William B. Langdon and Justyna Petke
Department of Computer Science, University College London (w.langdon@cs.ucl.ac.uk j.petke@ucl.ac.uk)

New Functionality via New Data

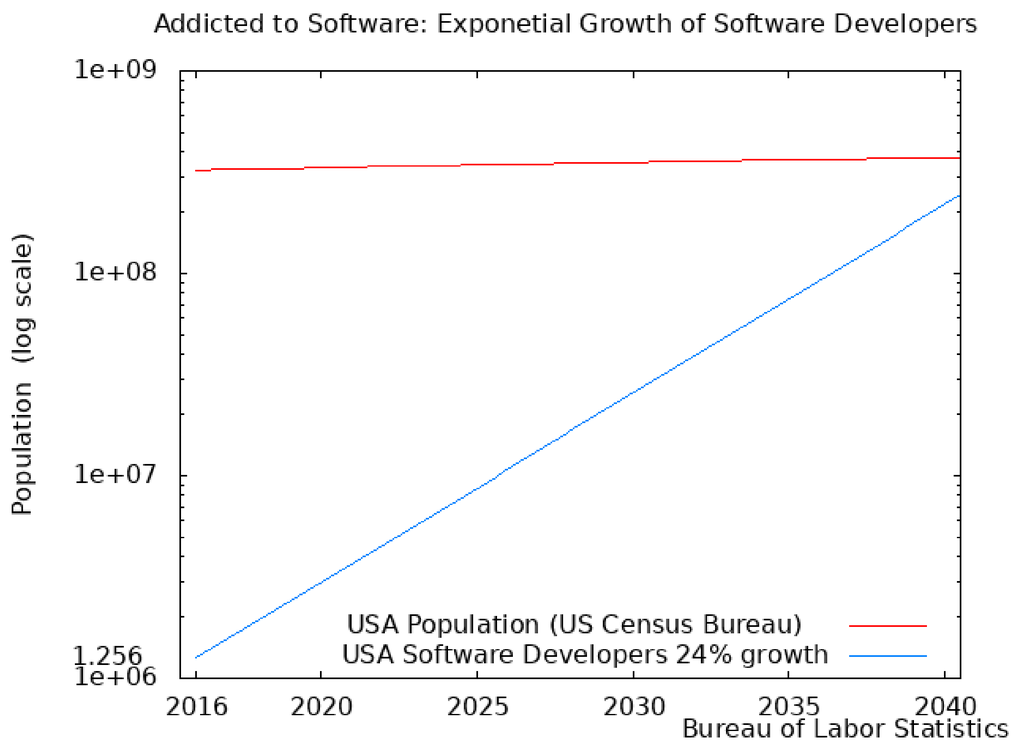

Figure 1: The world is addicted to software

1 Exponential growth of programmers is not possible

Fig. 1 shows the effect of extrapolating 24% growth in the number of software developers over the next 20 years. Clearly it is not feasible for 59% of the population to be full time software developers. Something has to change.

We propose increased automation of software development and maintenance.

2 How to automate software

Whereas Genetic Programming starts from a primodial ouze of random programs, Genetic Improvement starts from human written code.

3 Genetic Improvement via data

GI often mutates the program's source code. Here we use CMA-ES to evolve **data** (albeit with some small code changes) to give a new function (**Fig 2**).

4 Evolve sqrt into cube root

Start from a GNU C library table driven sqrt function. The table contains 512 start points for iterative solution to double precision sqrt. Minimal code changes to support cube root (mostly to do with double precision exponent). Evolve table contents by running CMS-ES 512 times.

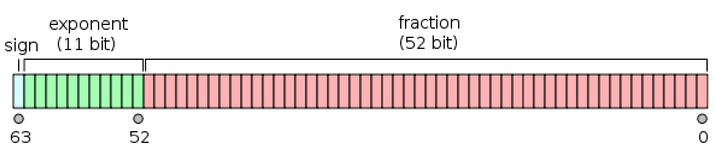

Figure 3: IEEE 754 Double precision number

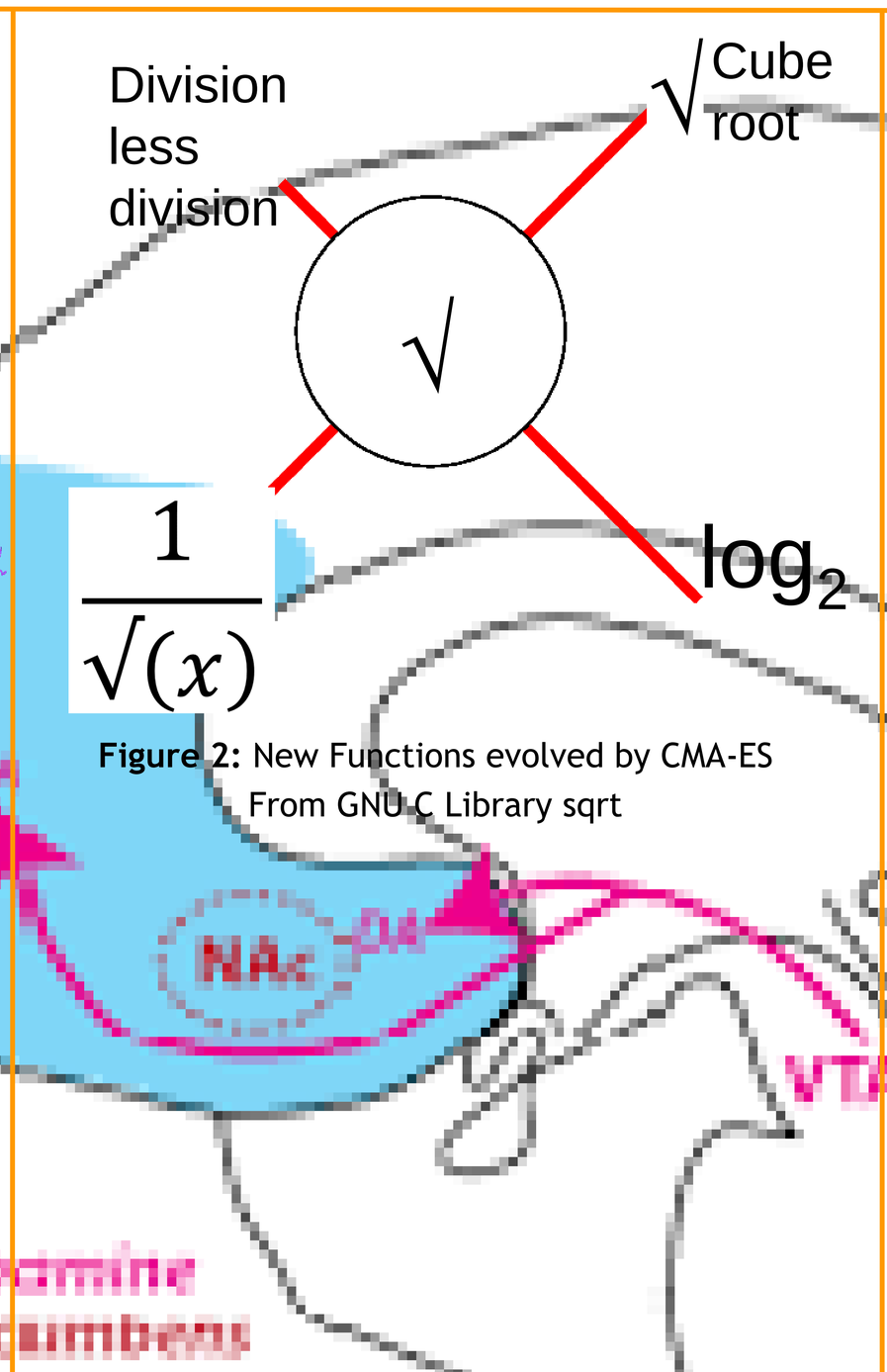

Figure 2: New Functions evolved by CMA-ES From GNU C Library sqrt

5 Fitness Function

Evolve table contents one at a time. For each we have 3 test points x: bottom, middle, top. So we have three subfitnesses, which are added together. For each x run code, cube its output. Fitness is absolute error between cube and input. To help CMA-ES |error| is log transformed.
Stop evolution when all three errors are zero.

6 Default CMA-ES parameters

All defaults except: force error to be zero before stopping, restart if fails to get perfect fitness.

7 Scientific Data RNAfold

GI on 50000 integers gave, on average, better predictions of the shape of RNA molecules, EuroGP 2018.

8 New Mathematical functions

Iterative technique (Newston-Raphson) relies on the new function being (in the range of interest) continuous and differentiable. It also need an objective measure of how close it is to the true solution.

Try evolving your favourite double precision function.

9 Code

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

William B. Langdon and Justyna Petke. Genetic Improvement of Data gives Binary Logarithm from sqrt. In GECCO 2019 Companion, 13-17 July, 2019, Prague, Czech Republic. ACM. https://doi.org/10.1145/3319619.3321954