# CONTENTS

# Introduction

This thesis connects two areas which, on the surface, seem unrelated. These areas are the asymptotic behavior of random planar matching problems and the average-case analysis of on-line bin packing algorithms. By applying theorems on planar matching, we can obtain better bounds for the expected behavior of several on-line bin packing algorithms than previously known. Planar matching and bin packing were first related in [KLM], where Karp et al applied the up-right matching problem to obtain bounds on the expected behavior of multi-dimensional bin packing algorithms. Since then, the up-right matching problem has been used in this thesis to obtain bounds on the expected behavior of the Best Fit bin packing algorithm and in related work by Coffman and Leighton [CL] to obtain bounds on the expected behavior of a dynamic allocation algorithm.

We will define four planar matching problems: average edge length matching, rightward matching, up-right matching and maximum edge length matching. Three of these problems have been previously considered. Up-right matching has arisen in studies of probability [Du1,Du2] and in the analysis of algorithms [CL,KLM]. Maximum edge length matching has arisen in work on VLSI [LL]. Average edge length matching was previously investigated as a possible statistical test [AKT].

The simplest random planar matching problem is average edge length matching. This matching problem is as follows: Suppose there are $n$ + points and $n$ − points, randomly distributed uniformly and independently in the unit square.

Match the + points to the − points points so as to minimize the average edge length. What is the expected length of the edges of an optimal matching?

The other three planar matching problems discussed in this thesis are similar. In rightward matching, the goal is to minimize the average length of an edge in a matching where every + is to the right of the − with which it is matched. In up-right matching, the goal is to minimize the average length of an edge where every + is both above and to the right of the corresponding −. In maximum edge length matching, the goal is to minimize the maximum length of an edge in a matching. Since there is a polynomial algorithm for weighted bipartite matching [PS], it is relatively easy to find the optimal solution in a particular instance of the problem. What we investigate in this thesis is the asymptotic behavior of the value of the optimal solution.

These problems are listed above in the order of increasing difficulty, i.e., the edge length is shortest in average edge length matching and longest in maximum edge length matching. Tight bounds are now known for three of these problems. Ajtai, Komlós and Tusnády [AKT] have proven that the optimum average edge length is $\Theta(\log^{1/2} n/\sqrt{n})$. In this thesis, we show that the optimum maximum edge length and the optimum edge length of an up-right matching are both $\Theta(\log^{3/4} n/\sqrt{n})$. This shows that the solution for rightward matching is between the bounds for average edge length and for up-right matching, i.e., it is $\Omega(\log^{1/2} n/\sqrt{n})$ and $O(\log^{3/4} n/\sqrt{n})$.

We also show that these problems are fairly robust. The conditions on the matching or the distribution of the points can be changed in several ways without changing the asymptotic behavior of the optimal solution. For instance, points may be allowed to be matched to the boundary of the square as well as to points of the opposite kind. Instead of taking $n$ + and $n$ − ponts, you can take $2n$ points, each of which has an equal probability of being a + or a − point. You

7

can also take the + points to be a fixed uniformly spaced $\sqrt{n} \times \sqrt{n}$ grid instead of randomly distributed in the square. None of these variations changes the asymptotic behavior of the problems.

Several of these problems have arisen before in various contexts. The most common one is up-right matching. This has appeared in a slightly different, but equivalent form: namely, the number of unmatched points in an optimum up-right matching. This differs from the average edge length in an up-right matching by a factor of $n$. Up-right matching was first investigated independently by Karp, Luby and Marchetti-Spaccamela [KLM], and by Dudley [Du1,Du2]. The bounds obtained by both investigations were very close; the best bounds obtained were $\Omega(\sqrt{n}\log^{1/2} n)$ and $O(\sqrt{n}\log n)$.

Karp, Luby and Marchetti-Spaccamela [KLM] investigated up-right matching because it arose in the analysis of a multi-dimensional bin packing algorithm. For two dimensions, they needed the asymptotic behavior of up-right matching. In the analogous problem for $d$ dimensions, a $-$ point can be matched to a + point only if all of its coordinates are less than the corresponding coordinates of the + point. This problem is easier to analyze in 3 or more dimensions than in 2 dimensions. Karp, Luby and Marchetti showed that the average edge length in $d$ dimensions was $\Theta(n^{-1/d})$.

Dudley investigated the equivalent dual problem to up-right matching. This dual problem is to find the expected maximum discrepancy of a lower set in the unit square. A lower set is a set $L$ such that if a point $x$ is in $L$, every point below and to the left of $x$ is also in $L$. If there are $n$ + points and $n$ $-$ points in the square, the discrepancy of a set $L$ is the excess of one kind of point, i.e.,

$$\Big| |\{+ \text{ points in } L\}| - |\{- \text{ points in } L\}| \Big|.$$

Although Dudley does not obtain bounds on the discrepancy of a lower set as tight as those obtained by Karp et al., he obtains bounds for a much more

8

general case. By turning the square 45°, a lower set becomes a set in which the first derivative of the boundary is at most 1. Dudley obtains bounds on the discrepancy of sets in $d$ dimensions in which the first $k$ derivatives are bounded. If $k \geq d$, he shows the expected discrepancy is $\Theta(\sqrt{n})$. When $k \leq d - 1$, this behavior changes. Dudley obtains bounds for the borderline case of $k = d - 1$ of $O(\sqrt{n} \log n)$ and $\Omega(\sqrt{n} \log^{1/2} n / (\log \log n)^{1+\epsilon})$ for any $\epsilon > 0$. In two dimensions, this case is precisely equivalent to the number of unmatched points in an up-right matching. In three dimensions, this is equivalent to the maximum discrepancy of a convex set.

The maximum edge length matching problem has arisen in VLSI design. Suppose there are $n$ working processors on a chip, distributed randomly, and that you wish to configure these into a $\sqrt{n} \times \sqrt{n}$ grid using short edges. One way to proceed is to match each of these processors to a grid point on a fixed $\sqrt{n} \times \sqrt{n}$ grid, and then configure them in the pattern of this grid. Study of this and related problems led Leighton and Leiserson [LL] to obtain a bound of $\sqrt{n} \log n$ on the maximum edge length problem.

The average edge length problem was solved in a paper of Ajtai, Komlós and Tusnády. They were interested in this problem because it makes a good statistical test. To test whether two sets of points come from the same distribution, they find the optimal matching between the two sets of points. If the two sets are from the same distribution, the matching will have average edge length $O(\log^{1/2} n / \sqrt{n})$. If the sets are from two different distributions, the edges are likely to be longer. Ajtai, Komlós and Tusnády show that if the points are taken from the uniform distribution, the expected average edge length is $\Theta(\sqrt{n}\sqrt{\log n})$. In this thesis, we will give a slightly different version of their proof which avoids appealing to some difficult theorems on probability that they use.

The first part of this thesis deals solely with these matching problems. We

9

first define all the problems. We then discuss the equivalent dual problems, which we will use in the analysis of these problems. We discuss several variations on these problems that do not affect the asymptotic behavior of the solutions. We finally give the proofs of the bounds for these problems, starting with average edge length matching and proceeding to up-right matching and maximum edge length matching.

In the second part of this thesis, we discuss bin packing. We will use the theorems on planar matching proved in the first part to prove results about the average case behavior of bin packing algorithms.

The problem of bin packing is: given a set of items with sizes between 0 and 1, pack them into the minimum number of bins of size 1 such that no bin contains items with sizes summing to more than 1. Finding a packing using the fewest possible bins is NP-complete. However, several algorithms can be shown to give fairly good packings. The best algorithm to date is that of Karmarkar and Karp [KK], which always uses $OPT + O(\log^2 OPT)$ bins, where $OPT$ is the number of bins used by the optimal packing.

We will study on-line algorithms for bin packing. These are algorithms that pack items in bins as soon as the items are received. On-line algorithms must make choices without knowing what size items they will receive, so it is not surprising that they cannot do as well as off-line algorithms. Brown and Liang showed independently that any on-line algorithm in the worst case will use at least $1.536\,OPT$ bins [Br,Li]

Two of the simplest on-line algorithms are Best Fit and First Fit. The algorithm Best Fit packs each item in the bin it fits "best" in, i.e., the one with the least empty space. The algorithm First Fit keeps the bins in order, and packs each item in the first bin it fits in. Both these algorithms use in the worst case $1.7\,OPT$ bins [Jo,JDUGG].

10

Instead of investigating the worst-case behavior of bin packing algorithms, we will investigate the average behavior. In order for the average behavior to be defined, we must assume some distribution on the item sizes. The simplest distribution, and one that has often been used, is the uniform distribution on $[0, 1]$. We will also use this distribution.

The measure we use to judge the performance of bin packing algorithms is the expected wasted space. The amount of wasted space in a packing is the amount of empty space in bins containing at least one item, or the number of bins used less the sum of the sizes of the items. Previously, the best known average-case bound for an on-line bin packing algorithm was an $O(n^{4/5})$ bound on the wasted space produced by the algorithm First Fit [BJLMM]. By using results on random planar matching problems, we show that when packing items uniformly distributed on $[0, 1]$, the expected wasted space produced by the algorithm First Fit is $\Omega(n^{2/3})$ and $O(n^{2/3} \log^{1/2} n)$, and the expected wasted space produced by the algorithm Best Fit is $\Theta(\sqrt{n} \log^{3/4} n)$. We also show that no on-line algorithms can have expected wasted space $o(\sqrt{n} \log^{1/2} n)$. The performance of Best Fit is very close to this theoretical lower bound.

Planar matching problems were first used to analyze average-case bin packing by Karp et al. [KLM]. They gave an algorithm for multi-dimensional bin packing. In multi-dimensional bin packing, the items are $d$-dimensional rectangles which are to be packed into $d$-dimensional hypercubes. Karp et al. assumed that each of the coordinates of the items was distributed uniformly on $[0, 1]$. They showed that the expected number of unmatched points in the $d$-dimensional up-right matching problem was equal to the expected wasted space in the packing achieved by their algorithm.

The result on Best Fit follows much the same pattern. We show that the amount of wasted space produced by the Best Fit algorithm is the number of

unmatched points in an optimal up-right matching. We do this by representing the items packed by Best Fit as points in the plane, with time as one coordinate and the size of the items as the other coordinate. We produce a matching by pairing items that were packed in the same bin by Best Fit. Unmatched points in this matching correspond to bins with only one item in them, which give rise to wasted space.

Coffman and Leighton use up-right matching to analyze the average-case performance of another algorithm [CL]. This algorithm is for dynamic allocation, i.e., items both arrive and depart, and they must be stored using the minimum storage space. The algorithm is a modification of the Best Fit dynamic allocation algorithm: the storage area is first partitioned into compartments, and then Best Fit is applied. This algorithm also produces $\Theta(\sqrt{n}\log^{3/4} n)$ wasted space. The bound does not depend on the distribution being uniform; however, the distribution must be known in advance so that the storage area can be partitioned properly.

As with Best Fit, the bounds on First Fit are proved by showing that it is equivalent to a planar matching problem. This matching problem is up-right matching with an added condition on pairs of edges. Roughly, this condition says that if one edge is directly to the left of another edge, the edge on the left must be longer. By using our result on maximum distance matching, we show that the number of unmatched points in an optimal matching with this condition is $O(n^{2/3}\log^{1/2} n)$, giving an upper bound on the performance of First Fit. The lower bound on First Fit is achieved by an argument which does not use planar matching. This bound comes a from close examination of the way in which a First Fit packing evolves.

The lower bound for the wasted space produced by an on-line algorithm is also obtained by relating it to a planar matching problem. This problem is

12

the rightward matching problem. The lower bound of $\Omega(\log^{1/2} n / \sqrt{n})$ on the average edge length in a rightward matching shows that no on-line bin packing algorithm can achieve $o(\sqrt{n} \log^{1/2} n)$ wasted space. It is interesting to note that this bound only applies to on-line algorithms that do not know how many items they will receive. If an on-line algorithm knows in advance that it will receive $n$ items, then it can pack bins with $\Theta(\sqrt{n})$ wasted space, which is optimal. In fact, there are algorithms for which the expected wasted space is $\Theta(\sqrt{n})$ when $n = 2^i$, for all integers $i$.

In the second part of this thesis, we first give a brief description of the history of bin packing. Then, in Chapter 5, we show the bounds on the algorithm Best Fit. In Chapter 6, we show the bounds on the algorithm First Fit. Finally, in Chapter 7, we show the lower bound for any on-line algorithm that does not know how many items it will receive, and give an on-line algorithm that produces $\Theta(\sqrt{n})$ wasted space by knowing the number of items in advance.