Statement of Purpose

Mihai Pătraşcu

My long and enjoyable interaction with computer science goes back one and a half decades, to the moment when I first learned to program. Before college, my main motivation in studying computer science was participating in computer olympiads. Among my achievements there were winning the first prize at the Romanian National Olympiad 9 years in a row, and obtaining 7 international medals (among which, 2 gold and 1 silver medal at IOI).

In college, I was naturally attracted to research in theoretical computer science. I have been working in this area for three years, under the supervision of Erik Demaine. Below, I touch on some of the contributions I made during this time. My main research interests are related to complexity in concrete models of computation (lower bounds), as well as advanced data structures and algorithms.

My plans for a PhD are centered around research in theoretical computer science. After the PhD, I will likely want to obtain a position in the academia. In this, I am motivated by my good experiences with teaching, including as a teaching assistant for a new graduate course at MIT.

Concrete Complexity. My most extensive contributions to date have been in the realm of dynamic cell-probe complexity, through a sequence of papers published in SICOMP, STOC, SODA and ICALP. The cell-probe model is a strong nonuniform model of computation, used for analyzing static or dynamic data-structure problems. For dynamic problems, lower bounds have been proved using the chronogram technique of Fredman and Saks, dating back to STOC'89. In that paper, a lower bound of $\Omega(\lg n / \lg\lg n)$ was derived, where n is the number of bits in the problem representation. Despite a flurry of works showing similar lower bounds for various problems, no higher lower bound could be proved *for 15 years*, and this limitation was recognized in papers and surveys as the *central open problem* of the field.

In papers with Erik Demaine appearing in SICOMP, STOC and SODA, we showed $\Omega(\lg n)$ lower bounds for maintaining partial sums and dynamic connectivity, breaking this long-standing barrier. Our bound demonstrates the optimality of the folklore solution to the partial-sums problem (augmented binary trees), which is quintessential of dynamic computation. Despite intense study, a tight bound was not even known in weaker algebraic models. Our bound for dynamic connectivity proves the optimality of several dynamic graph algorithms, including the famous dynamic trees of Sleator and Tarjan.

My work on these problems was recognized by the Computer Research Association's *award for best undergraduate research* in 2004. Interestingly, our original approach seemed entirely different from the chronogram technique. However, in joint work with Corina Tarniţă (Pătraşcu), we showed that this technique is equivalent to a subtle variation of the chronogram technique. Using this better understanding, we offered an almost quadratic improvement in the best lower bound for the bit-probe model, solving the *first open problem in a survey* by Miltersen. Our work received the *best student paper* award in ICALP.

In a recently submitted paper with Mikkel Thorup, we achieved a *break-through in static cell-probe complexity*. So far, there was essentially one known technique for proving space–time trade-offs for static data structures: reduction to asymmetric communication complexity. However, it is known that this approach cannot prove superconstant lower bounds for the most natural setting of parameters: the query and a machine word have $O(\lg n)$ bits. In addition, communication complexity cannot differentiate polynomial factors in the space, while for most natural problems, the interesting behavior occurs inside the polynomial domain. We prove the first lower bound which breaks the communication barrier, and does not suffer from these limitations. A fundamental implication of our result is the *first separation between polynomial and near-linear space* (any space $n^{1+o(1)}$). Our bounds give a complete understanding of predecessor search, one of the most fundamental and well-studied problems. A surprising conclusion is that the famous data structure of van Emde Boas is optimal for quasilinear space, and in the dynamic case. Another interesting conclusion applies to the external-memory model: it is always optimal to either use the classic, comparison-based B-trees, or use the best RAM solution, which ignores the benefits of external memory.

These results open the door to many interesting problems in cell-probe complexity, which I plan to investigate. In the dynamic case, one could hope to prove polylogarithmic lower bounds (e.g., for range queries in constant dimension) or even $n^{\Omega(1)}$ (e.g., for dynamic problems in directed graphs). In both cases, these problems have been studied extensively on the upper-bound side, but we cannot hope to understand them without progress on the lower bounds. In the static scenario, one could ask for much higher lower bounds, now that we are not limited to communication complexity. In particular, it would be interesting to prove bounds demonstrating the "curse of dimensionality" which is conjectured to hold for problems of critical importance.

Though my results so far have centered around the cell-probe model, I maintain an active interest in analyzing other strong models of computation, such as circuits and branching programs. It is quite likely that the information-theoretic tools and intuition that I have employed in the cell-probe model will also prove useful in other contexts. As an illustration, in joint work with Adler, Demaine and Harvey to appear in SODA, we used tools from communication complexity to analyze information transmission across asymmetric channels. This problem has been investigated extensively in sensor networks, and many protocols have been proposed. We proved the first lower bounds for this problem, which almost match the behavior of the best known solution.

Data Structures and Algorithms. My early training as programmer and competitor in computer olympiads have naturally given me a strong appreciation for algorithms. Despite my work in complexity, I find my instinctive patterns of reasoning are algorithmic.

One of my most influential papers, appearing in SICOMP and FOCS, concerns competitiveness of binary search trees. The famous dynamic optimality conjecture of Sleator and Tarjan asserts that splay trees are $O(1)$-competitive. However, no competitive ratio better than the trivial $O(\lg n)$ has been proved for splay trees or any other binary search tree, in *over two decades*. In joint work with Demaine, Harmon and Iacono, we described a new search tree which is provably $O(\lg \lg n)$-competitive. Of course, this result leaves two important open problems: are there $O(1)$-competitive search trees? and are splay trees $o(\lg n)$-competitive?

An important area of study in modern data structures is concerned with integer search problems. The van Emde Boas recursion is probably the most well-known result from the field, and its elegance helps motivate the field in general. For the predecessor problem, this algorithm is shown to be tight by my recent work with Mikkel Thorup mentioned above. However, for dynamic range reporting in one dimension, this turns out to not be the case. In joint work with Mortensen and Pagh appearing in STOC, we developed a *fundamental new recursive idea*, yielding a surprising exponential improvement in the query time. As opposed to van Emde Boas, who applies a binary search on paths of a trie, we apply a more complex recursion (similar to van Emde Boas search itself) on the paths. Nonetheless, the algorithm is remarkably clean and elegant.

Recently, I have been very interested in hashing and its applications. Our STOC paper on range reporting mentioned above needs to develop a *surprising hashing primitive*: a data structure which maintains a perfect hash function on a dynamic set, using sublinear memory (thus, without actually remembering the set). Tight upper and lower bounds for the space are developed in my subsequent LATIN paper with Demaine, Meyer auf der Heide and Pagh. An essential ingredient we develop is a dynamic dictionary which is simultaneously compact (uses asymptotically optimal space), and takes constant time per operation with high probability. Previous dictionaries could only achieve one of this desiderates. Using another set of hashing ideas, my WADS paper with Baran and Demaine achieves the first subquadratic algorithms for the famous 3SUM problem, exploiting the "parallelism" of the RAM or external memory models (given by bit-packing, respectively, larger memory pages).

There are many interesting open problems related to hashing, that I would like to investigate. Perhaps the most fundamental is the performance of deterministic dictionaries, which is one of the main uses of randomness in computation. Other interesting questions are related to permutation hash families, which also play a major role in cryptography. In the LATIN paper mentioned above, we develop an intriguing family of permutation hash functions, which fails to be k-wise independent for large k, yet has similar concentration bounds.

I also hold an interest in algorithmic number theory, and have three published results in the area. Additionally, in an ongoing collaborative research project, we are looking at the problem of counting primitive lattice points in planar shapes. This is an exciting question at the intersection of geometry and number theory, and has a long tradition on the mathematical side, dating back to Gauss. Our algorithm applies to polygons, and is significantly faster than previous methods for exact counting. In a paper with Corina Tarniţă (Pătraşcu) published in ANTS, we have already described a fast algorithm for a certain class of triangles. We used this to construct algorithms for rank and select queries in the Farey sequence, which are quadratically faster than enumerating the sequence.

Teaching. I believe teaching is an integral part of doing research. A discovery is far from complete if one does not also find a way to present it to others. Even more importantly, organizing a vast array of result for presentation is a crucial skill that a researcher must share with a teacher, since without it, the researcher cannot gain a clear sense of direction in his work.

My early experiences with students came as a member of the scientific committee in Romanian national olympiads and one Balkan olympiad. There, one has to create problems which are original and elegant, while gauging the level of difficulty to select the best from a group of talented students. This requires perhaps the most elusive skill of a teacher: entering the students' minds to judge difficulty based on their abilities and a 5-hour time frame. While this is not a skill I can ever hope to fully master, contest result showed my problems were relevant, and my contributions were praised by the senior members of the committees.

My most significant and enjoyable teaching experience was as a teaching assistant for a graduate course on advanced data structures, taught by Erik Demaine. I created and graded the problem sets, and taught four lectures. Yet, the most interesting aspect was working with Erik to "create" the course from scratch. We had to decide what broad topics should be covered, and how to best present each topic. Given such an old and diverse field, this was a very challenging, but intellectually rewarding task. It was especially encouraging to receive feedback from people at other universities, who declared themselves impressed by the simultaneous breadth and coherence of the course.

Conclusions. I look forward to continuing my research career as a PhD student. Above are a few of the open problems that motivate me, and I will continue to work on. In addition, as has happened frequently at MIT, interacting with members of your theory group will give me an invaluable opportunity to broaden my horizons, and work in many research areas that I cannot anticipate at the moment. Given my background, I believe I am in a good position to make crucial contributions in such pursuits.