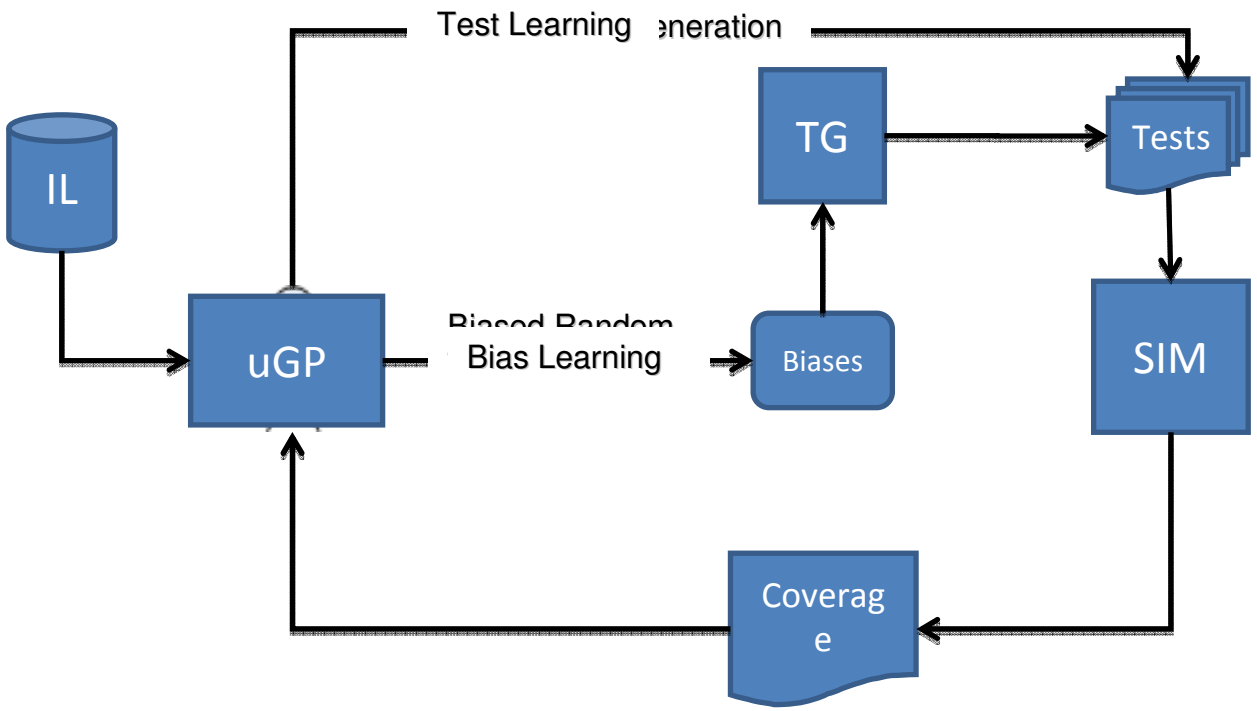
Feedback-based Coverage Directed Test Generation: An industrial evaluation

Charalambos Ioannides

Kerstin Eder, Geoff Barrett

Engineering and Physical Sciences Research Council

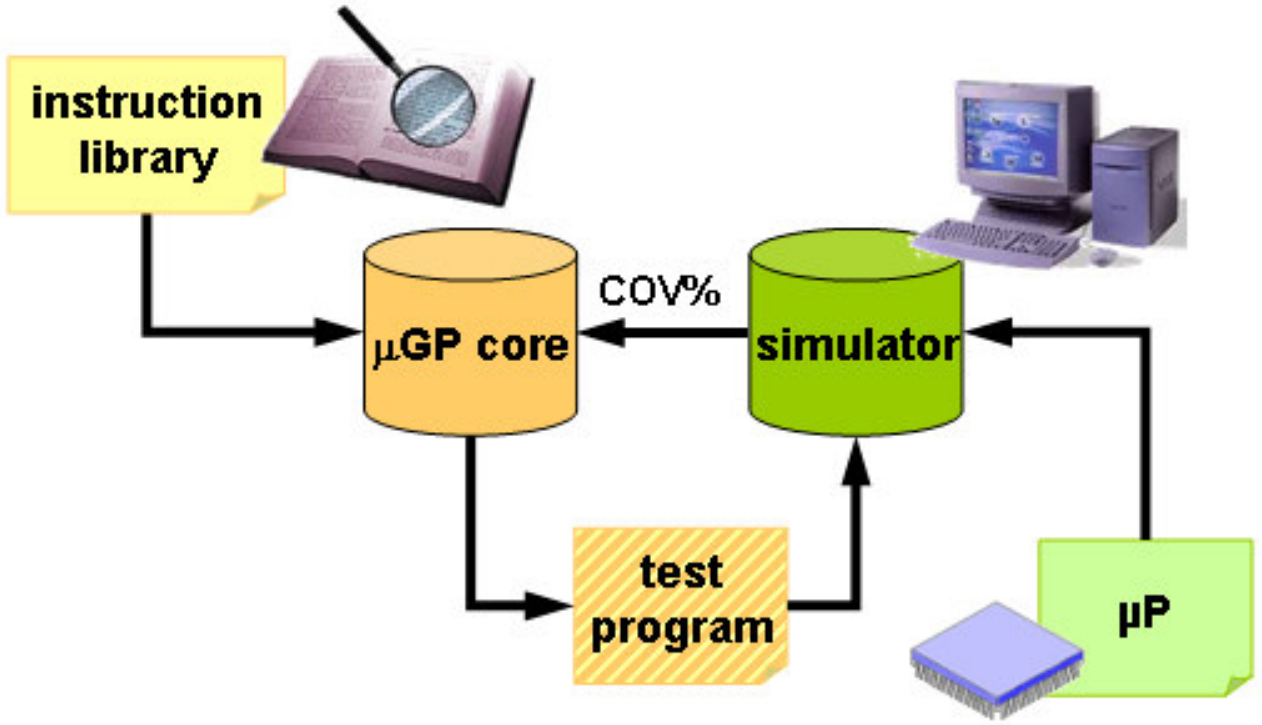
The Problem


Test Learning
IL
TG
Tests
uGP
Bias Learning
Biases
SIM
Coverag
e

Aims

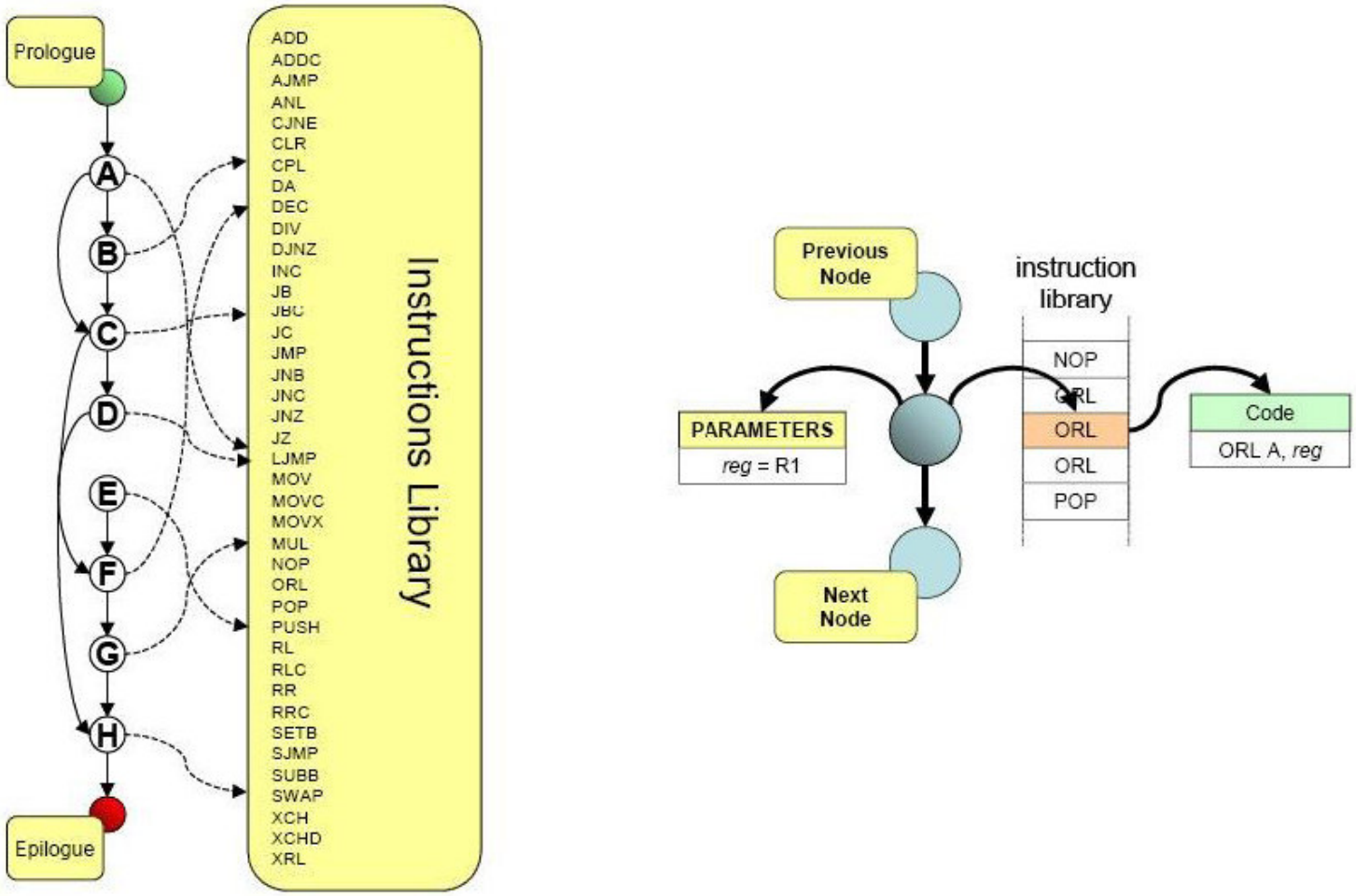
- Assess an academic CDG ML technique (MicroGP) on an in-house industrial design to:
  - Measure
    - Implementation Effort (man months)
    - Effectiveness (coverage)
    - Computational Effort (runtime)
  - Understand
    - Overall maturity of technology
    - Requirements for future research

MicroGP



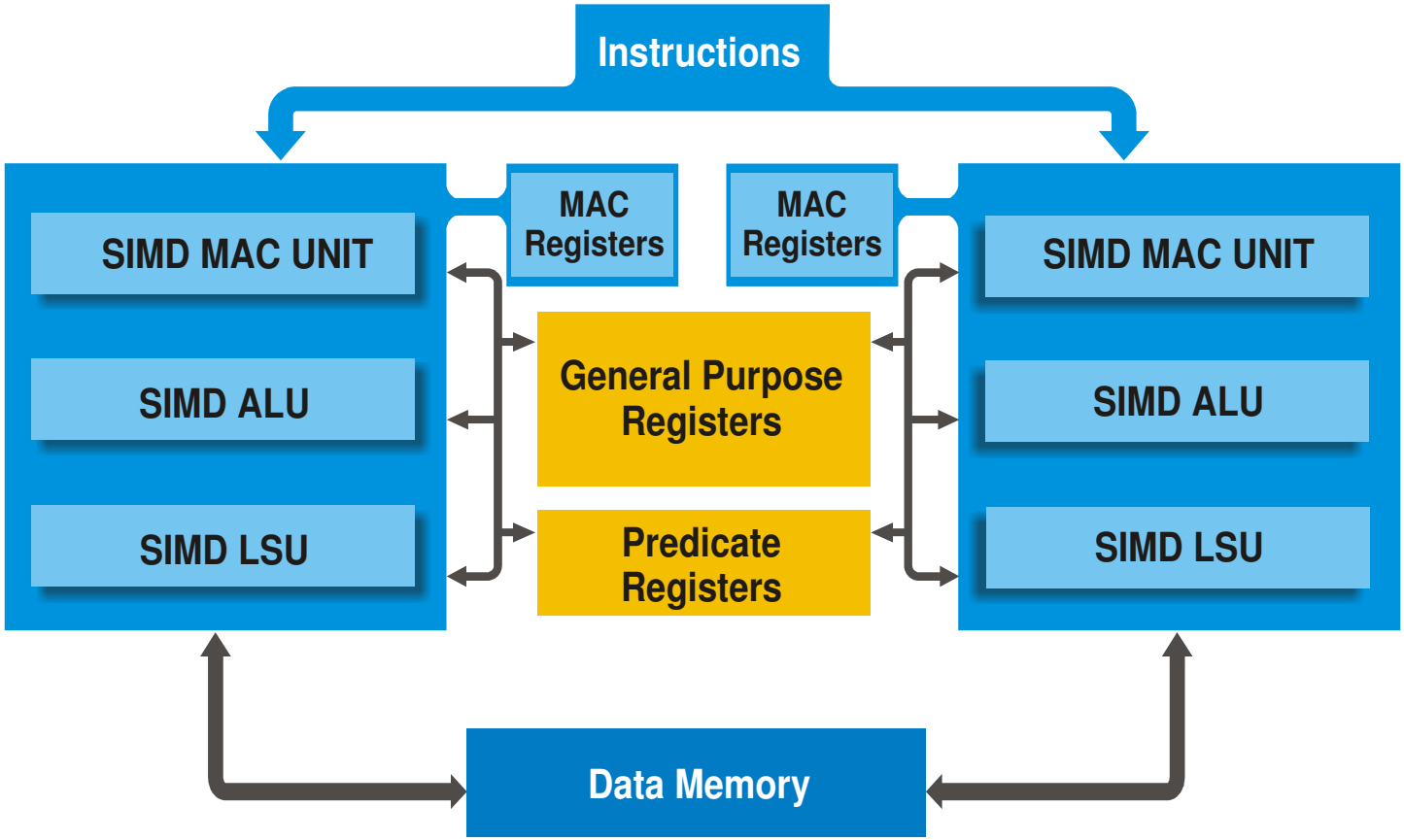
Source: http://www.cad.polito.it/research/Evolutionary%20Computation/MicroGP.html

MicroGP

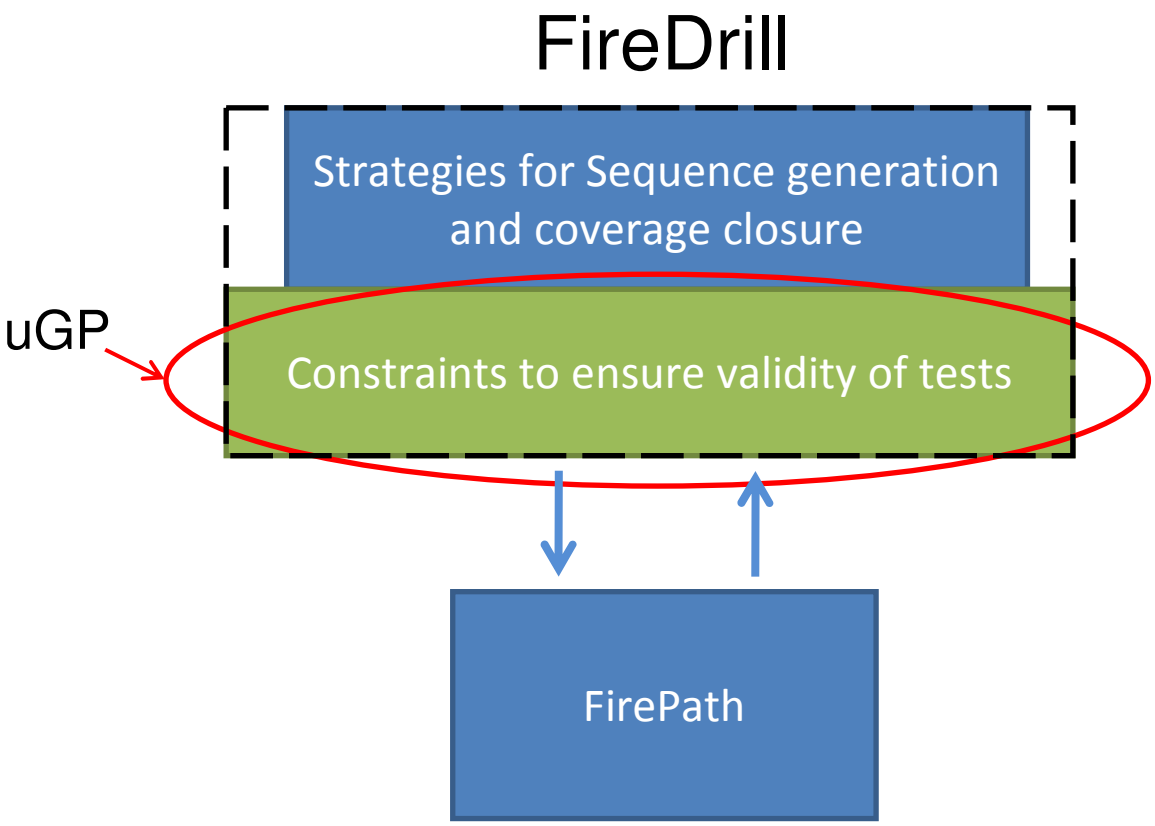


Source: F. Corno, G. Cumani, and G. Squillero, “Exploiting Auto-adaptive μGP for Highly Effective Test Programs Generation,” *Evolvable Systems: From Biology to Hardware,* Trondheim (Norway): Springer Berlin / Heidelberg, 2003, pp. 262—273.

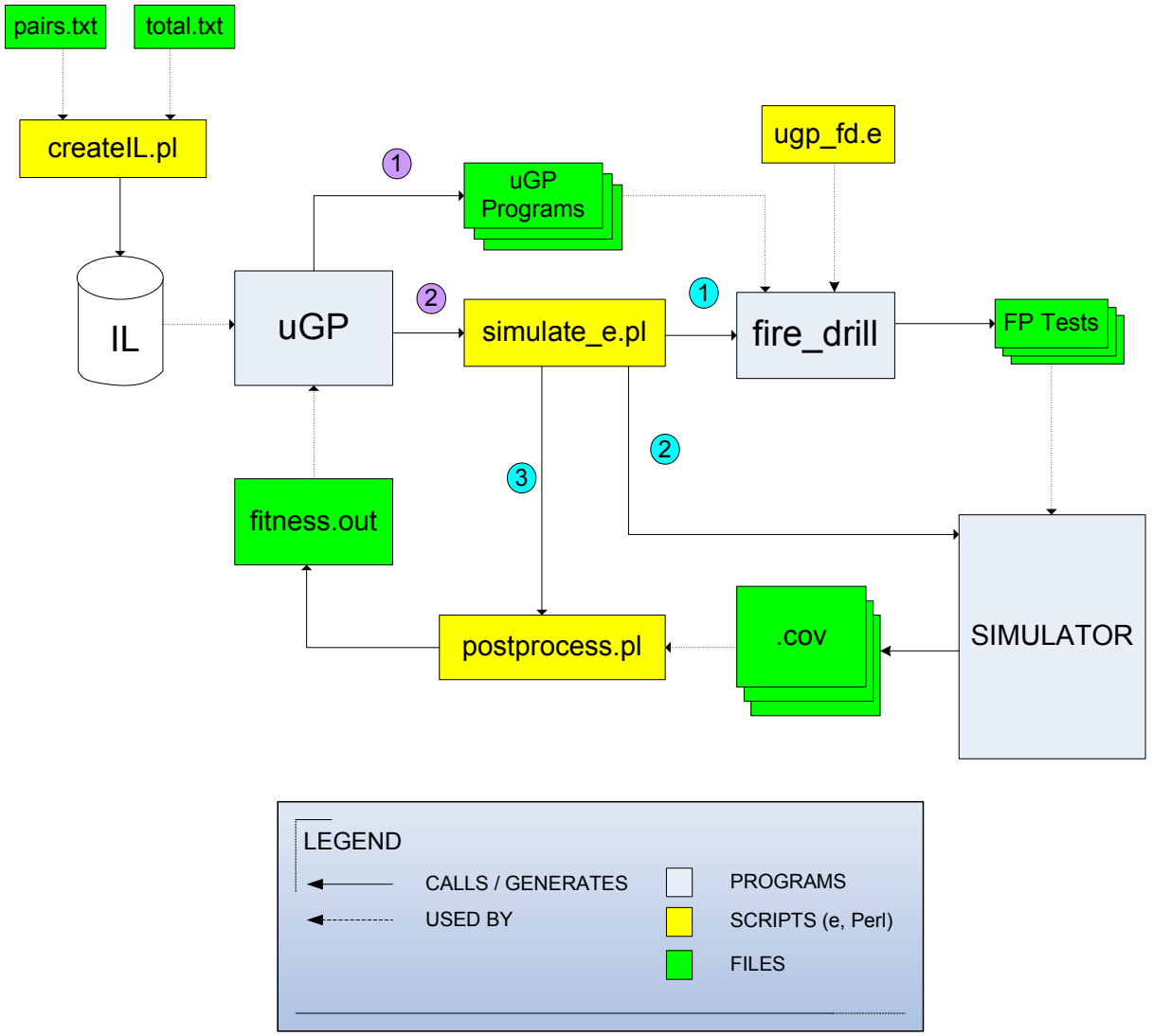
FirePath


Instructions
SIMD MAC UNIT
SIMD ALU
SIMD LSU
MAC Registers
MAC Registers
General Purpose Registers
Predicate Registers
SIMD MAC UNIT
SIMD ALU
SIMD LSU
Data Memory

FireDrill


FireDrill
Strategies for Sequence generation and coverage closure
uGP
Constraints to ensure validity of tests
FirePath

Experimental Framework



Experimental Setup

- Comparison between Evolutionary uGP (Evo), Random uGP (Rnd) and FireDrill (FD)
- Three alternative sorting methods for the population; by birth (BIRTH), in ascending (ASC), in descending (DSC) order
- Effectiveness comparison on Code Coverage Metrics, i.e. Expression, Branch, Toggle coverage
- Importance placed on Size of Tests evolved

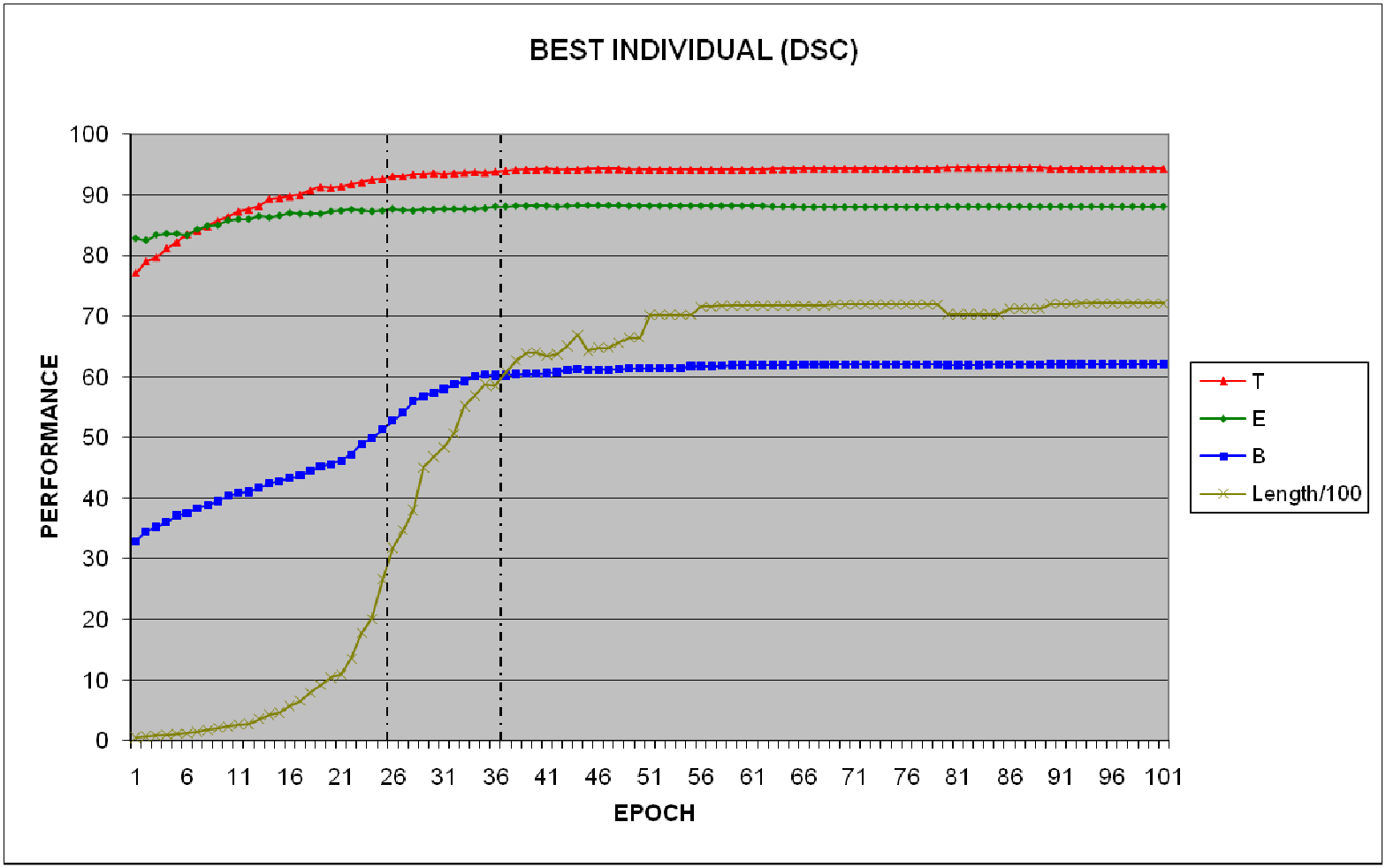
Results – Single Best

		E (%)	B (%)	T (%)	Best Size	Final Size
DSC	Evo	88.10	62.10	94.40	3344	7208
	Rnd	88.00	92.77	95.00		3344
	FD	86.16	65.08	90.62		3344
ASC	Evo	87.30	63.10	94.30	3743	7128
	Rnd	88.00	93.17	95.52		3743
	FD	85.42	58.73	89.17		3743
BIRTH	Evo	87.20	67.50	94.10	4277	7522
	Rnd	88.00	94.48	95.00		4277
	FD	86.58	70.70	91.54		4277

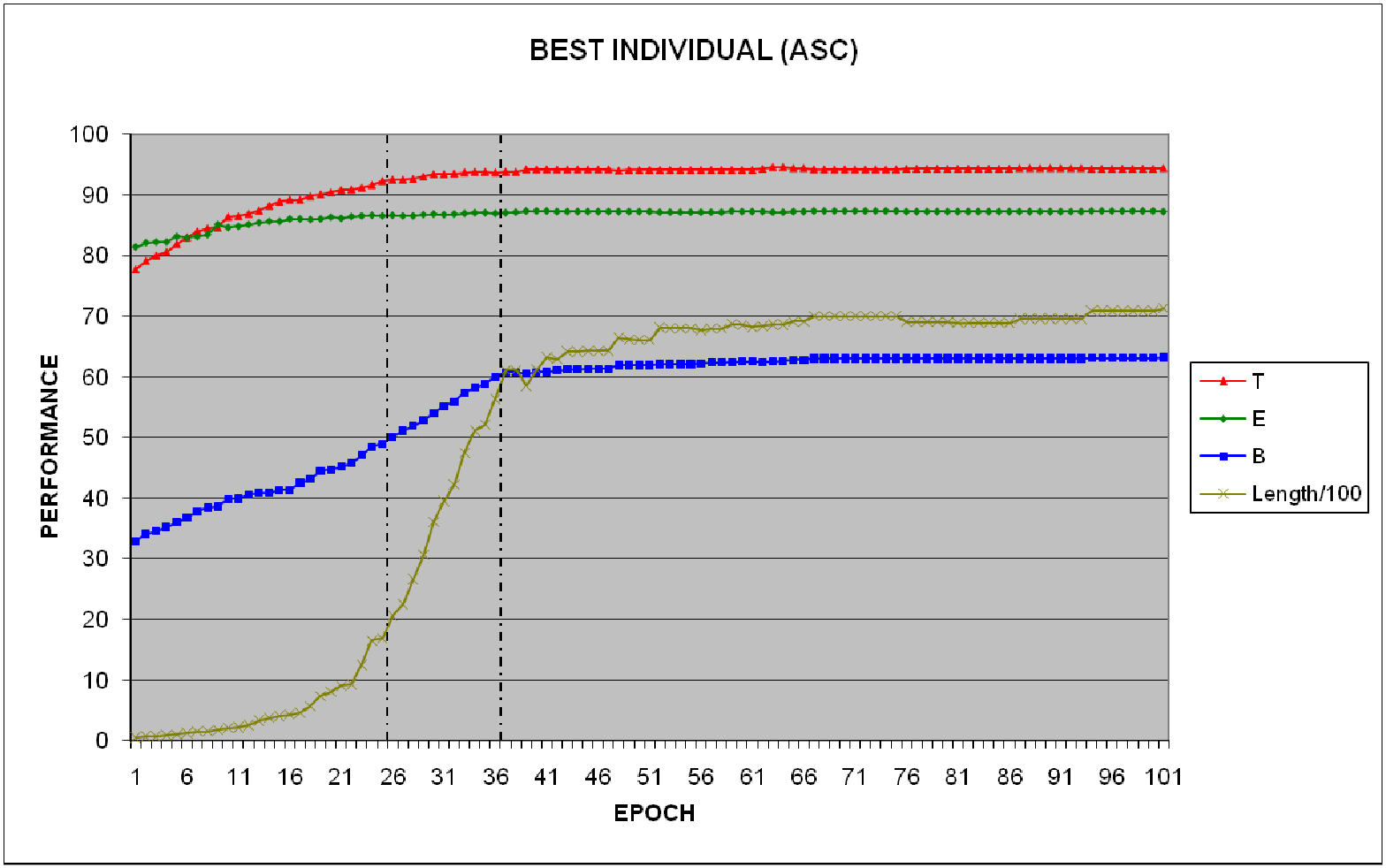
Results – Cumulative

		E (%)	B (%)	T (%)	Total Tests	Peak Value	Instr. At Peak
DSC	Evo	87	98	97	29	16	117326
	Rnd	89	99	98	2448	20	66880
	FD	89	99	93	1021	39	130416
ASC	Evo	88	98	97	25	20	141809
	Rnd	89	99	98	2740	61	228323
	FD	89	99	98	887	72	269496
BIRTH	Evo	88	98	96	18	11	82895
	Rnd	89	99	98	2373	126	538902
	FD	89	99	98	1644	65	278005

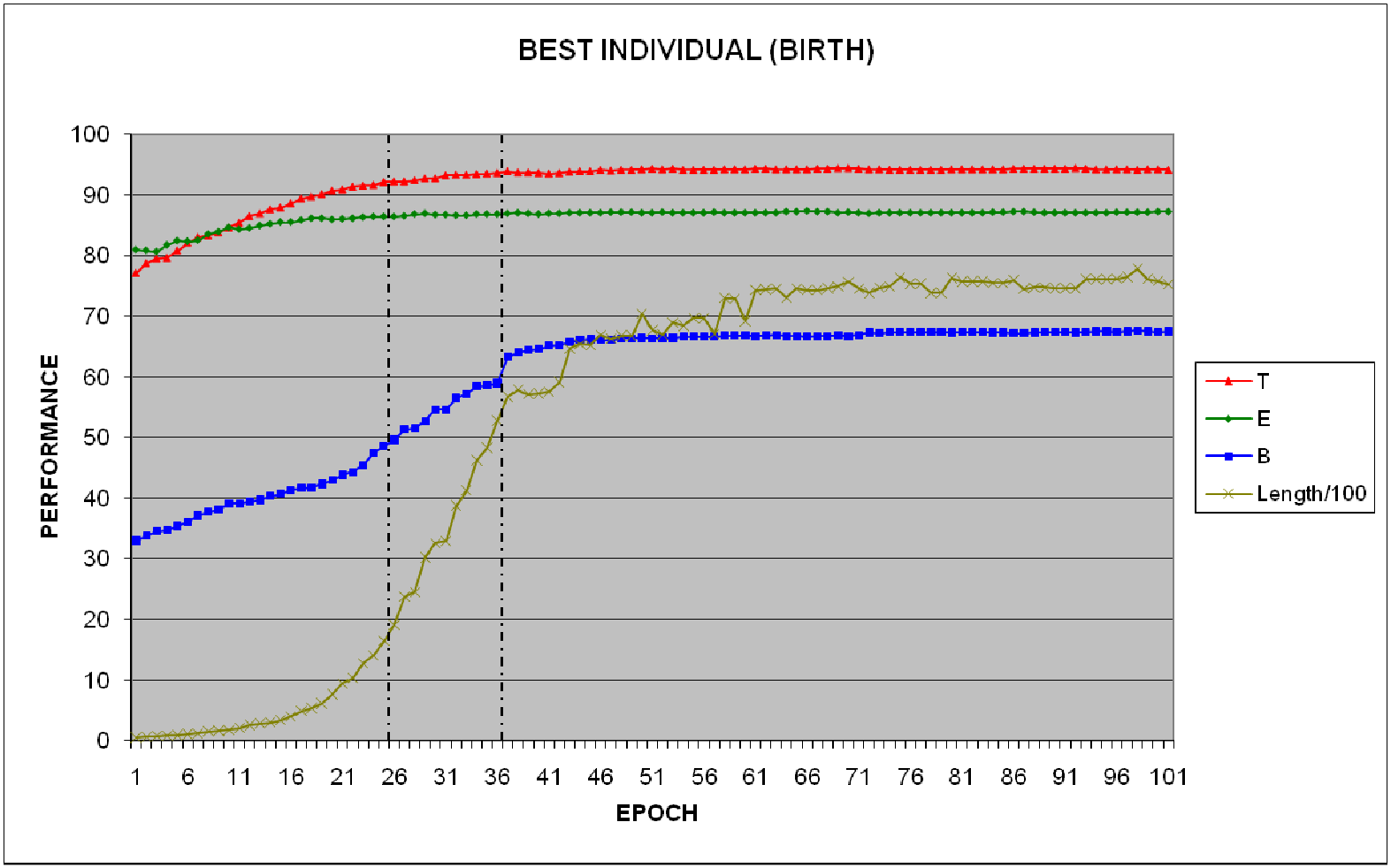
Results – Progression (DSC)



Results – Progression (ASC)



Results – Progression (BIRTH)



Academic Evaluation

- Non-optimal Evo behaviour
  - Lack of test diversity
  - Raised level of Abstraction using FD
- Random better due to greater code diversity
- FireDrill better as regression test suite than stand-alone test generator
- Little to no domain knowledge is required to achieve decent results (as in this research)
- uGP can act both as a stand-alone TG and a guide for another TG (like FireDrill)

Commercial Evaluation

- Implementation Effort:
  - Overall Implementation effort 5 man months
- Computational Effort:
  - Evo Runtime – Range 72-96 hours. Average ~76 hours for 100 epochs
  - Rnd Runtime – Range 18-24 hours. Average ~22 hours
  - FireDrill – Range 4-5 hours

Future Improvements

- Achieve better diversity of evolved tests
  - Use alternative GP selection strategies
  - Use tailored GP operators
- When guiding a TG, maximise GP potency by having 1-1 mapping between genotype and phenotype
- GP evolves single best solution. DV benefits from multi-path coverage

THANK YOU

Any Questions?

Engineering and Physical Sciences
Research Council