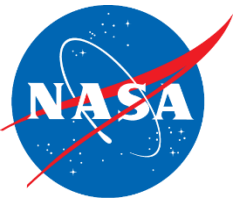


Data Mining in Computational Materials

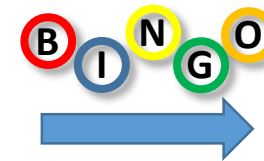
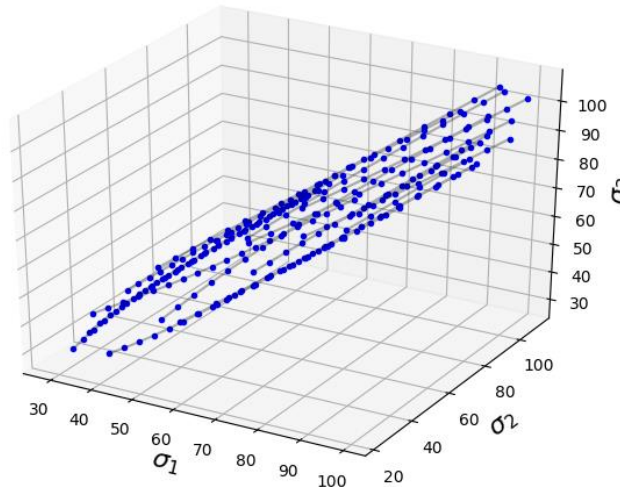
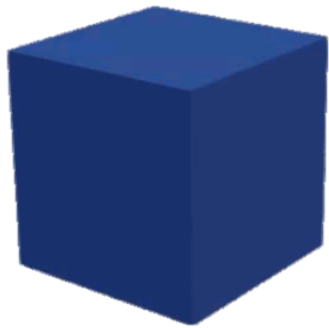
Tyler S. Townsend, University of Central Florida

Geoffrey F. Bomarito, NASA Langley

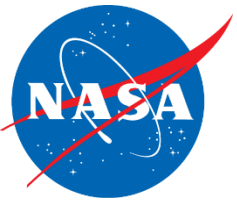


Motivation

- Push for certification by analysis of structural materials which relies on computational models and formulation
- Developing empirical equations is costly
- Proposed Solution: Self Formulating Computational Material Model
 - Bingo

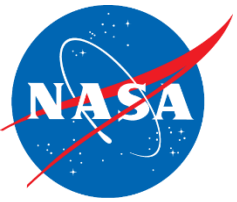


$$\Phi = f(\sigma)$$



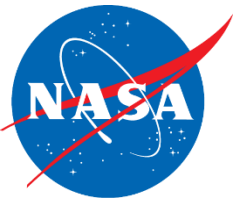
Outline

1. Symbolic regression and genetic programming basics
2. Verification of symbolic regression of plastic potentials
 - Von-Mises Equation
 - Effective Plastic Strain
3. Understand how much data is needed
4. Scaling study on processors
5. Fitness selection criteria



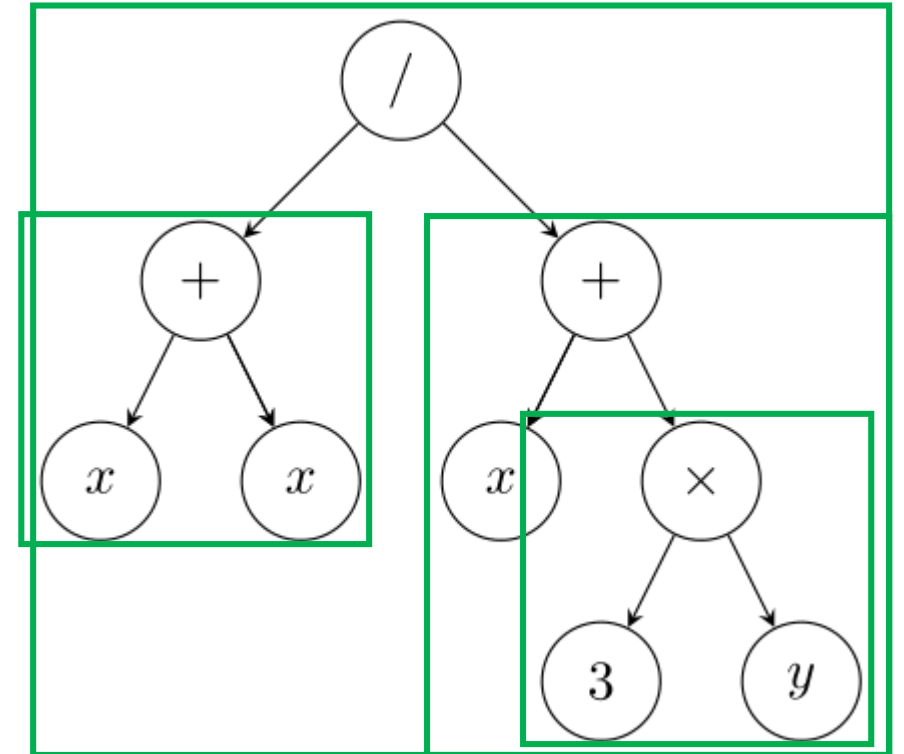
Genetic Programming

- An evolutionary computation (EC) technique
 - solves problems without knowing true form
 - based on how systems evolve
- Genetic operations
 - **Crossover:** Randomly combine parts of programs
 - **Mutation:** Randomly alter a part of a program
- Survival based on Fitness



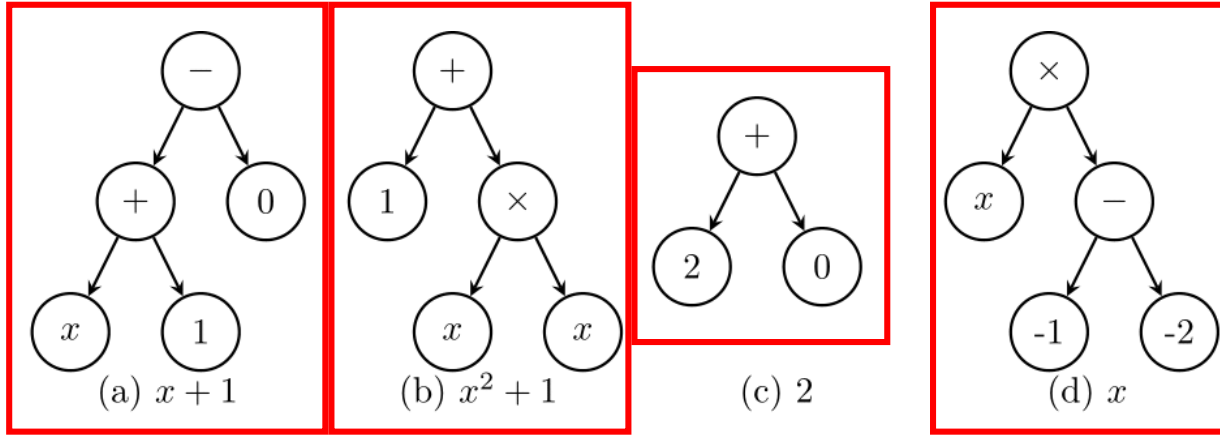
Symbolic Regression

- Method based on Genetic Programming
- Search for coefficients and functional form of equations
- Utilizes acyclic graph representation of functions

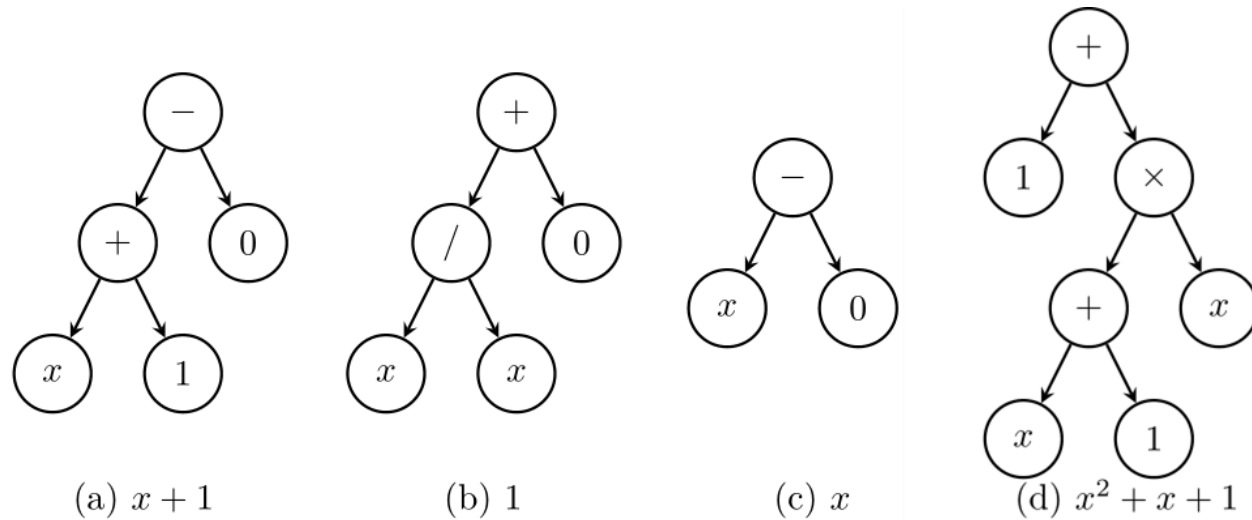


$$(x + x) / (x + (3 \times y)) = \frac{2x}{x + 3y}$$

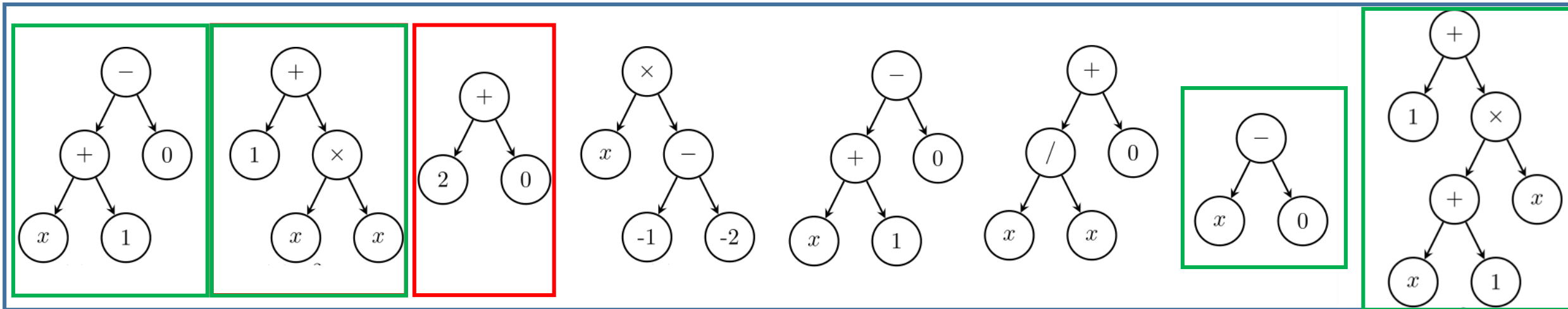
(Population of Solutions) Generation k



Children of k

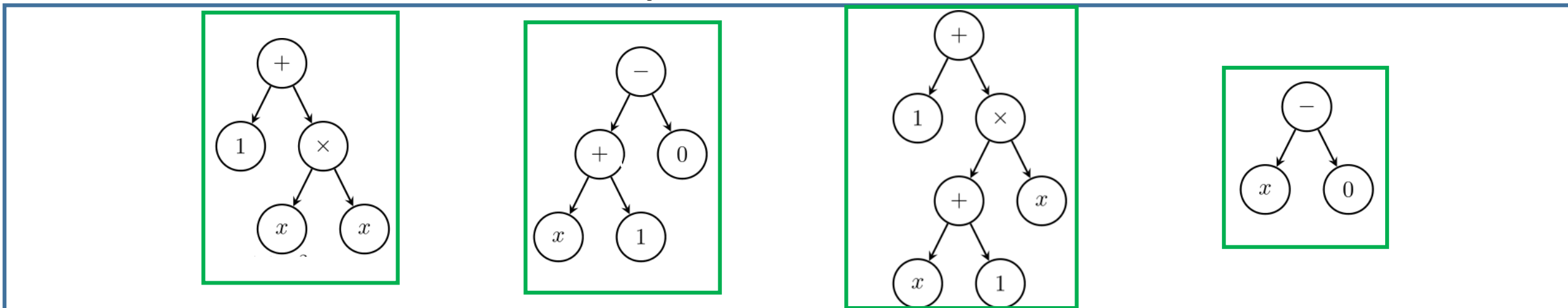


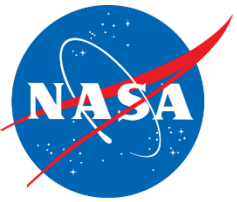
Candidate Solutions



Evaluate fitness from some training data for each Solution

Solution Population of Generation $k + 1$





Explicit vs Implicit Regression

- Labeled vs Unlabeled Data
- Supervised vs Un-Supervised Learning

Explicit Solution

$$y = f(x)$$

$$y = \sin(x)$$

$$y = f(x_1, \dots, x_{n-1})$$

Implicit Solution

$$f(x, y) = c$$

$$x^2 + y^2 = r^2$$

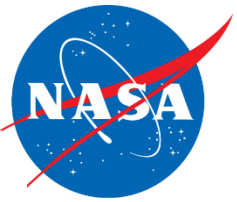
$$f(x_1, \dots, x_n) = c$$

Explicit Fitness Metric

$$\sum |f(x_1, \dots, x_{n-1}) - y| = 0$$

Implicit Fitness Metric

$$\frac{\Delta f}{\Delta t} = \sum_i \frac{\frac{\Delta x_i}{\Delta t} \cdot \frac{\partial f}{\partial x_i}}{\left\| \frac{\Delta x_i}{\Delta t} \cdot \frac{\partial f}{\partial x_i} \right\|} = 0$$



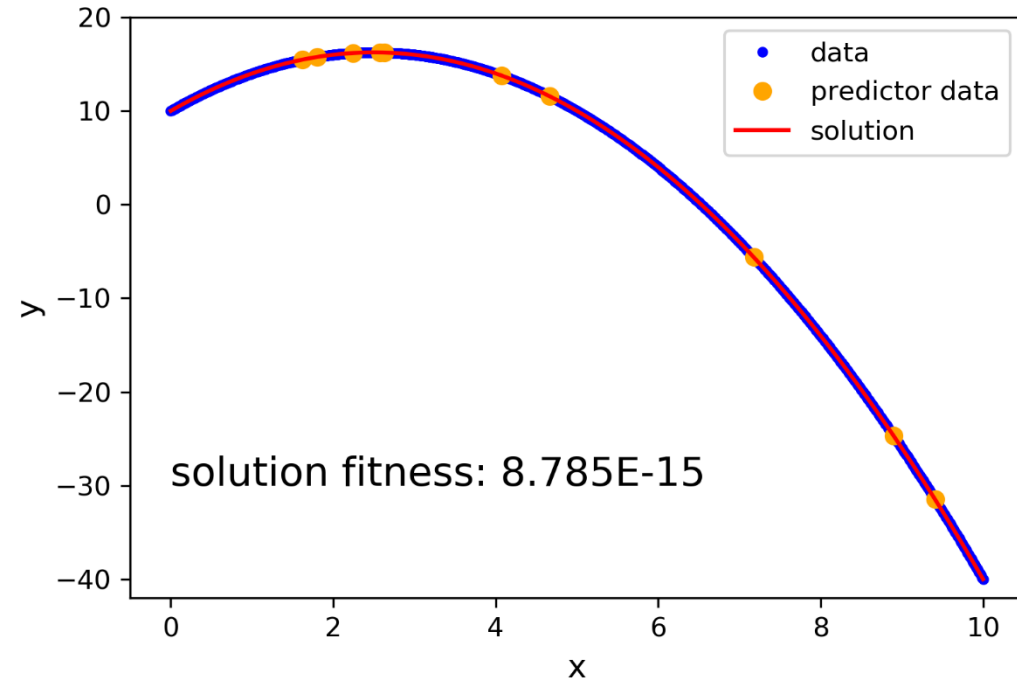
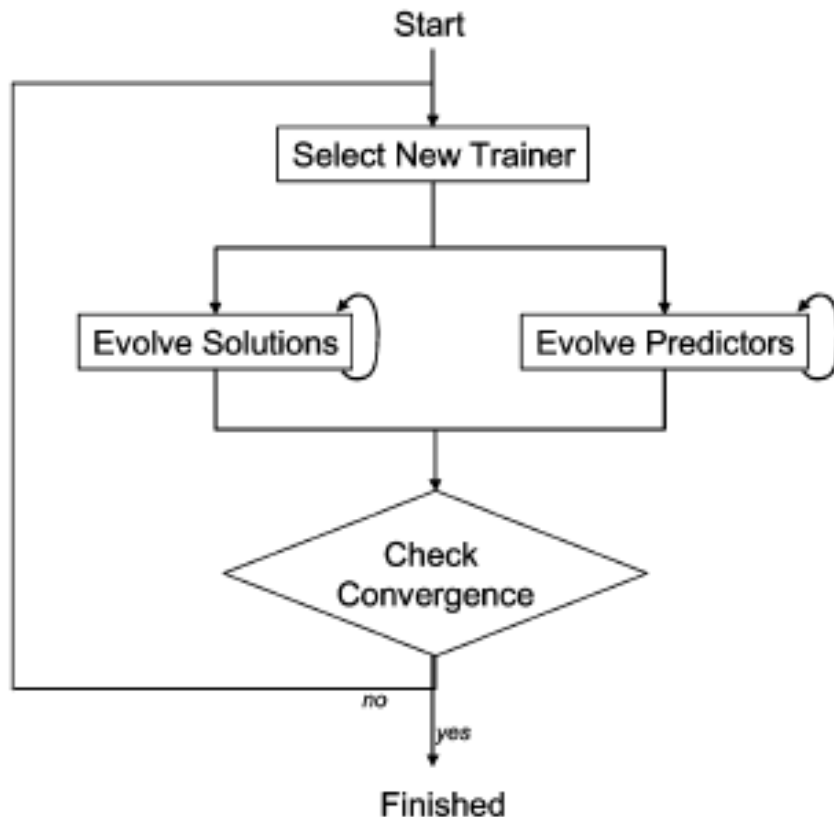
Co-Evolution

Predictors

Subset of training data

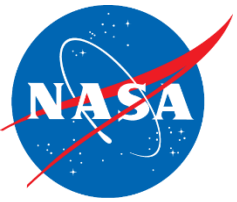
Trainers

Subset of solutions



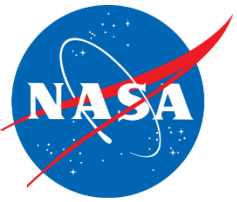
“Individuals that are known to be inferior still have some chance of being selected.”

- A Field Guide to Genetic Programming



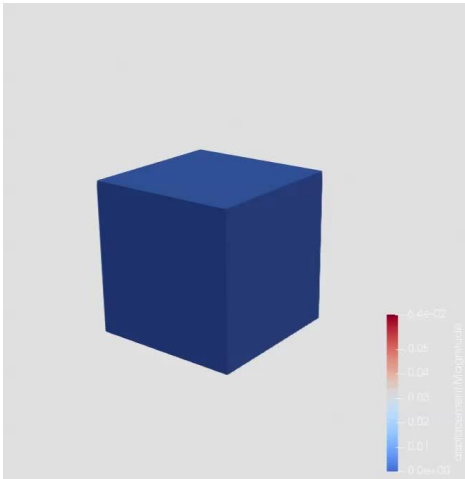
Wisecrack Analysis

- Self-deprecating humor
 - Louis C.K.
 - John Mulaney
 - Hannibal Burrell
- Not taking yourself *too* seriously is appreciated.
- Related selection of individual to genetic operations to receiving a position this summer
 - Mutated over summer?

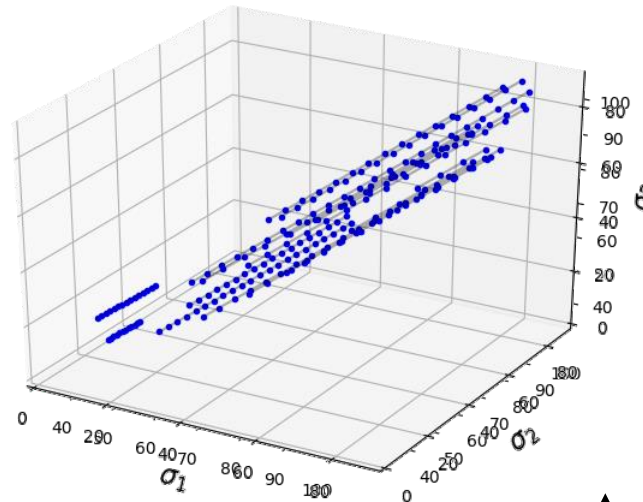


Data Generation for Von-Mises Yield Surface

Loading ratio: [0.675 1.0 0.8250]



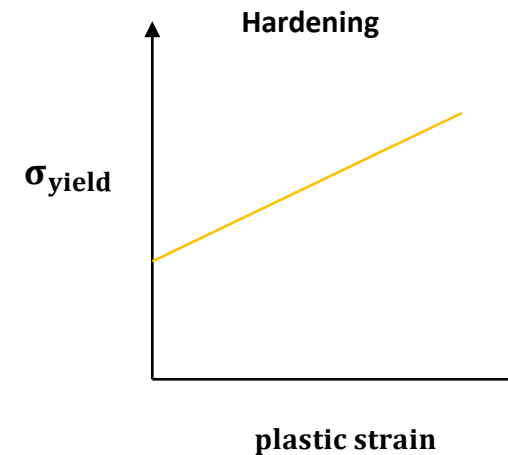
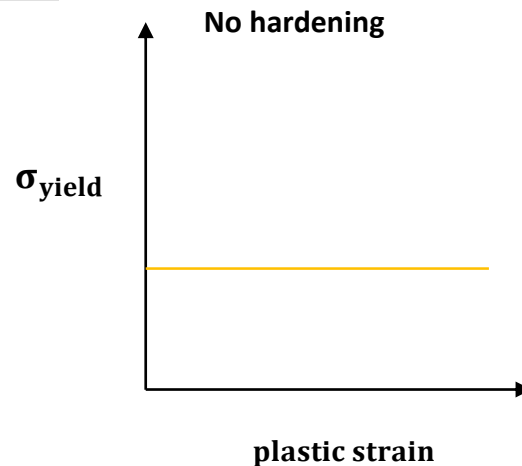
elastic
plastic

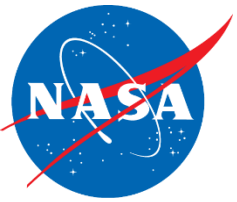


parameters

σ_i
 $\bar{\epsilon}_p$
 ϵ_i

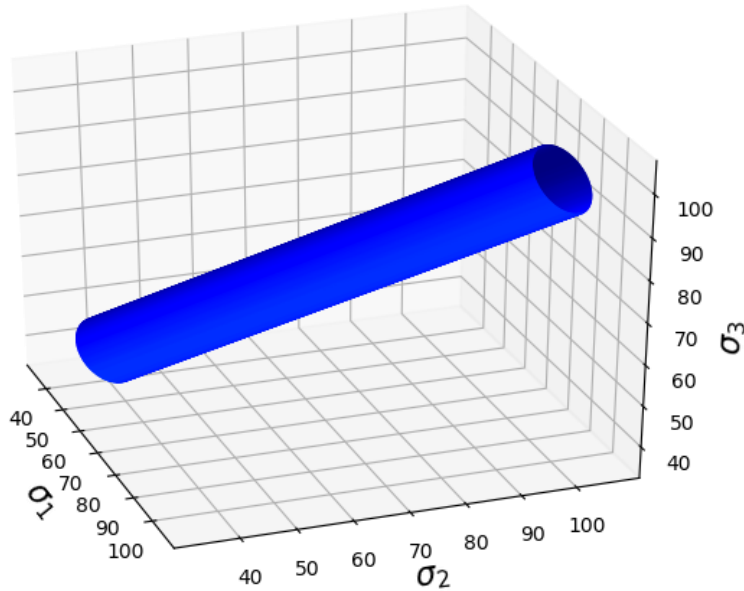
$$\Phi = f(\sigma)$$





Von-Mises Plasticity

Non-Hardening Case

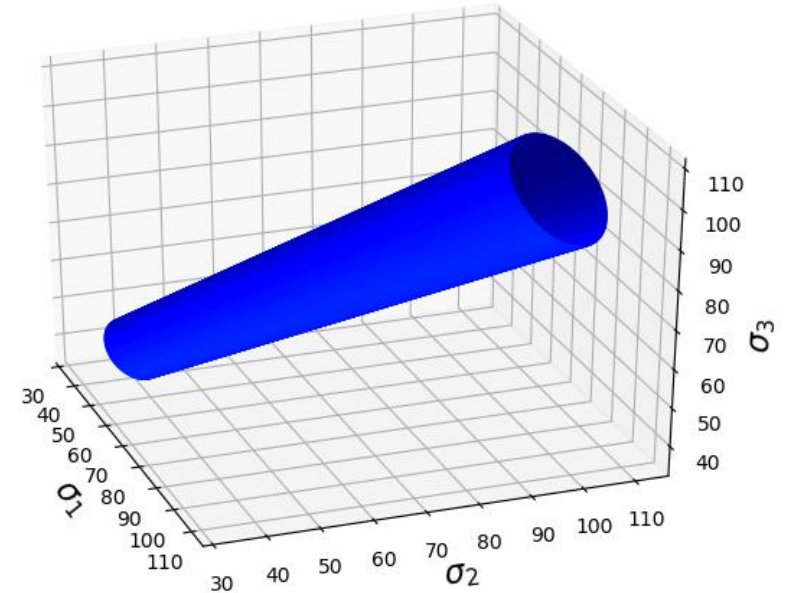


$$\Phi = (\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 - c = 0$$

Program terminates when solution is in desired fitness

Runtime: 32 seconds

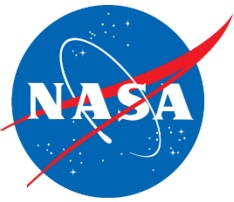
Hardening Case



$$\Phi_H = (\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 - c(\bar{\epsilon}_p) = 0$$

$$c(\bar{\epsilon}_p) = 2(c_1 + c_2 \bar{\epsilon}_p)^2$$

Runtime: 1.9 hours



Effective Plastic Strain Equation

Implicit Form

$$\frac{2}{9} [(\dot{\epsilon}_1 - \dot{\epsilon}_2)^2 + (\dot{\epsilon}_2 - \dot{\epsilon}_3)^2 + (\dot{\epsilon}_3 - \dot{\epsilon}_1)^2] - \dot{\epsilon}_p^2 = c$$

Trivial Solution

$$\dot{\epsilon}_1 + \dot{\epsilon}_2 + \dot{\epsilon}_3 = 0$$

Simplified Form

$$\frac{4}{3} (\dot{\epsilon}_1^2 + \dot{\epsilon}_1 \dot{\epsilon}_2 + \dot{\epsilon}_2^2) - \dot{\epsilon}_p^2 = c$$

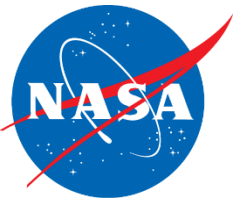
Output	$-4.006 \dot{\epsilon}_1^2 - 4.006 \dot{\epsilon}_2^2 + 3.006 \dot{\epsilon}_p^2 + \dot{\epsilon}_2(0. \times 10^{-11} \dot{\epsilon}_p - 4.006 \dot{\epsilon}_1) + 74.21$
---------------	--

Simplified Output	$-4 \dot{\epsilon}_1^2 - 4 \dot{\epsilon}_2^2 + 3 \dot{\epsilon}_p^2 + 4 \dot{\epsilon}_1 \dot{\epsilon}_2 + 74.21$
--------------------------	---

Runtime	1.8 minutes
----------------	-------------

Stress (Non-Hardening)
32 seconds

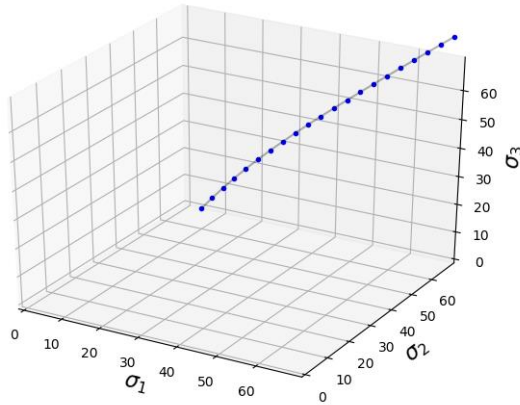
Stress (Hardening)
1.9 hours



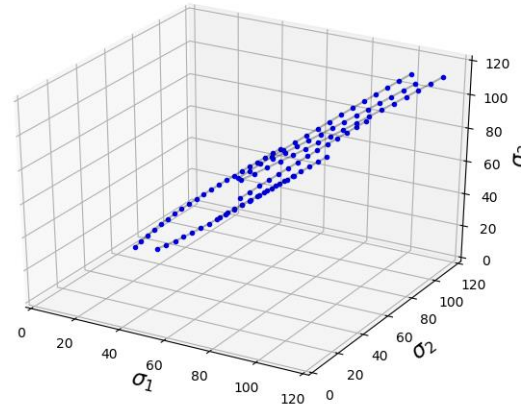
Data Variations on Non-Hardening Data

Scenarios

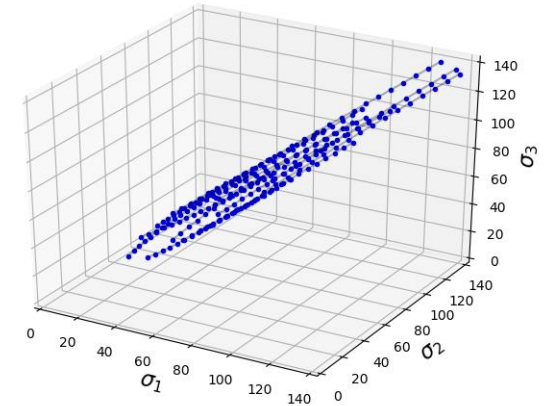
$N = 1$



$N = 8$

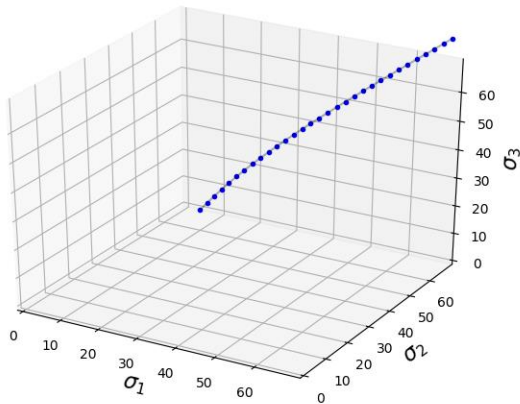


$N = 20$

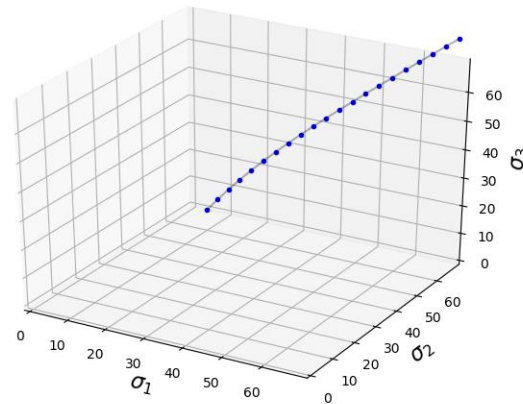


Steps (Strides through data)

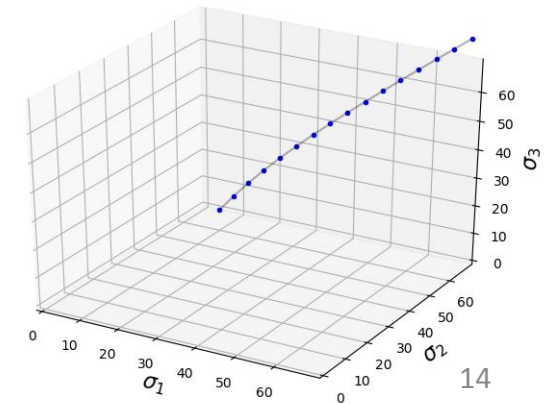
$n = 2$

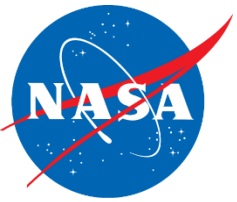


$n = 8$



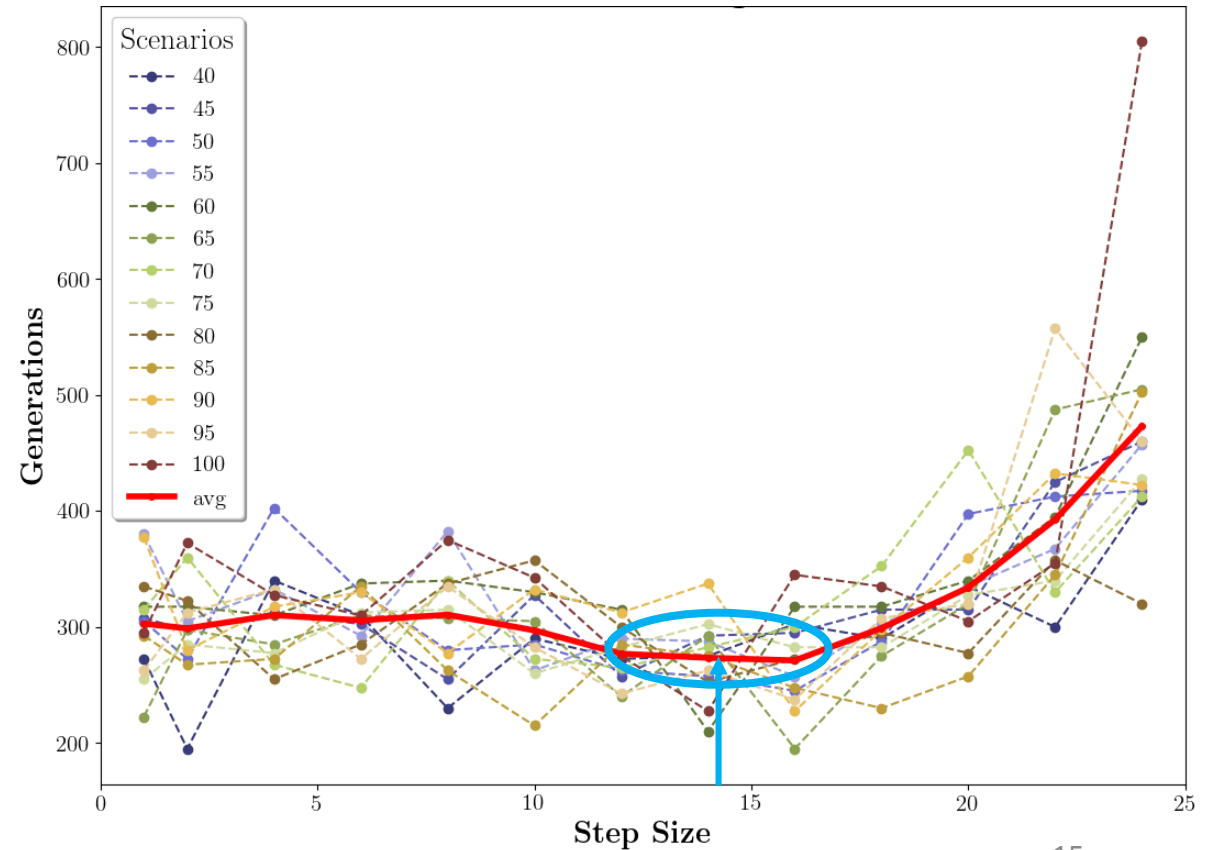
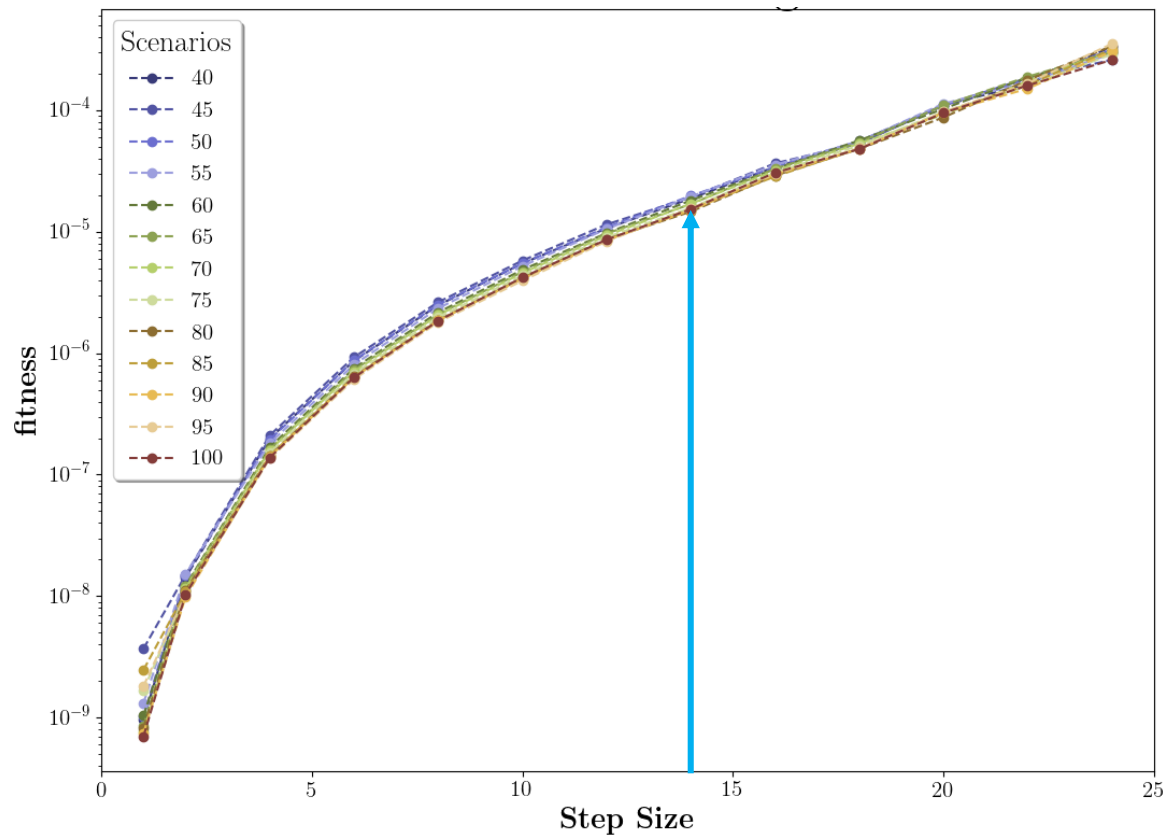
$n = 16$

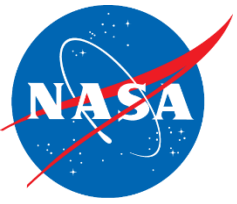




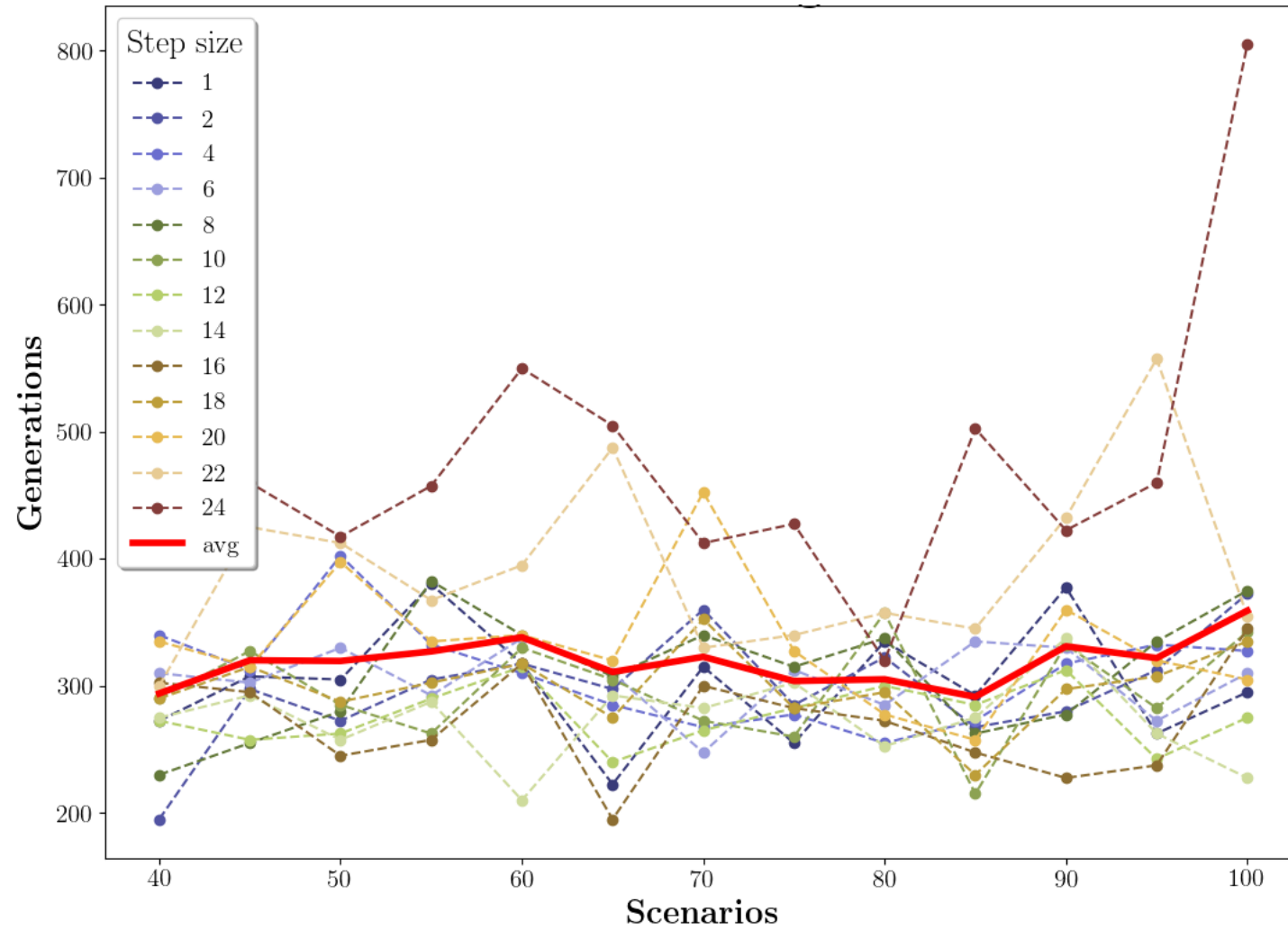
Step Size

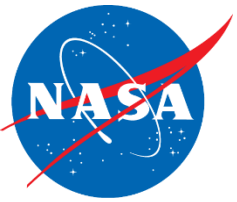
- Varied Step size and Scenarios on Non-hardening data
- Each test was run 20 times





Scenarios





Varying Solution Population Size

Bingo Test

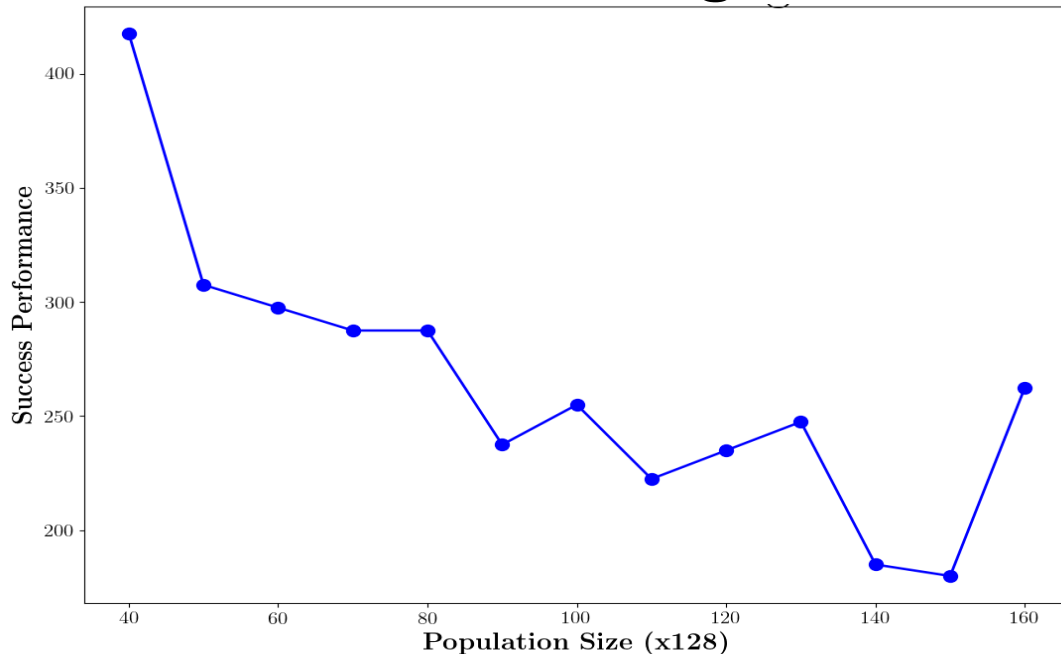
Varied Cores (128 Solutions per core)

Scenarios: 70

Steps: 12

Maximum generations to find solution: 175000

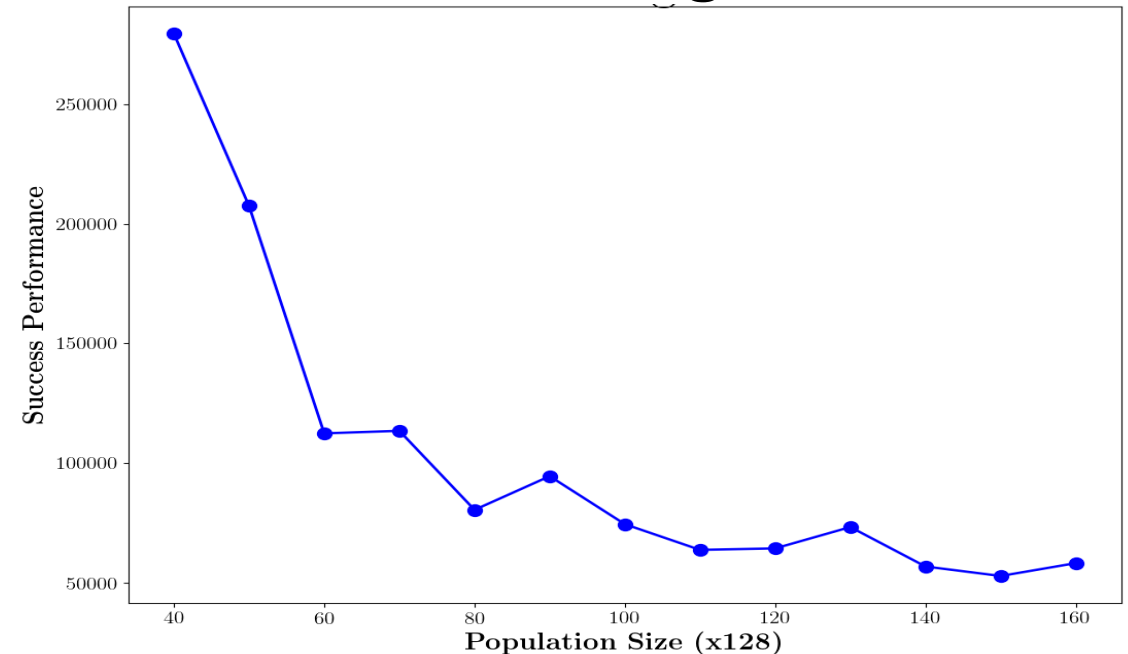
Non-Hardening Case

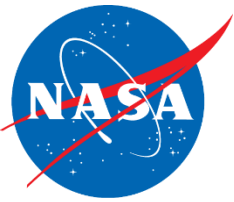


Success Performance

$$SP = \frac{\text{mean number of evaluations} \times \text{maxRuns}}{\text{number of successful runs}}$$

Hardening Case





Fitness Selection Criteria

Deterministic Crowding: Parents compete against children

Age Fitness: Random competition

Run Bingo on Hardening Data

Test Parameters

160 Core

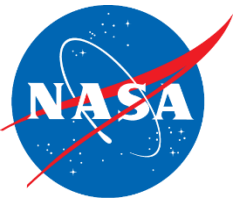
70 Scenarios

12 Steps

250,000 Maximum
Generations

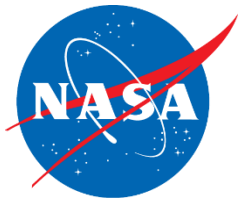
72 Hour Runtime

	Age Fitness	Deterministic Crowding
Correct Solutions	20	14
Finished Runs (out of 20)	20	16
SP (Generations)	76000	90000
SP (Hours)	1.9	4.3



Summary

- Verification of Von-Mises Stress and EPS
- Larger steps through data sets decrease maximum generations to obtain acceptable solution to data
- Population size greatly influences the performance of the algorithm for complex solutions
- Age Fitness selection criteria is the better than Deterministic Crowding



Questions?
