

# Evolving AVX512 Parallel C Code using Genetic Programming

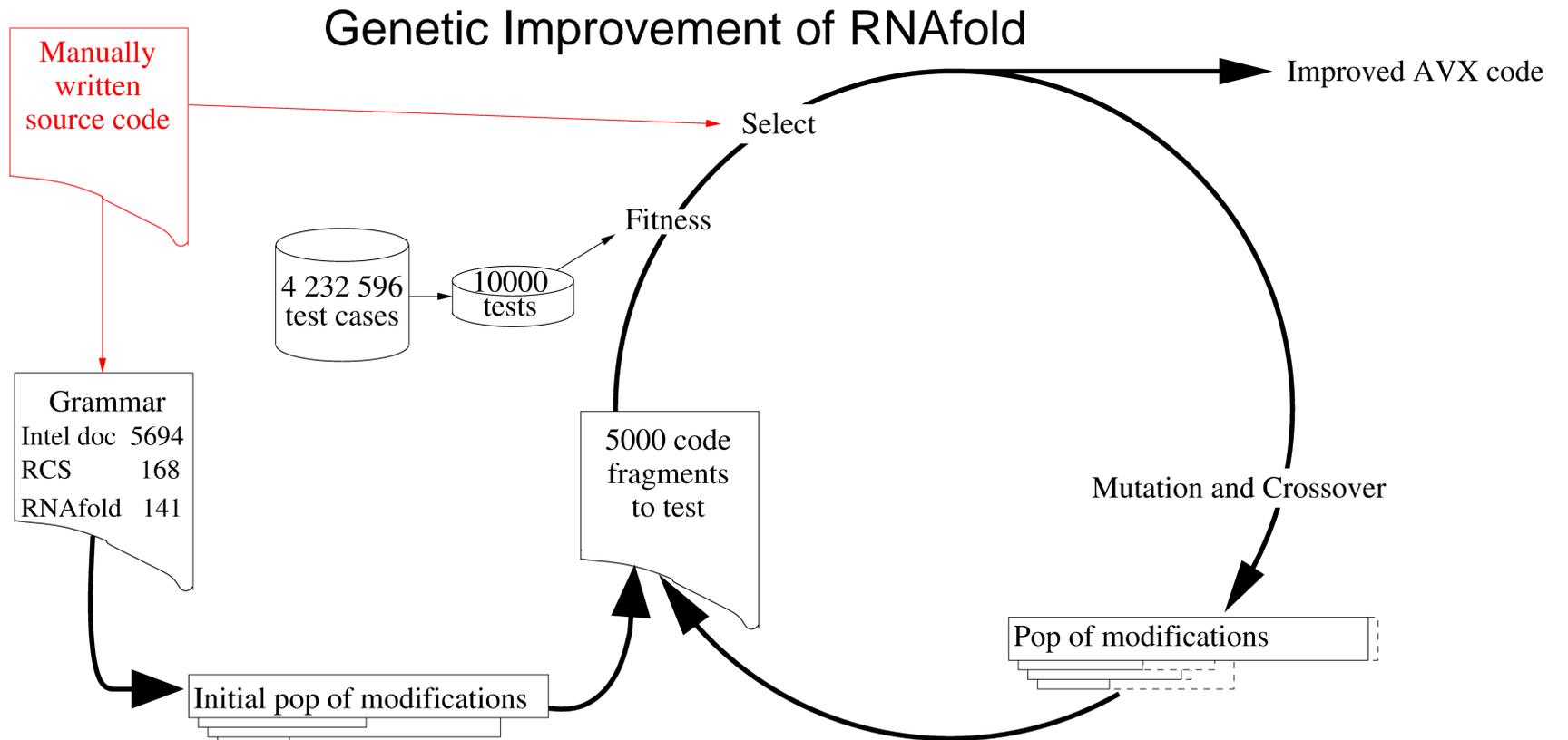


Figure 1: GISMO Evolutionary Cycle

## 1 Why

RNA is an important biomolecule. It is transcribed from DNA and translated into proteins, but is also important elsewhere, e.g. in gene regulation. Structure plays a large role in chemistry. RNAfold is the state-of-the-art predictor. It uses dynamic programming to find the optimal shape. Runtime is dominated by the inner dynamic programming cell update loop. The inner loop can be run in parallel.

## 2 Previously CUDA and Intel SSE

Running on nVidia GPUs, tech report RN/18/02. Convert to 128 bit Intel vectors, GI-2017.

## 3 Now 512 bit Intel AVX

Take existing SSE code, revision history, Intel documentation and test suite. Evolve 6.1 faster than sequential code. GI code incorporated into ViennaRNA package since Dec 2018.

## 4 RNA STRAND

Test on 4666 known RNA structures.

## 5 GGGP to improve RNAfold

**Representation:** variable list of replacements, deletions and insertions into BNF grammar (6122 rules).

**Fitness:** Compile (GCC 7.3.1), run on 10 000 random test cases. Two objectives: number of tests past and elapsed time.

**Population:** 5000, panmictic, elite 10, generational.

**Parameters:** Initial population of random single mutants. 50% truncation selection. 50% two point crossover, 50% mutation. No size limit. Stop after 50 generations.

## 6 Fitness Function

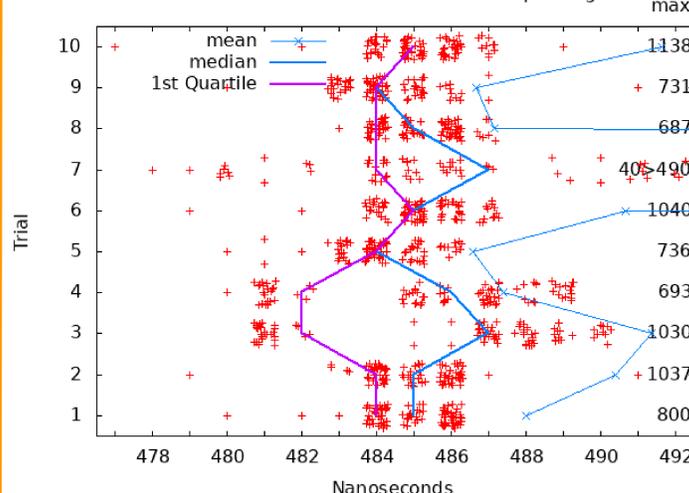
Before evolution Dynamic Programming cell update code called 4,232,596 times and inputs and results logged. Every generation 10,000 randomly chosen and GI code tested on them.  
Fitness<sub>a</sub> = does mutated code give same answer?  
Fitness<sub>b</sub> = how long does it take?

Using real (i.e. noisy) run time, so run 101 times and take first quartile.

## 7 Why 1<sup>st</sup> Quartile run time

	mean	SD	count
mean	491.387	19.9727	1000
1st quartile	483.028	2.40899	1000
median	485.085	2.43141	1000

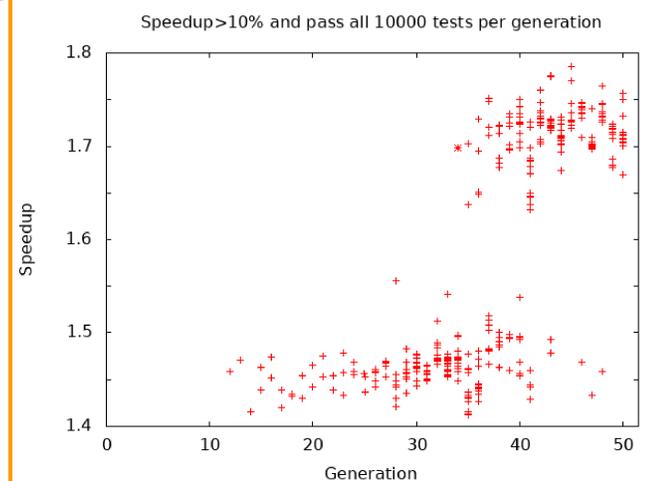
10 trials each with 101 runs of one example long test



## 8 Solution

Best individual in population at generation 34 contained 8 mutations, hill climbing leaves 3:  
•vecsize=16 converts generic code to 512 bits  
•Line 8506 of the Intel documentation is `_mm512_reduce_min_epi32`, so `en=_mm_cvtsi128_si32(min4)` becomes `en=_mm512_reduce_min_epi32(min4)`.  
•The remaining mutation increases efficiency.

## 9 Evolution of fitness



## 10 Summary

On a standard Intel Xeon Gold 6126 2.60GHz server, with an unmodified version of Linux, evolution can optimise a critical function written in C. Automatically creating AVX instructions, genetic improvement gives almost a doubling in speed (1.71 fold) on top of hand written SSE instructions (6.1 times the sequential code). This GI code in ViennaRNA since release 2.4.11.

## 11 Code

Reference: W.B. Langdon and R.Lorenz. Evolving AVX512 Parallel C Code using GP. In L.Sekanina, T.Hu and N.Lourenco editors, EuroGP 2019, Leipzig. Springer Verlag.

GISMO

<http://www.cs.ucl.ac.uk/staff/W.Langdon/ftp/gp-code/rnafoldAVX.tar.gz>

Background predicted structure of RNA from Spinach 23S Ribosome CRW\_01456