MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS

Aarhus University

ANNUAL REPORT 2008

CENTER FOR MASSIVE DATA ALGORITHMICS

MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS

Aarhus University

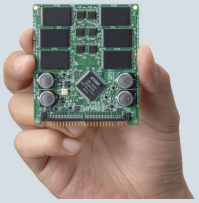
2008 Highlights

Research team

Center for Massive Data Algorithmics (MADALGO – see www.madalgo.au.dk) was established in 2007. At the end of 2008 the center research team consisted of six senior researchers (2 at AU), five Post Docs (4 at AU) and fifteen PhD students (11 at AU). Additionally, two further Post Docs and 4 PhD students (who obtained degrees during the year) were part of the center in 2008. All center Post Docs are internationals and so is a good deal of the PhD students.

Research collaboration and results

In 2008 MADALGO researchers published 35 peer reviewed research paper within the center research areas. Several of these papers have appeared in highly ranked journals and conference proceedings. Some of the results in the papers have been obtained with the many international researchers that have visited MADALGO in 2008. The center also has extensive multidisciplinary and industry collaboration, e.g. on issues in connection with massive terrain data.



Flash memory devices are quickly becoming popular, e.g. as replacements for hard-disks in laptop computers. These devices have very different access characteristic than disks (and other kinds of memory).

In 2008 center researchers – including MPI PhD student and subsequently AU Post Doc Deepak Ajwani – have performed an extensive experimental characterization of flash devices, and proposed several new theoretical models of them. The goal is to develop algorithms that are particular efficient on flash memory devices.

Center events

Apart from a large number of smaller research seminars and workshops, as well as a retreat for center employees, MADALGO also arranged a four day international summer school on “Cache-Oblivious Algorithms” in 2008, where four international experts lectured for around 60 participants (mostly PhD students).

In 2008 center researchers also gave more than 25 presentations at international research conferences, participated in more than 15 invitation only international research workshops, and gave more than 10 invited presentations at University seminars.

Awards and acknowledgments

The senior center researchers received a number of awards and acknowledgments in 2008. Mehlhorn received an honorary doctor degree from University of Aarhus, and Arge was elected member of the Royal Danish Academy of Sciences and Letters. Demaine received a number of awards, such as Universite Libre de Bruxelles gold medal and the Katayanagi emerging leadership price, and Indyk was invited to contribute a research highlight to Communication of the ACM.

Also the junior researchers received awards and acknowledgments. Center MIT PhD student Patrascu won the best student paper award at the top conference FOCS, and the two teams with center PhD student participation finished in fourth and fifth place (out of 47) in the northwestern Europe programming contest (NWERC).

Content

This report describes the 2008 activities at the Danish National Research Foundation *Center for Massive Data Algorithmics* (MADALGO). The outline of the report more or less follows the agenda of the annual review meeting for the center held on February 24, 2009 (except that external funding and teaching is only covered in the appendices). The report is accompanied by a number of appendices (covering external relations, conferences, educational activities, academic services, external funding, awards, public outreach, patents and applications, publications, list of personel) as specified by the foundation; note that *some* of these appendices only cover the employees at University of Aarhus (and not the participants at Max Planck Institute for Informatics, Massachusetts Institute of Technology and Frankfurt University). Finally, note that the 2008 accounts for the center with appendices are also important in order to obtain a complete overview of the 2008 activities of the center.

Center director statement

By signature it is confirmed that the annual report and accounts/budget with accompanying notes and appendices contain all relevant information regarding the annual primary activities in the Danish National Research Foundation Center for Massive Data Algorithmics.

14 *April, 2009* Lars Arge

Lars Arge
Center Director

1 Center background

Center for Massive Data Algorithmics (MADALGO) is motivated by the spectacular increase in the amount of data being collected and processed in many modern applications, as well as the inadequacy of traditional algorithms theory in many such applications – in particular that traditional theory has used a simplistic machine model that does take the hierarchical memory organization of modern machines into account. The memory system of a typical computer system is made up of several levels of cache, main memory and disk, where the access times of the different levels can vary by several orders of magnitude, and where data is transferred between the levels in large contiguous blocks. This means that often it is much more important to minimize the number of block transfers than it is to minimize CPU computation, as is done in the simplistic models used in traditional algorithmics. Thus, the inadequacy of the simplistic machine models often translates into software inadequacies when processing massive data.

MADALGO strives to become a world-leading center in algorithms for handling massive data, where massive is interpreted broadly to cover computations where the data is large compared to the resources of the computational device. The high-level objectives of the center are:

- To significantly advance the fundamental algorithms knowledge in the area of efficient processing of massive datasets
- To train the next generation of researchers in a world-leading and international environment
- To be a catalyst for multidisciplinary collaborative research on massive dataset issues in commercial and scientific applications

To meet these objectives the center builds on existing research strength, international research collaboration, multidisciplinary and industry collaboration, and in general on a vibrant international environment at the main center site. Another key is the focus on three related but also very different research focus areas, namely I/O-efficient, cache-oblivious and streaming algorithms, as well as algorithm engineering. Please refer to the original research plan for a more thorough discussion of the center motivation, objectives and core research areas. The four focus areas will also be described further in Section 3.

2 Center organization

The main center site is at Aarhus University (AU) in Denmark (with Brodal and Arge as core faculty) and the center also included researchers at the Max Planck Institute for Informatics (MPI) and at Frankfurt University (FRA) in Germany (with Mehlhorn and Meyer as core faculty), as well as at Massachusetts Institute of Technology (MIT) in the US (with Demaine and Indyk as core faculty). The sites continue to complement each other nicely in terms of research strength within the core focus areas.

In terms of research organization, the main event during 2008 was the formation of an international advisory board consisting of six world-leading researchers in the center research areas. The board will provide feedback on center research and activities, and advise the center on scientific planning. The board consists of *Pankaj K. Agarwal* (Chair and RJR Nabisco Professor of Computer Science at Duke University), *Giuseppe F. Italiano* (Professor of Computer Science and Chair of the Department of Computer Science, Systems and Production at University of Rome "Tor Vergata"), *J. Ian Munro* (University Professor and Canada Research Chair in Algorithms Design in the School of Computer Science at the University of Waterloo), *S. Muthu Muthukrishnan* (Research Scientist at Google New York), *Peter Sanders* (Professor of Computer Science at Karlsruhe University) and *Jeff S. Vitter* (Provost and Executive Vice President for Academics and Professor of Computer Science and Engineering at Texas A&M University). More information about the advisory board can be found on the centers web-page (www.madalgo.au.dk/html_sider/2_2_people/Advisory_board).
Other than the formation of the advisory board the center has been quite stable organizationally throughout 2008. Logistically the main center site continues to be managed by center manager Else Magård and accountant Ellen Lindstrøm (half time), with secretarial support from the Department of Computer Science. Four programmers (mostly funded by other grants) have also been associated with the center in 2008. The center also continues to benefit from the nice and connected facilities (meeting and conference rooms, as well as laboratory, lounge and office space) obtained at the center start. These facilities support a dynamic and collaborative environment and contribute to making the center attractive for visiting researchers.
On the research personnel side the center developed more or less as planned in 2008. Three post docs where hired at AU and two additional Post Docs also worked at the center in 2008. No less than six new PhD students were also hired at AU in 2008, in addition to the six other PhD students also associated with the center at AU in 2008. Center Post Docs and PhD-students are discussed further in Section 6.

3 Center research activities

3.1 Research plan

The original center research plan discusses a number of main research challenges in each of the four focus areas, and lists a number of concrete open geometric and graph problems. However, it is also noted that the list of problems is non-exhaustive, and that the outcome of current and future research of course should influence the exact directions taken in the center. Furthermore, the plan discusses that as the center matures, other interesting methodologies for massive data processing than I/O-efficient, cache-oblivious and streaming algorithms should be considered. In the 2007 annual report we described four such new methodologies/directions. Below we quickly recall all of these areas and highlight some main problems.

Main focus areas

I/O-efficient algorithms are algorithms designed in a two-level memory-disk *external memory* (or *I/O-*) model, where the memory hierarchy consists of a main memory of limited size M and an external memory (disk) of unlimited size; the goal is to minimize the number of times a block of B consecutive elements is read (or written) from (to) disk (an *I/O-operation*, or simply *I/O*). The model is motivated by the fact that transfers between main memory and disk, rather than e.g. CPU computation, is the bottleneck when processing massive datasets residing on disk. The I/O-efficient algorithms area is quite developed but a number of interesting problems remain open. The research plan outlines a number of such problems, including a number of fundamental geometric data structure and graph traversal problems, as well as some very practically motivated terrain data processing problems.

Cache-oblivious algorithms are algorithms designed in the I/O-model – but without knowledge of M and B – and then analyzed as I/O-model algorithms. The beauty of the model is that since the I/O-model analysis holds for any block and memory size, it holds simultaneously on *all* levels of *any* multi-level memory hierarchy. Thus the cache-oblivious model is effectively a way of modeling a complicated (maybe even unknown and/or changing) multi-level hierarchy using the simple two-level I/O-model. The cache-oblivious algorithms area is relatively new, with many even very fundamental problems remaining open. The research plan outlines a number of fundamental problems to be considered, mainly on geometric data structure and batched problems (including terrain algorithms).

Streaming algorithms are algorithms designed in a model where only one (or a small constant number of) sequential pass(es) over the data is (are) allowed. The goal is to solve a given problem while using significantly smaller space than the input data size, and while processing each data object as fast as possible. The model is motivated by the fact that when processing truly massive datasets, solutions requiring more than one sequential pass over the data are often infeasible, since random access to disk blocks are much slower than sequential access. Moreover, in some applications data simply *has* to be processed sequentially as it is generated. While the streaming algorithms area has flourished in recent years, fundamental gaps still remain in the understanding of what problems can be solved in the streaming model. The research plan outlines a number of fundamental problems to be considered, including investigation of the general applicability of developed techniques, as well as of fundamental geometric problems and of graph problems in variants of the streaming model.

Algorithm engineering covers the design and analysis of practical algorithms, efficient implementation of these algorithms, as well as experimentation that provide insight into their applicability and further improvements. Algorithm engineering is naturally an integrated part of the center both because a main motivation for the center is the inadequacy of traditional algorithms theory in providing practically efficient algorithms, and because engineering work naturally supports multidisciplinary and industry collaboration. Additionally, algorithm engineering work often provide valuable input to theoretical model work, and can sometime even lead to practical runtime breakthroughs. The research plan outlines a number of algorithm engineering challenges in the I/O-efficient and cache-oblivious algorithms areas. In the cache-oblivious area focus is on engineering simple algorithms for very fundamental problems, whereas in the I/O-efficient algorithms area focus is on leveraging existing basic I/O-algorithms libraries to further engineer algorithms for fundamental problems (e.g. graph problems) and to (further) develop software for efficient processing of massive terrain data.

Annual report 2007 new directions

Chip manufactures are increasingly producing chips with several CPUs (or cores) on a single chip. Thus there is a need for parallel algorithms that can use all of these cores. In last years annual report we outlined some challenges and results in a *parallel private-cache model,* which is basically a parallel extension of the I/O-model where P processors have a main memory of limited size M each and share an external memory of unlimited size. The goal is to minimize the number of *parallel I/Os*, in which blocks of B elements can simultaneously be transferred between the external memory and each of the main memories.

Another hardware trend is that flash memory devices are becoming increasingly large and cheap. Flash memory may eventually replace disks as the external storage in e.g. mobile computing. Since flash memory appears to have very different characteristics than both internal memory and disks, we in last years report outlined work on characterizing the new devices in order to develop *models of flash memory* and ultimately efficient algorithms for these models/devices.

Modern memory is not always fully reliable. Sometimes the content of a memory location may be temporarily or permanently corrupted, and error rates are expected to increase as memory is getting smaller and more complex. When processing massive data one usually performs a large number of memory accesses over a long period of time. Thus memory errors can become a serious problem. While error handling has been addressed in various ways, both at the hardware and software level, we in the 2007 annual report discussed planned work in a recently proposed interesting model for modeling and handling memory errors called the *faulty-memory RAM model.*

When handling massive data it is often important to use really space-efficient data structures. *Succinct data structures* are data structures that support efficient queries while occupying an amount of space that is provably close to the information-theoretic minimum. In last years report we described how we had hired a Post Doc (Rao) that is an expert in the area of succinct data structures, and how we planned to pursue the area further.

3.2 Overview of selected 2008 results

In general 2008 was a busy year with steady progress on both the original research plan problems and on problems in the new direction areas outlined above. In this section we briefly discuss some of the obtained results. Note that due to space limitations, we will not be able to give a complete overview of all the results obtained by center researchers (see Appendix I for a full list of papers published in 2008).

I/O-efficient algorithms

During 2008 we have obtained a number of interesting results in relation to the I/O-model problems outlined in the research plan.

In terms of geometric problems, we presented work on both a fundamental data structure problem (point location) and terrain data problems (contour line computation) at the 2008 ACM Symposium on Computational geometry [C29, C35].[1] This work was discussed in last year's annual report. We have continued both our data structure and terrain data work. We have e.g. developed new improved I/O-efficient structures for three and higher dimensional orthogonal range searching, that is, the problem of storing a set of points such that the points in a query hyper-rectangle can be reported efficiently. In three dimensions our new structures improve upon previously know structures and in higher dimensions they provide the first known non-trivial structures. Furthermore, we also prove a lower bound that show that our main d-dimensional data structure is space optimal. In fact, using the techniques we have developed, we were also able to improve an internal memory (pointer machine) variant of the problem, which has been studied extensively for decades without seeing any improvements. In terms of terrain data, we have developed an I/O-optimal algorithm for simplifying the contour tree of a terrain, which is a representation of the topological changes of a height h contour as h is increased. We have also used the techniques used to obtain this result to obtain I/O-efficient algorithms for removing small depressions in a terrain (the so-called flooding problem) based on any of a number of local geometric measures. Previously, only algorithms based on the depth of the depressions were known. Papers describing both of the above results are under submission.

We have also continued our work on fundamental graph traversal problems. During 2008 we presented results (which was discussed briefly in the 2007 annual report) on both approximate diameter computation [C27] and dynamic breadth-first search [C26] at international conference. The last result is based on graph

[1] Numbers in brackets are the number of the relevant papers in Appendix I.

clustering, but unfortunately it requires some randomization. However, recently we obtained a simple deterministic algorithm, which – on certain graph classes – offers additional benefits for static breadth-first search. A paper with these results is in submission. The PhD thesis [T4] of center student Ajwani gives an overview of the I/O-efficient graph algorithms area, and contains some new results on I/O-efficient algorithms for online topological ordering.

Cache-oblivious algorithms

In the cache-oblivious area we have also obtained a number of results in 2008 along the lines outlined in the research plan. In the batched geometric areas we have obtained the first results on red-blue line segment intersection [C34] (as discussed in last years report), as well as on efficient batched construction of a so-called quad-tree on a set of points in the plane (and for each quad-tree leaf compute the points in all neighboring leaves). The latter result has direct applications in the computation of a grid terrain model from an irregular set of terrain sample points; thus we are moving from only considering very fundamental problems in the cache-oblivious model to also considering problems with direct practical applications. We are in the process of experimenting with (and I/O-efficient version of) this algorithm, and plan on finalizing a paper with our results soon. The recently hired center Post Doc Peyman Afshani has also obtained some very interesting results on cache-oblivious geometric data structures. In a paper to be presented at this years ACM Symposium on Computational Geometry it is shown that any query-efficient cache-oblivious data structure for a special case of planer orthogonal range searching, as well as a number of other problems, have to use more than linear space. This provides a very interesting separation between the cache-oblivious and I/O models, since linear space I/O-model structures exist for the problem. In another papers to be presented at the same conference, query and space efficient structures are then developed for the problems, all use an interesting unified approach based on approximate counting.

Streaming algorithms

One of the most important characteristic of a data set is its entropy. Several near space optimal algorithms have previously been designed for approximating the entropy of a data stream. However, the known algorithms for the scenario where the stream elements can be inserted as well as deleted are much weaker. During 2008, we managed to develop an algorithm with a bound matching that for the insertions-only case [C42].

As already reported in the 2007 annual report, we have previously provided evidence that it is not a coincidence that almost all know streaming algorithms use efficient sketch construction [C39]. Our entropy algorithm, as many other streaming algorithms, relies on the technique of linear sketching, which has also found many other applications e.g. in compressed sensing and efficient experiment pooling. For a high-dimensional vector x, its *sketch* is equal to Ax, where A is an m by n matrix. Although typically the sketch length m is much smaller than the number of dimensions n, the sketch contains plenty of useful information about x; in particular, it enables the recovery of an approximation to x. Recently we showed that the two major sketching approaches – geometric and combinatorial – are essentially different manifestations of the same phenomenon, and as a result we obtained algorithms with the "best of both worlds" properties, having efficient sketch length and faster recovery time [C53]. In follow-up work [C21, C52], we further improve the recovery time to nearly linear in the dimension of x. These are the only known algorithms with these properties.

Our sketching results follow the line of research on applicability of various streaming algorithm techniques outlined in the research plan. Similarly, we have also e.g. investigated how properties of the input data influence algorithms development. The vast majority of streaming algorithms (including the ones mentioned above) assume that the data elements arrive in arbitrary order. However, in many applications one can make stronger assumptions, for example that the elements arrive in a random order. In this case one could expect to obtain algorithms with better performance. However, we managed to show that in many cases random order does not help (much) [C46].

Finally, it should be mentioned that Post Doc Mohammad Abam was hired as a Post Doc at the main center site in 2008. Abam has done some work on (geometric) streaming algorithms, and will thus support the effort to build up streaming algorithm expertise at AU, that started with the streaming algorithm summer school in 2007.

Algorithm engineering

Apart from algorithm engineering being an integrated part of much of our work described above, and also in the work in the 2007 opportunities areas discussed below, we have done specific engineering work on our

libraries for efficient implementation of I/O-model algorithms, on implementation of I/O-efficient graph algorithms, as well as on algorithms for massive terrain data. So far we have not had much progress on cache-oblivious algorithm engineering.
We have done library work on both the TPIE and STXXL libraries for efficient implementation of I/O-efficient algorithms. The TPIE work has included a general modernization of the library, as well as addition of features such as efficient external priority queues (see www.madalgo.au.dk/Trac-tpie/). The STXXL work has included a parallelization of the internal computation, as well as an augmentation of the STXXL pipelining feature to enable automatic task parallelization. The STXXL parallelization work is described in a paper to be presented at this years Parallel and Distributed Processing Symposium.
Our work on engineering of I/O-efficient graph algorithms – based on STXXL – has included work on shortest path algorithms. In the recent Workshop on Algorithm Engineering and Experimentation we reported on experimental results for an algorithm on undirected sparse graphs where the ratio between the largest and the smallest edge weight is reasonably bounded, and were the (realistic) assumption holds that main memory is big enough to keep one bit per vertex. While our implementation only guarantees average-case efficiency, i.e., assuming randomly chosen edge-weights, it turns out that its performance on real-world instances with non-random edge weights is even better than on the respective inputs with random weights. On a real-world 24 million node street graph, our implementation was over 40 times faster than previous implementations.
Our work on engineering I/O-efficient (geometric) algorithms for massive terrain data has been performed within the TerraSTREAM software project using TPIE. During 2008, a lot of work has been done on the TerraSTREAM software and 4 new releases[2] have been made. Many of the changes in the new releases – new smaller feature and bug fixes – have been made as a result of feedback from users (most notably from environmental researchers at DJF and industry experts at COWI). Refer to the TerraSTREAM webpage (www.madalgo.au.dk/Trac-TerraSTREAM/) for details. Some of the larger new features include erosion modeling (so-called LS-factor computation) and a tool for predicting where flooding is likely to occur in case of a rise in ocean level. To our knowledge, our software is the only software capable of performing these computations on massive terrains, and especially our flooding software has attracted a lot of attention from industry. We have also worked extensively on algorithms and software for constructing terrain models in the standard so-called grid and TIN formats from massive terrain point samples. Some of the grid model work was performed in collaboration with Assistant Professor Andrew Danner from Swarthmore College in the US while he was visiting the main center site for a month in 2008, and some of the work includes a new method for estimating the quality of the computed model. We are in the process of preparing two papers on this work. The TIN model work has been performed by center Post Doc Peter Hachenberger who has hired in 2008. Finally, following our theoretical results on contour line computation (discussed in the 2007 report and published in 2008) and on removing small depressions in a terrain based on any of a number of local geometric measures (the flooding problem) described above, we have also worked on engineering practical algorithms for these problems. Our solution to the flooding problem has proven particularly efficient and interesting, for example in connection with simplification of contour lines; we are in the process of preparing a paper on our results.

New direction areas
Among the new areas we described in the 2007 annual report, the *succinct data structures* area is probably the one we have had most success in. The fundamental and well-studied problem of encoding a bit-vector space-efficiently forms a basic building-block in many succinct data structures. As mentioned in the 2007 annual report, we have managed to improve the space redundancy of the existing bit-vector encodings by a logarithmic factor [C30]. Using a non-trivial recursive procedure, center PhD student Patrascu has further improved this space redundancy by a poly-logarithmic factor [C23]. The result solves a long-standing open problem in succinct data structures, and immediately improves the space redundancy of several other succinct data structures. Patrascu received the best student paper award at the 2008 Symposium on Foundations of Computer Science (one of the absolute top conferences in theoretical computer science) for the paper describing the results. The bit-vector problem is closely related to another well-studied problem called the predecessor problem. In a paper to appear in the 2009 International Symposium on Theoretical

[2] Only limited releases have been made publically available since the possibilities for commercialization of the software are currently being explored. Still more than 25 different institutions and companies have requested download information (download password). The full release has been made available to the participants in the strategic research council (NABIIT) project also supporting the work.

Aspects of Computer Science, we further established tight connections between these two problems. By developing new space-efficient substructures and combining them with the existing ones, we gave an improved structure for the bit-vector problem in the general case. We have also developed succinct versions of structures used in verification [C51], and in recent work to be presented at the Symposium on Principles of Database Systems (PODS) we considered succinct structures in the I/O-model. More precisely, we developed the first space- and query-optimal external memory search tree structure. Earlier known structures, (namely B-trees and compressed bitmap indexes) did not achieve both space and query-time optimality simultaneously.

Based on our experimental work on the characteristics of *flash memory devices* [C25] mentioned in last years report, we have proposed two models of flash memory: The *general flash model* and the *unit-cost flash model*. The general flash model is similar to the I/O model, except that read and write block sizes are different and incur different costs. The unit-cost flash model augments the general flash model with the assumption that the throughput provided by sequential reads and writes is equal. Our models are generic (as they abstract away many device-specific characteristics) and simple enough for meaningful algorithm design. In particular, we have shown that a large body of existing merging based I/O-model algorithms can be easily adapted to the unit-cost flash model. The work is to be presented in the upcoming Workshop on Experimental Algorithms. We are currently investigating if efficient algorithms can also be obtained for problems that are solved using distribution in the I/O model.

While our previous work in the *fault tolerant RAM model* concentrated on fault tolerant data structures for fundamental problems, our recent work has considered a combination of fault tolerance and I/O-efficiency. Previous fault tolerant algorithms and data structures do not scale beyond internal memory. In recent (submitted) work, we have proposed a faulty memory version of the I/O-model and developed optimal sorting algorithms, as well as dictionary and priority queue data structures, in the model. Our work has revealed some interesting dependencies – the optimal bounds e.g. depends on weather faults can happen both in internal and external memory and on if the algorithms are deterministic or randomized.

As mentioned in the 2007 annual report, our *parallel private-cache model* work is performed in collaboration with UC Irvine Professor Mike Goodrich and his PhD student Nodari Sitchinava who, as in 2007, visited us during the summer of 2008 (Sitchinava visited for an extended period of time). Although we have developed efficient or even optimal algorithms for a number of fundamental problem such as sorting and selection [C33], as well as for a several fundamental graph problems (paper in submission), our parallel private-cache model work has not been as successful as we had hoped. One reason might be that several different parallel private-cache models have been proposed recently, and it is still unclear which model will be most successful. However, we have also worked on parallel algorithms in other contexts, e.g. in the STXXL software library work described above. Recently, we have also worked on the intriguing connections between provable efficient parallelizable problems and their streaming- and I/O-model solution. The results include both results on classes of problems that have both efficient parallel and (multi-pass) streaming solutions, and on the lack of a complete correspondence between the class of problem that have efficient parallel solutions and the class of problems that have I/O-efficient solutions.

3.3 Updated research plan

As outlined in the previous section, center researchers have obtained a number of interesting and important results on the problems outlined in the original research plan, as well as on problems in the new direction areas described in the 2007 annual report. In general we will continue to follow the research plan, while of course letting the outcome of current and future research, as well as the interests of center researchers, influence the exact research directions taken by the center. We will also continue to pursue the new direction areas to varying degree. This work will also be guided by progress and personnel: In 2008 we had most success in the area of succinct data structures, whereas we expect less progress in 2009 since the expert Post Doc in the area (Rao) has left the center. On the other hand, the center has retained Post Doc expertise in the flash model area (Ajwani), and this Post Doc is also interested in parallel algorithm issues, just as we also still have local expertise in fault tolerant algorithms.

As mentioned in the original research plan and also reflected in the new direction areas discussed above, we plan to continue to consider new and interesting methodologies for massive data processing as the center matures. One development we are already seeing in our work, and which we plan to pursue, is increasing

work on models that cuts across individual models. As described above, we have already considered succinct data structures in the I/O-model, the combination of fault tolerance and I/O-efficiency, as well as connections between parallel, streaming and I/O-models solutions for the same problem (or classes of problems). Similar to cache-oblivious algorithms, we are also e.g. considering the development of "flash-oblivious" models, that is, models for designing efficient algorithms on solid-state disks without knowing the hardware parameters. The overall goal in this "*model combination*" work is of course to develop better and more realistic models for efficient algorithm design.

Another crosscutting "theme" that has emerged in the centers work is data structures, that is, efficient ways of storing (massive) data using small space, such that queries on the data can be answered efficiently. Much of the work in the various models/areas described above is on data structuring problems, and so is much of the rest of the centers work published in 2008 (e.g. [C22, C24, C28, C44, C45, C51, J2, J4]). We see an opportunity in the broad data structure expertise present in the center, and plan to intensify our data structure work. For example, our recently hired Post Doc Peyman Afshani is a data structures expert, just as Associate Professor John Iacono from NYU who visited the main center site for an extended period of time in the summer of 2008. Afshani and other center researchers have recently obtained very interesting results on classical data structure problems in both internal memory models and in the I/O and cache-oblivious models (in particular on higher-dimensional range searching problems as already discussed above). Iacono has recently worked with center researchers on a more than 20 year old data structure challenge, namely the so-called dynamic optimality conjecture. In this work, presented at the recent Symposium on Discrete Algorithms, the conjecture is addresses by studying the combinatorial properties of search trees as a geometric problem. Although the conjecture is still to be proved, the work gives a new perspective on the problem. We hope an increased center focus on data structures can lead to even more results on classical data structure problems.

4 Collaboration

The center continues to try to maintain a vibrant, world-class and international environment at the main center site, e.g. through emphasis on hosting international visitors (faculty as well as PhD students) at AU. All core MIT, MPI and FRA faculty have visited AU during 2008, and many MIT, MPI and FRA PhD-students and Post Docs have also visited. Additionally, several non-center faculty – Bradford Nickerson (New Brunswick), John Iacono (NYU) and Andrew Danner (Swarthmore) – visited AU for a longer period of time (more than 3 weeks) in 2008, along with PhD students Nodari Sitchinava (UC Irvine), Muriel Dulieu (NYU) and Jeremy Fineman (MIT). The list of shorter term non-center faculty visiting AU include Rajeev Raman (Leicester), Herman Haverkort (Eindhoven), Norbert Zeh (Dalhausie), Alex Lopez-Ortiz (Waterloo), Ian Munro (Waterloo), Seth Pettie (Michigan), Alexander Wolff (Eindhoven), Mikkel Thorup (AT&T labs), Mike Goodrich (UC Irvine), Spyros Sioutas (Ionian), Kostas Tsichlas (Thessaloniki), Jeremy Barbay (Chile) and Rasmus Pagu (ITU); the list of non-center students include Arash Farzan (Waterloo), Dmitriy Morozov (Duke), Jeff Philips (Duke) and Ana Krulec (Primorska). In general, center researchers have extensive international research collaboration with other computer scientists; ongoing research collaborations – many with joint publications – are listed in Appendix A.

Through collaborative efforts, the center also continues to try to be a catalyst for multidisciplinary and industry collaboration. For instance, center researchers continue to collaborate extensively with environmental and agricultural researchers (e.g. at the Department of Biology, the Faculty of Agricultural Sciences and National Environmental Research Institute), as well as with industry (e.g. COWI A/S, EIVA A/S, and DHI) on issues in connection with massive terrain data. Much of this collaboration is centered around the TerraSTREAM software package, also supported through several other sources such as a NABIIT grant from the Danish strategic research council and co-financed ("samfinancierede") PhD scholarships (one of which was obtained during 2008). See again Appendix A for a more comprehensive list of multidisciplinary and industry collaborators.

The center is continuously seeking to extend existing collaborative efforts, as well as initiate new ones. For example, the center already participate in a Nordic Algorithms Network (supported by NorForsk) and is exploring the possibility of forming a European (EU funded) massive data algorithms research network, just as new multidisciplinary collaborations are being explored with e.g. the state library and the (intelligence) data mining company Prevail. Very recently, the possibility of establishment of a research collaboration with

Institute for Theoretical Computer Science (ITCS) at Tsinghua University in Beijing (the leading research institution in China) has also been explored (along with other AU theoretical computer science faculty) during a weeklong joint ITCS and AU workshop in Beijing (see www.itcs.tsinghua.edu.cn/CTACC2009/).

5 Events

During 2008 the center participated in and/or organized a large number of research events. These included internal weekly seminars at AU, as well as a PhD student (and Post Doc) only seminar series in the fall of 2008. The center also e.g. hosted a number of informal workshops around the TerraSTREAM software (with COWI A/S and Faculty of Agricultural Sciences researchers). Externally, center researchers gave more than 25 presentations at international research conferences, participated in more than 15 invitation only international research workshops, and gave more than 10 invited presentations at University seminars.

In August 2008 the center organized a four day international summer school on "Cache-Oblivious Algorithms" (see www.madalgo.au.dk/cacheschool08). The goal of the summer school was to provide an in-depth introduction to some of the key issues in cache-oblivious algorithms and data structures. The lecture program was coordinated by core researcher Brodal and Demaine, who also lectured at the school. Other lectures were Norbert Zeh (Dalhousie University) and Nodari Sitchinava (UC Irvine – as a stand-in for Mike Goodrich). The school had around 60 participants (mostly PhD students), and judging from evaluation conducted at the end of the school the school was a great success. In 2008, center researchers (Brodal and Indyk) also lectured at summer schools at the IT University in Copenhagen and at Kent State University.

Center researchers have also participated in several public outreach activities. Arge and Brodal have for example lectured on massive dataset processing and algorithms in general at industry society, high-school and company events, just as center researchers participated in the research day ("forskningens døgn") with an exhibition on flood prediction. Demaine also e.g. gave a lecture entitled "Origami, Linkages and Polyhedra: Folding with Algorithms" for a group of female high-school students in MITs woman in technology program.

6 Research education

As mentioned, one key goal of the center is to train the next generation of researchers in a world-leading and international environment. In general, focus is on people and on maintaining a large vibrant environment at the main center site at AU. Thus PhD-students and Post Docs are a very important part of the center, and the center strives to have a large population of international PhD students and Post Docs at AU, just as AU center students stay 6 months abroad. Currently, the center houses 17 PhD students (13 at AU, 3 internationals) and 6 Post Docs (5 at AU, all internationals); 2 of these PhD students and 1 Post Doc (all at AU) have been hired in 2009 (and thus do not appear on the 2008 list of personnel). Furthermore, three center PhD students (two at MIT and one at MPI) finished their PhD studies in 2008, and one AU PhD student decided to leave the PhD program with a MS degree after his qualification exam. Below we give a brief overview of the 2008 center Post Docs and PhD students, with emphasis on the AU Post Docs and PhD students. Apart from the listed PhD students, 7 MS students have also been associated with the center in 2008.

Post Docs:

- *Henrik Blunck*, PhD Munster 2006; employed November 2006-October 2008
  Works on I/O-efficient GIS algorithm. Still associated with the center even though he is now working on the Galileo project funded by the high-tech foundation (Højteknologifonden).
- *Srinivas Rao*, PhD Chennai 2002; employed August 2007-January 2009
  Mainly works on succinct data structures. Now an Assistant Professor at Seoul National University.
- *Mohammad Adam*, PhD Eindhoven 2007; hired January 2008
  Mainly works in computational geometry but has also done some streaming algorithm work.
- *Peter Hachenberger,* PhD MPI 2006; hired September 2008
  Mainly works on implementations of computational geometry algorithms.
- *Deepak Ajwani,* PhD MPI 2008; hired October 2008
  Mainly works on I/O-efficient graph algorithms and algorithms for flash memory.

Furthermore, *Payman Afsahni* (PhD Waterloo 2008) joined the center (at AU) in February 2009. He is mainly working on I/O-efficient and cache-oblivious geometric data structures. Additionally, *Gabriel Moruz* (PhD AU, 2007) continues to be a Post Doc at FRA. He mainly works on fault tolerant algorithms and data structures. Finally, *Kevin Chang* (PhD Yale 2006) finished his Post Doc at MPI in August 2008 and is now working in industry in the US. Note that just like in 2007 (Moruz from AU to FRA), a newly educated PhD from one center site obtained a Post Doc at another center site (Ajwani from MPI to AU).

PhD students:

- *Martin Olsen* (Advisor: Brodal).
  Works on algorithms for analyzing the web-graph; will finish PhD in July 2009.
  Spent the fall of 2008 at ICT, Sidney.
- *Thomas Mølhave* (Advisor: Arge).
  Mainly work on I/O-efficient and cache-oblivious algorithms, as well as algorithm engineering; will finish in August 2009.
  Spent 2 months at AT&T research and 4 months at Duke University.
- *Allan G. Jørgensen* (Advisor: Brodal).
  Mainly works on fault tolerant algorithms; will finish in January 2010.
  Spent the spring of 2008 at MIT.
- *Anders H. Jensen* (Advisor: Arge).
  Mainly worked on I/O-efficient algorithms; decided to leave with an MS degree in July 2008.
- *Lasse Deleuran* (Advisor: Arge).
  Works on line simplification (computational geometry) algorithms; will finish PhD in July 2011.
- *Kostas Tsakalidis* (Advisor: Brodal).
  Works on I/O-efficient interpolation search. Came from Patras, Greece, and will finish PhD in July 2011.
- *Mark Greve* (Advisor: Brodal).
  Started in February 2008; will finish PhD in January 2012.
- *Morten Revsbæk* (Advisor: Arge).
  Started in February 2008; will finish PhD in January 2012.
- *Jesper Eshøj* (Advisor: Arge along with Biology and Agricultural Sciences faculty)
  Started in February 2008; will finish PhD in January 2012.
- *Pooya Davoodi* (Advisor: Brodal).
  Started in May 2008; will finish PhD in April 2011.
- *Jacob Truelsen* (Advisor: Brodal).
  Started in August 2008; will finish PhD in July 2012.
- *Kasper D. Larsen* (Advisor: Arge).
  Started in August 2008; will finish PhD in July 2013.

In general, the hiring of PhD students at AU proceeds more or less according to plan (actually a little ahead of the original plan) with no less than 6 PhD students hired in 2008. So far the hiring in 2009 also proceeds according to the plan with two PhD students being hired in February: *Freek van Walderveen* (Advisor: Arge) and *Casper Kejlberg-Rasmussen* (Advisor: Brodal). Additionally, *Andreas Beckmann* (Advisor: Meyer) is a center PhD student at FRA, and *Oren Weimann* (Advisor: Demaine), *Kahn Do Ba* (Advisor: Indyk) and *Jelain Nelsen* (Advisor: Demaine) are center PhD students at MIT.
During 2008 center PhD students *Mihai Patrascu* (Advisor: Demaine) and *Anastasios Sidiropoulos* (Advisor: Indyk) obtained their PhD from MIT; Patrascu is now a Post Doc at IBM Almaden and Sidiropoulos a Post Doc at University of Toronto. Furthermore, as mentioned above, MPI student *Deepak Ajwani* (Advisor: Mehlhorn and Meyer) obtained his PhD from MPI in 2008 and joined AU as a Post Doc. Thus center PhDs manager to obtain good international Post Doc positions. AU student Mølhave who will finish his PhD this summer has also already accepted a Post Doc offer from Duke University.

Recruiting and research education environment

So far the centers Post Doc and PhD student recruitment efforts have been relatively successful. All center Post Docs are internationals and the center is experiencing high interest in its advertised Post Doc positions. For this reason we are working on obtaining additional Post Doc funding from other sources. PhD hiring is also progressing nicely at AU. Actually, the number of AU PhD students has now reached the anticipated maximum in the 5 year center period. In the spring of 2008 as well as in the spring of 2009 hiring rounds we have managed to recruit an international student (Davoodi from Iran and van Walderveen from Holland,

respectively). We would like to increase the international recruitment further; we especially believe there is an opportunity to recruit more highly skilled Iranian students – now that we have already recruited three Iranians (two Post Docs and a PhD student) – as well eastern European students. Although difficult, we would also like to recruit female PhD students. It is also worth highlighting that one of the AU PhD students (Eshøj) is a Biology student (with some Computer Science background) who works on using massive data algorithms techniques to improve spatial modeling of the plant diversity of Denmark, co-advised by Arge and Biology and Agricultural Sciences Faculty. We believe that such co-advised PhD students are a good way of fulfilling the centers objective of being a catalyst for multidisciplinary collaboration. Finally, two of the PhD students hired in the last year (Larsen and Kejlberg-Rasmussen) were actually admitted to the PhD program based on only three years of study (a bachelor degree) rather than the normal four years of study. The center supports this new special honors masters study ("elitekandidat uddannelse") possibility introduced by the Faculty of Science; it allows for the recruitment (and thus retention) of highly skilled students early in their study, as well as potentially better recruitment possibilities of skilled bachelor students from other institutions. In general, the center (with success) has hired PhD students at almost all levels: Among the AU students hired in the last year there has been students admitted using all of so-called 3+5, 4+4 and 5+3 admissions possibilities.

As mentioned, the centers focus on research education includes exchange of Post Doc and PhD students, as well as a 6 months stay abroad for AU PhD students. As discussed in Section 5, the center also arranged an international summer school in 2008 (as in 2007), just as center researchers (Brodal and Indyk) have lectured at other summer schools. Furthermore, two of the center Post Docs (Adam and Rao) organized a special 5 EATCS PhD class on Advanced Computational Geometry in the fall of 2008. Another class is planned for this spring, and we plan to continue such intensive classes on selected subjects.
Finally, the center continues to emphasize initiatives designed to create as sense of community at the main center site and among the center sites. This includes a yearly two day fall retreat (at "Sandbjerg" in both 2007 and 2008) attended by most of the AU staff and several MPI, MIT and FRA Post Docs and PhD students, monthly center lunches at AU, as well as number of social events (Christmas dinner, weekly breakfast, weekly soccer, go-carting trips, etc).

7 Milestones

As discussed in the research plan, it is often very hard to establish very concrete milestones and goals for the kind of theoretical research conducted in the center. In Section 3, we have discussed results obtained in 2008 and their relation to the research plan, as well as outlined a research agenda for 2009 (and beyond).

A significant research production (published in the major conference and journals) and high visibility in the research community is obviously one of the major goals of the center. We believe both of these goals are being meet. The original research plan and the 2007 annual report identified a number of high-level 2008 milestones. Most of these milestones have (at least partially) been met: Major release of TPIE software, summer school on cache-oblivious algorithms, (small) multidisciplinary workshop on massive terrain data handling, formation of an international advisory board, and the establishment of a formal weekly seminar. However, mainly due to lack of (human) resources some milestones, mainly publishing of a quarterly center newsletter and development of educational material/lecture notes on I/O-efficient algorithms, have had to be abandoned. Rather than publishing a newsletter we are concentrating on maintaining an up-to-date website. We still plan on developing educational material I/O-efficient algorithms, but are reluctant to set it as a milestone for 2009.

As already discussed in the 2007 report, milestones for 2009 include organization of the first Workshop on Algorithms for Massive Datasets, as well as the 25th Annual ACM Symposium on Computational Geometry. We are also still considering the possibility of establishing a Danish advisory board, the establishment of video conference and remote collaboration facilities between the centers cites, as well as the establishment of more formalized international PhD student and summer student programs. Finally, milestones for 2009 also include further work on commercialization of the TerraSTREAM software (e.g. through a license agreement with an independent company to be established), development of further PhD classes, organization of a larger multidisciplinary workshop on terrain data handling and analysis, as well as a renewed focus on collaborative projects, most notably with Chinese researchers.

Appendix to MADLGO Annual report 2008

Appendix A: External Relations
Please list collaborations with danish and foreign research groups, universities, companies and/or institutions and state subject and/or title and output/results if any.

List **includes** relevant 2008 collaboration for AU as well as MIT, MPI and FRA researchers.

			Collaboration with: (Please tick the appropriate box)			
Collaborator Name (person and/or institution) Country	**Collaboration subject and/or title**	**Output/ results of collaboration, if any**	**Danish universities, research groups and institutions**	**Foreign universities, research groups and institutions**	**Danish companies**	**Foreign companies**
Norbert Zeh (Dalhausie), Canada	I/O-efficeint and cache-oblivious algorithms	Publications		X		
COWI A/S (incl. Kristian Keller, Johnny Koust Rasmussen, Michael Schultz Rasmussen) and The Faculty of Agricultural Science (incl. Peder Klith Bøcher), Denmark	Efficient Handling of Massive Heterogenious Terrain Data	Efficient terrain processing algorithms and TerraSTREAM software modules	X		X	
Pankaj K. Agarwal and Bardia Sadri (Duke University), Andrew Danner (Swarthmore College), USA	I/O-efficient terrain algorithms and TerraSTREAM	Publications, as well as the TerraSTREAM software package		X		
Jan Vahrenhold (Technical University, Dortmund), Germany and Andrew Danner (Swarthmore College), USA	TPIE	TPIE software package		X		
Peder Klith Bøcher (Faculty of Agricultural Science, AU), Jens-Christian Svenning (Institute of Biological Sciences, AU), and National Environmental Research Institute (NERI), Denmark	Collaborators and co-advisors of PhD student Jesper Moeslund Eshøj ("Fintopløselig, geospatial modellering af Danmarks nuværende og potentielle fremtidige plantediversitet")	Publication	X			
Mike Goodrich and Nodari Sitchinava (UC Irvine), USA	Algorithms for private-cache chip multiprocessors	Publications		X		

Ian Munro and Alex Lopez-Ortiz (University of Waterloo), Norbert Zeh (Dalhousie University), Bradford G. Nickerson (New Brunswick), Mark Masry (CARIS Geomatics Software Solutions), Canada	Data Structures for Efficient Organization and Retrieval of Massive Spatial Data			X		X
Mark de Berg and H. Haverkort (TU Eindhoven), Netherlands	Spatial data structures	Publications		X		
S. Muthu Muthukrishnan (Google and Rutgers University), USA	Identification of data analysis collaboration project between MADALGO and MassDAL			X		X
Researcher at Aarhus and Aalborg University, Terma A/S, Systematic Software Engineering A/S, Dansk Landbrugsrådgivning, Alexandra Instititute, Denmark	Hightech foundation project "A platform for Galileo based pervasive computing"		X		X	
Andy Brodnik (University of Kope), Slovenia	MS student (Ana Krulec) advising			X		
EIVA A/S, Denmark	Fast construction of quad-tree based indexes				X	
The Faculty of Agricultural Sciences, AU (incl. Jørgen E. Olesen), National Environmental Research Institute, AU (incl. Erik Jeppesen), Institute of Biological Sciences, AU (incl. Niels Peter Revsbech), Denmark	iCLIMA (integrated Climate-Ecosystem research: Impact and Adaptation). AU UNIK proposal		X			
Kostas Tsichlas (Aristotle University of Thessaloniki), Spyros Sioutas (Ionian Univeristy), Alexis Kaporis (University of Patras), Greece	Full Persistent B-trees, 3-sided 2D Ortogonal Range Reporting in Double-logarithmic Time	Publication		X		
Thomas Mailund, Christian Nørgaard Storm Pedersen (Bioinformatics Research Center, Aarhus University), Denmark	Computing the All-Pairs Quartet Distance on a set of Evolutionary Trees	Publication	X			

Rolf Fagerberg (University of Southern Denmark), Denmark, Alex López-Ortiz (University of Waterloo), Canada	Online Sorted Range Reporting	Publication	X	X		
Rasmus Pagh (IT University Copenhagen), Denmark	Secondary Indexing in One Dimension	Publication	X	X		
Alexander Golynski (Google), USA, Rajeev Raman (University of Leicester), Great Britain	Redundancy of Succinct Data Structures	Publication		X		
Jyrki Katajainen (University of Copenhagen), Denmark	Compact Dynamic Dictionaries for Representing Multisets	Publication	X			
Roberto Grossi, Alessio Orlandi (University of Pisa), Italy, Rajeev Raman (University of Leicester), Great Britain	More Haste Less Waste: Lowering the Redundancy in Fully Indexable Dictionarie	Publication		X		
Henning Meyerhenke (University of Paderborn), Germany	Realistic Computer Models	Publication		X		
John Iacono (NYU), Stefan Lagnerman (Universtite Libre de Bruxelles), Ian Munro (University of Waterloo), USA, Belgium and Canada, Jeremy Finman (MIT), USA	Cache oblivious arrays	Publications		X		
Joachim Gudmundsson (NICTA), Taso Viglas (University of Sydney), Australia	Graph algorithms	Publication		X		
Mark de Berg (TU Eindhoven), The Netherlands, Alireza Zarei (Sharif University of Technology), Iran	Streaming Algorithms for Line Simplification	Publication		X		
Mark de Berg (TU Eindhoven) The Netherlands, Mohammad Farshi, Michiel Smid (Carleton University), Canada, Joachim Gudmundsson (NICTA), Australia	Geometric spanners	Publications		X		

Mark de Berg (TU Eindhoven), The Netherlands, Sheung-Hung Poon (Tsing Hua University), Taiwan	Fault-Tolerant Conflict Free Coloring	Publication		X		
Hee-Kap Ahn (Pohang University of Science and Technology, Sang Won Bae (KAIST) South Korea and Marc Scherfenberg (Freie Universität, Berlin) Germany	Rectangle Covering of Imprecise Points	Publication		X		
Xavier Molinero, Maria Serna, Josep Freixas (Universitat Politècnica de Catalunya), Spain	Simple Games	Publication		X		
Riko Jacob (TU Munich), Germany, Norbert Zeh (Dalhousie University), Canada, Itay Malinger, Sivan Toledo (Tel-Aviv University), Israel	Flash Memory Models	Publications		X		
Timothy Chan (University of Warerloo), Canada	Range reporting data structures	Publication		X		
Group of Peter Sanders (Karlsruhe), Germany	(1) Libraries for Parallel/External Compuation (2) Implementation of an EM SSSP algorithm	Publications		X		
Group of Sivan Toledo (Tel Aviv), Israel	Algorithms and Experiments for Flash Memories	Publication		X		
Lars Engebretsen (Google Zurich)	External-Memory Graph Clustering					X
Several EU partners	DELIS: Dynamically Evolving Large-Scale Information Systems			X		
Stefan Langerman (Universite Libre de Bruxelles), Belgium	Confluently Persistent Tries	Publication		X		
Morteza Zadimoghaddam (Sharif University), Iran	Ordinal Embedding	Publication		X		

Greg Aloupis, Sébastien Collette, Stefan Langerman (Universite Libre de Bruxelles), Belgium and Vera Sacristán (UPC Barcelona), Spain and Stefanie Wuhrer (Carleton University), Canada	Reconfiguration of Crystalline Robots	Publications		X		
Takehiro Ito (Tohoku University), Ryuhei Uehara (JAIST), Yushi Uno (Osaka), Japan and Nicholas J. A. Harvey (Waterloo), Canada	Complexity of Reconfiguration Problems	Publication		X		
Ken-ichi Kawarabayashi (Tohoku University), Japan	Graph List-Coloring	Publication		X		
Hamid Mahini, Morteza Zadimoghaddam (Sharif Univeristy), Iran	Network Creation Games	Publication		X		
Glencora Borradaile (Waterloo), Canada, Siamak Tazari (Berlin), Germany	Graph Subset-Connectivity	Publication		X		
Danny Hermelin, Gad M. Landau (Heifa), Shir Landau (Bar-Ilan University), Israel	Accelerating Edit-Distance Computation via Text-Compression	Publications		X		
Robert Krauthgamer (Weizman), Israel, Jiri Matousek (Prague), Chech Republich	Metric embeddings	Publications		X		
Anna Gilbert, Martin Strauss (U Michigan), USA	Streaming algorithms, sparse recovery, compressed sensing	Publications, software libraries		X		
Milan Ruzic (ITU), Denmark	Compressed sensing	Publication	X			

Appendix B: Conferences

Please outline the centres disseminations of results and networking through participation in large internationale conferences, symposia etc. by a) list the number of conferences, symposia, seminars etc. the centre has arranged or participated in the planning of and by b) list the number of conferences, symposia, seminars etc. the centre has participated in with contributions.

List **includes** relevant 2008 information for AU as well as MIT, MPI and FRA researchers.
(Note: Last year only talks listed - this year also conference participation with paper or as PC member listed)

a) Organisation of international conferences, symposia, seminars etc.

Title of event	**Number of participants, estimated**	
	Danish	**International**
Summer school on cache-oblivious algorithms	30	30
Dagstuhl seminar on Data Structures	4	45

b) Participation in international conferences, symposia, seminars etc.

Title of event	**Venue**	**Name(s) of participant(s)**	**Contribution (talk, abstract, paper, poster, other)**	**Invited talk (please check)**
European Symposium of Algorithms	Karlsruhe, Germany	Mølhave, Arge, Mehlhorn	Talk, PC	
Dagstuhl seminar on Data Structures	Dagstuhl, Germany	Arge, Abam	Talks, PC	
Korean workshop on Computational geometry	Pohang, South Korea	Arge	Talk	X
International Symposium on Algorithms and Computation	Surfers paradise, Australia	Arge	PC	
Computer science seminar	University of Koper, Slovenia	Arge	Talk	X
Computer science seminar	University of Copenhagen	Arge	Talk	X
Symposium on Computational Geometry	Maryland, USA	Rao, Abam, Sidiropoulos, Demaine	Talks	
Scandinavian Workshop on Algorithm Theory	Gothenburg, Sweden	Rao, Demaine, Meyer	Talks	
International Symposium on Algorithms and Computation	Gold Coast, Australia	Grønlund Jørgensen, Arge	Talk, PC	

Northwest European Programming Contest 2008	Utrecht University, The Netherlands	Truelsen, Greve, Dalgaard Larsen, Mølhave, Ehlerns NyholmThomsen	Participation	
International Computing and Combinatorics Conference	Dlian, China	Olsen	Talk	
International Workshop on Experimental Algorithms	Massachusetts, USA	Ajwani	Talk	
Canadian Conference in Computational Geometry	Vancouver, Canada	Abam	Talk	
Research visit	Paderborn University, Germany	Abam	Talk	
Research visit	Carleton University, Canada,	Abam	Talk	
Seminar	Institute of Business and Technology, Aarhus University	Olsen	Talk	
Seminar	NICTA, Sydney, Australia	Olsen	Talk	
Dagstuhl Seminar on Sublinear Algorithms	Dagstuhl, Germany	Meyer	Talk	X
Seminar of the German Research Cluster Algorithm Engineering (SPP 1307)	Saarbrücken, Germany	Meyer	Talk	X
Algorithm Engineering Workshop @ Google	Google Offices, Zurich,	Meyer, Moruz	Talks	
Symposium on Theoretical Aspects of Computer Science	Bordeaux, France	Meyer	Talk	
Matching under Preferences, Satellite Workshop of ICALP	Reykjavik, Iceland	Mehlhorn	Talk	X
Kyoto RIMS Workshop on Computational Geometry and Discrete Mathematics	Kyoto, Japan	Mehlhorn	Talk	X
The Design and the Elastic Mind Symposium	New York, New York, USA	Demaine	Talk	X
International Workshop on Parameterized and Exact Computation	Victoria, Canada	Demaine	Talk	X
7th International Conference on Information Processing in Sensor Networks	St. Louis, Missouri	Demaine	Paper	
IEEE Conference on Computational Complexity	College Park, Maryland, USA	Demaine	Paper	

MathFest 2008	Madison, Wisconsin	Demaine	Talks	X
International Workshop on Approximation Algorithms for Combinatorial Optimization Problems	Boston, Massachusetts	Demaine	Paper	
Advances in Architectural Geometry	Vienna, Austria	Demaine	Paper	
DARPA InfoChemistry meeting	Cambridge, MA, USA	Demaine	Talk	X
International Workshop on the Algorithmic Foundations of Robotics	Guanajuato, México	Demaine	Papers	
Symposium on Combinatorial Pattern Matching	Pisa, Italy	Demaine	Talk	
IEEE Symposium on Foundations of Computer Science	Philadelphia, PA, USA	Nelson, Sidiropoulos, Pătraşcu	Talks	
ACM/SIAM Symposium on Discrete Algorithms	San Francisco, CA, USA	Pătraşcu	Talk	
Seminar	Stanford University	Indyk	Talk	X
Seminar	University of Illinois at Urbana-Champaign	Indyk	Talk	X
Seminar	University of Michigan at Ann Arbor	Indyk	Talk	X
Tutorial	Princeton University	Indyk	Talk	X
Allerton Conference	Allerton	Indyk	Talk	X
Seminar	State University of New York at Buffalo	Indyk	Talk	X
Seminar	Dartmouth College	Indyk	Talk	X
Summer School	Kent State University	Indyk	Talk	X

Appendix C: Educational activities

Please list all educational activities the centre is envolved in. Including PhD-courses, courses on bachelor- and master-level and other activities on academic level as well as summer schools and courses teached abroad. Please state ECTS points (if possible) and length of the course (in hours)

List only **includes** 2008 information for AU employees (as well as relevant information for MIT, MPI and FRA employees **taught outside** their home institution). PhD student TA'ing is **not** included.

Title and date of activity	**ECTS**	**Length of course (number of hours)**	**Number of participants**
BSc course: Algorithms and Data Structures 1, Spring 2008	5	28	~130
BSc course: Algorithms and Data Structures 2, Spring 2008	5	28	~130
BSc course: Computer Science in Perspective - 1 out of 7 lecturers, Spring 2008	5	10	~130
BSc course: C++/C programming, Institute of Business and Technology, Spring 2008	5	60	~20
MSc course: I/O-efficient algorithms, spring 2008	10	39	17
MSc course: Computational Geometry, Fall 2008	10	42	23
PhD course on Advanced Computational Geometry, Fall 2008	5	18	6
Algorithms for Advanced Processor Architectures, PhD course, IT University in Copenhagen, Summer 2008	5	24	~30
MADALGO Summer School on Cache-Oblivious Algorithms, August 2008		~25	60
Explicit Constructions in High-Dimensional Geometry, Summer School on Probabilistic and Fourier Methods in Geometric Functional Analysis, Kent State University,		4	60

Students surpervised at the centre by staff members

List **only** includes 2008 information for AU employees

Number of graduated PhD-students	Number of graduated Master-students
3	6

Appendix D: Academic services

Please list the activities for the entire group of centre staff members.

List **only includes** 2008 information for AU employees.
Conference reviewing is also included since conferences play a large role in computer science; the lists are **probably incomplete** and numbers **only rough estimates**, since computer scientists typically do a very large number of (especially conference) reviews each year.

a) Scholarly communication

Title of the journal	**Editorial work (please check)**	**Number of peer reviews and pre-assess-ments done for the journal**
Journal of Graph Algorithms and Applications	X	2
Journal of Discrete Algorithms	X	0
Algorithmica	X	~8
Theoretical Computer science	X	4
International Journal of Computational Geometry and Applications		1
Journal of ACM		1
ACM Transactions on Algorithms		1
Theoretical Computer Science		1
Journal of Algorithms - Algorithms in Cognition, Informatics and Logic		1
Algorithmica		2
Information Processing Letters		2
Journal of Mathematical Modelling and Algorithms		1
SIAM Journal on Discrete Mathematics		1
International Journal of Computational Geometry		1
International Symposium on Algorithms and Computation	X	3
ACM International Symposium on Advances in Geographic Information Systems	X	~10
International Symposium on Algorithms anc Computation	X	~15
Symposium on Combinatorial Pattern Matching	X	9
International Workshop on Experimental Algorithms	X	6
International Workshop on. Experimental Algorithms	X	~10
Other reviewing for major conferences (where not on Program Committee)		~25

b) Assessments, international panels, membership and supervision

Number of staff conducting academic services	**Number of grant and/or career/tenure assessments plus international panels**	**Number of positions and council seats in learned societies**	**Number of Ph.d. and doctor evaluations, national and international**	**Number of primary supervisions of Ph.D and master students**
2	1	3 conference steering committees, one project advisory board, one membership of academy of science and letters	2	12 PhD,6 MS

Appendix E: External funding

Please list all external funding gained by the centreleader or centre members. List only funding gained for centre activities including external PhD-grants. List the full amount and the part of the total amount allocated to reported year.

List **only includes** information for AU employees; it includes all active/new funding in 2008

	Funding body	**Purpose**	**Applicant**	**Activity period**	**Granted amount in DKK**	**Part of total amount allocated to reported**
Public Danish funds	Danish National Advanced Technology Foundation (Højteknologi-fonden)	A platform for Galileo based pervasive positioning	AU (incl Arge), AAU, Danish agricultural advisory service and several companies including Alexandra A/S, Terma A/S, and Systematic	2007-2010	~11.700.000 (AU part)	?
	Strategic Research Council (NABIIT program)	Efficient Handling of Massive Heterogeneous Terrain Data	AU (Arge), DJF and COWI A/S	2006-2010	~5.400.000 (AU part)	~1.700.000
	Science Research Council	Algorithms for Processing Massive Datasets: Theory and Practice	Arge	2004-2008	~5.800.000	~900.000
Private Danish funds						
International funds	NordForsk	Establishing research network between algorithm groups in the Nordic Region	Fedor Fomin (University of Bergen, Norway)	2008-2009	500000	250000

Appendix F: Awards

Please list awards or other acknowledgements gained by centre members in reported year.

List **includes** relevant 2008 information for AU as well as MIT, MPI and FRA researchers.

Awards	**Recipient**	**Granted amount in DKK, if relevant**
Elected member of Royal Danish Academy of Science and Letters	Arge	
Honorary doctor (”doctor honoris causa”) in Science at Aarhus University	Mehlhorn	
International Francqui Chair of Belgium and Francqui Gold Medal	Demaine	~400.000
Katayanagi Emerging Leadership Prize, Carnegie Mellon University and Tokyo University of Technology	Demaine	~60.000
Earle Raymond Hedrick Lecturer, Mathematical Association of America	Demaine	
Featured in 56-minute documentary *Between the Folds*	Demaine	
Université Libre de Bruxelles Gold Medal	Demaine	
Paper on “Near-optimal Hashing...” invited to the “Research Highlights” section of the Communications of the ACM, January 2008.	Indyk	

Appendix G: Public outreach

Please list public outreach activities in media, press, hign schools etc.

List **only includes** 2008 information for AU employees.

a)Electronic media

Specific media (TV, radio, other)	**Date**	**Type of communication (inter-view, commen-tary,**	**Subject/ Title**	**Contributor from the centre**
Videnskab.dk	June 16, 2008	Article	Forskere vil redde os fra data-oversvømmelse	Arge

b)Press

Specific media (Daily newspapers, journals, magazines, other)	**Date**	**Type of communication (inter-view, commen-tary, debate, feature etc)**	**Subject/ Title**	**Contributor from the centre**
Newletter	November 27, 2008	Interview	Daimi topper i programmeringskonkurrence	Mølhave
Weekendavisen	April 25 to 29, 2008	Article	Logaritmen af n	Arge

c)Other (talks at secondary educational institutions etc)

Specific type of communication (presentation/lecturing at open university, high school etc)	**Date**	**Subject/Title**	**Contributor from the centre**
Presentation (Geoforum dagseminar om højdedata)	April 24, 2008	Algoritmer til håndtering af store datamængder	Arge
Presentation (Nordic Geodetic Commission taskforce workshop)	December 11, 2008	Algorithms for Handling Massive (Height) Datasets	Arge
Exhibition: Forskningens Døgn 2008	April 24, 2008	"Oversvømmelse! Hvor? Hvor meget?"	Mølhave, Blunck, Revsbæk
Company lecture, Danske Bank	January 17, 2008	Massive Data Udfordringer	Brodal
Gymnasiepraktik	October 21, 2008	Exercises in Algorithms	Brodal

Appendix H: Patents and applications

List the number of submitted patent applications, granted patents etc. gained by the centre in reported year. Also list possible spin-off companies and collaborations/partners.

Number of submitted patent applications	Number of granted patents	Number of mutually agreed licence, sale and option agreements	Names of spin-off companies established	Other application (please specify)	Partner (s), if any

Appendix I: Publications

Please enclose, in a seperate appendix, a full publication list including all authors dating back from the centre start. The publication list shall be arranged in cronological order and numbered so the most recent publication gets the highest number and is placed at the bottom of the list. Divide the list in types of publications including conference proceedings/presentations as well as master- and PhD-thesis made at the centre. List only accepted publications and don't include publications that are "submitted" or "in press". List only publications authored or co-authored by centre members. For each publication state number, author(s), place and year of publication. If the publication have been peer reviewed please note in brackets "PR". Also note in brackets "CO" if the publiation is co-authored by another national or international research institution.

List **includes** relevant publications for AU as well as MIT, MPI and FRA researchers

Conference proceedings

1	2007	B. Escoffier, G. Moruz and A. Ribichini	Adapting Parallel Algorithms to the W-Stream Model, with Applications to Graph Problems	Proc. International Symposium on Mathematical Foundations of Computer Science (MFCS)	(PR)(CO)
2	2007	S. Guha, P. Indyk and A. McGregor	Sketching Information Divergences	Proc. Annual Conference on Learning Theory (COLT)	(PR)(CO)
3	2007	G. S. Brodal and A. G. Jørgensen	A Linear Time Algorithm for the k Maximal Sums Problem	Proc. International Symposium on Mathematical Foundations of Computer Science (MFCS)	(PR)(CO)
4	2007	G. S. Brodal, L. Georgiadis, K. A. Hansen and I. Katriel	Dynamic Matchings in Convex Bipartite Graphs	Proc. International Symposium on Mathematical Foundations of Computer Science (MFCS)	(PR)(CO)
5	2007	G. Jørgensen, G. Moruz and T. Mølhave	Resilient Priority Queues	Proc. International Workshop on Algorithms and Data Structures (WADS)	(PR)
6	2007	G. S. Brodal, R. Fagerberg, I. Finocchi, F. Grandoni, G. Italiano, A. G. Jørgensen, G. Moruz and T. Mølhave	Optimal Resilient Dynamic Dictionaries	Proc. European Symposium on Algorithms (ESA)	(PR)(CO)

7	2007	P. K. Agarwal, L. Arge, A. Danner, H. Mitasova, T. Mølhave and K. Yi	TerraStream: From Elevation Data to Watershed Hierarchies	Proc. ACM International Symposium on Advances in Geographical Information Systems (ACM-GIS)	(PR)(CO)
8	2007	M. Patrascu and Mikkel Thorup	Planning for Fast Connectivity Updates	Proc. IEEE Symposium on Foundations of Computer Science (FOCS)	(PR)(CO)
9	2007	G. Franceschini, S. Muthukrishnan, and M. Patrascu	Radix Sorting With No Extra Space	Proc. European Symposium on Algorithms (ESA)	(PR)(CO)
10	2007	E. D. Demaine, S. Mozes, B. Rossman and O. Weimann	An Optimal Decomposition Algorithm for Tree Edit Distance	Proc. International Colloquium on Automata, Languages and Programming (ICALP)	(PR)(CO)
11	2007	M. A. Bender, M. Farach-Colton, J. T. Fineman, Y. Fogel, B. C. Kuszmaul and J. Nelson	Cache-Oblivious Streaming B-trees	Proc. ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)	(PR)(CO)
12	2007	E. D. Demaine, M. Ghodsi, M. Hajiaghayi, A. S. Sayedi-Roshkhar and M. Zadimoghaddam	Scheduling to Minimize Gaps and Power Consumption	Proc. ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)	(PR)(CO)
13	2007	M. Patrascu	Lower Bounds for 2-Dimensional Range Counting	Proc. ACM Symposium on Theory of Computing (STOC)	(PR)
14	2007	G. M. Landau, D. Tsur and O. Weimann	Indexing a Dictionary for Subset Matching Queries	Proc. Symposium on String Processing and Information Retrieval (SPIRE)	(PR)(CO)
15	2007	T. Friedrich and D. Ajwani	Average-Case Analysis of Online Topological Ordering	Proc. International Symposium on Algorithms and Computation (ISAAC)	(PR)
16	2007	K. Chang	Multiple pass streaming algorithms for learning mixtures of distributions in R^d	Proc. Algorithmic Learning Theory (ALT)	(PR)

17	2007	M. Westergaard, L. M. Kristensen, G. S. Brodal and L. Arge	The ComBack Method - Extending Hash Compaction with Backtracking	Proc. International Conference on Applications and Theory of Petri Nets and Other Models of Concurrency (ICATPN)	(PR)
18	2007	M. A. Bender, G. S. Brodal, R. Fagerberg, R. Jacob and E. Vicari	Optimal Sparse Matrix Dense Vector Multiplication in the I/O-Model	Proc. ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)	(PR)(CO)
19	2007	A. Golynski, R. Grossi, A. Gupta, R. Raman and S. S. Rao	On the Size of Succinct Indices	Proc. European Symposium on Algorithms (ESA)	(PR)(CO)
20	2007	M. Olsen	Nash Stability in Additively Separable Hedonic Games is NP-hard	Proc. Conference on Computability in Europe (CiE)	(PR)
21	2008	M. Ruzic and P. Indyk	Near-Optimal Sparse Recovery in the L1 norm	Proc. Symposium on Foundations of Computer Science (FOCS)	(PR)(CO)
22	2008	M. Patrascu	(Data) STRUCTURES	Proc. Symposium on Foundations of Computer Science (FOCS)	(PR)
23	2008	M. Patrascu	Succincter	Proc. Symposium on Foundations of Computer Science (FOCS)	(PR)
24	2008	E. Demaine, S. Langerman and E. Price	Confluently Persistent Tries for Efficient Version Control	Proc. Scandinavian Workshop on Algorithm Theory (SWAT)	(PR)(CO)
25	2008	D. Ajwani, I. Malinger, U. Meyer and S. Toledo	Characterizing the Performance of Flash Memory Storage Devices and Its Impact on Algorithm Design	Proc. Workshop on Experimental Algorithms (WEA)	(PR)(CO)
26	2008	U. Meyer	On Dynamic Breadth-First Search in External-Memory	Proc. Symposium on Theoretical Aspects (STACS)	(PR)
27	2008	U. Meyer	On Trade-Offs in External-Memory Diameter Approximation	Proc. Scandinavian Workshop on Algorithm Theory (SWAT)	(PR)

28	2008	G. S. Brodal and A. G. Jørgensen	Selecting Sums in Arrays	Proc.International Symposium on Algorithms and Computation (ISAAC)	(PR)
29	2008	L. Arge, G. S. Brodal and S. S. Rao	External Memory Planar Point Location with Logarithmic Updates	Proc. Symposium on Computational Geometry (SoCG)	(PR)
30	2008	A. Golynski, R. Raman and S. S. Rao	On the Redundancy of Succinct Data Structures	Proc. Scandinavian Workshop on Algorithm Theory (SWAT)	(PR)(CO)
31	2008	M. Olsen	The Computational Complexity of Link Building	Proc. International Conference on Computing and Combinatorics (COCOON)	(PR)
32	2008	M.A. Abam, M. de Berg and J. Gudmundsson	A Simple and Efficient Kinetic Spanner	Proc. Symposium on Computational Geometry (SoCG)	(PR)(CO)
33	2008	L. Arge, M.T. Goodrich, M. Nelson and N. Sitchinava	Fundamental Parallel Algorithms for Private-Cache Chip Multiprocessors	Proc. Symposium on Parallelism in Algorithms and Architectures (SPAA)	(PR)(CO)
34	2008	L. Arge, T. Moelhave and N. Zeh	Cache-Oblivious Red-Blue Line Segment Intersection	Proc. European Symposium on Algorithm (ESA)	(PR)(CO)
35	2008	P.K. Agarwal, L. Arge, T. Moelhave and B. Sadri	I/O-efficient Algorithms for Computing Contour Lines on a Terrain	Proc. Symposium on Computational Geometry (SoCG)	(PR)(CO)
36	2008	J. Feldman, S. Muthukrishnan, A. Sidiropoulos, C. Stein and Z. Svitkina	On Distributing Symmetric Streaming Computations	Proc. Symposium on Discrete Algorithms (SODA)	(PR)(CO)
37	2008	P. Indyk	Explicit Constructions for Compressed Sensing of Sparse	Proc Symposium on Discrete Algorithms (SODA)	(PR)
38	2008	A. Andoni, P. Indyk and R. Krauthgamer	Earth Mover Distance over High-Dimensional Spaces	Proc. Symposium on Discrete Algorithms (SODA)	(PR)(CO)
39	2008	P. Indyk and A. McGregor	Declaring Independence via the Sketching of Sketches	Proc. Symposium on Discrete Algorithms (SODA)	(PR)(CO)
40	2008	K. Onak and A. Sidiropoulos	Circular Partitions with Applications to Visualization and Embeddings	Proc. Symposium on Computational Geometry (SoCG)	(PR)(CO)

41	2008	J. Matousek and A. Sidiropoulos	Inapproximability for metric embeddings into R^d	Proc. Symposium on Foundations of Computer Science (FOCS)	(PR)(CO)
42	2008	N. J. A. Harvey, J. Nelson and K. Onak	Sketching and Streaming Entropy via Approximation Theory	Proc. Symposium on Foundations of Computer Science (FOCS)	(PR)(CO)
43	2008	A. Andoni, D. Croitoru and M. Patrascu	Hardness of Nearest Neighbor under L-infinity	Proc. Symposium on Foundations of Computer Science (FOCS)	(PR)(CO)
44	2008	T. Chan, M. Patrascu and L. Roditty	Dynamic Connectivity: Connecting to Networks and Geometry	Proc. Symposium on Foundations of Computer Science (FOCS)	(PR)(CO)
45	2008	S. Mozes, K. Onak and Oren Weimann	Finding an Optimal Tree Searching Strategy in Linear Time	Proc. Symposium on Discrete Algorithms (SODA)	(PR)(CO)
46	2008	A. Chakrabarti, T.S. Jayram and M. Patrascu	Tight Lower Bounds for Selection in Randomly Ordered	Proc. Symposium on Discrete Algorithms (SODA)	(PR)(CO)
47	2008	E. Demaine, T. Ito, Ni. J. A. Harvey, C. H. Papadimitriou, M. Sideri, R. Uehara and Yushi Uno	On the Complexity of Reconfiguration Problems	Proc. International Symposium on Algorithms and Computation (ISAAC)	(PR)(CO)
48	2008	E. Demaine, G. Aloupis, S. Collette, S. Langerman, V. Sacristan and S. Wuhrer	Reconfiguration of Cube-Style Modular Robots Using O(log n) Parallel Moves	Proc. International Symposium on Algorithms and Computation (ISAAC)	(PR)(CO)
49	2008	E. Demaine, M. Buadoiu, M. Hajiaghayi, A. Sidiropoulos and M. Zadimoghaddam	Ordinal Embedding: Approximation Algorithms and Dimensionality Reduction	Proc. International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)	(PR)(CO)
50	2008	E. Demaine, T. G. Abbott, Z. Abel, D. Charlton, M. L. Demaine and S. D. Kominers	Hinged Dissections Exist	Proc. Symposium on Computational Geometry (SoCG)	(PR)(CO)
51	2008	E. R. Hansen, S. S. Rao and P. Tiedemann	Compressing Binary Decision Diagrams	European Conference on Artificial Intelligence (ECAI)	(PR)(CO)
52	2008	R. Berinde, P. Indyk and M. Ruzic	Practical Near-Optimal Sparse Recovery in the L1 Norm (invited	Proc. Allerton Conference	(CO)

53	2008	R. Berinde, A. Gilbert, P. Indyk, H. Karloff and M. Strauss	Combining Geometry and Combinatorics: A Unified Approach to Sparse Signal Recovery (invited paper)	Proc. Allerton Conference	(CO)
54	2008	M.A. Abam, M. de Berg, and S-H. Poon	Fault-Tolerant Conflict-Free Coloring	Proc. Canadian Conference on Computational Geometry	(CO)

Journals

1	2007	G. S. Brodal, R. Fagerberg and G. Moruz	On the Adaptiveness of Quicksort	ACM Journal of Experimental Algorithmics, 12	(PR) (CO)
2	2008	D. Ajwani, T. Friedrich and U. Meyer	An O(n^{2.75}) Algorithm for Incremental Topological Ordering	ACM Transactions on Algorithms, 4(4)	(PR)
3	2008	M. Stissing, T. Mailund, C. N. S. Pedersen, G. S. Brodal and R. Fagerberg	Computing the All-Pairs Quartet Distance on a set of Evolutionary Trees	Journal of Bioinformatics and Computational Biology, 6(1)	(PR)(CO)
4	2008	L. Arge, M. de Berg, H. J. Haverkort and K. Yi	The Priority R-Tree: A Practically Efficient and Worst-Case Optimal R-Tree	ACM Transactions on Algorithms, 4(1)	(PR)(CO)

Thesis

1	2007	I. Brudaru	Heuristics for Average Diameter Approximation with External Memory	MPI	MS Thesis
2	2007	G. Moruz	Hardware-Aware Algorithms and Data Structures	AU	PhD Thesis
3	2008	M. Patrascu	Lower Bound Techniques for Data Structures	MIT	PhD Thesis
4	2008	A. Sidiropoulos	Computational metric embeddings	MIT	PhD Thesis
5	2008	D. Ajwani	Traversing large graphs in realistic settings	MPI	PhD Thesis
6	2008	K. Do Ba	Testing closeness of distributions under the EMD metric	MIT	MS Thesis
7	2008	K. Lai	Complexity of Union-Split-Find Problems	MIT	MS Thesis

8	2008	J. M. Larsen og M. Nielsen	En undersøgelse af algoritmer til løsning af generalized movers problem i 3D	AU	MS Thesis
9	2008	C. Andersen	An optimal minimum spanning tree algorithm	AU	MS Thesis
10	2008	M. Revsbæk	I/O-efficient Algorithms for Batched Union-Find with Dynamic Set Properties and its Applications to Hydrological Conditioning	AU	MS Thesis
11	2008	A. H. Jensen	I/O-efficient Processing of LIDAR Data	AU	MS Thesis

Other

1	2008	E. Demaine, B. Gassend, J. O'Rourke, and G. T. Toussaint	All Polygons Flip Finitely ... Right?	In "Surveys on Discrete and Computational Geometry: Twenty Years Later", Contemporary Mathematics 453	(CO)
2	2008	A. Andoni and P. Indyk	Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions	Communications of the ACM, 51(1)	(CO)
3	2008	K. Mehlhorn and P. Sanders	Algorithms and Data Structures: The Basic Toolbox	Springer Verlag	(CO)

Total number of publications in reported year	**Peer Rev.**	**Non-peer Rev.**
Number of journal articles	3	0
Number of conference series	32	3
Number of monographs	0	1
Number of book chapters	0	0
Others	0	11

List the 10 most prestigious channels of publication within the centres research area and prioritise in relation to importance and influence if possible.

Since conferences play an equally important role as journals in Computer Science, we indicate 10 conferences as well as 10 journals (the ranking should not be taken too seriously)

Conferences

ACM Symposium on Theory of Computing (STOC)
IEEE Symposium on Foundations of Computer Science (FOCS)
ACM-SIAM Symposium on Discrete Algorithms (SODA)
Symposium on Computational Geometry (SoCG)
International Colloquium on Automata, Languages, and Programming (ICALP)
European Symposium on Algorithms (ESA)
ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)
International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)/ International Workshop on Randomizationand Computation (RANDOM)
Scandinavian Workshop on Algorithm Theory (SWAT)/Workshop on Algorithms and Data Structures (WADS)
Workshop on Algorithm Engineering and Experiments (ALENEX)

Journals

Journal of the ACM
SIAM Journal on Computing
ACM Transactions on Algorithms
Discrete & Computational Geometry
Algorithmica
Journal of Computer and System Sciences
Computational Geometry: Theory and Applications
ACM Journal of Experimental Algorithmics
Theoretical Computer Science
Journal of Discrete Algorithms

State information about any bibliometric analysis the centre or an external part has made in accordance to standards within the centres field of research.

None

MADALGO List of Personel 2008 *01.01.-31.12.2008*

Personel		Hiring	Finansing (fraction of year) *)			Forign	For PhD and
Name	**Position**	**period in 2008 **)**	**Foundation**	**AU**	**Other finansing ***)**	**employee**	**Post Doc: Previous education**
Centerleder							
Lars Arge (AU)	Professor	all period	0,1	0,3	0,6		
Faculty							
Gerth S. Brodal (AU)	Associate Professor	all period		1			
Piotr Indyk (MIT)	Associate Professor	all period	0,1		0,1	x	
Eirk Demaine (MIT)	Associate Professor	all period	0,1		0,1	x	
Kurt Mehlhorn (MPI)	Professor	all period			0,1	x	
Ulrich Meyer (FRA)	Professor	all period			0,3	x	
Henrik Blunck (AU)	Post Doc	-31.10			0,8	x	MS, PhD
Srinivas Rao (AU)	Post Doc	all period	1			x	MS, PhD
Mohammad Abam (AU)	Post Doc	all period	1			x	MS, PhD
Peter Hachenberger (AU)	Post Doc	01.07-	0,5			x	MS, PhD
Deepak Ajwani (AU)	Post Doc	01.10-	0,3			x	M.TECH, PhD
Kevin L. Chang (MPI)	Post Doc	-30.09			0,8	x	MS, PhD
Gabriel Moruz (FRA)	Post Doc	all period			1	x	MS, PhD
Guests							
Bradford G. Nickerson	Professor	-30.06			0,5	x	
Nodari Sitchinava	PhD student	01.8-15.09	0,1		0,1	x	
Muriel Dulieu	PhD student	01.7-31.08			0,2	x	
Jeremy T. Fineman	PhD student	01.06-30.06			0,1	x	
John Iacono	Associate Professor	01.07-31.08	0,1		0,1	x	
Andrew Danner	Assistant Professor	01.08-31.08	0,1			x	
Tehnical staff							
Kasper D. Larsen	Programmer	all period			1		
Anders H. Jensen	Programmer	01.09-30.09	0,1				
Adam E.N. Thomasen	Programmer	01.04-			0,8		
Thor Siiger Prentow	Programmer	01.11-	0,2				
Administrative staff							
Else Magård	Center manager	all period	1				
Ellen Lindstrøm	Regnskabsmedarbejder	all period		0,5			
Ph.d.-studerende							
Thomas Mølhave (AU)	PhD student	all period			1		BS + 1 år
Martin Olsen (AU)	PhD student	all period		0,4	0,6		Cand. Scient
Allan G. Jørgensen (AU)	PhD student	all period	0,9		0,1		BS + 1 år
Anders H. Jensen (AU)	PhD student	-30.07			0,6		BS + 1 år
Lasse Deleuran (AU)	PhD student	all period		1			BS + 1 år
Kostas Tsakalidis (AU)	PhD student	all period		1		x	BS + 1 år
Jesper Eshøj (AU)	PhD student	01.02-		0,9			BS + 1 år
Morten Revsbæk (AU)	PhD student	01.02-			0,9		Cand. Scient
Mark Greve (AU)	PhD student	01.02-		0,9			BS + 1 år
Pooya Davoodi (AU)	PhD student	01-05-	0,7			x	MS
Jacob Truelsen (AU)	PhD student	01.08-		0,4			BS + 1 år
Kasper D. Larsen (AU)	PhD student	01.08-		0,4			BS
Deepak Ajwani (MPI)	PhD student	-30.09			0,8	x	M.TECH.
Andreas Beckmann (MPI/FRA)	PhD student	all period	1			x	MS
Oren Weimann (MIT)	PhD student	all period			1	x	MS
Mihai Patrascu (MIT)	PhD student	-31.05	0,4			x	BS
Anastasio Sidiropoulos (MIT)	PhD student	-31.05			0,4	x	BS
Khan Do Ba (MIT)	PhD student	all period	0,3		0,7	x	BS
Jelani Nelsen (MIT)	PhD student	all period			1	x	MS

*) Approximation. Max one decimal.

**) More then three weeks.

***) Including no financing.