CREST

Grow and Graft a better CUDA pknotsRG

for RNA pseudoknot free energy calculation

Genetic Improvement workshop GECCO 2015

12th July 2015

W. B. Langdon

Department of Computer Science

CREST

Grow and Graft a better CUDA pknotsRG

for RNA pseudoknot free energy calculation

W. B. Langdon

Department of Computer Science

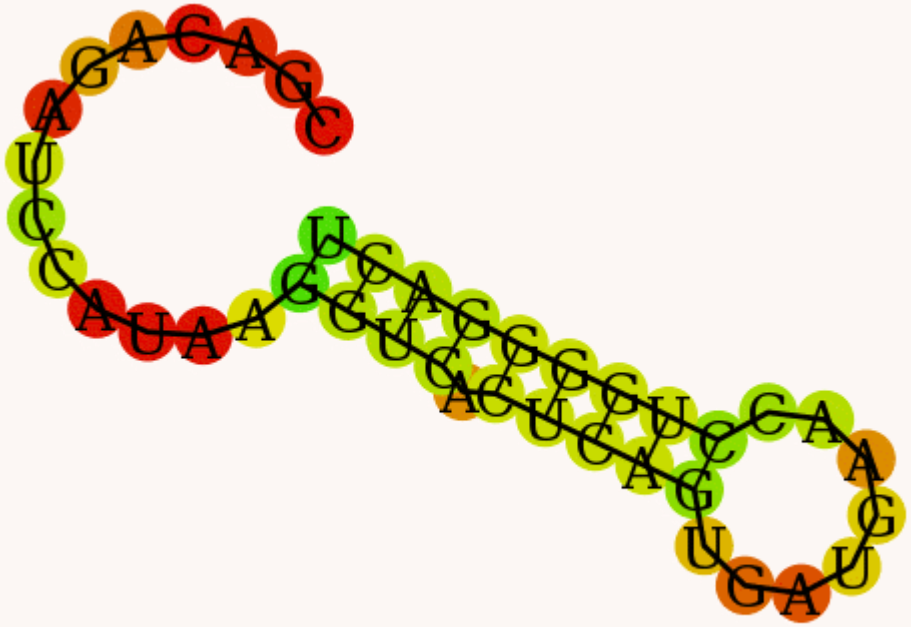
GGGP and CUDA pknotsRG

- CUDA pknots
  - Calculate shape of RNA molecules
  - 11 000 lines of C and CUDA code
- Grow and Graft Genetic Programming
- 10 thousand fold speedup

pknots: RNA sequence → folding

Input → CGACAGAUCCAUAAGGUCACUCAGUGAUGAACCUGGGGACU

Output → .....((((.(((((.....)))))))))

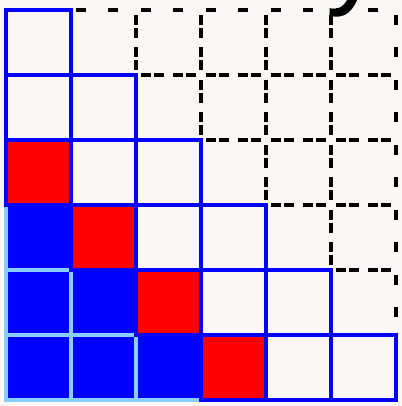


free energy of
-9.30 kcal/mol

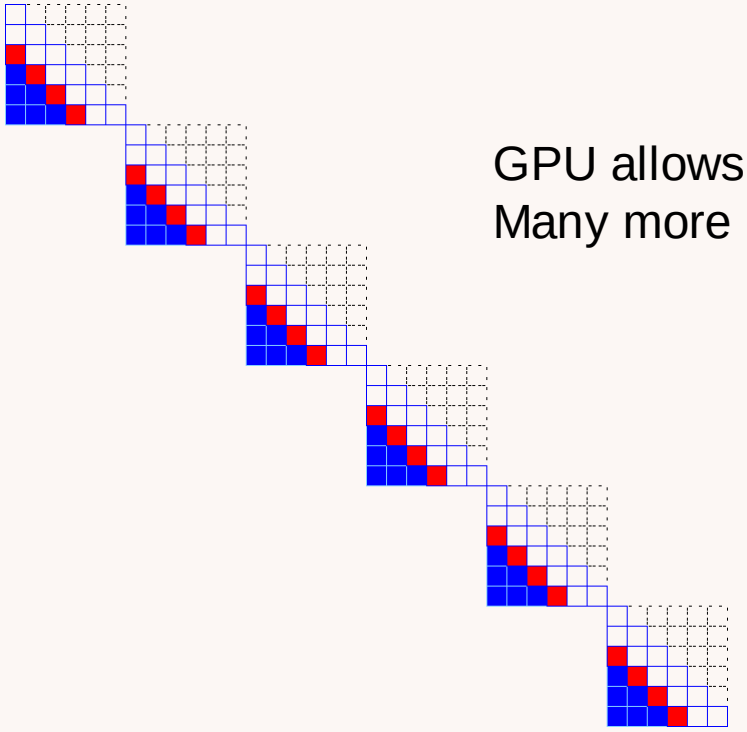
Dynamic programming

- pknots uses dynamic programming to find minimum energy of folding
- One molecule at a time
- Not enough parallelism
- Run Dynamic Programming algorithm on 200,000 matrices in parallel
- GGGP convert CUDA from 1 to n matrices

Dynamic programming



(n+1) by (n+1) matrix.
Only lower half used.
Active front can be calculated in parallel
(needs 1 to n+1 threads)

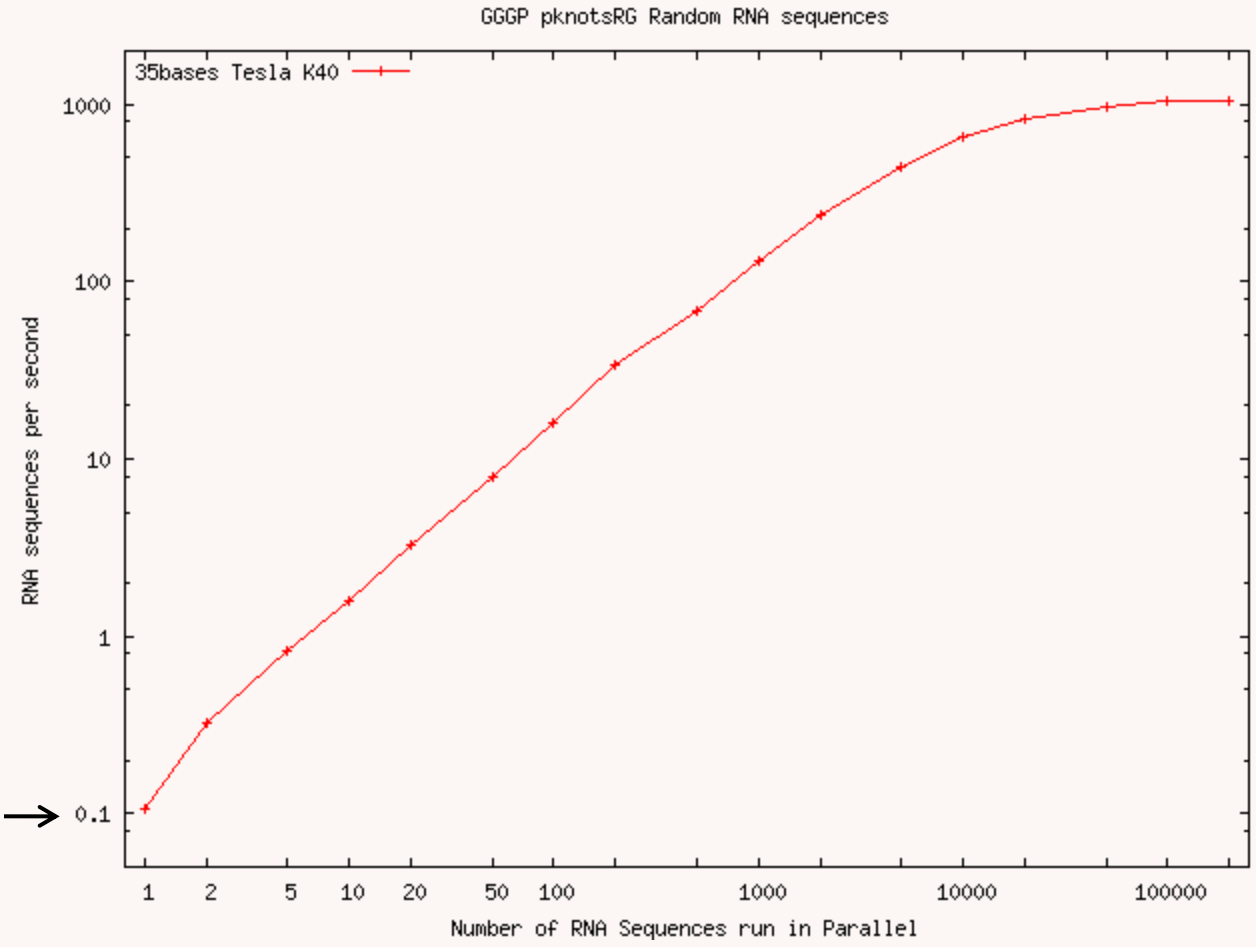


GPU allows many matrices to be calculated in parallel
Many more GPU threads can be used.

Grow and Graft Genetic Programming

- Manually written CUDA code automatically converted to BNF grammar
- GP uses grammar to create population of syntactically correct mutants.
  - All compile
  - Array and pointer protection
  - Time out infinite loops
- Fitness by testing and comparing against outputs and speed of original code

GGGP CUDA pknotsRG speed



Original →

Up to 10000 times faster

Conclusions

- Genetic programming and human work together.
- On 11 000 lines of C/CUDA code
- Give spectacular speed up

Talk GP3 Tuesday 11:35 Patio 2

END

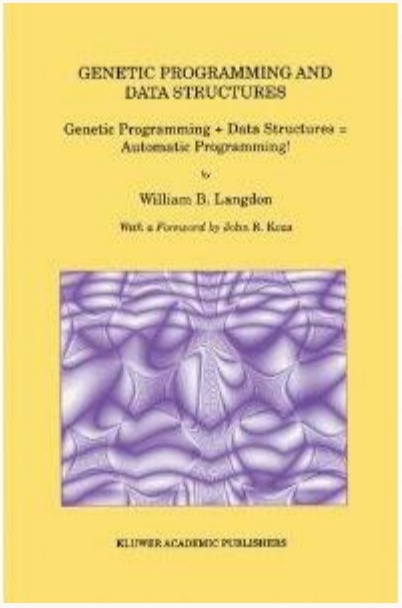
http://www.cs.ucl.ac.uk/staff/W.Langdon/ http://www.epsrc.ac.uk/ EPSRC

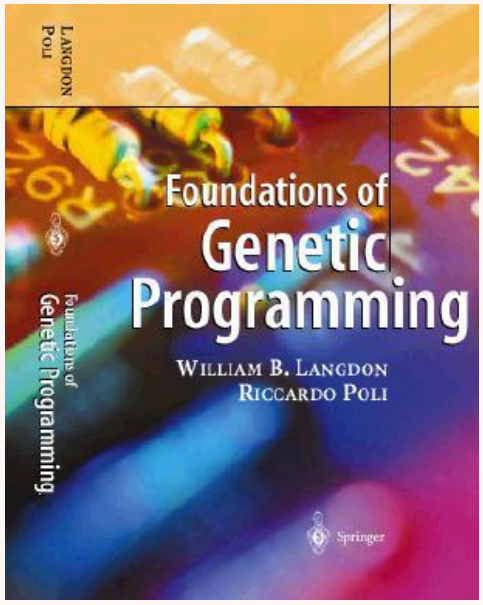
Genetic Improvement

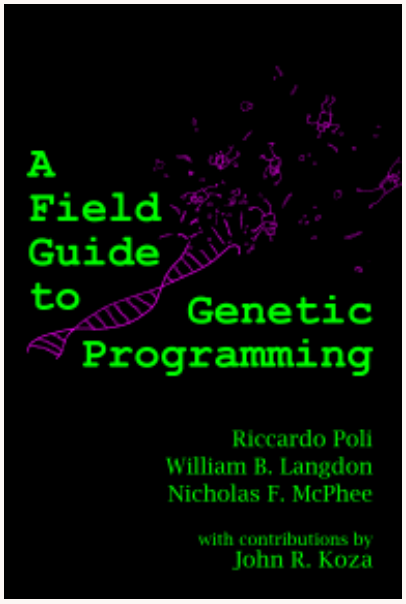
W. B. Langdon

CREST

Department of Computer Science



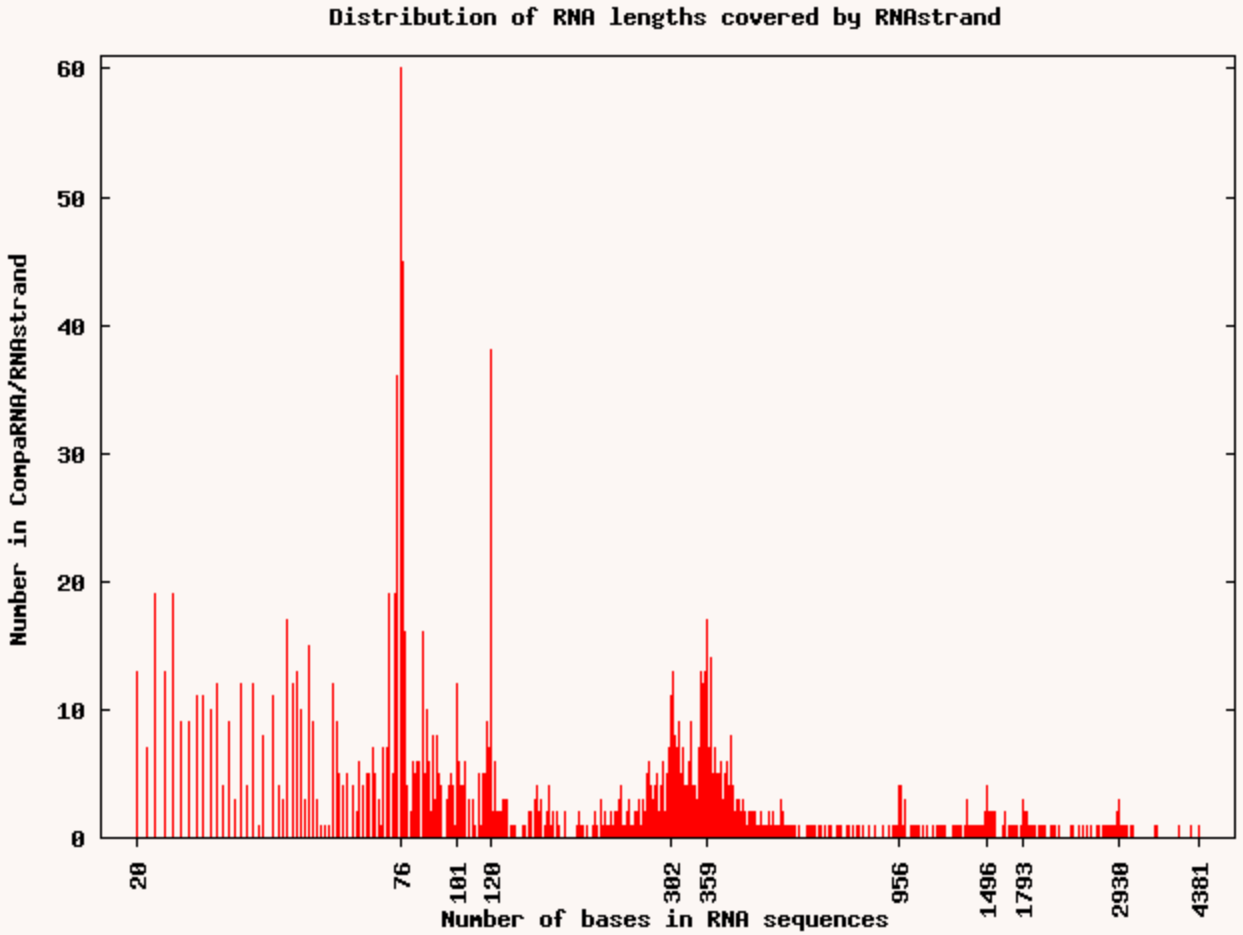




GPUs

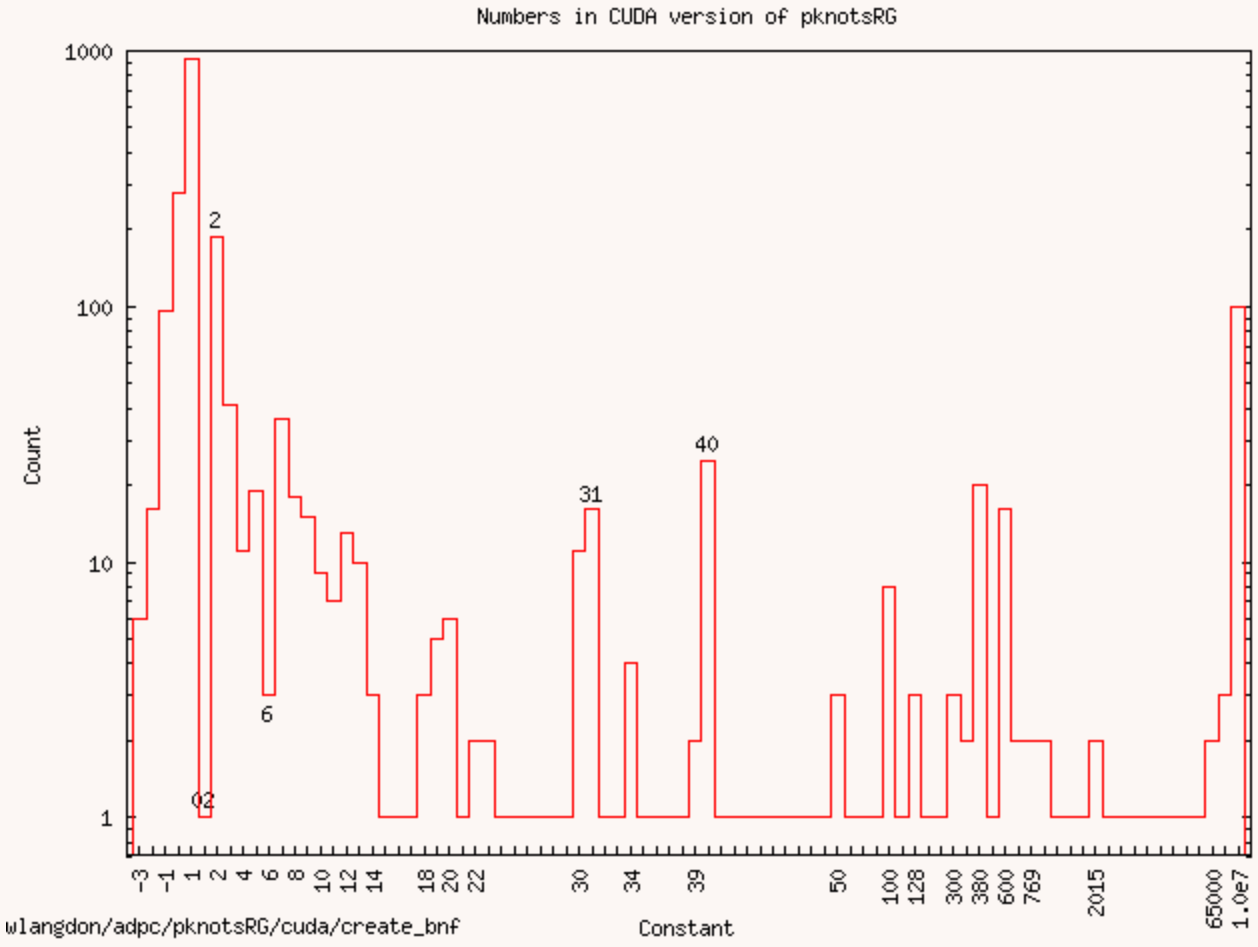
GPU	Total cores	clock	Bandwidth Giga Bytes/sec
Tesla K20	2496	0.71 GHz	140
Tesla K40	2880	0.88 GHz	180

Lengths of RNA in CompaRNA



Number of bases in RNA sequences

Constants in pknotsRG sources



Conclusions

- Genetic programming can automatically re-engineer source code. E.g.
  - hash algorithm
  - Random numbers which take less power, etc.
  - mini-SAT (Humie award)
- fix bugs (>10^6 lines of code, 16 programs)
- create new code in a new environment (graphics card) for existing program,gzip WCCI '10
- new code to extend application (GGGP) SSBSE'14
- speed up GPU image processing EuroGP'14 GECCO'14
- speed up 50000 lines of code IEEE TEC

10000 speed up GI-2015

The Genetic Programming Bibliography

http://www.cs.bham.ac.uk/~wbl/biblio/

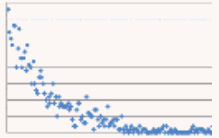
10318 references

RSS Support available through the Collection of CS Bibliographies. XML RSS

A web form for adding your entries. Co-authorship community. Downloads

A personalised list of every author's GP publications.

blog.html



Search the GP Bibliography at
http://liinwww.ira.uka.de/bibliography/Ai/genetic.programming.html