

# Sublinear Algorithms

17.08.08 - 22.08.08, Seminar 08341

organized by

A. Czumaj,  
S. Muthukrishnan,  
R. Rubinfeld,  
C. Sohler

## Abstract

This report summarizes the content and structure of the Dagstuhl seminar 'Sublinear Algorithms', which was held from 17.8.2008 to 22.8.2008 in Schloss Dagstuhl, Germany.

## Summary

With the increasing role of information technologies we are often confronted with a huge amount of information that is generated without pace by distributed sources or by large scale complex information systems. In many scenarios, it is not possible to entirely store this information on standard storage devices. Examples include the World Wide Web, data accumulated in network traffic monitoring, or sensor network data. One of the key challenges in this context is to efficiently process these massive data sets and to extract knowledge by summarizing and aggregating their major features. Most of the time, it is impossible to use traditional algorithms for this purpose. Even linear time algorithms are typically too slow because they require random access to the input data. We require algorithm that either look only at a small random sample of the input or process the data as it arrives extracting a small summary. Algorithms of this type are called *sublinear algorithms*.

The purpose of this workshop was to bring together leading researchers in the are of sublinear algorithms to discuss recent advances in the area, identify new research directions, and discuss open problems.

The area of sublinear algorithms can be split into three subareas: *property testing*, *sublinear time approximation*, and *data streaming algorithms*. These areas are not only connected by the fact that they require algorithms with sublinear resources but also that they heavily rely on randomization and random sampling. Researchers from all three areas came to attend this workshop and we believe that it helped to exchange ideas between these different areas.

During the seminar one could obtain a good overview of the current state of sublinear algorithms. In many interesting talks new algorithms and models as well as solutions to well-known open problems were presented. To give an idea about the topics of the seminar we present a few examples of topics that were discussed in a number of talks at the seminar. These examples are not meant to be exhaustive.

## Property Testing

Property testing is a relaxation of a standard decision problem. Instead of deciding whether a given input satisfies a certain property or not, a property testing algorithm only has to distinguish between input that have the property and inputs that are far away from the property. In property testing, one assumes that the algorithm is given oracle access to the input and an important complexity measure

is the query complexity, i.e. the number of queries to the oracle that are needed for the algorithm to succeed. In the workshop one new direction of work that has been discussed was the question, how much adaptivity helps in property testing and for which properties we can get very efficient property testing algorithms. Other work included the question whether a Boolean function is a half-space, testing of subgraph properties, and property testing for algebraic properties.

### **Sublinear time approximation**

In sublinear approximation the goal is to develop algorithms that read only a part of the input and compute an approximate solution to a given problem using the classical notion of approximation. In this workshop, a new technique for developing sublinear time approximation algorithms was presented that under certain conditions turns greedy algorithms into sublinear time approximation algorithms. Also, a sublinear time algorithm for hypergraph partitioning was presented.

### **Data streaming algorithms**

In data streaming we assume that we access the input in the form of a data stream. Further, the memory is assumed to be sublinear (typically, polylogarithmic) in the length of the stream. One focus of the data streaming talks has been the development of new techniques to prove lower bounds in the streaming scenario. Other talks discussed upper and lower bounds on the longest increasing subsequence problem.

### **Concluding remarks**

The seminar was attended by 46 researchers from six countries (18 USA, 9 Israel, 9 Germany, 5 UK, 4 Canada, 1 India). The inspiring atmosphere at Schloss Dagstuhl and the great working and living environment as well as interesting talks and many discussions between researchers contributed a very successful workshop.