Parallel Techniques for Computational Geometry

Mikhail J. Atallah
Purdue University, mja@cs.purdue.edu

Report Number:
90-1020
PARALLEL TECHNIQUES FOR
COMPUTATIONAL GEOMETRY

Mikhail J. Atallah

Department of Computer Sciences
Purdue University
West Lafayette, IN 47907

CSD TR #1020
Parallel Techniques for Computational Geometry

Mikhail J. Atallah

Department of Computer Science
Purdue University
West Lafayette, IN 47907.

Abstract

A survey of techniques for solving geometric problems in parallel is given, both for shared memory parallel machines and for networks of processors. Open problems are also discussed, as well as directions for future research.

*This work was supported by the Office of Naval Research under Contracts N00014-84-K-0502 and N00014-86-K-0689, the Air Force Office of Scientific Research under Grant AFOSR-90-0107, the National Science Foundation under Grant DCR-8451393, and the National Library of Medicine under Grant R01-LM05118.
1 Introduction

Many of the problems in computational geometry come from applications in pattern recognition, computer graphics, statistics, operations research, computer-aided design, robotics, etc. The problems which arise in these areas can come from real-time applications, and hence need to be solved as fast as possible. For many of these problems, however, we are already at the limits of what can be achieved through sequential computation. Such sequential methods can be inadequate for situations in which the input consists of a large number of geometric objects. Thus, it is natural to study what kinds of speed-ups can be achieved through parallel computing. As an indication of the importance of this research direction, we note that four of the eleven problems used as benchmark problems to evaluate parallel architectures for the DARPA Architecture Workshop Benchmark Study of 1986 were computational geometry problems.

Unfortunately, many of the techniques used to find efficient sequential algorithms for computational geometry problems do not translate well into a parallel setting. That is, while providing elegant paradigms for designing sequential algorithms, these techniques use methods which seem to be inherently sequential. Therefore, one needs to develop new paradigms for computational geometry, paradigms better suited for a parallel processing environment. This article is a survey of the main known algorithmic techniques for solving computational geometry problems efficiently in parallel. Since the focus is on general algorithmic techniques rather than on specific problems, no attempt is made to list all of the known parallel complexity bounds for geometric problems (there are too many of them).

The rest of the paper is organized as follows. Section 2 briefly reviews parallel models, Section 3 discusses basic subproblems that tend to arise in the solution of geometric problems on any parallel model, Section 4 discusses PRAM techniques, Section 5 discusses techniques for mesh-connected arrays of processors, Section 6 deals with the hybrid RAM/ARRAY model and its connection to I/O complexity, Section 7 mentions some experimental work, and Section 8 concludes.

2 Parallel Models

This section briefly reviews the models of parallel computation for which parallel geometric algorithms have been designed.
2.1 PRAM Models

The PRAM (Parallel Random Access Machine) model of parallel computation is the shared-memory model where the processors operate synchronously. A *step* in a PRAM consists of each processor reading the content of a cell in the shared memory, writing something in a cell of the shared memory, or performing a computation within its own registers. Thus all communication is done via the shared memory. The PRAM comes in many flavors. The CREW (Concurrent Read Exclusive Write) version of this model allows many processors to simultaneously read the content of a memory location, but forbids any two processors from simultaneously attempting to write in the same memory location (even if they are trying to write the same thing). The CRCW (Concurrent Read Concurrent Write) version of the PRAM differs from the CREW one in that it also allows many processors to write simultaneously in the same memory location: in any such common-write contest, only one processor succeeds, but it is not known in advance which one. (There are other versions of the CRCW-PRAM but we shall not concern ourselves with these here.) The EREW-PRAM is the weakest version of the PRAM: it forbids both concurrent reading and concurrent writing.

The PRAM has so far been the main vehicle used to study the parallel algorithmics of geometric problems, and much of this survey (Section 4) will deal with PRAM techniques.

2.2 Networks of Processors

A network of processors is modeled as a graph where the nodes represent processors and the edges represent communication lines. All the network models we consider are synchronous, and a *step* of such a network of processors consists either of each processor communicating with a neighbor by sending/receiving the contents of a register (a *data movement step*), or of each processor performing a computation within its own registers (a *computation step*). We next briefly review some network models.

2.2.1 The Mesh

In a $d$-dimensional mesh of processors, the processors operate synchronously and are positioned on an $h_1 \times \ldots \times h_d$ grid, one processor per grid point. A processor is denoted by its position in the grid, a typical one being denoted by $(i_1, \ldots, i_d)$ where $1 \leq i_k \leq h_k$ for every $k \in \{1, \ldots, d\}$. Processors $(i_1, \ldots, i_d)$ and $(j_1, \ldots, j_d)$ are neighbors if and only
if \(|i_1 - j_1| + |i_2 - j_2| + \cdots + |i_d - j_d| = 1\). Note that a processor cannot have more than 2d neighbors (processors at the boundary have fewer). A processor has a fixed (i.e., \(O(1)\)) number of storage registers. Some researchers assume that a register can store up to \(\log n\) bits, while others limit the size of a register to \(O(1)\) bits: here we assume the former model.

2.2.2 The Hypercube

Every processor in the \(k\)-dimensional hypercube \(H\) is labeled as \(b_0b_1\ldots b_{k-1}\), where \(b_i \in \{0,1\}\) for \(0 \leq i \leq k-1\). A processor with label \(b_0b_1\ldots b_{k-1}\) is connected to \(k\) processors, having labels \(b_0b_1\ldots \bar{b}_s\ldots b_{k-1}\), for \(0 \leq s \leq k-1\) (where \(\bar{b}_s\) denotes the complement of \(b_s\)). An edge \((v_1, v_2)\) of \(H\) is said to be of dimension \(s\), if \(v_1\) and \(v_2\) differ in bit position \(s\), i.e., \(v_1 = b_0b_1\ldots b_s\ldots b_{k-1}\) and \(v_2 = b_0b_1\ldots \bar{b}_s\ldots b_{k-1}\).

2.2.3 Other Network Models

Some geometric algorithms were designed for a number of other networks, which we shall not cover in any detail. These include the tree of processors, the pyramid, and the mesh of trees. Although the general algorithmic techniques for solving geometric problems on these models can be quite similar to the techniques used for other models (like the mesh), there are significant differences in the way processors communicate (these networks have smaller diameter than the mesh). Generally speaking, the tree of processors has been used more for parallel information storage and retrieval than for solving geometric problems. The pyramid has been used mostly for image processing applications. For the reader interested in learning more about these and other network models, see the forthcoming books by Leighton [94] and by Miller and Stout [98].

2.3 Hybrid Models

These are models consisting of more than one type of machine, and the main one for which geometric problems have been considered consists of a sequential computer to which a mesh is attached. We postpone the description of this model until Section 6.

Although geometric algorithms have been designed for all of the network models mentioned above, there are far fewer geometric algorithms for network models than for PRAMs. Furthermore, among the network models, more geometric algorithms have been designed for the mesh than for any of the other network models, perhaps because the parallel complexity of such basic operations as sorting and list ranking is well understood for the mesh.
Other models like the hypercube are just as important, but the complexity of the most basic operations on them is still open. For this reason, among all the network models, we shall focus on the mesh (in Section 5) and on hybrid variants of the mesh (in Section 6). We also briefly discuss the connections between these and the I/O complexity of geometric problems (in Section 6).

3 Basic Subproblems

This section reviews some basic subproblems that are ubiquitous in the design of parallel geometric algorithms, no matter what parallel model is used. In many models the complexity of these basic subproblems is well understood, but for some models (like the hypercube) the complexity of some of these (like sorting and list-ranking) is still open, and in such situations no final statement about the complexity of the most common geometric problems can be made until these issues are resolved (especially since many geometric problems are related to sorting). These basic operations are reviewed below.

3.1 Sorting, Merging

Sorting is probably the most frequently used subroutine in parallel geometric algorithms. Fortunately, for PRAM models and for the mesh, we know how to sort optimally: $O(\log n)$ time and $n$ processors on the EREW-PRAM [41, 8], $O(\sqrt{n})$ time on a $\sqrt{n} \times \sqrt{n}$ mesh [95, 105, 87]. The parallel complexity of sorting on the hypercube is not known (the current best hypercube bound is $O(\log n(\log \log n)^2)$ with $n$ processors [49]). On the mesh, the complexity of merging is the same as that of sorting, but on the hypercube and PRAM it is easier than sorting [117, 127, 29, 79]; it is $O(\log n)$ time on an $n$-processor hypercube, and on the PRAM it is $O(\log \log n)$ time with $n$ processors or, alternatively, $O(\log n)$ time with $n/\log n$ processors.

3.2 Parallel Prefix

Given an array $A$ of $n$ elements and an associative operation $+$, the parallel prefix problem is that of computing the array $B$ of $n$ elements such that $B(i) = \sum_{k=1}^{i} A(k)$. Parallel prefix can be solved in $O(\log n)$ time and $n/\log n$ processors on an EREW-PRAM [90], $O(\log n/\log \log n)$ time and $n\log \log n/\log n$ processors on a CRCW-PRAM [45], $O(\sqrt{n})$ time on the mesh (trivial), and in $O(\log n)$ time on an $n$-processor hypercube (trivial). Computing the smallest element in the $A$ array is a special case of parallel prefix; for the
CRCW model it can be done faster than general parallel prefix—in $O(\epsilon^{-1})$ time with $n^{1+\epsilon}$ processors for any positive constant $\epsilon$ or, alternatively, in $O(\log \log n)$ time with $n/\log \log n$ processors [117].

3.3 List Ranking

List ranking is a more general version of the parallel prefix problem: the elements are given as a linked list, i.e., we are given an array $A$ each entry of which contains an element as well as a pointer to the entry of $A$ containing the predecessor of that element in the linked list. The problem is to compute an array $B$ such that $B(i)$ is the "sum" of the first $i$ elements in the linked list. This problem is considerably harder than the previous one, and most tree computations as well as many graph computations reduce, via the Euler tour technique [122], to solving that problem. EREW-PRAM algorithms that run in $O(\log n)$ time and $n/\log n$ processors are known [44, 101]. An $O(\sqrt{n})$ time mesh algorithm is also known [22]. Its complexity on the hypercube is still an open problem.

3.4 Tree Contraction

Given a (not necessarily balanced) rooted tree, the problem is to reduce it to a single node by a sequence of node removals, where a node $v$ can be removed if it is not the root and either (i) it is a leaf, or (ii) it has only one child. In case (ii) the removal of $v$ is accomplished by "bypassing it", i.e., making $v$'s child the child of $v$'s parent. In a parallel setting, many nodes can be removed simultaneously so long as they are independent, in the sense that the parent of a node being removed cannot be removed at the same time. This problem is an abstraction of many other problems, including that of evaluating an arithmetic expression tree [102]. Many elegant optimal EREW-PRAM algorithms for it are known [1, 44, 71, 84], running in $O(\log n)$ time with $n/\log n$ processors. It is easy to implement these in $O(\sqrt{n})$ time on a $\sqrt{n} \times \sqrt{n}$ mesh by using the techniques in [22].

The above list of basic subproblems is not exhaustive in that (i) many techniques that are basic for general combinatorial problems were omitted (we have focused only on those most relevant to geometric problems rather than to general combinatorial problems), and (ii) among the techniques applicable to geometric problems we have postponed covering the more specialized ones (they tend to be model-dependent).
4 PRAM Techniques

The PRAM has been extensively used in theoretical studies as a uniform vehicle for designing parallel algorithms. The PRAM is generally considered to be a rather unrealistic model of parallel computation. However, although there are no PRAMs commercially available, algorithms designed for PRAMs can often be efficiently simulated on some of the more realistic parallel models. The PRAM enables the algorithm designer to focus on the structure of the problem itself, without being distracted by architecture-specific issues. Another advantage of the PRAM is that, if one can give strong evidence (in the sense explained in the next subsection) that a problem has no fast parallel solution on the PRAM, then there is no point in looking for a fast solution to it on more realistic parallel models (since these are weaker than the PRAM).

4.1 Inherently Sequential Geometric Problems

A parallel algorithm is said to run in polylogarithmic time if its time complexity is $O(\log^k n)$, where $n$ is the problem size and $k$ is a constant independent of $n$ (i.e., $k = O(1)$). A problem solvable in polylogarithmic time using a polynomial number of processors is said to be in the class NC. It is strongly believed (but not proved) that not all problems solvable in polynomial time sequentially are solvable in polylogarithmic time using a polynomial number of processors (i.e., it is believed that $P \neq NC$). As in the theory of NP-completeness, there is an analogous method for showing that a particular problem is probably not in NC: by showing that the membership of that problem in NC would imply that $P = NC$. Such a proof consists of showing that each problem in $P$ admits an NC reduction to the problem at hand (an NC reduction is a reduction that takes polylogarithmic time and uses a polynomial number of processors). Such a problem is said to be $P$-complete. For a more detailed discussion of the class NC and parallel complexity theory, see (for example) [108] or [83]. A proof establishing P-completeness of a problem is viewed as strong evidence that the problem is "inherently sequential". Most of the problems shown to be P-complete to date are not geometric (most are graph or algebra problems). This is no accident: geometric problems in the plane tend to have enough structure to enable membership in NC. Even the otherwise P-complete problem of linear programming [65, 66] is in NC when restricted to the plane. In the rest of this subsection we mention the (very few) planar geometric problems that are known to be P-complete, and also a problem that is conjectured to be
P-complete. Each of the problems known to be P-complete involves a collection of line segments in the plane.

- **Plane-sweep triangulation.** One is given a simple $n$-vertex polygon $P$ (which may contain holes) and asked to produce the triangulation that would be constructed by the following sequential algorithm: sweep the plane from left to right with a vertical line $L$, such that each time $L$ encounters a vertex $v$ of $P$ one draws all diagonals of $P$ from $v$ that do not cross previously drawn diagonals. This problem is a special case of the well-known polygon triangulation problem (see [67, 110]), and it clearly has a polynomial time sequential solution.

- **Weighted planar partitioning.** One is given a collection of $n$ non-intersecting segments in the plane, such that each segment $s$ is given a distinct weight $w(s)$, and asked to construct the partitioning of the plane produced by extending the segments in order of their weights. The extension of a segment “stops” at the first segment (or segment extension) that is “hit” by the extension. This problem has applications to “art gallery problems” [48, 106], and is P-complete even if there are only 3 possible orientations for the line segments. It is straightforward to solve it sequentially in $O(n \log^2 n)$ time (using the dynamic point-location data structure of [111]), and in $O(n \log \log n)$ time by a more sophisticated method [48].

- **Visibility layers.** One is given a collection of $n$ non-intersecting segments in the plane, and asked to label each segment by its “depth” in terms of the following layering process (which starts with $i = 0$): find the segments that are (partially) visible from $(0, +\infty)$, label each such segment as being at depth $i$, remove each such segment, increment $i$, and repeat until no segments are left. This is an example of a class of problems in computational geometry known as layering problems or onion peeling problems [35, 92, 107], and is P-complete even if all the segments are horizontal.

The P-completeness proofs of the above problems were given in [13]; for the third problem see also [78]. The proofs consist of giving NC reductions from the monotone circuit value problem and planar circuit value problem, which are known to be P-complete [70, 88, 103]. These reductions typically involve the use of geometry to simulate a circuit, by using the relative positions of objects in the plane.
Perhaps the most famous open problem in the area of geometric P-completeness is that of the convex layers problem [35]: given $n$ points in the plane, mark the points on their convex hull as being layer zero, then remove layer zero and repeat the process, generating layers 1, 2, ..., etc. In view of the P-completeness of the above-mentioned visibility layers problem, it is reasonable to conjecture that the convex layers problem is also P-complete.

### 4.2 "Fast" and "Efficient"

Once one has established that a geometric problem is in NC, the next step is to design a PRAM algorithm for it that runs as fast as possible, while being efficient in the sense that it uses as few processors as possible. Ideally, the parallel time complexity should match the lower bound (assuming such a lower bound is known), and the time $\times$ processors product should match the sequential time complexity of the problem. A parallel lower bound for a geometric problem is usually established by showing that it can be used to solve some other (perhaps non-geometric) problem having that lower bound. For example, it is well known [47] that computing the logical OR of $n$ bits has an $\Omega(\log n)$ time lower bound on a CREW-PRAM. This can easily be used to show that detecting whether the boundaries of two convex polygons intersect also has an $\Omega(\log n)$ time lower bound on that same model, by encoding the $n$ bits whose OR we wish to compute in two concentric regular $n$-gons such that the $i$th bit governs the relative positions of the $i$th vertices of the two $n$-gons. Interestingly, if the word "boundaries" is removed from the previous sentence then the lower bound argument falls apart and it becomes possible to solve the problem in constant time on a CREW-PRAM, even using a sublinear number of processors [20, 128].

Before reviewing the techniques that have resulted in many PRAM geometric algorithms that are fast and efficient in the above sense, a word of caution is in order. From a theoretical point of view, the class NC and the requirement that a "fast" parallel algorithm should run in polylogarithmic time, are eminently reasonable. But from a more practical point of view, not having a polylogarithmic time algorithm does not entirely doom a problem to being "non-parallelizable". One can indeed argue [126] that a problem of sequential complexity $\Theta(n)$ and that is solvable in $O(\sqrt{n})$ time by using $\sqrt{n}$ processors is "parallelizable" in a very real sense, even if no polylogarithmic time algorithm is known for it.
4.3 Divide and Conquer

The sequential divide and conquer algorithms that have efficient PRAM implementations are those for which the "conquer" step can be done extremely fast (e.g., in constant time). Take, for example, an \( O(n \log n) \) time sequential algorithm that works by recursively solving two problems of size \( n/2 \) each, and then combining the answers they return in linear time. In order for a PRAM implementation of such an algorithm to run in \( O(\log n) \) time with \( n \) processors, the \( n \) processors must be capable of performing the "combine" stage in constant time. For some geometric problems this is indeed possible (these include the convex hull problem [20, 128]). The time and processor complexities then obey the recurrences

\[
T(n) \leq T(n/2) + c_1,
\]

\[
P(n) \leq \max\{n, 2P(n/2)\},
\]

with boundary conditions \( T(1) \leq c_2 \) and \( P(1) = 1 \), where \( c_1 \) and \( c_2 \) are constants. These imply that \( T(n) = O(\log n) \) and \( P(n) = n \).

But for many problems, such an attempt at implementing a sequential algorithm fails because of the impossibility of performing the "conquer" stage in constant time. For these, the next two approaches often work.

4.4 "Rootish" Divide and Conquer

By "rootish" we mean partitioning the problem into \( n^{1/k} \) subproblems to be solved recursively in parallel, for some constant integer \( k \) (usually, \( k = 2 \)). For example, instead of dividing the problem into two subproblems of size \( n/2 \) each, we divide it into (say) \( \sqrt{n} \) subproblems of size \( \sqrt{n} \) each, which we recursively solve in parallel. That the conquer stage takes \( O(\log n) \) time (assuming it does) causes no harm with this subdivision scheme, since the time and processor recurrences in that case would be

\[
T(n) \leq T(\sqrt{n}) + c_1 \log n,
\]

\[
P(n) \leq \max\{n, \sqrt{n}P(\sqrt{n})\},
\]

with boundary conditions \( T(1) \leq c_2 \) and \( P(1) = 1 \), where \( c_1 \) and \( c_2 \) are constants. These imply that \( T(n) = O(\log n) \) and \( P(n) = n \).

The problems that can be solved using rootish divide and conquer include the convex hull [2, 19], the visibility of nonintersecting planar segments from a point [28], and the
visibility of a polygonal chain from a point [15]. The scheme is useful in various ways and forms, and sometimes with recurrences very different from the above-mentioned ones. For example, it was used in the form of a fourth-root divide and conquer to obtain (in a rather involved way) an optimal EREW algorithm for the visibility of a simple polygon from a point [15] (that is, \(O(\log n)\) time with \(n/\log n\) processors).

There are instances where one has to use a hybrid of two-way divide and conquer and rootish divide and conquer, in order to obtain the desired complexity bounds. For example, in [15], the recursive procedure takes two parameters (one of which is problem size) and uses either fourth-root divide and conquer or two-way divide and conquer, depending on the relative sizes of these two input parameters.

4.5 Cascading

This sampling and iterative refinement method was introduced by Cole [41] for the sorting problem, and was further developed in [16, 72] for the solution of geometric problems. It has proved to be a fundamental technique, one that enables optimal solutions when most other techniques fail. Its details are intricate even for sorting, but the gist of it can easily be described. Since the technique works best for problems that are solved sequentially by divide and conquer, we use such a hypothetical problem to illustrate the discussion: consider an \(O(n \log n)\) time sequential algorithm that works by recursively solving two subproblems of size \(n/2\) each, followed by an \(O(n)\) time conquer stage. Let \(T\) be the tree of recursive calls for this algorithm, i.e., a node of this recursion tree at height \(h\) corresponds to a subproblem of size equal to the number of leaves in its subtree (= \(2^h\)). A "natural" way of parallelizing such an algorithm would be to mimic it by using \(n\) processors to process \(T\) in a bottom up fashion, one level at a time, completing level \(h\) before moving to level \(h + 1\) of \(T\) (where by "level \(h\)" we mean the set of nodes of \(T\) whose height is \(h\)). Such a parallelization will yield an \(O(\log n)\) time algorithm only if processing each level can be done in constant time. It can be quite nontrivial to process one level in constant time, so this natural parallelization can be challenging. However, it is frequently the case that processing one level cannot be done in constant time, and it is precisely in these situations that the cascading idea can be useful. In order to be more specific when sketching this idea, we assume that the hypothetical problem being solved is about a set \(S\) of \(n\) points, with the points stored in the leaves of \(T\).

In a nutshell, the general idea of cascading is as follows. The computation proceeds
in a logarithmic number of stages, each of which takes constant time. Each stage involves activity by the n processors at more than one level, so the computation diffuses up the tree T, rather than working on only one level at a time. For each node v ∈ T, let h(v) be the height of v in T, L(v) be the points stored in the leaves of the subtree of v in T, and let I(L(v)) be the information we seek to compute for node v (the precise definition of I(·) varies from problem to problem). The ultimate goal is for every v ∈ T to compute the I(L(v)) array. Each v ∈ T lies “dormant” and does nothing until the stage number exceeds a certain value (usually h(v)), at which time node v “wakes up” and starts computing, from stage to stage, I(L′) for a progressively larger subset L′ of L(v), a subset L′ that (roughly) doubles in size from one stage to the next of the computation. I(L′) can be thought of as an approximation of the desired I(L(v)), an approximation that starts out being very rough (when L′ consists of, say, a single point) but gets repeatedly refined from one stage to the next. When L′ eventually becomes equal to L(v), node v becomes inactive for all future stages (i.e., it is done with its computation, since it now has I(L(v))). There are many (often intricate) implementation details that vary from one problem to the next, and many times the scheme substantially deviates from the above rough sketch, but our purpose was only to give the general idea of cascading.

The cascading technique has been used to solve many problems (not just geometric ones). Some of the geometric problems for which it has been used are:

- **Fractional cascading.** Given a directed graph G in which every node v contains a sorted list C(v), construct a linear space data structure (that is, one whose size is at most a constant factor larger than the space taken by the input) that enables one processor to quickly locate any x in all the sorted lists stored along a given path (v1, v2, . . . , vk) in G (by “quickly” we mean in O(log |C(v1)| + k) time). This problem was introduced by Chazelle and Guibas [37] who gave an elegant optimal sequential algorithm. An optimal O(log n) time and n/ log n processor parallel algorithm for this problem was given in [16].

- **Trapezoidal decomposition.** Given a set S of n planar line segments, determine for each segment endpoint p the first segment encountered by starting at p and walking vertically upward (or downward). An O(log n) time and n/ log n processor CREW-PRAM algorithm is known [16]. This implies similar bounds for the polygon triangulation problem [72, 74, 130].
• Topological sorting of \( n \) nonintersecting line segments. This is the problem of ordering the segments so that, if a vertical line \( l \) intersect segments \( s_i \) and \( s_j \) and \( i < j \), then the intersection between \( l \) and \( s_i \) is above the intersection between \( l \) and \( s_j \). An \( O(\log n) \) time, \( n \)-processor CREW-PRAM algorithm is easily obtained by implementing the main idea of the mesh algorithm of [23] (which reduces the problem to a trapezoidal decomposition computation followed by a tree computation).

• Planar point location. Given a subdivision of the plane into polygons, build a data structure that enables one processor to quickly locate, for any query point, the face containing it. Using \( n \) processors, cascading can be used to achieve \( O(\log n) \) time for both construction and query [16, 121, 46]. The planar point location problem itself tends to arise rather frequently, even in geometric problems apparently unrelated to it.

• Intersection detection, three-dimensional maxima, two-set dominance counting, visibility from a point, all nearest neighbors. For all of these problems, cascading can be used to achieve \( O(\log n) \) time with \( n \) processors [16, 42].

Alternative approaches to cascading have been proposed for some of the above problems; for example, see [28, 113, 129] and also the elegant parallel hierarchical approach of Dadoun and Kirkpatrick, which is discussed next.

4.6 Geometric Hierarchies

This paradigm has proved extremely useful and general in computational geometry, both sequential [85, 62, 63, 64] and parallel [51, 52]. Generally speaking, the method consists of building a sequence of descriptions of the geometric object under consideration, where an element of the sequence is simpler and smaller (by a constant factor) than its predecessor, and yet "close" enough that information about it can be used to obtain in constant time information about the predecessor. This "information" could be, for example, the location of a query point in the subdivision, assuming the elements of the sequence are progressively simpler subdivisions of the plane (in that case pointers exist between faces of a subdivision and those of its predecessor — these pointers are part of the data structure representing the sequence of subdivisions). The technique turns out to be useful for other models than the PRAM (see Subsection 5.2).

13
4.7 Brent's Theorem

This technique is frequently used to reduce the processor complexity without any increase in the time complexity.

Theorem 1 (Brent) Any synchronous parallel algorithm taking time $T$ that consists of a total of $W$ operations can be simulated by $P$ processors in time $O((W/P) + T)$.

There are actually two qualifications to the above Brent's theorem [34] before one can apply it to a PRAM: (i) at the beginning of the $i$-th parallel step, we must be able to compute the amount of work $W_i$ done by that step, in time $O(W_i/P)$ and with $P$ processors, and (ii) we must know how to assign each processor to its task. Both (i) and (ii) are generally (but not always) easily satisfied in parallel geometric algorithms, so that the hard part is usually achieving $W$ operations in time $T$.

4.8 From CREW to EREW

In order to turn a CREW algorithm into an EREW one, one needs to get rid of the read conflicts, the simultaneous reading from the same memory cell by many processors. Such read conflicts often occur in the conquer stage, and can take the form of concurrent searching of a data structure by many processors (see, e.g., [15]). To avoid read conflicts during such concurrent searching, the scheme of [109] can be helpful:

Lemma 4.1 (Paul, Vishkin, Wagener [109]) Suppose $T$ is a 2-3 tree with $m$ leaves, suppose $a_1, a_2, \ldots, a_k$ are data items that may or may not be stored in (the leaves of) $T$, and suppose each processor $P_j$ wants to search for $a_j$ in $T$, $j = 1, 2, \ldots, k$. Then in $O(\log m + \log k)$ time, the $k$ processors can perform their respective searches without read conflicts.

Many types of searches can be accommodated by the above lemma. The following tend to occur in geometric applications:

- Type 1: Searching for a particular item in the tree, and
- Type 2: Searches of the type "find the $i$-th item starting from item $p$".

The search tree need not be a 2-3 tree: the requirements for the concurrent searching scheme of [109] to be applicable are that (i) each node of the tree has $O(1)$ children, and (ii) the $k$ searches should be "sortable" according to their ranks in the sorted order of the leaves.
of the tree. (The scheme of [109] has other requirements, but they are needed only for the concurrent insertions and deletions that it can also handle, not for searching.) Requirement (i) is usually satisfied in geometric applications. Requirement (ii) is also clearly satisfied for the searches of Type 1. It can be made to be satisfied for searches of Type 2 by sorting the tree. (The scheme of [109] has other requirements, but they are needed only for the
concurrent insertions and deletions that it can also handle, not for searching.) Requirement (i) is usually satisfied in geometric applications. Requirement (ii) is also clearly satisfied for the searches of Type 1. It can be made to be satisfied for searches of Type 2 by sorting the queries according to the leaf orders of their targets (this requires first doing a search of Type 1 to determine the leaf order of p).

4.9 Matrix Searching Techniques

A significant contribution to computational geometry (both sequential and parallel) is the formulation of many of its problems as searching problems in monotone matrices [5, 3]. Geometric problems amenable to such a formulation include the largest empty rectangle [6], various area minimization problems [5] (such as finding a minimum area circumscribing d-gon of a polygon), perimeter minimization [5] (finding a minimum perimeter triangle circumscribing a polygon), the layers of maxima problem [5], and rectilinear shortest paths in the presence of rectangular obstacles [14]. Many more problems are likely to be formulated as such matrix searching problems in the future. We briefly review these matrix searching formulations next.

4.9.1 Row Minima

An important matrix searching technique for solving geometric problems was introduced by Aggarwal et al. in [3], where a linear time sequential solution was also given. The technique, which we review next, has myriads of applications to geometric and combinatorial problems [5, 3].

Suppose we have an $m \times n$ matrix $A$ and we wish to compute the array $\theta_A$ such that, for every row index $r$ ($1 \leq r \leq m$), $\theta_A(r)$ is the smallest column index $c$ that minimizes $A(r, c)$ (that is, among all $c$'s that minimize $A(r, c)$, $\theta_A(r)$ is the smallest). If $\theta_A$ satisfies the following sorted property:

$$\theta_A(r) \leq \theta_A(r + 1),$$

and if for every submatrix $A'$ of $A$, $\theta_A'$ also satisfies the sorted property, then matrix $A$ is said to be totally monotone [5, 3].

Given a totally monotone matrix $A$, the problem of computing the $\theta_A$ array is known as that of "computing the row minima" of that matrix [5]. The best EREW-PRAM algorithm

15
for this problem runs in $O(\log n)$ time and $n$ processors [24] (where $m = n$). Any improvement in this parallel complexity bound will also imply corresponding improvements on the parallel complexities of the many geometric applications of this problem.

4.9.2 Tube Minima

In what can be viewed as the three-dimensional version of the above row minima problem [5], one is given an $n_1 \times n_2 \times n_3$ matrix $A$ and one wishes to compute, for every $1 \leq i \leq n_1$ and $1 \leq j \leq n_3$, the $n_1 \times n_3$ matrix $\theta_A$ such that $\theta_A(i, j)$ is the smallest index $k$ that minimizes $A(i, k, j)$ (that is, among all $k$'s that minimize $A(i, k, j)$, $\theta_A(i, j)$ is the smallest).

The matrix $A$ is such that $\theta_A$ satisfies the following sorted property:

$$\theta_A(i, j) \leq \theta_A(i, j + 1),$$

$$\theta_A(i, j) \leq \theta_A(i + 1, j).$$

Furthermore, for any submatrix $A'$ of $A$, $\theta_{A'}$ also satisfies the sorted property.

Given such a matrix $A$, the problem of computing the $\theta_A$ array is called by Aggarwal and Park [5] "computing the tube minima" of that matrix. Many geometric applications of this problem are mentioned in [5]. There are many nongeometric applications to this problem as well. These include parallel string editing [11], constructing Huffman codes in parallel [25], and other tree-construction problems. (In [25] the problem was given the name "multiplying two concave matrices".)

The best CREW-PRAM algorithms for this problem run in $O(\log n)$ time and $n^2 / \log n$ processors [5, 11], and the best CRCW-PRAM algorithm runs in $O(\log \log n)$ time and $n^2 / \log \log n$ processors [12] (where $n = n_1 = n_2 = n_3$). Both the CREW and the CRCW bounds are easily seen to be optimal.

4.10 Randomization

Reif and Sen [113, 114, 115] have modified and applied to parallel geometric computation the randomization techniques that had proved their worth in sequential geometric computing (cf. the work of K. Clarkson, and also Haussler and Welzl, Mulmuley) as well as in areas other than computational geometry. Recall that a randomized algorithm is one which bases some of its decisions on the outcomes of coin flips. Thus for a particular input, there are many possible executions of a randomized algorithm (which one actually happens depends on the outcomes of the coin flips). A good randomized algorithm must ensure that the number of "bad" possible executions (e.g., those that take too long to terminate) is a small fraction of
all the possible executions. Algorithms that are not randomized are deterministic (although this adjective is usually omitted when the context does not leave room for confusion). Some deterministic algorithms (such as the 2-dimensional parallel Voronoi diagram algorithm given in [91]) have efficient expected time behavior for a randomly chosen set of input points, whereas randomized algorithms make no assumption about the input distribution.

Randomized algorithms have the disadvantage that they might fail, but if the probability of failure is made small enough then they can have advantages over deterministic ones: they are typically very simple (which makes them easy to program and to comprehend), and the multiplicative constant in their time complexity is usually small. For example, the algorithms given by Reif and Sen in [113] have a running time of $O(\log n)$ with $n$ processors, with high probability (i.e., a probability that approaches one for very large $n$). The problems they deal with include planar point location and trapezoidal decomposition. The techniques they use there (and also in [115]) are somewhat reminiscent of the Flashsort algorithm of Reif and Valiant [116]. The polling technique of Reif and Sen [114] has yielded optimal randomized parallel bounds for two problems that continue to frustrate deterministic approaches, namely, the problems of computing the 3-dimensional convex hull and 2-dimensional Voronoi diagram.

4.11 Other PRAM Techniques

There are other techniques that we did not describe in detail because of their somewhat specialized nature. One such technique is the “array of trees” parallel data structure, originally designed in a non-geometric framework [21] but later used in [76] to establish geometric parallel bounds for such problems as hidden-line elimination, CSG evaluation, and computing the contour of a collection of rectangles. Another technique is the “stratified decomposition tree” used in [77] in the parallel solution of visibility and path problems in polygons.

5 Mesh Techniques

In this section, for convenience, we limit the discussion to 2-dimensional (i.e., $\sqrt{n} \times \sqrt{n}$) meshes, but most of the results and techniques mentioned are known to easily generalize to higher dimensional meshes as well. The geometric objects under consideration (e.g., points) are initially stored in the mesh, one object per processor. Therefore we are implicitly assuming that the mesh has enough processors to store the problem description — the
important case where the problem size is too large to fit in the mesh, is discussed in the next section.

Since it is known how to sort \( n \) items optimally (i.e., in \( O(\sqrt{n}) \) time) on a \( \sqrt{n} \times \sqrt{n} \) mesh, sorting is not a bottleneck when trying to design \( O(\sqrt{n}) \) time solutions to geometric problems on the mesh (contrast this with the situation for the hypercube, a network in which the complexity of sorting is still unknown). In fact many of the classical problems of computational geometry have been shown to be solvable on the mesh within the optimal \( O(\sqrt{n}) \) time bound (we mention some of these later). Most of these problems have an \( O(n \log n) \) sequential time complexity, and since the mesh time \( \times \) processors product is proportional to \( n^{\sqrt{n}} \), one might think that the word "optimal" is being abused here. However, this is not the case: any nontrivial problem on a \( \sqrt{n} \times \sqrt{n} \) mesh requires \( \Omega(\sqrt{n}) \) time (since it can take that long for two processors to communicate), and there is usually no hope of using \( o(n) \) processors because of the already mentioned \( O(1) \) storage limitation per processor (it takes \( \Omega(n) \) space, and hence \( \Omega(n) \) processors, just to store the input).

5.1 Mesh Divide and Conquer

Many geometric algorithms on the mesh use some form of divide-and-conquer: the problem gets partitioned into (e.g.) four pieces of size \( n/4 \) each, then each piece is moved into one of the four quadrants of the mesh where it is solved recursively by the \( (\sqrt{n}/2) \times (\sqrt{n}/2) \) quadrant, after which the answers returned by the four recursive calls are combined to obtain the overall solution. The "conquer" stage as well as the various bookkeeping steps usually involve sorting and take \( O(\sqrt{n}) \) time. Thus the time recurrence of this scheme generally ends up being of the form

\[
T(n) = T(n/4) + c\sqrt{n},
\]

where \( c \) is a constant, which implies \( T(n) = O(\sqrt{n}) \). An example of this is the convex hull algorithm of [100]:

1. If \( n \) is small (say, \( n \leq 4 \)) then solve the problem directly by brute force, otherwise proceed to Step 2 below.

2. Sort the \( n \) points whose convex hull we seek by \( z \) coordinates. Put those with the smallest \( n/4 \) \( x \)-coordinates in one of the four quadrants, those with the next \( n/4 \) smallest \( x \)-coordinates into another quadrant, etc. In fact the sorting itself can be
done so that each quadrant automatically contains the appropriate $n/4$ points, i.e., no separate data movement is needed other than sorting (see [100] for details).

3. Recursively solve the problem for each of the four quadrants.

4. Combine the solutions returned by the four recursive calls into the hull of the whole point set. This involves finding the common tangents between pairs of disjoint convex polygons in $O(\sqrt{n})$ time.

The nontrivial part is usually the "combine" part (i.e., Step 4 in the above example). The data movement techniques of [105] often play a role in that stage, and sometimes the tree computation technique of [22] is needed (e.g., in [82] and [23]).

5.2 Multisearching

The following problem is often the bottleneck in the parallel solution of geometric problems on a network of processors. It is a generalization of the problem described in Subsection 4.8: given a search structure modeled as a graph $G$ with $n$ constant-degree nodes, and given $O(n)$ search processes on that structure, the multisearch problem is that of performing as fast as possible all of the search processes on that structure. The searches need not be processed in any particular order, and can simultaneously be processed in parallel by using, for example, one processor for each. However, the path that a search query will trace in $G$ is not known ahead of time, and must instead be determined "on-line": only when a search query is at (say) node $v$ of $G$ can it determine which node of $G$ it should visit next (it does so by comparing its own search key to the information stored at $v$ — the nature of this information and of the comparison performed depend on the specific problem being solved).

The multisearch problem is a useful abstraction that can be used to solve many problems (more on this later). It is a challenging problem both for EREW-PRAMs and for networks of processors, since many searches might want to visit a single node of $G$, creating a "congestion" problem (with the added complication that one cannot even tally ahead of time how much congestion will occur at a node, since one does not know ahead of time the full search paths, only the nodes of $G$ at which they start). When the parallel model used to solve the problem is a network of processors, the graph $G$ is initially stored in the network in the natural way, with each processor containing one node of $G$ and that node's adjacency list. It is important to keep in mind that the computational network's topology is not the same as the search structure $G$, so that a neighbour of node $v$ in $G$ need not be stored in a
processor adjacent to the one containing \( v \). Each processor also contains initially (at most) one of the search queries to be processed (in which case that search does not necessarily start at the node of \( G \) stored in that processor).

In the EREW-PRAM, the difficulty comes from the "exclusive read" restriction of the model: if \( k \) processes were to simultaneously access node \( v \)'s information, the \( k \) processors assigned to these \( k \) search processes are apparently unable to simultaneously access \( v \)'s information. We have already mentioned, in Lemma 4.1, an elegant way around this problem, designed by Paul, Vishkin and Wagener [109] for the case where \( G \) is a 2-3 tree (although they assume a linear ordering on the search keys, something which usually does not hold in a geometric framework involving multidimensional search keys).

The multisearch problem is even more challenging for networks of processors. In such models, data is not stored in a shared memory, but is distributed over a network and requires considerable time to be permuted to allow different processors access to different data items. Furthermore, each memory location can be accessed by only \( O(1) \) query processes at a time, since a processor containing (say) node \( v \)'s information would be unable to simultaneously store more than a constant number of search queries.

In [125] the multisearch problem is solved \( O(\sqrt{n} + r\frac{\sqrt{n}}{\log n}) \) time on a \( \sqrt{n} \times \sqrt{n} \) mesh-connected computer, where \( r \) is the length of the longest search path associated with a query. For most geometric data structures, the search path traversed when answering a query has length \( r = O(\log n) \), and hence the time complexity is \( O(\sqrt{n}) \) time, which is asymptotically optimal. The classes of graphs for which this result holds contain most of the important cases of \( G \) that arise in practice, ranging from simple trees to the powerful Kirkpatrick hierarchical search DAG [85] that is so important in both sequential and parallel computational geometry (see Subsection 4.6). Applications include interval trees and the related multiple interval intersection search, as well as hierarchical representations of polyhedra and its many applications including lines-polyhedron intersection queries, multiple tangent plane determination, three-dimensional convex hull and intersecting convex polyhedra.

A special case of the multisearching problem for hypercube multiprocessors was studied in [59].
5.3 Mesh Prune and Search

Some mesh algorithms use the parallel equivalent of what has been called, in sequential computation, the "prune and search" paradigm [92]. This paradigm consists of throwing away a subset of the input (after determining that it does not contribute to the answer) and then recursively searching the surviving portion of the input. The portion of the input thrown away is a fixed fraction of the input (i.e., a subset of size \( cn \) of an input of size \( n \), where \( c \) is a positive constant). Mesh implementations of this idea have the intriguing feature of advantageously keeping many of the processors idle during much of the computation. This is because, after doing the "pruning" (= decreasing problem size by a constant factor), the resulting (smaller) problem is compressed into a smaller submesh of the original mesh, where it is recursively solved while the processors not in this smaller submesh remain idle.

Mesh algorithms might involve a sequence of many recursive calls (occurring after one another rather than in parallel) and still run in \( O(\sqrt{n}) \) time. So long as these successive recursive calls are on problems of sizes \( c_1n, c_2n, \ldots, c_kn \) where \( k \) as well as the \( c_i \)'s are constants and \( \sqrt{c_1} + \sqrt{c_2} + \ldots + \sqrt{c_k} < 1 \), the time complexity is \( O(\sqrt{n}) \) (assuming that setting up each recursive call is done in \( O(\sqrt{n}) \) time, and that the other bookkeeping and combining of subsolutions also takes \( O(\sqrt{n}) \) time). As an example, see the algorithm in [100] for computing the closest pair among a set of \( n \) planar points.

The parallel version of the prune and search technique has been far more useful for the mesh than for the PRAM, because in the mesh we can afford to prune in \( O(\sqrt{n}) \) time and still end up with an optimal algorithm, whereas in the PRAM model the technique typically yields superlogarithmic time bounds ([61] is one of the few instances where it was used for a PRAM geometric algorithm, and that was for the CRCW model).

5.4 Some Known Bounds and Open Problems for the Mesh

We now mention some problems for which \( O(\sqrt{n}) \) time mesh algorithms are known, as well as some open problems for the mesh.

The following problems have known \( O(\sqrt{n}) \) time solutions on the mesh (the list is not exhaustive).

- Convex hull and all nearest neighbour problems for planar point sets [100].

- Voronoi diagram of a planar set of \( n \) points [82]. This remained an open problem for
a while, until Jeong and Lee gave their elegant algorithm achieving an optimal time bound.

- Minimum distance spanning tree for planar point sets. This follows from the above-mentioned Voronoi diagram result of Jeong and Lee, and the fact that a minimum spanning tree of an $e$-edge undirected graph can be computed in $O(\sqrt{e})$ time on a $\sqrt{e} \times \sqrt{e}$ mesh [112].

- Trapezoidal decomposition of a set of $n$ (possibly intersecting) segments [82]. Note that visibility is a special case of trapezoidal decomposition.

- Polygon triangulation. This follows from the above-mentioned trapezoidal decomposition result of Jeong and Lee, and the fact that polygon triangulation can be solved by two calls to the trapezoidal decomposition procedure [130].

- Topological sorting of nonintersecting line segments. This follows from [23] and the above-mentioned trapezoidal decomposition result.

- The area of the union of iso-oriented rectangles [100].

- Intersection detection between $n$ planar line segments [82, 100].

- Computing the largest empty rectangle [53]. This is the problem of computing the largest-area iso-oriented rectangle that is constrained to lie in a given iso-oriented rectangular region and not to contain any of $n$ given points.

- 3-dimensional convex hull, computing the intersection of two 3-dimensional convex polyhedra [125, 80, 93].

Other geometric problems considered in the literature include the computation of robot configuration space [55], visibility and separability [54], ECDF searching [60], multipoint and planar point location [82], and others.

The following problems remain open on the mesh, in the sense that no $O(\sqrt{n})$ time algorithm on a $\sqrt{n} \times \sqrt{n}$ mesh is known for them.

- Convex layers in the plane. (See the end of Subsection 4.1 for a definition of this problem.)
• The layers of maxima in the plane. This is defined in a similar way to convex layers, but with the words “convex hull” replaced by “maximal elements”.

The most interesting open geometric problems on the mesh are in the hybrid model described in the next section, and will be mentioned there.

Another framework in which geometric problems have been considered on the mesh is that in which the input geometric figure is a binary image stored in the mesh in the natural way (the \((i, j)\)th pixel is stored in the processor at row \(i\) and column \(j\)). The techniques needed in this image processing framework can be quite different from those we mentioned above and are not within the scope of this survey (see, for example, [101, 56]).

6 A Hybrid Model: The RAM/ARRAY

The main justification for the hybrid model that is the subject of this section, is that many existing parallel machines have a “front end” that is a conventional sequential computer, and that the number of processors in the parallel machine itself is typically the fixed number purchased rather than a function of the problem size \(n\).

Suppose we have a parallel machine (like a \(d\)-dimensional mesh-connected array of \(p\) processors) that can solve a problem of size \(p\) in time \(O(p^{1/d})\) (this includes the time to input the data to the array as well as the actual computation time, a standard assumption in the literature of mesh-connected arrays, and certainly a reasonable one for the case \(d = 1\)). Suppose such a mesh-connected array of processors is attached to a conventional random access machine (RAM) that wishes to solve a problem of size \(n > p\). We call such a machine a RAM/ARRAY\((d)\). It is important to realize that the mesh alone cannot even store the description of the geometric problem, because of the limitation that each processor has \(O(1)\) storage registers, and hence the sequential “front end” must play a role in the solution process. If the problem’s sequential time complexity is, say, \(\Theta(n \log n)\), then the mesh gives a factor of \(s(p) = p^{1-1/d} \log p\) speedup for a problem of size \(p\). However, if the RAM/ARRAY\((d)\) is trying to solve a problem of size \(n\), \(n > p\), then it is not clear how it should use the mesh to achieve the factor of \(s(p)\) speedup and obtain \(O(n \log n/s(p))\) time performance. Actually, it is not even clear whether maintaining the \(s(p)\) speedup is at all possible. Identifying the problems for which this optimal \(O(n \log n/s(p))\) time can be achieved is an interesting question that was originally posed, for sorting in the case \(d = 1\), by Mueller [104] who also gave a partial solution. The question has been answered
in the affirmative in [17, 27]. This result immediately implies an affirmative answer on a
RAM/ARRAY(1) for the geometric problems that can be solved in linear time after a pre-
processing sorting step, like the planar convex hull and maximal elements problems. The
$O(n\log n/s(p))$ time bound can also be achieved on a RAM/ARRAY(1) for the following
geometric problems [26]: all nearest neighbors of a planar set of points, the measure and
perimeter of a union of rectangles, the visibility of a set of non-intersecting line segments
from a point. 3-dimensional maxima, dominance counting between two sets of points (and
hence the related problem of counting intersections between rectilinear rectangles). Essentially the same method as for the RAM/ARRAY(1) establishes that all these problems can
be solved in $O(n\log n/s(p))$ on a RAM/ARRAY(d), with $s(p) = p^{1-1/d}\log p$ [26].

We illustrate the technique for the case $d = 1$ and for geometric problems whose se-
quential time complexity is $\Theta(n\log n)$, i.e., when the task is to design $O(n\log n/\log p)$ time
algorithms on a RAM/ARRAY(1). In that case the algorithm usually follows the $p$-way
divide-and-conquer paradigm (there is an alternative method, using a "lazy B-tree" ap-
proach [17], which we do not discuss). That is, the algorithm divides the problem into $p$
subproblems. Then it recursively solves each of the $p$ subproblems, one after the other.
After the $p$ recursive calls return, it combines the subsolutions to form the final solution.
The main difficulty is how to perform the combining step in $O(n)$ time. If the combining
step can be performed in $O(n)$ time, then the overall time complexity $T(n)$ satisfies the
recurrence $T(n) = p \cdot T(n/p) + O(n)$, which implies that $T(n)$ is $O(n\log n/\log p)$. In the
case of a RAM/ARRAY(d) where $d > 1$, instead of partitioning the problem into $p$
subproblems, the problem gets partitioned into $p^{1/(d+1)}$ subproblems. In that case the $p^{1/(d+1)}$
subsolutions must be "combined" in $O(n/p^{1-1/d})$ time.

The following result from sequential computation, due to Frederickson and Johnson [69],
is often useful in the RAM/ARRAY(d) framework. Given an $a \times b$ matrix ($b \leq a$) whose
columns are sorted, the $k$th smallest element can be selected in time $O(b + m\log(k/m))$,
where $m = \min\{k, b\}$, if the matrix is already in the memory, or if any element of the matrix
can be produced in constant time. This implies that the $b$th element can be selected from
the matrix in $O(b)$ time. This selection algorithm has been used in one of the two schemes
given in [17] to establish the optimal sorting algorithm for this model, and it turned out to
be a crucial tool for many geometric problems [26].

The question of whether the speedup of $s(p)$, that the $d$-dimensional array makes possible
for a problem of size $p$, can be carried over to larger problems is really dealing with the
fundamental issue of the parallel-decomposability of the problem at hand: given that a problem of size \( p \) can be solved on a parallel machine faster by a factor of (say) \( s(p) \) than on a RAM alone, then that problem is *fully parallel-decomposable* if a RAM to which the parallel machine is attached can solve arbitrarily large instances of that problem with a speedup of \( s(p) \) when compared to a RAM alone. Although it is known that for some geometric problems the speedup of \( s(p) \) for a problem of size \( p \) can indeed be translated into a speedup of also \( s(p) \) for problems of size \( n \), \( n > p \), this question remains open for many other classical geometric problems, such as:

- Topological sorting of nonintersecting line segments.
- Trapezoidal decomposition. However, if the nonintersecting line segments \( s_1, s_2, \ldots, s_n \) are given in topologically sorted order, then it is known how to solve it in \( O(n \log n/s(p)) \) time on a RAM/ARRAY\((d)\) [26]. What makes the problem easier in that case is the fact that, if one partitions the problem into \( p \) equal-sized subproblems according to their topological order, then the "interaction" between subproblems is encapsulated by their visibilities from a point at infinity. In particular, it is known for the case when the line segments are horizontal because sorting them by \( y \) coordinates is like sorting them topologically.
- Voronoi diagram of a planar point set.
- 3-dimensional convex hull, computing the intersection of two 3-dimensional convex polyhedra.

Negative results would also be interesting: which problems are inherently such that it is impossible to maintain the same speedup for \( n > p \) as for \( n = p \) ?

The techniques developed for RAM/ARRAY\((d)\)s have also been used in [124] to achieve linear speedups on several hypercube-related computers which consist of \( p \) processors each containing \( O(n/p) \) local memory, provided that \( n > p^{1+\epsilon} \) for some constant \( \epsilon > 0 \). The same speedup is known for sorting [4, 50].

Finally, there are close connections between the work on parallel-decomposability and the work on I/O complexity [7, 81]. In the study of I/O complexity, one is given a sequential computer which has a small main memory and a large secondary storage, and one is interested in solving problems of arbitrarily large size. The input of the problem is initially stored in the secondary storage and the output has to be written to the secondary storage.
The limitation that the size of the main memory is small, is similar to the limitation that the size of the attached parallel machine is small. The major concern in the study of I/O complexity is to minimize the amount of I/O between the main memory and the secondary storage. To achieve the best I/O performance, the algorithm is allowed arbitrarily long computation times for scheduling the I/Os (i.e., only the amount of I/O matters). On the other hand, the time to decompose the computation into subcomputations and to schedule the subcomputations must be counted in the study of parallel-decomposability. The techniques developed for the study of geometric parallel-decomposability can be used to obtain I/O complexity bounds for the geometric problem considered [124].

7 Experimental Work

Much of the work in parallel computational geometry has been theoretical in nature, but some researchers have implemented geometric algorithms on various parallel architectures and reported interesting results.

Guy Blelloch [30, 31] has implemented parallel geometric algorithms on the Connection Machine (CM)\(^1\), including convex hull (the \(\sqrt{n}\)-divide-and-conquer method we mentioned earlier). Blelloch argued that in the CM architecture, scan operations (essentially, parallel prefix) are implemented so efficiently that one should solve problems on the CM architecture by using, whenever possible, calls to these built-in routines. In fact he went as far as assuming the cost of a parallel prefix to be \(O(1)\), and gave a detailed study of the implications of such an assumption on solving various problems. The experimental data obtained by Blelloch and by other researchers seems to confirm that Blelloch's assumption is quite reasonable.

Cohen, Miller, Sarraf and Stout have implemented parallel geometric algorithms on hypercube architectures like the iPSC, including convex hulls and domination [40], and convex hulls of digitized pictures [97].

The above-mentioned experimental work demonstrates, among other things, that algorithmic ideas developed for abstract parallel models can be useful when programming "real" parallel machines.

Generally speaking, work in parallel computational geometry continues to be mostly theoretical, with experimental work being the exception rather than the rule. Perhaps this

---

\(^1\) "Connection Machine" and "CM" are registered trademarks of Thinking Machines Corporation.
will change as researchers gain increased access to parallel machines.

8 Further Remarks

In view of the importance of the hypercube, surprisingly few geometric algorithms have been designed for this parallel model (see [33, 58, 59, 97, 99, 119] for some of these). We believe that, once the complexity of such basic operations as sorting and list ranking is settled for the hypercube model, algorithm design for geometric problems on that model will probably receive increased attention. An important step in this direction has recently been taken in the new sorting algorithm of Cypher and Plaxton [49]. One way around the "sorting bottleneck" for the hypercube would be to take the randomization approach, the way Reif and Sen did [113] (sorting is then no longer a bottleneck, since there is an optimal randomized sorting algorithm for the hypercube [116]).

In addition to the open problems in parallel computational geometry that we already mentioned earlier, the following open problems are likely to receive considerable attention in the future:

- Optimal deterministic PRAM construction of Voronoi diagrams in the plane. The current best bounds are, in the CREW-PRAM model, $O(\log n \log \log n)$ time and $n \log n / \log \log n$ processors or, alternatively, $O(\log^2 n)$ time and $n / \log n$ processors (see [43]).

- Optimal deterministic PRAM construction of 3-dimensional convex hull.

- Optimal EREW-PRAM solution to linear programming in the plane (an algorithm exists in the CRCW model [61]).

The following are additional promising directions for future research:

- Output-sensitive PRAM algorithms — where the complexity depends on the size of the output (for example, in [75], the number of processors needed depends on the number of intersections). Most geometric problems remain open when looked at from this perspective (even the planar convex hull problem).

- Robust parallel algorithms. Recall that robust algorithms are such that their correctness is not destroyed by roundoff error. Most existing parallel geometric algorithms misbehave if implemented with rounded arithmetic. There has been recently a flurry
of activity in designing efficient and robust sequential algorithms for geometric problems (see [96] for a list of references), and we expect this important activity to spread to the design of parallel geometric algorithms as well.

Acknowledgement. It is a pleasure to acknowledge the helpful comments of Danny Chen on an earlier draft of this survey.

References


31


[112] J. H. Reif and Q. F. Stout, manuscript.


