

RNA-Folding - From Hardness to Algorithms

Virginia Vassilevska Williams

Stanford University, USA

Abstract

This is an overview of the invited talk delivered at the 41st International Symposium on Mathematical Foundations of Computer Science (MFCS-2016).

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases RNA folding, matrix multiplication

Digital Object Identifier 10.4230/LIPIcs.MFCS.2016.5

Category Invited Talk

1 Overview of the Talk

A fundamental problem in computational biology is predicting the base-pairing of an RNA secondary structure. Most algorithms for this rely on an algorithm for a simplified version of this problem, RNA-folding, defined as follows: given a sequence S of letters over the alphabet $\{A, U, C, G\}$ where A can only be paired with U and C can only be paired with G , determine the best “folding” of S , i.e. a maximum size *nested* pairing of the symbols of S . For instance, in the sequence $ACUG$ the best pairing is either matching A with U , or matching C with G , but not both as that pairing wouldn't be nested.

A dynamic programming algorithm from 1980 by Nussinov and Jacobson [1] solves the RNA-folding problem on an n letter sequence in $O(n^3)$ time. Despite many efforts, until recently, the best algorithms for RNA-folding only shaved small logarithmic factors over this cubic running time. In this talk I will discuss our recent research on RNA-folding and related problems.

Our first result attempts to explain why it has been so difficult to obtain faster algorithms. We show that if one can solve RNA-folding on n length strings faster than one can currently multiply n by n matrices, then the Clique problem would have surprisingly fast algorithms. The current fastest algorithm to multiply n by n matrices runs in $O(n^{2.373})$ time and the fastest known Clique algorithms use this result. Obtaining an $O(n^{2.36})$ time algorithm for RNA-folding would thus be potentially difficult as it would imply a breakthrough for Clique algorithms and potentially also for matrix multiplication.

While this hardness result is appealing, it does not explain the seeming n^3 barrier. No better hardness seemed possible to us, and thus it became increasingly more plausible that RNA-folding should have a faster algorithm and in fact one using fast matrix multiplication. Indeed, this turned out to be true: we were recently successful in obtaining the first truly subcubic time algorithm for the problem. My talk will strive to give some insights into the hardness result and the new algorithm.

References

- 1 Ruth Nussinov and Ann B. Jacobson. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 77(11):6309–6313, 1980.



© Virginia Vassilevska Williams;
licensed under Creative Commons License CC-BY

41st International Symposium on Mathematical Foundations of Computer Science (MFCS 2016).

Editors: Piotr Faliszewski, Anca Muscholl, and Rolf Niedermeier; Article No. 5; pp. 5:1–5:1

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany