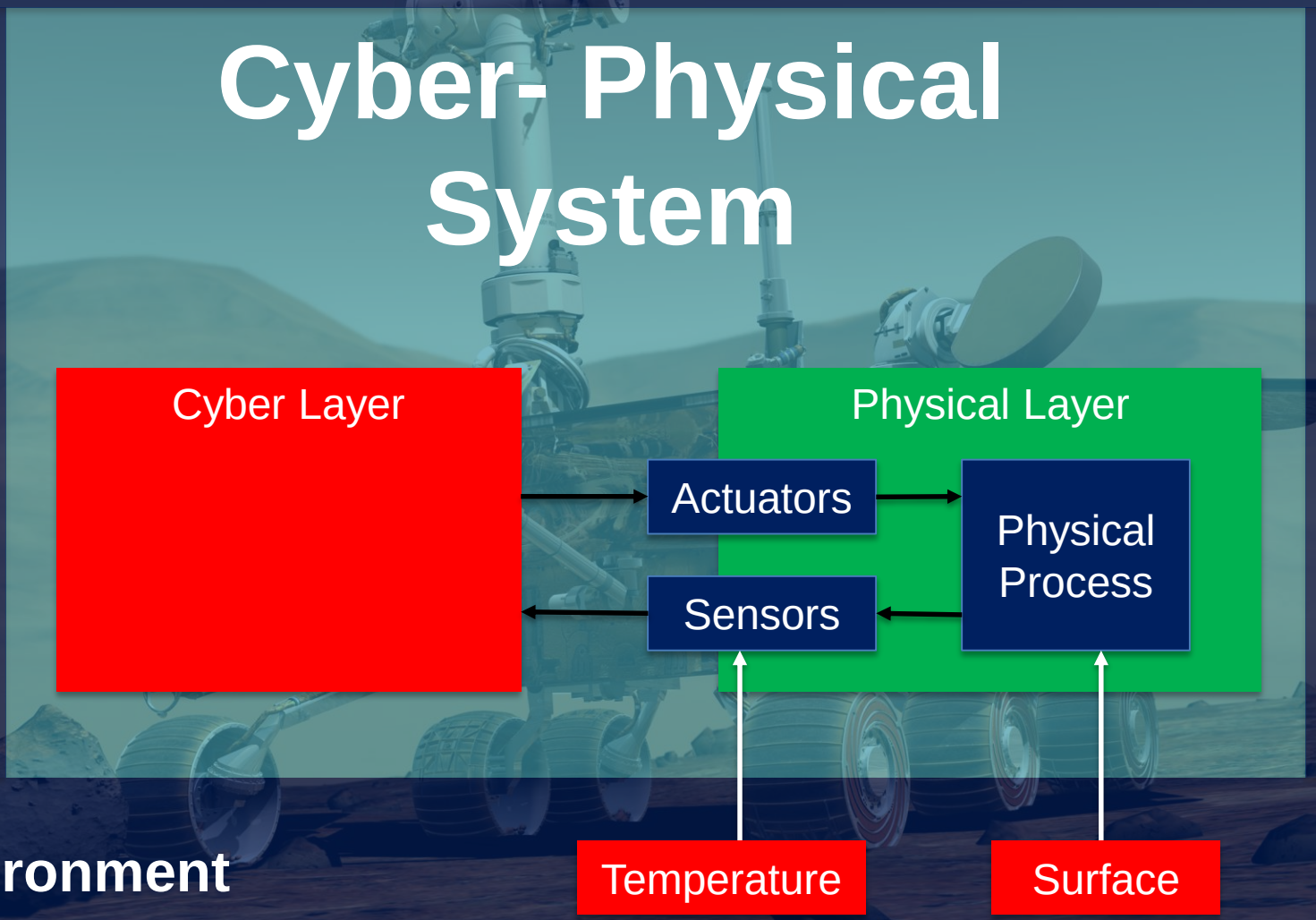
Test Optimization of Configurable Cyber-Physical Systems with Search Algorithms

Aitor Arrieta

Collaborators:
Goiuria Sagardui,
Leire Etxeberria
Shuai Wang

Summer School in
Search Based Software Engineering
Cádiz
30/06/2016

Cyber-Physical Systems



Testing configurable CPS vs configurable SW/SPLs

Configurable CPSs	Configurable SW/SPLs
Cost of the prototype	No prototype needed
Higher simulation time (Simulation of SW + Physical Layer)	Lower simulation time
Many domains (often co-simulation needed)	Mainly software domain
Faults: software, interaction, sensors, actuators, communication systems, etc.	Faults: software, interaction faults, etc.
More test levels: Model-, Software-, Processor- and Hardware- in the Loop	Model-,Software- in the Loop
High Variability: Many configurations to test Infeasible to test all possible configurations Unclear notion of test coverage	

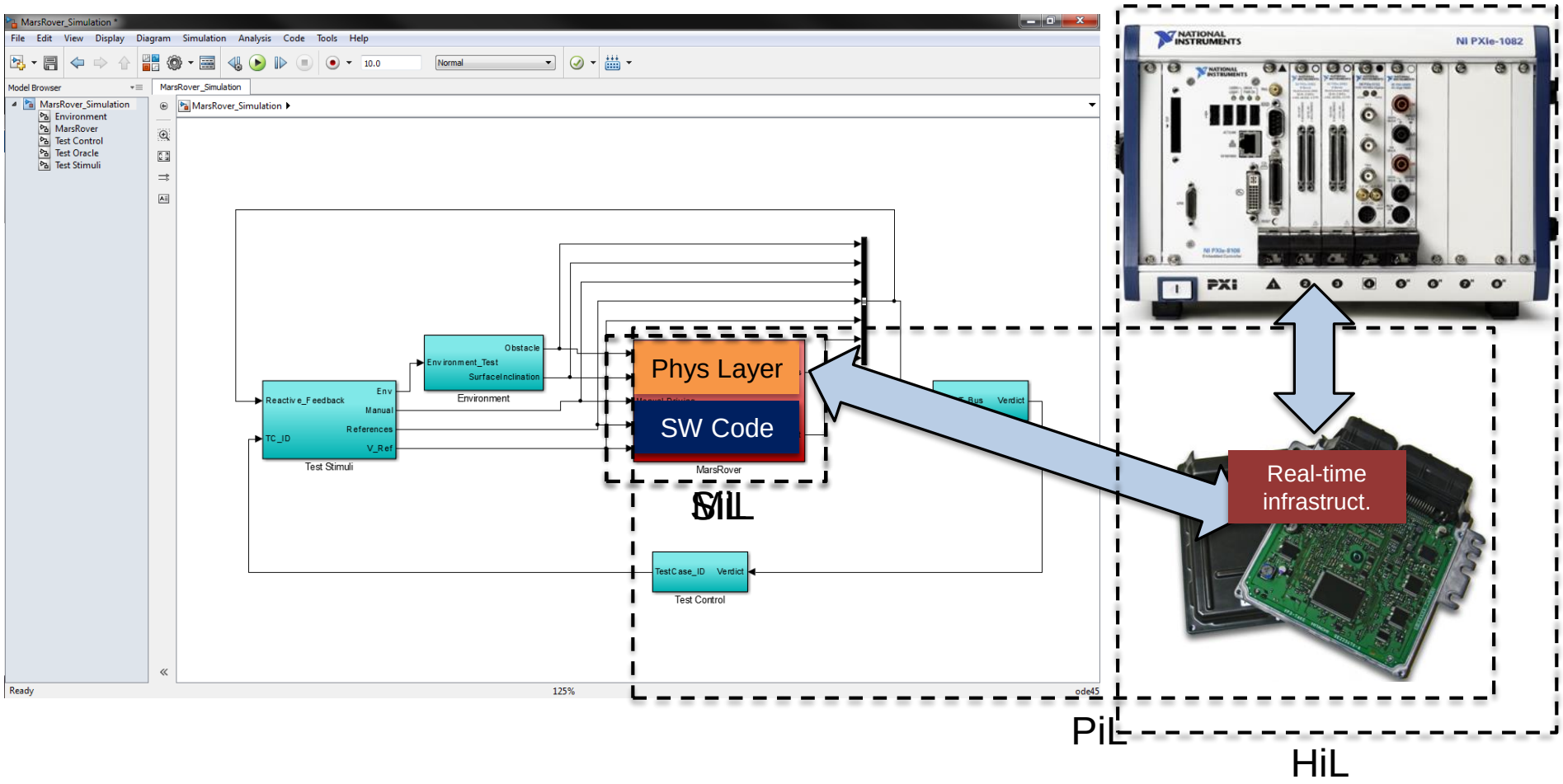
Our research focus

- Test Automation
  - Automatic generation of the test infrastructure
  - Automatic generation of test cases
- Test Optimization
  - Test case selection
  - Test case prioritization
- Variability Modeling
- Simulation
  - Heterogeneous model simulations with the FMI standard
  - Co-simulation engines

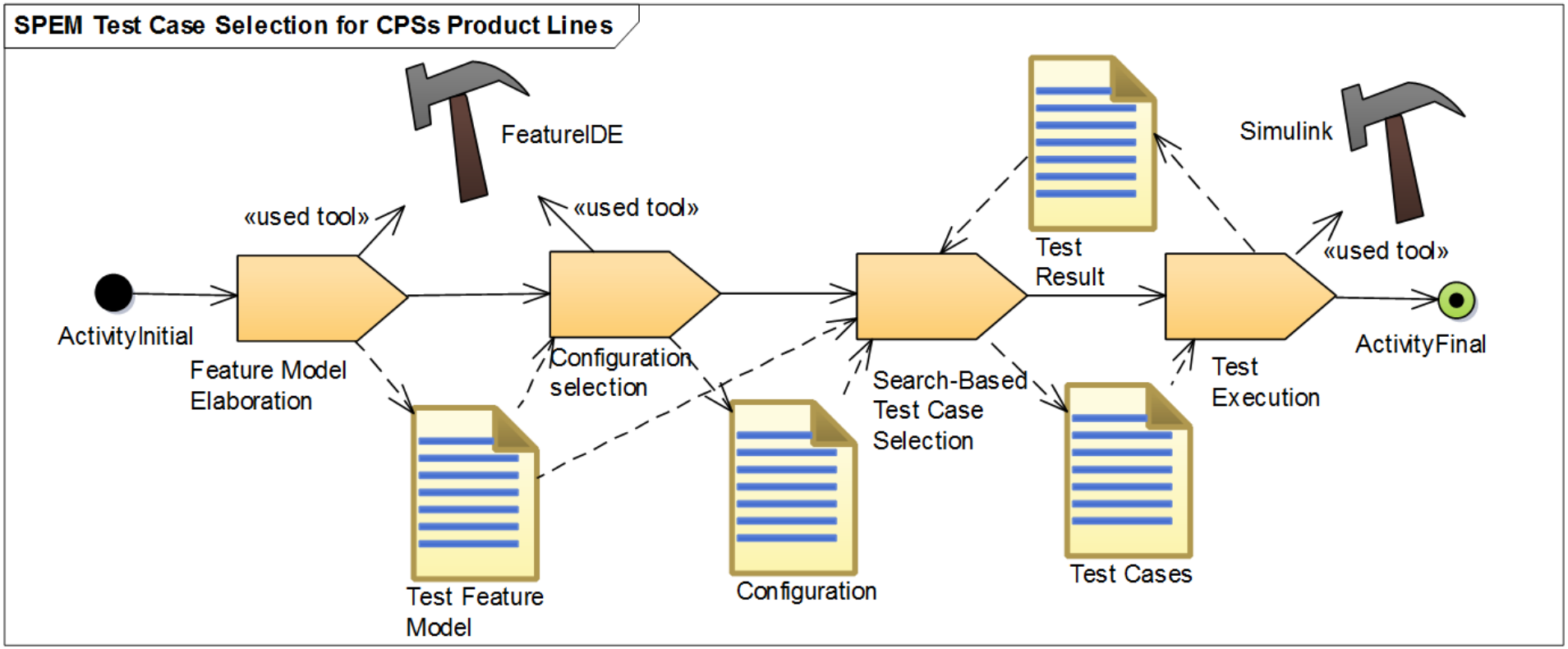
Testing challenges

- Multi-Level Test Case Selection
- Test Case Prioritization with Environmental Unpredictability
- Test Case Generation
- Environmental Testing of CPSs

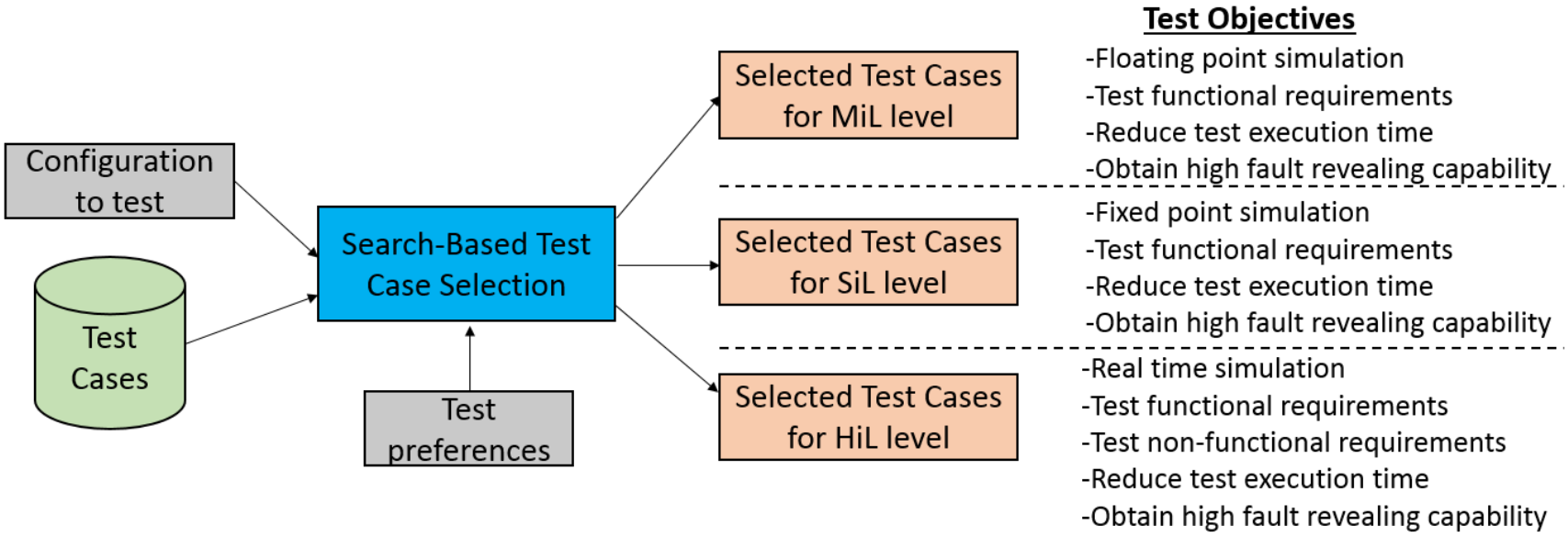
Simulation-based testing CPSs



Overview of the Search-Based Test Case Selection Process



Overview of the Search-Based Test Case Selection Process



Corresponding test objectives for each test level

- Cost Measures
  - Test Execution Time: MiL, SiL, HiL
  - Test Suite Similarity: SiL
- Quality Measures
  - Fault Detection Capability: MiL, SiL, HiL
  - Functional Requirements Coverage: MiL, SiL, HiL
  - Pairwise Functional Requirement Coverage: MiL, SiL, HiL
  - Non-Functional Requirements Coverage: HiL
  - Pairwise Non-Functional Requirements Coverage: HiL
- Weight-Based Search Algorithms were employed: GA, AVM, GRE and RS

Research Questions

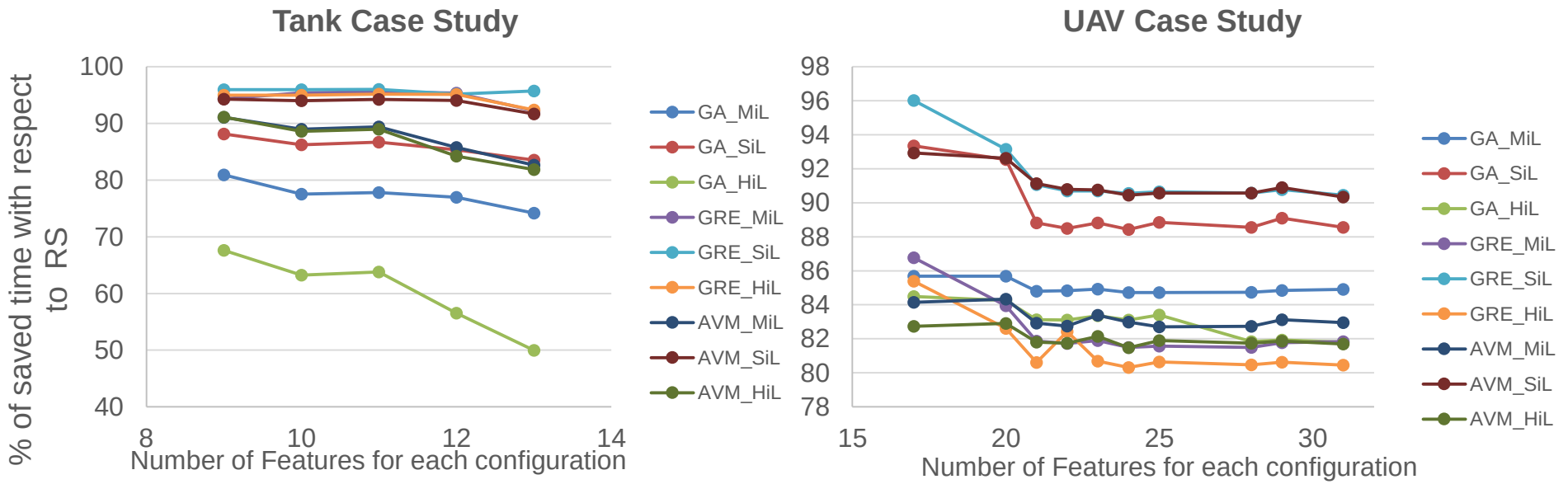
- RQ 1: Are the selected algorithms cost-effective as compared to RS?
- RQ 2: Which algorithm shows best performance?

Experiment Design

- Two case studies:
  - Industrial tank (5 configurations)
  - Configurable Drone (10 configurations)
- Four Search Algorithms: GA, AVM, GRE, RS
- Five Test Suites: (80, 90, 100, 110 and 120 tests)
- 20 mutants (50 % physical layer, 50 % software)
- 100 algorithm runs
- In total:
  - 25 artificial problems for industrial tank
  - 50 artificial problems for configurable drone
- Evaluation Metrics:
  - Test Execution Time, Mutation Score, Achieved Test Coverage (4 types)

Results – Statistical Analysis

- Test Execution Time
  - Industrial Tank : GRE
  - UAV: GA, GRE (at SiL level)



Results – Statistical Analysis

- % of Detected Faults
  - Industrial Tank: GA slightly better
  - UAV: similar
- Requirements Coverage
  - Industrial Tank : Similar
  - UAV: Similar
- Pairwise Requirement Coverage
  - Industrial Tank : GA

Results – Statistical Analysis

- Non-Functional Requirement Coverage
  - Industrial Tank: GA
  - UAV: Similar

- Pairwise Non-Functional Requirement Coverage
  - Industrial Tank: GA

Concluding Remark

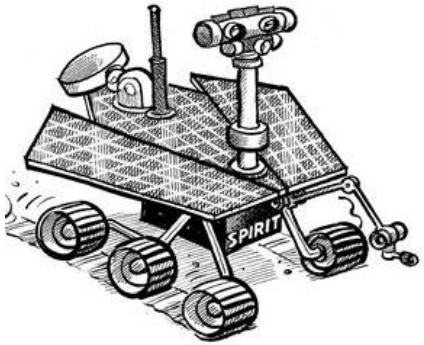
- All the algorithms show better performance than RS
- The GA gives most of the importance to test quality
- If pairwise coverage is taken into account
  - Test quality is important?
    - Yes: GA is recommended
    - No: Time is highly important?
      - YES: GRE is recommended
      - NO: GA is recommended
- If pairwise coverage is not taken into account: GA is recommended

Detailed Information

A. Arrieta, S. Wang, G. Sagardui und L. Etxeberria, “Search-Based Test Case Selection of Cyber-Physical Systems Product Lines for Simulation-Based Validation," in SPLC 2016: Systems and Software Product Line Conference,2016.

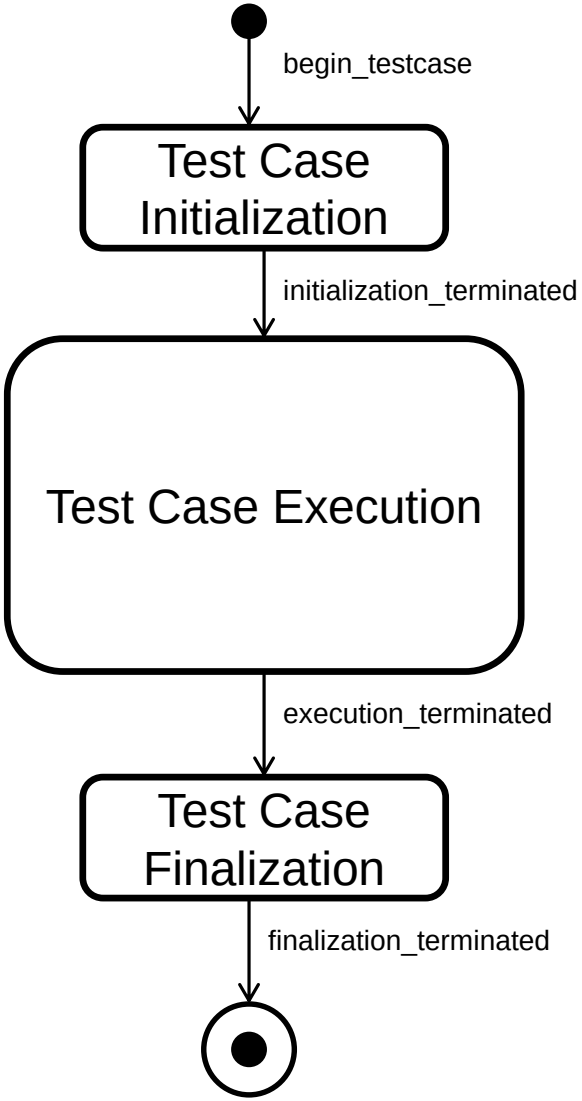
Testing challenges

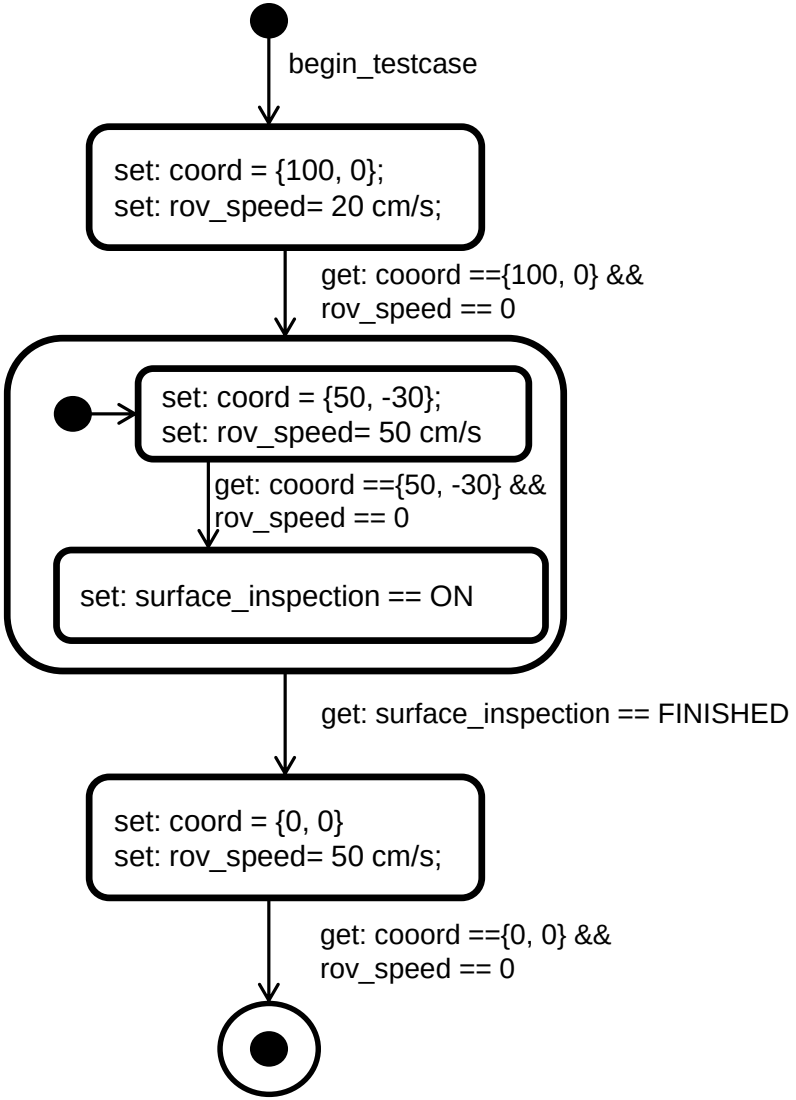
- Multi-Level Test Case Selection
- Test Case Prioritization with Environmental Unpredictability
- Test Case Generation
- Environmental Testing of CPSs



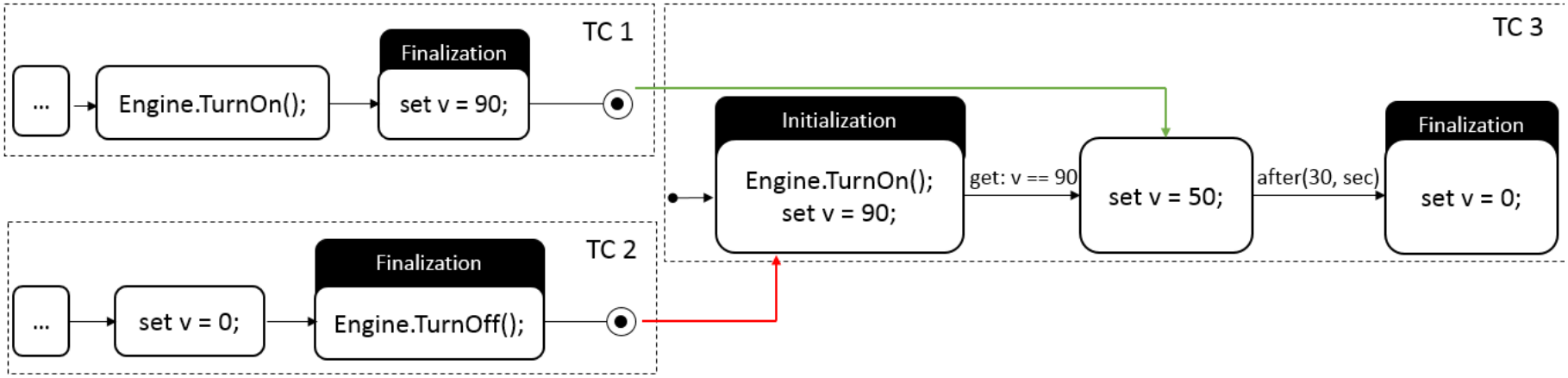
Come Here!

Reactive Test Cases





Variable Test Execution Time



Objective Functions

- Test Case Execution Time
  - Test Initialization Time
  - Test Execution Time
  - Test Finalization Time
- Fault Detection Capability
  - Number of times the test detected faults with respect to number of times the test was executed
- Weight-Based Search Algorithms were employed: WBGA, RWGA, AVM, GRE and RS

Evaluation – Research Questions

- RQ 1: Are the selected search algorithms cost-effective as compared to Random Search?
- RQ 2: Which search algorithm achieves the best performance among the selected search algorithms?
- RQ 3: How does the increment of test cases impact the performance of the selected search algorithms?

Experiment Design

- Two Case Studies:
  - Cruise Control (Daimler AG)
  - Synthetic Case Study: Industrial Tank
- 5 Search Algorithms
  - WBGA
  - RWGA
  - AVM
  - Greedy
  - RS
- 20 Test Suites (25, 30, 35,…, 115, 120)
- 100 repetitions per test suite
- Evaluation: Fault Detection Time – Mutation testing with 4, 10 and 20 mutants

Results – RQ 1 & RQ 2 for Cruise Control

RQ	Pair of Algorithms	A<B	A>B	A=B
1	Greedy vs RS	60/60	0/0	0/0
	RWGA vs RS	60/60	0/0	0/0
	WBGA vs RS	60/60	0/0	0/0
	AVM vs RS	60/60	0/0	0/0
2	Greedy vs RWGA	50/50	10/8	0/2
	Greedy vs WBGA	54/54	6/5	0/1
	Greedy vs AVM	51/51	9/9	0/0
	RWGA vs WBGA	41/38	18/11	1/11
	RWGA vs AVM	20/19	26/19	14/22
	WBGA vs AVM	17/14	40/31	3/15

Results – RQ 1 & RQ 2 for Industrial Tank

RQ	Pair of Algorithms	A<B	A>B	A=B
1	Greedy vs RS	57/56	3/2	0/2
	RWGA vs RS	56/56	4/2	0/2
	WBGA vs RS	55/51	5/2	0/7
	AVM vs RS	56/56	4/2	0/2
2	Greedy vs RWGA	19/10	39/18	2/32
	Greedy vs WBGA	29/19	27/27	4/11
	Greedy vs AVM	8/8	39/39	13/13
	RWGA vs WBGA	43/27	16/0	1/33
	RWGA vs AVM	6/0	52/41	2/19
	WBGA vs AVM	6/0	53/43	1/17

Results – RQ 3

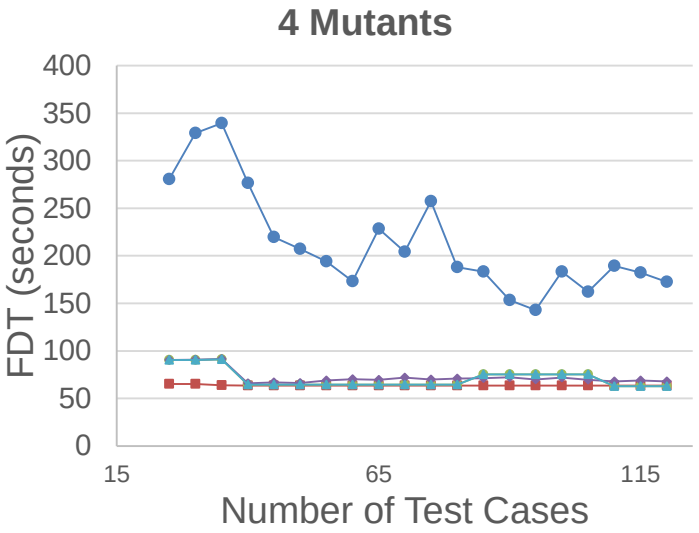
- Cruise Control:

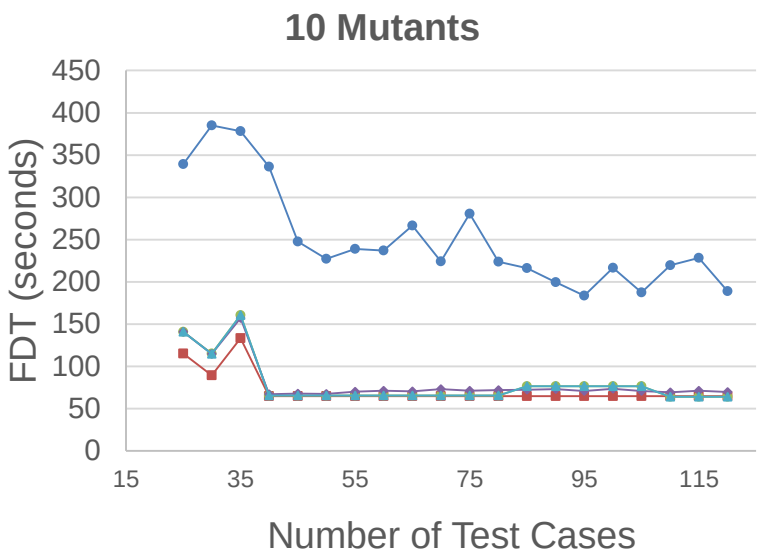
Algorithm	4 mutants	10 mutants	20 mutants
RS	ρ = -0,212	ρ =-0,235	ρ =-0,156
Greedy	ρ = -0,621	ρ =-0,614	ρ =-0,614
RWGA	ρ = -0,912	ρ = -0,923	ρ =-0,924
WBGA	ρ = -0,457	ρ = -0,453	ρ =-0,434
AVM	ρ = -0,457	ρ =-0,438	ρ =-0,438

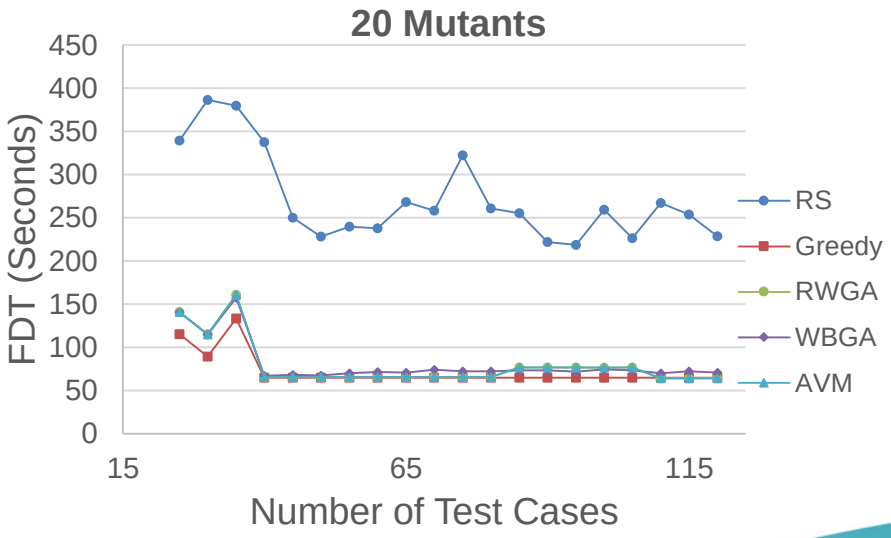
- Industrial Tank

Algorithm	4 mutants	10 mutants	20 mutants
RS	ρ =-0,045	ρ =-0,194	ρ =-0,058
Greedy	ρ =-0,297	ρ =-0,297	ρ =-0,297
RWGA	ρ =-0,277	ρ =-0,393	ρ =-0,393
WBGA	ρ =-0,115	ρ =-0,204	ρ =-0,205
AVM	ρ =-0,698	ρ =-0,697	ρ =-0,697

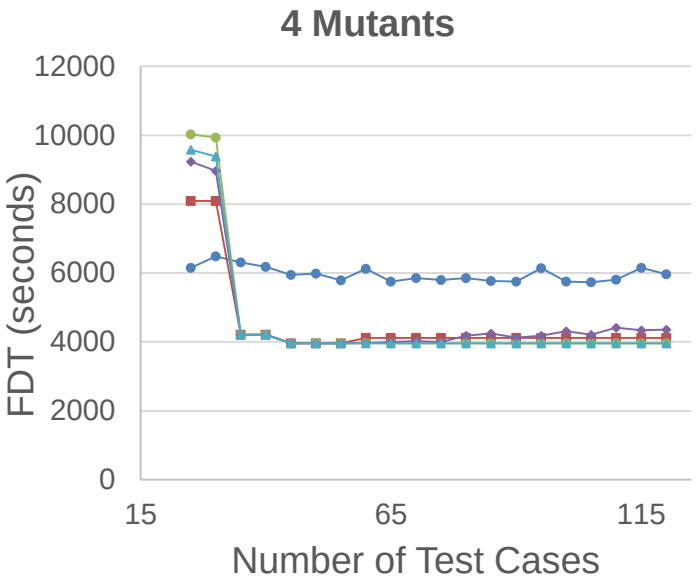
Results – Cruise Control

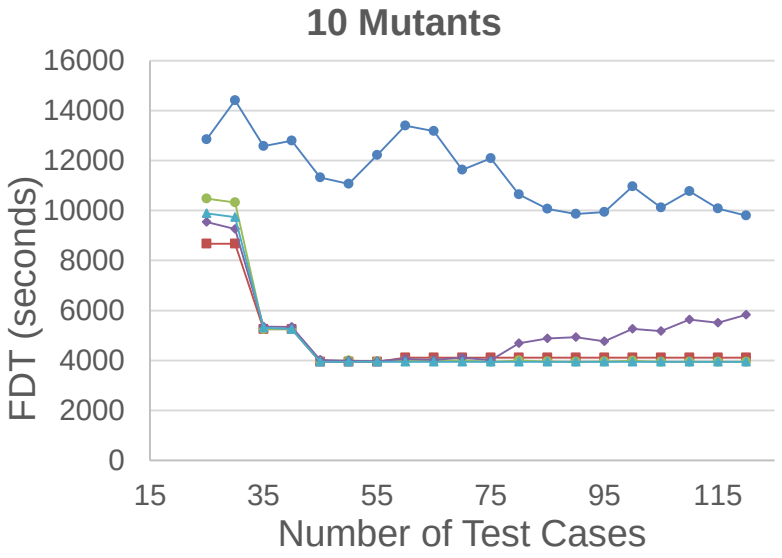


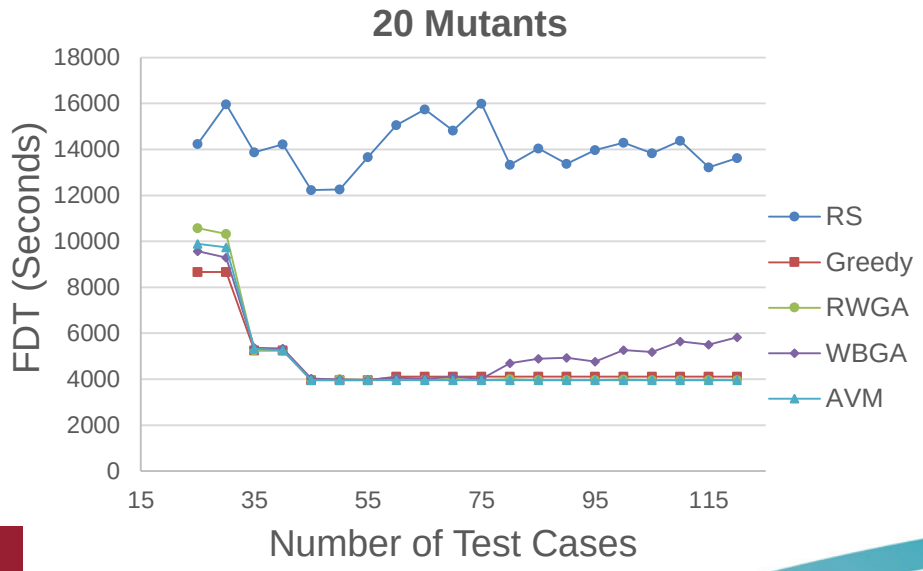




Results – Industrial Tank







Concluding Remarks

- Local search algorithms are better
- < 40 test cases: Greedy is recommended
- Test Case Execution Time < 100 seconds and < 110 test cases: Greedy is recommended
- > 40 test cases and Test Case Execution time > 100 seconds: AVM is recommended
- > 110 test cases: AVM is recommended

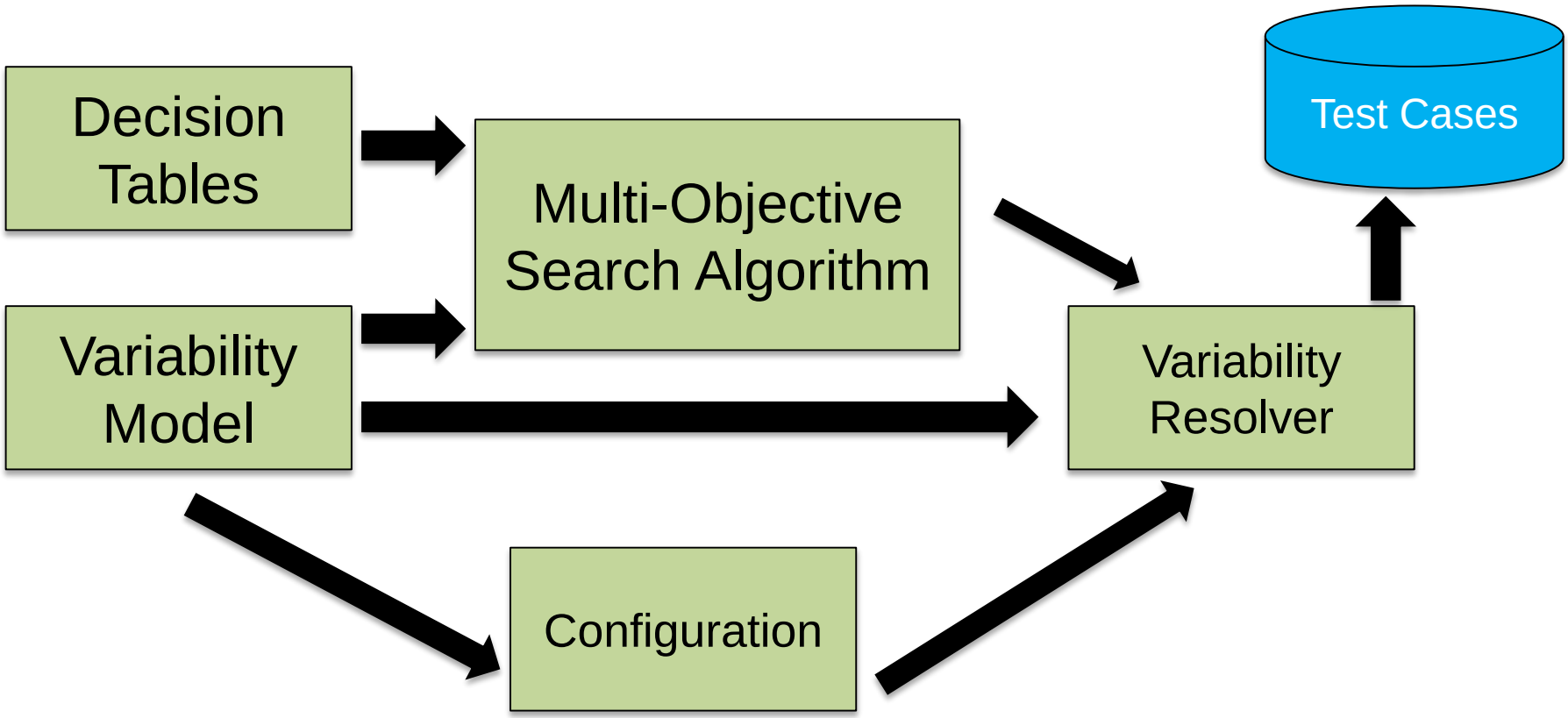
Detailed Information

A. Arrieta, S. Wang, G. Sagardui und L. Etxeberria, “Test Case Prioritization of Configurable Cyber-Physical Systems with Weight-Based Search Algorithms," in GECCO 2016: Genetic and Evolutionary Computation Conference, 2016.

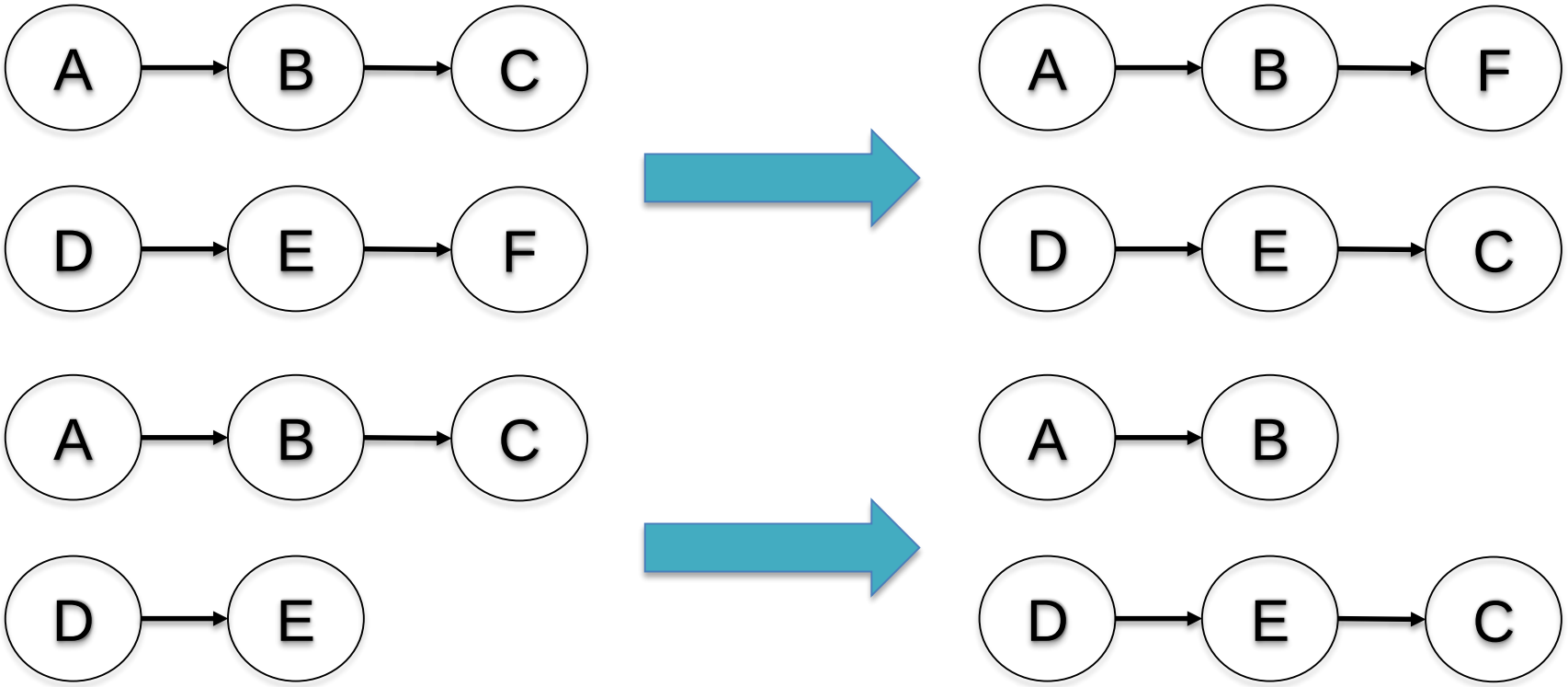
Testing challenges

- Multi-Level Test Case Selection
- Test Case Prioritization with Environmental Unpredictability
- Test Case Generation
- Environmental Testing of CPSs

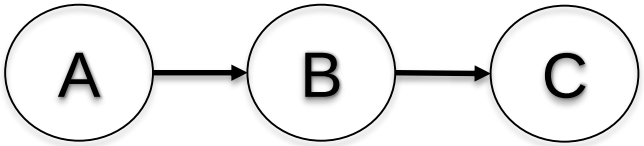
Automatic Generation of Reactive Test Cases



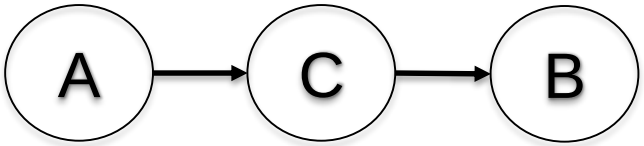
Crossover Operator



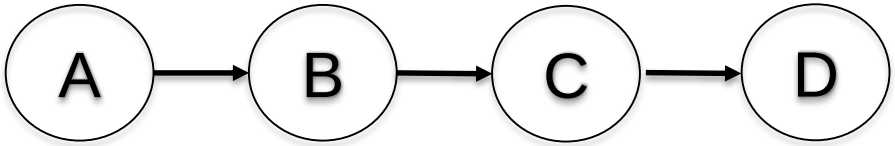
Mutation Operator



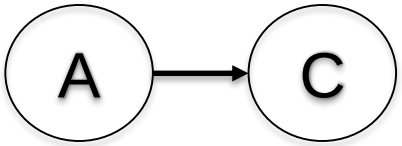
- Replace



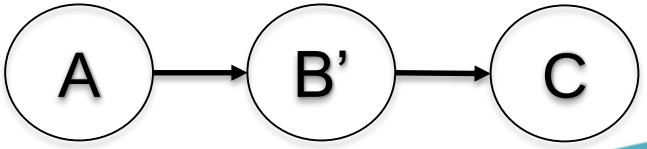
- Addition



- Remove

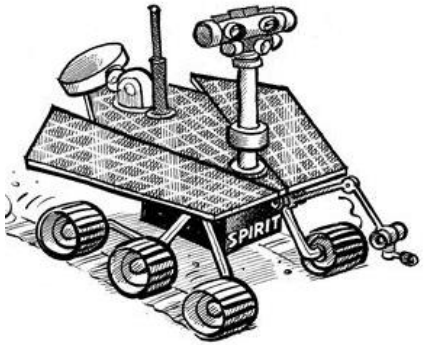


- Change of variable



Testing challenges

- Multi-Level Test Case Selection
- Test Case Prioritization with Environmental Unpredictability
- Test Case Generation
- Environmental Testing of CPSs



Come Here!

Conclusion

- Search algorithms optimize test processes in the context of CPSs
- Still many challenges and a lot of work to do
- Many industries interested:
  - Automotive
  - Railway
  - Maritime
  - …

Thank you for your attention!

Aitor Arrieta
aarrieta@mondragon.edu