MŰEGYETEM 1782

Combinatorial Algorithms in VLSI Routing

PhD Dissertation

Written by: Dávid Szeszlér

Supervisor: Professor András Recski

2005.

Nyilatkozat

Alulírott Szeszlér Dávid kijelentem, hogy ezt a doktori értekezést magam készítettem és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalommal, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Contents

1 Introduction **4**

2 The Routing Problem **6**
2.1 Setting of the Problem 6
2.2 Phases of the Solution 7
2.3 Circuit Partitioning 7
2.4 Placement 8
2.5 Global Routing 9

3 Detailed Routing **9**
3.1 The Switchbox Routing Problem 10
3.2 Single Row Routing 13
3.3 Channel Routing 15
3.3.1 2 layers, Manhattan model 16
3.3.2 2 layers, fixed length Manhattan model 20
3.3.3 2 layers, dogleg free Manhattan model 26
3.3.4 2 layers, unconstrained model 27
3.3.5 k layers, Manhattan model 28
3.4 Switchbox Routing 29
3.5 Gamma Routing 32
3.6 Edge-Disjoint Routing 33

4 Switchbox Routing in the Multilayer Manhattan Model **38**
4.1 The Square-Shaped Switchbox 38
4.2 The Switchbox of Arbitrary Shape 46

5 Single Active Layer Routing **49**
5.1 Definitions and Straightforward Bounds 50
5.2 The $s_n = 1$ Case 52
5.2.1 Two Simple Steps 52
5.2.2 The Real Routing 56
5.3 The $s_n, s_w \geq 2$ Case 59
5.3.1 2-terminal Nets 60
5.3.2 Multiterminal Nets 65
5.3.3 Algorithmic Aspects 68
5.3.4 Comparison of the Bounds 69
5.3.5 Final Remark 70

1 Introduction

The design of very large scale integrated (VLSI) circuits is one of the broadest areas in which the methods of combinatorial optimization can be applied. There are plenty of results in this topic from the last few decades. However, the list of NP-complete problems arising in this field is also very long and there is an abundance of heuristics, many of them with a very good performance, to handle these.

The term VLSI usually covers not only a single problem but a range of substantially different problems that arise during the design of circuits. In the next section we give a very short overview of these problems. However, the main focus of this dissertation is going to be *detailed routing*, one of the last phases of the design process.

Assume that the devices of an electric equipment to be designed have already been placed on the four boundaries of a rectangular circuit board. In the detailed routing problem our task is to interconnect certain given subsets (or *nets*) of the pins (or *terminals*) of these devices by wires. Wires belonging to different nets must never get closer to each other than a given distance. To this end, the wires are usually embedded in a rectangular grid. However, this grid is not planar (this would make the problem unsolvable in most cases), it consists of a number of planar layers, each of them parallel with the circuit board. Wires can leave a layer for a consecutive one at any gridpoint. To sum it up from a graph-theoretical viewpoint, the detailed routing problem consists of finding vertex-disjoint Steiner-trees (trees with a given terminal vertex set) in a 3-dimensional rectangular grid. The problem is more precisely formulated and a short survey of related results is given in Section 3. We also prove a new result on *channel routing*, a much investigated subproblem of detailed routing: we characterize all solvable specifications (if the width of the circuit board can be arbitrarily large, but the length is fixed) and provide a linear time algorithm that solves every solvable specification with a width that is at most constant times the length (see Theorem 5, published in [52]).

Traditionally, detailed routing was considered a 2-dimensional problem because the number of layers was very small compared to the length and the width of the board. (Originally, in the ancient times of printed circuit technology there were only two layers: the two sides of the board. Later the number of layers was gradually extended to 3, 4, ...) Since recent technology permits more and more layers (6, 8 or even more) a 'real' 3-dimensional approach becomes reasonable. In the last two sections we aim at exploiting this possibility.

Section 4 is dedicated to *switchbox routing* in the *Manhattan model*, another subproblem of detalied routing. We prove that every such problem instance can be solved in linear time on at most 6 layers if the circuit board is square-shaped. In the general case (where the length and the width of the circuit board can be different), we provide a linear time algorithm that approximates the minimum number of layers needed with an additive constant of 5. As a corollary, we improve a previous result of E. Boros, A. Recski and F. Wettl [5] (see Theorems 17, 18, 19 and Corollary 20, published in [51]).

In Section 5 we consider the *single active layer routing problem* which is a real 3-dimensional problem as the 3 dimensions are thought of as equal: terminals occupy the vertices of a planar layer and the routing is to be realized in a 3-dimensional cubic grid above this layer. To ensure solvability, it is allowed to extend the length and the width of the circuit board by introducing extra columns and rows into the grid. First we consider the case where only the width w is extended by a constant factor c and the length n is preserved; we prove that for any fixed value of w a routing with height $O(n)$ exists (if c can be at least 8) (see Theorem 23, published in [44]). Finally, we prove various results for the case where both the length and the width can be extended by a constant factor c. For example, we show that a routing of height $6\max(n, w)$ exists if $c \geq 2$ (and this can be improved to $3\max(n, w)$ if each net contains two terminals only). Each construction of this section can be realized in polynomial time. (See Theorems 25, 26, 29, 30, 32 and Corollaries 27, 28, 31, published in [45]).

I would like to express my gratitude to my supervisor, András Recski for his support in my work. Without his advice and experience this dissertation would have never come into existence.

2 The Routing Problem

2.1 Setting of the Problem

The problem of designing integrated circuits is very complex and it includes a wide range of subproblems that require substantially different approaches and solution techniques. Therefore it is not worth trying to formulate the routing problem in its full generality, because such a definition would necessarily be so technical that it would be useless from any practical point of view.

However, a rough setting of the problem could be the following. A set of circuit elements is given, each circuit element has a few terminals (or pins). Furthermore, a description of the circuit to be designed is also given, which is nothing else but a list of pairwise disjoint subsets of terminals, called *nets*. We aim at interconnecting the terminals of each net by wires such that the circuit elements and the interconnecting wires are embedded into the plane or – more typically – into a few planar layers. Furthermore, certain requirements of the fabrication technology must be fulfilled. The most essential of these requirements is that a minimum distance between any two distinct wires has to be kept. The easiest and most common way to achieve this is that the wires must adhere to the edges of a given rectangular grid.

There is a range of cost functions that have to be minimized (jointly, or according to a priority order) in various subproblems. The most important of these is the *area*, that is, the number of grid vertices on a layer. (It is widely known that miniaturization is essential in the design of computer chips because it affects the performance fundamentally.) Further cost functions, like the *total wire length*, the *length of the longest wire segment*, the *number of vias* (the number of transitions between two consecutive layers) can also be of importance (but these are mostly disregarded in this dissertation).

Naturally, the above description should only be regarded as a framework. It can be refined or modified in many ways to obtain the various subcases of the routing problem.

2.2 Phases of the Solution

Practically all versions of the above described general form of the routing problem are NP-hard. Therefore a widely accepted solution scheme for the routing problem developed as the result of long-term experience with the design of integrated circuits. According to this scheme, the problem is broken up into a sequence of phases.

The first one of these is the *placement phase* in which the elements of the circuit to be designed are placed on the circuit board. In the *global routing* phase the approximate course of the wires connecting the devices is determined. The final course of the wires is determined in the *detailed routing* phase. Both global and detailed routing can be followed by a *compaction* phase to reduce the area of the routing.

A further, widely applied method is that the circuit is broken up into smaller parts (before the routing is started or within a phase) and the parts are dealt with separately. This is performed by *circuit partitioning* algorithms.

In the following few sections we give a sample of the flavour of optimization problems that arise during the phases of the routing process preceding detailed routing.

2.3 Circuit Partitioning

In a typical partitioning problem we are given a hypergraph $\mathcal{H} = (V, E)$; a weight function $w : V \mapsto \mathbf{N}$ on the set of its vertices; a cost function $c : E \mapsto \mathbf{N}$ on the set of its edges; the desired number of the partition classes $r \in \mathbf{N}$; and the minimal and maximal partition class weights $b(i) \in \mathbf{N}$ and $B(i) \in \mathbf{N}, i = 1, \ldots, r$. We are looking for a partition $\bigcup_{i=1}^{r} V_i$ of the set V for which $b(i) \leq w(V_i) \leq B(i)$ holds for all $i = 1, \ldots, r$ and the total cost of all hyperedges that intersect more than one partition class is minimum. (The problem is NP-hard even if the value of r is fixed to $r = 2$ and we restrict ourselves to graphs instead of hypergraphs [8].)

A possible application of this problem (or a heuristic algorithm solving it) is to decompose the circuit into very small parts, lay out these parts first and then combine them to obtain bigger ones using similar methods, etc. There are lots of applications of the case in which r is small (for example 2), among others as a tool for the placement phase.

2.4 Placement

The difficulties of defining the placement problem lie in the fact that at the time of placing the circuit elements on the board we still do not know the course of the wires connecting them. Therefore it is not easy to find appropriate optimization criteria. The only feasible approach is to assign a cost to each placement which probably more or less describes its area demand and then minimize this cost function.

In accordance with the above, a typical placement problem is the following. Again, we are given a hypergraph $\mathcal{H} = (V, E)$ and the integers $r, s \in \mathbf{N}$ which describe the dimensions of the circuit board. By a *placement* we mean an injective function $p : V \to \{1, 2, \ldots, r\} \times \{1, 2, \ldots, s\}$. For any given placement p, we assign to each edge e of $\mathcal{H}$ the half-perimeter $c(e, p)$ of the smallest rectangle containing all $p(v)$-s for each vertex v belonging to e. We define the cost $c(p)$ of the placement p as $c(p) = \sum_{e \in E} c(e, p)$. Now we are looking for the minimum cost placement. (It may not be surprising, but this problem is again NP-hard [18].) Naturally, depending on the various applications, completely different cost functions can also be of interest.

Another type of placement problem with a completely different flavour is *floorplanning*. This arises if the circuit is decomposed into a number of blocks which are to be routed separately. We first want to place these blocks on the circuit board. We assume that each block is going to occupy a rectangular area. Furthermore, to each block we have assigned a function which gives an upper bound on the length of the rectangle accomodating the block as a function of the width of the rectangle. We aim at choosing the dimensions of the blocks and then placing them on the board such that the total area is minimum.

2.5 Global Routing

This phase of routing is useful for those technologies in which the circuit board can be decomposed into smaller parts in a natural way (for example, in the placement phase). In these cases global routing only determines the way in which the wires manoeuvre between these parts (for example, whether a ‘mainly horizontal’ longer wire segment goes round an obstacle from below or from above). The final course of the wires will only be determined in the detailed routing phase.

In the formulation of a typical global routing problem a graph $G = (V, E)$ is given whose vertices represent the smaller parts of the circuit board and edges represent adjacencies between these parts. (That is, G is the planar dual of the graph representing the smaller parts of the cicruit board; see [30] for the details.) A capacity function $c : E \to \mathbf{R}_+$ is also given together with a set system $\mathcal{N} \subseteq 2^V$, the elements of which correspond to the nets. We are looking for a Steiner-tree for each net such that for each edge e of G the number of Steiner-trees containing e does not exceed the capacity $c(e)$ of e. Furthermore, usually the minimization of certain cost functions is also a requirement, we omit the details. (However, the decision version of this problem, that is, the version without minimization requirements is already NP-complete [34].) In applications the capacities of the edges are drawn from estimates on the number of wires that can cross the corresponding area.

3 Detailed Routing

In the detailed routing phase the circuit elements have already found their final position on the board and the course of wires connecting them is to be determined. Since the remaining parts of this dissertation deal with various subproblems of detailed routing, we now give a formal definition of this problem.

Definition 1 *Assume that a graph G (usually called the* routing graph*) is given. A* net *is a set of vertices of G (of cardinality at least 2). A* detailed routing problem *is a family* $\mathcal{N} = \{N_1, N_2, \ldots, N_t\}$ *of pairwise disjoint nets. The vertices of the nets* N_i *are called terminals.*

The routing graph G can simply be a 3-dimensional rectangular grid graph, but in practical applications it is usually obtained is some way from a 3-dimensional grid (for example, to leave space for the circuit elements, or to let the terminals be accessible from any layer; see Definition 3 below).

Definition 2 *A* solution *(or* routing*, or* layout*) of a detailed routing problem* $\mathcal{N} = \{N_1, \ldots, N_t\}$ *is a set* $\mathcal{H} = \{H_1, \ldots, H_t\}$ *of pairwise vertex-disjoint, connected subgraphs of the routing graph* G*, such that* $N_i \subset V(H_i)$*, that is,* H_i *connects the vertices of* N_i*. The subgraphs* H_i *are called* wires*.*

As we have already mentioned, the wires are usually chosen to be minimal, thus they are Steiner-trees. We also mention that, in contrast to the above definition, in some cases the wires are required to be edge-disjoint only, see Section 3.6.

According to Definition 2, the detailed routing problem is a decision problem. However, as we will see in the forthcoming sections, its various subproblems are usually formulated as minimization problems. This is achieved by letting the underlying routing graph G depend on a parameter (for example, the number of layers or the number of rows in the grid), and looking for the optimum value of this parameter such that the corresponding detailed routing problem is solvable.

3.1 The Switchbox Routing Problem

We mentioned above that the structure of the routing graph can be much more complicated than a 3-dimensional grid. Therefore in practical applications the routing area is very often broken up into smaller parts which are routed one after the other in a specified order. These smaller parts are of much simpler structure (usually channels or switchboxes, see the following definitions). By this method the problem can be reduced to the very special case in which the terminals are situated on the boundaries of a rectangular board.

The following definitions are formulated in accordance with the above and with a few further technological requirements. One is that there are no terminals on the 'corners' of the board and routings must not use them either. The other one is that wires can access the terminals on any layer. The definition of the *k-layer rectangular grid graph* is therefore suitably altered.

Definition 3 *Let the vertex set of a graph be the set* $\{0, \ldots, n+1\} \times \{0, \ldots, w+1\} \times \{1, \ldots, k\}$*. Let two vertices be adjacent if and only if they differ in exactly one coordinate and by exactly one. Delete the 'corner' vertices from the graph, that is, the vertices* $(0, 0, l)$*,* $(n+1, 0, l)$*,* $(0, w+1, l)$ *and* $(n+1, w+1, l)$*, where* $l = 1, \ldots, k$*. Now in the remaining graph contract all the subsets of vertices* $\{(i, j, l) : l = 1, \ldots, k\}$ *into a single vertex, where*

- $i = 0$ *or* $i = n+1$ *and* $j = 1, \ldots, w$*; or*
- $j = 0$ *or* $j = w+1$ *and* $i = 1, \ldots, n$*.*

Call the thus obtained graph G_k k-layer rectangular grid graph*.* w *and* n *are the* width *and the* length *of the grid, respectively. Denote the vertex obtained from the contraction of the set* $\{(i, j, l) : l = 1, \ldots, k\}$ *by* $t_{i,j}$*. The vertices* $t_{i,j}$ *are called* terminals*. The terminal* $t_{i,j}$ *is called*

- northern*, if* $j = w+1$*;*
- southern*, if* $j = 0$*;*
- western*, if* $i = 0$*;*
- eastern*, if* $i = n+1$*.*

The sets of all non-terminal vertices with a common z*-coordinate are called* layers*. The sets of all vertices of a layer with a common* x*-coordinate or* y*-coordinate are called* columns *or* tracks*, respectively.*

Figure 1 shows the 2-layer grid graph with $w = n = 3$.

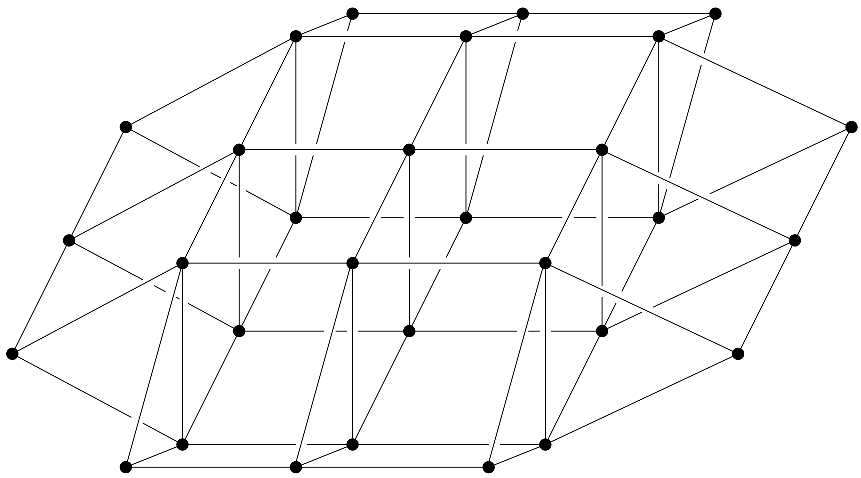

Figure 1

Definition 4 *A* switchbox routing problem *is the special case of the detailed routing problem (see Definition 1) in which the routing graph is the k-layer rectangular grid graph (for some value of k) and each net is a subset of the set of terminals $t_{i,j}$ (see Definition 3).*

Many restrcitions of the above defined problem are considered in the literature, which are called *routing models*. If no further restriction is imposed on the solution of a swithcbox routing problem then it is said to belong to the *unconstrained model*. In this dissertation we only consider one of the above mentioned routing models, the *Manhattan model*. Its definition is motivated by the fact that for certain technologies it is advantageous not to have long parallel wire segments on two consecutive layers. Therefore there are many results that provide routings in this model. (We also consider edge-disjoint routing in Section 3.6, where the wires are required to be edge-disjoint only, but that is not a restriction of switchbox routing according to the definition we gave above.)

Definition 5 *A solution of a switchbox routing problem is said to belong to the* Manhattan model *if consecutive layers contain wire segments of different directions only. That is, layers with horizontal (east-west) and vertical (north-south) wire segments alternate.*

3.2 Single Row Routing

By the *single row routing problem* we mean the special case of the switchbox routing problem in which all the terminals of each net are, say, northern. In this case the specification of a routing problem only fixes the length n. Therefore the usual formulation of the problem is to fix the number of layers and ask for the minimum width routing.

The first classic result in the topic of VLSI routing is probably Gallai's linear time algorithm that solves the single row routing problem with optimal width in the 2-layer Manhattan model. For every vertical line e that cuts the grid into two we define its *congestion* $c(e)$: it is the number of nets that are divided into two by e (that is, the number of nets that have terminals both left to e and right to e). For example, the congestion of the line e in Figure 2 is $c(e) = 3$. The maximum congestion of all vertical lines that cut the grid into two is called the *density* of the problem. It is straightforward that the density is a lower bound on the width of any routing (again, in the 2-layer Manhattan model).

Theorem 1 (T. Gallai) [14] *The minimum width of a solution of a single row routing problem in the 2-layer Manhattan model is equal to the density of the problem. Moreover, such a minimum width routing can be found in linear time.*

Proof: The proof makes use of the fact that interval graphs are perfect. A horizontal interval is associated with every net, stretching from its leftmost terminal to its rightmost terminal. The corresponding interval graph is defined in the following way: the intervals form the vertex set and two vertices are adjacent if and only if the corresponding intervals intersect. The clique number of this graph is equal to the density of the routing problem. A colouring with an equal number of colours exists by the perfectness of interval graphs. Such a colouring can easily be transformed into an optimal width routing: nets belonging to a common colour class can be routed in a common track.

An example is shown in Figures 2 and 3. The interval graph of Figure 3, which corresponds to the routing problem of Figure 2, is coloured using three colours; the solution of the routing problem obtained from this colouring is shown in Figure 2. (In Figure 2 solid dots denote the terminals and sets of terminals marked with a common number form the nets. Wire segments of the two layers are denoted by continuous and dashed lines, respectively. Empty dots denote the vias. We adopt the same notation for all figures of Sections 3 and 4.)

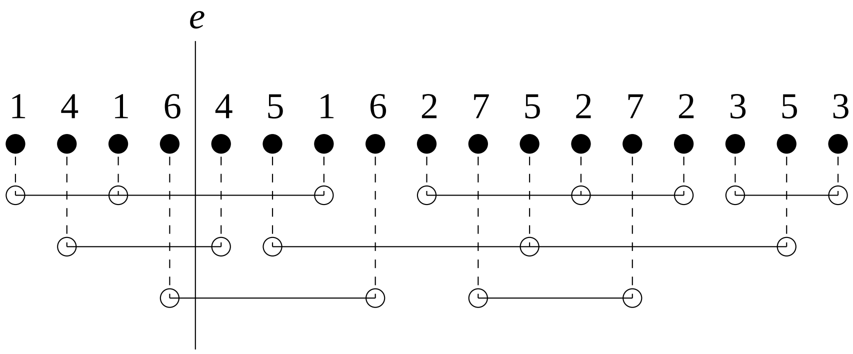

Figure 2

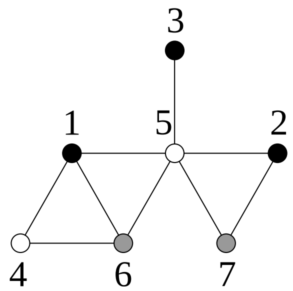

Figure 3

To show that that the above described routing can be realized in linear time, we only have to check that an optimal colouring of the corresponding interval graph can be found in linear time, that is, in $O(n)$ steps, where n is the length of the given problem instance. (The rest of the work requires constant time for each terminal if wire segments are output by the coordinates of the endpoints.) We assume that the interval graph to be coloured is given by an ordered list $\mathcal{L}$ containing the positions of all the endpoints of the intervals. (That is, both left and right endpoints are included in a common list $\mathcal{L}$.) $\mathcal{L}$ can trivially be obtained from the specification of the single row routing problem by scanning through the row of terminals. A colouring of the interval graph will be obtained by scanning through the elements of $\mathcal{L}$. We use positive integers as colours. We maintain a list $\mathcal{C}$ of free colours and the largest colour M that has already been used. At the beginning we set $\mathcal{C} = \emptyset$ and $M = 0$. If during the scanning of $\mathcal{L}$ we reach a left endpoint of an interval I such that $\mathcal{C} = \emptyset$ then we introduce the new colour $M + 1$, we assign it to I and increase the value of M to $M + 1$. If we reach a left endpoint such that $\mathcal{C} \neq \emptyset$ then then we choose a colour k from $\mathcal{C}$, we assign k to I and delete k from $\mathcal{C}$. If, on the other hand, we reach a right endpoint of an interval I then we simply put the colour of I into $\mathcal{C}$. It is easy to verify that the obtained colouring is optimal: the vertical line going through the left endpoint of the interval for which the largest one of the colours M was used for the first time is intersected by M intervals, thus the clique number of the interval graph is also M. $\square$

We mention that no polynomial time algorithm is known to find an optimum width solution for a single row routing problem in the unconstrained 2-layer model.

3.3 Channel Routing

By the *channel routing problem* we mean the special case of the switchbox routing problem in which all the terminals of each net are situated on two opposite boundaries of the grid (say, they are all northern or southern). Again, the usual setting of the problem is to fix the number of layers and ask for the minimum width routing (if at all a routing exists).

3.3.1 2 layers, Manhattan model

Channel routing in the 2-layer Manhattan model is one of the most popular and most investigated problems in VLSI routing. Trivial examples show that some specifications are unsolvable in this model with an arbitrary width (although the range of these is very limited), see Section 3.3.2. Therefore the usual problem setting, which is adopted in many papers, allows an arbitrary number of extra empty columns to be added to the western and the eastern ends of the given channel routing instance.

We have seen that in case of single row routing finding the optimum width solution in the 2-layer Manhattan model is always possible in linear time. However, channel routing is much more complicated than single row routing as it is shown by the following theorem.

Theorem 2 (T.G. Szymanski, 1985) [55] *It is NP-complete to decide whether a channel routing problem is solvable in the 2-layer Manhattan model with width at most w (where w is part of the input).*

Several extensions of the above result are also known. For example, assume that the nets of a channel routing problem are all ‘one-sided’, that is, the terminals of each net occur either only on the northern boundary or only on the southern boundary. A straightforward approach to such a problem is to use Gallai’s algorithm (Theorem 1) twice (for both boundaries). It is easy to verify that the width of the thus obtained routing is at most twice the optimum width. Therefore it is worthwhile mentioning that finding the minimum width for such a problem is still NP-hard, even if each net contains two terminals only [37].

With respect to the above facts, plenty of practically effective heuristic algorithms have been proposed, which can solve ’difficult’ problems (of length 150–200) with width around 20 [11, 23, 31, 47]. Furthermore, there exist a few positive results concerning 2-layer Manhattan channel routing.

There is a linear time algorithm to decide whether a channel routing problem is solvable in this model with a given width w [53]. (That is, the algorithm runs in linear time in the length n for all constant width w, but the running time depends superpolynomially on w.)

The most well-known positive result on Manhattan channel routing is the approximation algorithm due to Baker, Bhatt and Leighton. To state the corresponding theorem, the notion of density (see page 13) must be extended to channel routing as well to obtain a lower bound on the minimum width. To each net, a horizontal interval is assigned stretching from its leftmost terminal to its rightmost terminal (where the terminals of each net are considered on both boundaries). The *congestion* of a vertical line e is the number of intervals intersecting e. The *density* of a routing problem is defined as the maximum congestion. It is worth mentioning that the congestion of a vertical line belonging to the grid can be larger by at most 1 than the maximum congestion of all vertical lines not belonging to the grid. Hence, in contrast to single row routing, vertical lines of the grid should be considered when determining the density of a channel routing problem. For example, the congestion of the line e on Figure 4 is 3 (which proves that the solution shown is optimal), while all vertical lines not belonging to the grid have congestion at most 2.

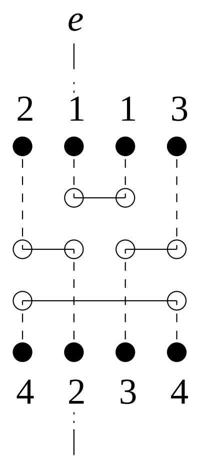

Figure 4

Since the density is a lower bound on the minimum width, the natural question arises if it is possible to give an upper bound on the minimum width as a function of the density. However, the answer is no. It was observed by Brown and Rivest [7] that the minimum width solution of the channel routing problem of Figure 5 (which is often referred to as *shift-right-1 problem*) requires $\Theta(\sqrt{n})$ tracks, while the density is constant 2.

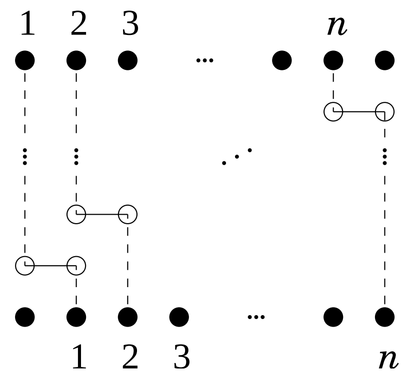

Figure 5

This idea was further extended by Baker, Bhatt and Leighton [3]. They defined a lower bound, called *flux*, on the minimum width which is independent from the density. The following proposition serves as a preparation to the somewhat technical definition of this notion. A net is called *trivial* if it consists of two terminals situated in the same column.

Proposition 3 (S. B. Baker, S. N. Bhatt and F. T. Leighton, 1984) *[3] Choose k consecutive terminals on one boundary of a given channel routing problem instance; denote the set of these terminals by S. Denote the number of non-trivial nets that have terminals both in S and outside S by l. If a solution with width w exists then*

$$w(k-l) + w(w+1) \geq l.$$

Proof: Assume that the terminals of S are on the northern boundary. We can assume without loss of generality that each net is non-trivial and each net contains exactly two terminals: one terminal in S and one outside S. (If this is not the case then choose one terminal in S and one outside S from all non-trivial nets that have a terminal in S and delete all other terminals from each net.) Thus the number of nets is l.

Since each net contains two terminals, each wire can be chosen to be a path. We will think of these paths as being directed from inside S to outside S. The routing of each net consists of an alternating sequence of vertical and horizontal wire segments (disregarding the vias between the two layers). Since there are no trivial nets, each path contains at least one horizontal segment. We will say that a net N is *routed in track* j if the first one of the horizontal segments belonging to the routing of N is in track j.

Denote the set of the k consecutive columns belonging to S by C. Number the tracks of the channel from north to south. Assume that a net N is routed in track 1. Then its horizontal segment in track 1 either leaves C or it ends in a column whose northern terminal is not occupied by any net. Since at most two wires can leave C in track 1 and the number of unoccupied terminals in S is $k - l$, the number of nets routed in track 1 is at most $k - l + 2$.

Now choose a track j for any $2 \le j \le w$. Since at most two wires could have left C in each one of the tracks $1, 2, \ldots, j-1$, the routings of at least $l - 2(j-1)$ nets must enter track j in C. These entering points occupy at least $l - 2j + 2$ gridpoints in track j (on the layer accomodating vertical wire segments). Therefore if a net is routed in track j then the corresponding horizontal segment in track j either leaves C (the number of such nets is at most 2) or it ends in one of the remaining $k - l + 2j - 2$ columns. Hence the number of nets routed in track j is at most $k - l + 2j$ (for each $j = 1, 2, \ldots, w$).

Since each net is routed in one of the tracks, we get that

$$\sum_{j=1}^{w} (k - l + 2j) \ge l,$$

which is equivalent to the statement of the proposition. $\square$

In accordance with the above proposition, the flux f of a channel routing problem instance is defined as the the minimum (integer) value of w for which the inequality of Proposition 3 holds for any choice of a number of consecutive terminals on either boundary. (Here we follow the definition of [30]; the original definition in [3] differs from the present one only by a constant.) For example, the flux of the shift-right-1 problem (Figure 5) is $\lfloor \sqrt{n} \rfloor$ or $\lceil \sqrt{n} \rceil$. It is straightforward from the definition that the flux is always at most $\sqrt{t}+1$, where t is the number of nets.

The flux f serves as a proof for the fact that there is no upper bound on the minimum width as a function of the density d only. However, Baker, Bhatt and Leighton [3] proved the following theorem.

Theorem 4 (S. B. Baker, S. N. Bhatt and F. T. Leighton, 1984) *[3] Every channel routing problem can be solved in the 2-layer Manhattan model with width at most $2d + O(f)$. Furthermore, the bound on the width is at most $d + O(f)$ if each net consists of two terminals only.*

The proof of the above theorem involves an algorithm which runs in linear time in the area of the obtained routing. The disadvantage of the method is that the number of extra columns that have to be added to the original problem instance can be as large as $O(f)$ (which can be as large as $O(\sqrt{t})$ in the worst case). On the other hand, the authors of the above theorem claim that the flux of practical channel routing problem instances appears to be bounded by a small constant.

Essentially the same result as Theorem 4 is obtained in [56] with better constants. The upper bound on the width is substantially improved to $\frac{3}{2}d + O(\sqrt{d \log d}) + O(f)$ in [16, 17].

3.3.2 2 layers, fixed length Manhattan model

In this section we restrict the problem of 2-layer, Manhattan channel routing in the sense that, in contrast to Section 3.3.1, we do not allow extra empty columns to be added to the grid. We mentioned that some problem instances become unsolvable with this restriction. The following theorem characterizes all solvable specifications and provides an upper bound on the minimum width if a solution exists.

A channel routing problem is called *bipartite* if each net contains exactly two terminals, one on the northern, and one on the southern boundary. A channel routing problem is *dense* if each terminal on the northern and southern boundaries belongs to some net. (We always assume that each net contains at least two terminals.) Recall that a net is called trivial if it consits of two terminals which are situated in the same column.

Theorem 5 (Sz. D., 2003) [52] *A channel routing problem is* not *solvable in the 2-layer Manhattan model (with an arbitrary width) if and only if it is bipartite, dense and has at least one non-trivial net. Moreover, if a specification is solvable then it can be solved in linear time with width at most* $\frac{3}{2}n$ *in the bipartite, and* $\frac{7}{4}n$ *in the general case (where* n *is the length of the channel).*

To prove the above theorem, we need the following definition.

Definition 6 *Assume that a channel routing problem* $\mathcal{N} = \{N_1, \ldots, N_t\}$ *is given. The* vertical constraint graph *of* $\mathcal{N}$ *is a directed graph* G_v *whose vertex set is* $\mathcal{N}$ *and a directed edge* (N_i, N_j) $(1 \leq i, j \leq t,\ i \neq j)$ *is introduced whenever there exists a column in which* N_i *has a northern terminal and* N_j *has a southern terminal.*

Proof of Theorem 5: Throughout the proof we assume that the top and bottom layers are reserved for vertical and horizontal wire segments, respectively.

In a dense, bipartite specification the length n of the channel equals the number of nets. If a non-trivial net N_i exists, then the routing of N_i requires at least one horizontal wire segment on the bottom layer in track j. The remaining $n-1$ nets must also cross track j. However, the two vias at the endpoints of the horizontal wire segment belonging to N_i occupy two columns on the top layer, therefore at most $n-2$ further nets can cross track j. Hence such a specification cannot be routed.

To show the converse implication, first assume that a given specification $\mathcal{N} = \{N_1, \ldots, N_t\}$ is bipartite. In this case each vertex of its vertical constraint graph G_v has both in-degree and out-degree at most one, therefore each connected component is either a directed path or a directed cycle. If each net is trivial then a trivial routing exists (with width 1). Otherwise we construct a routing by processing the components of G_v one after the other. (The routing of trivial nets can be solved by a single vertical segment on the top layer, we disregard these henceforth.) We first process directed paths, then directed cycles (otherwise the order of the components is arbitrary).

Assume that the directed path $P = (N_{i_1}, N_{i_2}, \ldots, N_{i_k})$ is a component of G_v. Proceeding from North to South, we assign a separate track on the bottom layer to each net of P in the above order and route them in the most straightforward way: in the corresponding track of the bottom layer we introduce a horizontal wire segment connecting the columns of the two terminals, we switch to the top layer at both ends of this segment and connect to the two terminals. Thus we have consumed k consecutive tracks to route the nets of P. Now assume that the directed cycle $C = (N_{i_1}, N_{i_2}, \ldots, N_{i_k}, N_{i_1})$ is a component of G_v. We choose a net of C, say N_{i_1}, arbitrarily; we are going to refer to the chosen net as an *exceptional net.* Again, proceeding from North to South we assign a track to each net of C in the above order. However, the exceptional net N_{i_1} receives two tracks: the topmost and the bottommost track of the altogether $k+1$ tracks assigned to C. Routing of the nets $N_{i_2}, \ldots, N_{i_k}$ is performed in the same straightforward way as in the case of directed paths. To route N_{i_1}, we first find a column in which the southern terminal does not belong to any net. (Since the length n is larger than the number of nets t by the assumption of the theorem, the existence of such a column is guaranteed.) We place a vertical wire segment in this column on the top layer to connect the two tracks assigned to N_{i_1}. At both ends of this segment we switch to the bottom layer and introduce two horizontal wire segments in the upper and lower tracks assigned to N_{i_1} to reach the column of the northern and southern terminal of N_{i_1}, respectively. We complete the routing of N_{i_1} by joining the endpoints of the above horizontal segments with the two terminals using the top layer.

The routing of a bipartite specification, together with the corresponding vertical constraint graph is shown in Figure 6. (Exceptional nets are 1 and 4.)

It is easy to verify that no conflict can occur between the routings of any two nets. Since each track of the bottom layer is assigned completely to one of the nets, no two wires intersect on this layer. Assume that both terminals of a column are occupied by a net; then these two (not necessarily distinct) nets belong to the same component of G_v. If neither of these two nets are exceptional then the track assigned to the net of the northern terminal lies

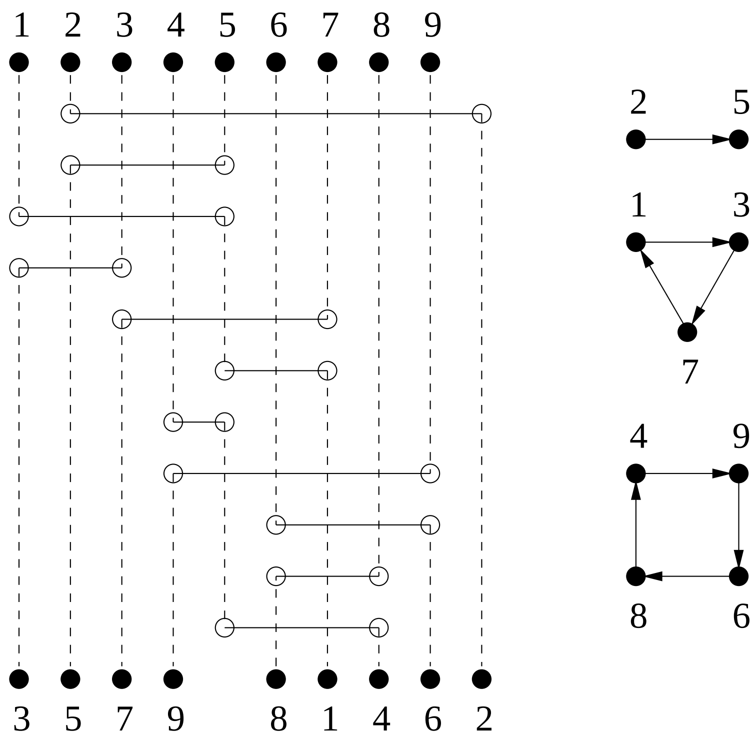

Figure 6

one unit north to the track assigned to the net of the southern terminal, therefore the two vertical wire segments of this column are disjoint. This is true even if, say, the net of the northern terminal is exceptional since in this case this terminal is joined to the topmost of the tracks assigned to the component of its net. If the northern terminal of a column is empty, but the southern one is occupied by a net then only one vertical segment was placed in this column, thus no conflict can occur. If the southern terminal or both terminals of a column are empty then vertical segments belonging to exceptional nets may have been placed in this column. If the northern terminal is occupied by a net then this net is the endpoint of a directed path in G_v, therefore it was processed prior to all exceptional nets. Thus no two vertical segments can intersect even in such columns.

Finally, we consider the general (non-bipartite) case. The upper and lower boundaries can each be regarded as a single row routing problem. These can be solved using Gallai's algorithm (Theorem 1). This way the problem is reduced to the bipartite case: from each net that appears on both boundaries one northern and one southern terminal is chosen arbitrarily. Obviously, the thus obtained bipartite problem is not dense, hence it can be routed using the above described algorithm.

Now we show that the above constructed routing uses not more than $\frac{3}{2}n$ tracks in the bipartite, and not more than $\frac{7}{4}n$ tracks in the general case.

If the problem is bipartite then the track consumption of the routing of each non-exceptional and exceptional net is 1 and 2, respectively. Since each connected component of G_v has at least two vertices, at most half of the nets are exceptional. Hence, if t denotes the number of nets, we have $w \leq t + \frac{t}{2} = \frac{3}{2}t \leq \frac{3}{2}n$.

In the general case, denote by x_i and y_i the number of non-exceptional and exceptional nets containing exactly i terminals, respectively ($i \geq 2$). Denote by e the number of terminals that are not occupied by any net. Thus we have

$$e + \sum_{i \geq 2} i \cdot (x_i + y_i) = 2n. \tag{1}$$

The following table shows the maximum number of tracks consumed by the construction to route a net, depending on the number of terminals (correctness of the values can easily be verified).

	2 terminals	3 terminals	$\geq$ 4 terminals
Non-exceptional nets	1	2	3
Exceptional nets	2	3	4

It follows that for the width w of the routing

$$w \leq x_2 + 2y_2 + 2x_3 + 3y_3 + 3\sum_{i \geq 4} x_i + 4\sum_{i \geq 4} y_i \tag{2}$$

holds. Furthermore,

$$\sum_i x_i \geq \sum_i y_i \tag{3}$$

follows from the fact that each connected component of G_v has at least two vertices. From (1) and (2) we have

$$w \leq 2n - s, \tag{4}$$

where

$$s = e + x_2 + x_3 + \sum_{i \geq 4} ((i-3)x_i + (i-4)y_i)\,. \tag{5}$$

Finally, $2n \leq 8s$ is easy to verify from (1), (3) and (5), which proves $w \leq \frac{7}{4}n$ by (4).

To conclude the proof, we show that the above presented construction can be realized by a linear time algorithm. Since the two single row routing problems for the two boundaries can be solved in linear time (Theorem 1), it suffices to consider the bipartite case. The size of the input is $O(n)$ since the problem is specified by the two sequences of terminals on the two boundaries. The vertical constraint graph G_v can be generated by a scanning of the columns, that is, in $O(n)$ time. With the same scanning x-coordinates of the two terminals of each net can be recorded. By a scanning of the vertices of G_v starting points of the directed path components of G_v (that is, vertices with in-degree 0) can be found. Once a separate track is assigned to a net, its routing requires constant time (since only the coordinates of the endpoints of the straight wire-segments – together with the positions of the necessary vias – are to be output). The same holds even for exceptinal nets if a column with an empty southern terminal was previously recorded. Once a net is routed, the next one to route can also be found in constant time using the information provided by G_v. Hence the nets can be processed one-by-one, yielding a total running time of $O(n)$. □

We also mention a straightforward corollary of the above theorem.

Corollary 6 *Every channel routing problem can be solved in linear time in the 2-layer Manhattan model if it is allowed to extend the length of the channel by at most one (by introducing an extra column).*

Proof: If an extra column is inserted, the obtained specification cannot be dense. Hence it is solvable in linear time by the above theorem. □

The construction of the above proof does not approximate the minimum width within a constant factor. The natural question arises if the statement of Theorem 4 remains to be true even if it is not allowed to introduce extra columns. However, the answer is no: since the density d of the shift-right-1 problem of Figure 5 is constant 2 and the flux f of the same problem is at most $\lceil \sqrt{n} \rceil$, the following proposition shows that it is impossible to bound the minimum width by $O(d + f)$.

Proposition 7 *The minimum width of a solution of the shift-right-1 problem of Figure 5 is exactly* n *in the 2-layer Manhattan model (if no extra columns are allowed to be added to the grid).*

Proof: The idea is very similar to the proof of unsolvability in Theorem 5. If a solution of the shift-right-1 problem is given, choose a track j in which a horizontal wire segment belonging to a net N is placed (on the bottom layer). The vias at the endpoints of this segment block two columns in the top layer, so the number of remaining columns is $n - 1$. Since the wires of the remaining $n - 1$ nets must also cross track j, it follows that there is no other horizontal segment in track j than the one belonging to N. Since the routing of each net requires at least one horizontal segment, it follows that the required number of tracks is at least n. On the other hand, a solution with width n is shown in Figure 5. □

We mentioned that the minimum width of a solution of the shift-right-1 problem is $\Theta(\sqrt{n})$ if the length of the channel is not fixed. Thus the above proposition proves the fact that allowing extra columns to be added to the grid does not only affect solvability, it can also modify the minimum width significantly.

3.3.3 2 layers, dogleg free Manhattan model

A further, much investigated restriction of 2-layer Manhattan routing is *dogleg free routing*. If each wire in the solution of a channel routing problem in the 2-layer Manhattan model contains a single horizontal wire segment only, then the solution is called *dogleg free*. The importance of such a routing lies in the fact that such a solution clearly minimizes the number of necessary vias.

Gallai's algorithm (Theorem 1) provides a dogleg free solution, hence it also proves that there is no difference between the complexity of Manhattan routing and dogleg free routing in case of the single row routing problem. However, in case of channel routing this is not true any more. For example, the minimum width solution of the channel routing problem of Figure 7 requires 3 and 4 tracks in the Manhattan model (left hand side of Figure 7) and in the dogleg free model (right hand side of Figure 7), respectively.

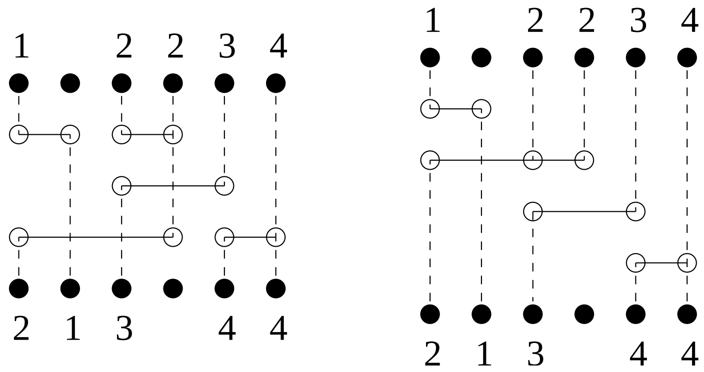

Figure 7

The class of unsolvable specifications is also much wider in the dogleg free model than in the Manhattan model. However, solvable ones can still easily be recognized: a specification is solvable in a dogleg free way if and only if its vertical constraint graph G_v (see Definition 6) is acyclic. Moreover, if this is true then the number of vertices in the longest directed path in G_v is a lower bound on the minimum width. However, determining the minimum width of a solution in the same model remains to be NP-hard [26]. Again, this is still true even if each net is one-sided and contains two terminals only [20].

3.3.4 2 layers, unconstrained model

In the unconstrained model it is true that, in contrast to the Manhattan model, every channel routing problem is solvable in polynomial time on 2 layers with a sufficiently large width (even without extending the length of the channel). This was first proved by M. Marek-Sadowska and E. Kuh [35].

Later A. Recski and F. Strzyzewski found a linear time algorithm.

Theorem 8 (A. Recski and F. Strzyzewski, 1990) [43] *Every channel routing problem can be solved in linear time in the 2-layer unconstrained model.*

Their algorithm does not give an optimal width solution (however, the obtained width can be upper-bounded by n in the bipartite, and by $\frac{3}{2}n$ in the general case). The complexity of the naturally arising question of finding a minimum width routing is not known, but according to the widely accepted conjecture of D. S. Johnson [24] it is NP-hard.

3.3.5 k layers, Manhattan model

It is also true that every channel routing problem is solvable in the k-layer Manhattan model for every $k \geq 3$. This can be proved by a simple modification of Gallai's algorithm (Theorem 1).

Denote the number of horizontal and vertical layers (that is, layers reserved for horizontal and vertical wire segments) by h and v, respectively. Obviously, $h + v = k$ and $|h - v| \leq 1$. Then $\lceil d/h \rceil$ is a trivial lower bound on the minimum width (where d still denotes the density). The following proposition is folklore.

Proposition 9 *Every channel routing problem can be solved with width* $\lceil d/(v-1) \rceil$ *in the Manhattan model (if* $v \geq 2$*). Such a routing can be found in linear time.*

Proof: Number the layers $1, 2, \ldots, k$. We assume that k is odd; if this is not the case then one layer will not be used. In the following construction layers numbered with an odd number will contain vertical segments, while the remaining layers will contain horizontal segments.

The intervals assigned to the nets (see page 17 for a definition) can be packed into d colour classes using Gallai's algorithm (Theorem 1). Use the tracks of layers numbered with an even number to accomodate the colour classes. (More precisely, each colour class corresponds to a set of horizontal wire segments and these sets of segments are placed into the tracks mentioned.) Since the number of horizontal layers used is $v-1$, a width of $\lceil d/(v-1)\rceil$ is indeed sufficient to accomodate all colour classes.

Place a single vertical wire segment in each column of every layer numbered with an odd number; in layers $1, 5, 9, \ldots$ start this single segment from the northern terminal, while in layers $3, 7, 11, \ldots$ the segment should start from the southern terminal. Now each net can be routed simply by introducing the appropriate vias connecting the vertical segments starting from the terminals with the single horizontal segment corresponding to the interval of the net. (Note that most of the vertical wire segments used in the above construction are unnecessary; they were introduced only to make the description of the construction simpler.) □

The above construction gives an optimal solution only if k is odd and $v = h+1$ is a further constraint on the routing model. In general we only get that

$$\left\lceil \frac{d}{\lceil \frac{k}{2} \rceil} \right\rceil \leq \min w \leq \left\lceil \frac{d}{\lceil \frac{k}{2} \rceil - 1} \right\rceil .$$

Thus the above algorithm approximates the optimum width with a factor of $\dfrac{\lceil \frac{k}{2} \rceil}{\lceil \frac{k}{2} \rceil - 1}$. However, the complexity of finding a minimum width routing is known to be NP-hard [40].

3.4 Switchbox Routing

In this section we consider the general switchbox routing problem, that is, in which terminals are placed on all four boundaries of the grid. (In the literature it is common to use the name "swithcbox routing" only for this general case.)

As opposed to single row routing and channel routing, both the length n and the width w are fixed by the specification of such a problem instance. Therefore the objective is to minimize the number of layers needed (or to decide solvability on a given number of layers).

Since switchbox routing is a generalization of channel routing, it follows from Theorem 2 that deciding solvability in the 2-layer Manhattan model is

NP-complete. Again, many heuristics with a good performance have been proposed, see [9, 21, 25, 33] for example.

We have seen that in case of single row and channel routing two layers were always sufficient to solve any problem (and if we restrict ourselves to the Manhattan model, three layers were needed in case of channel routing). This, however, is not true for the general switchbox routing problem. Moreover, no fixed number of layers suffice, which is shown by the following theorem.

Theorem 10 (S. E. Hambrusch, 1985) [22] *For every positive integer k there exists a switchbox routing problem that cannot be solved on k layers in the unconstrained model.*

Proof: Consider the switchbox routing problem of Figure 8. The congestion of the line e is $n + w$, that is, each of the $n + w$ nets have terminals on both sides of e. Therefore the existence of a routing on k layers implies $n+w \leq kw$ since there are w rows on every layer. From this we have $\frac{n}{w} + 1 \leq k$. The value of n and w can be chosen such that this inequality does not hold, which proves the theorem. □

Obviously, if k becomes very large then the construction of the above proof becomes merely theoretical because the length n cannot be very large compared to the width w in practical problems. Denote the ratio $\max(\frac{n}{w}, \frac{w}{n})$ by m. The proof of the above theorem also includes the following statement: $\lceil m \rceil + 1$ is a lower bound on the minimum number of layers of a solution in the worst case. A slight modification of the proof shows that if we restrict ourselves to the Manhattan model then at least $2\lceil m \rceil + 1$ layers are needed in the worst case if $m > 1$ and 4 layers are needed if $m = 1$ (since in the latter case both lines e and f have congestion $2n = 2w$). The following observation, which seems to be new, claims that 4 is also a lower bound in the unconstrained case (and thus the above lower bound of $\lceil m \rceil + 1$ can be improved if $m \leq 2$).

Proposition 11 *Any solution of the switchbox routing problem of Figure 8 uses at least 4 layers (in the unconstrained model).*

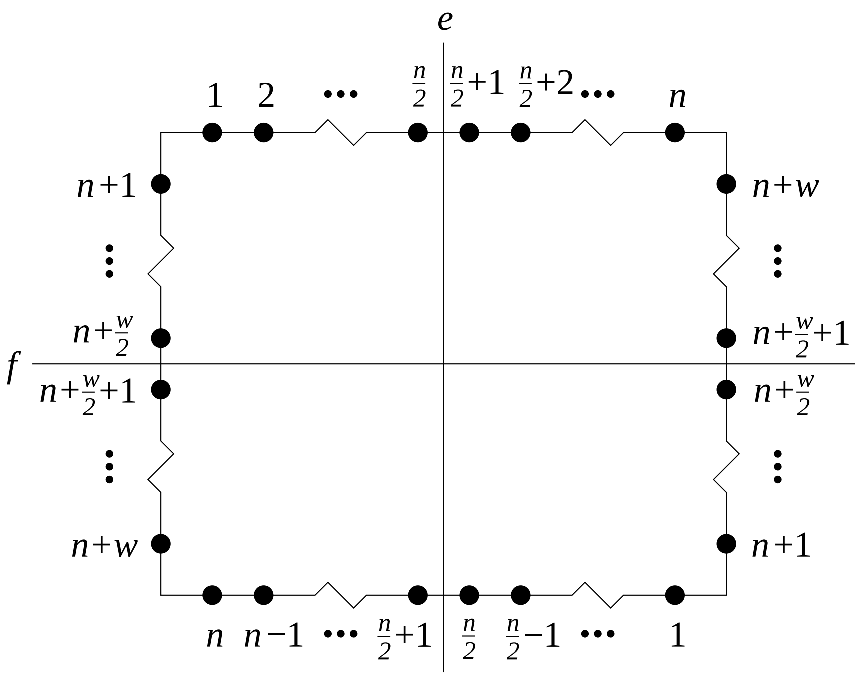

Figure 8

Proof: Assume indirectly that a routing is given on $k \leq 3$ layers. Since each net has only two terminals, it can be assumed that the wires are paths. Denote by N the number of non-terminal vertices of the k-layer grid graph that are used by the routing. It is easy to verify that for every $1 \leq i \leq \frac{n}{2}$ the number of (non-terminal) vertices of a shortest path connecting the terminals of the net i or the net $n+1-i$ is $n+w+1-2i$. Similarly, for every $1 \leq i \leq \frac{w}{2}$ the number of (non-terminal) vertices of the wire routing the net $n+i$ or $n+w+1-i$ is also at least $n+w+1-2i$. Therefore $N \geq 2\sum_{i=1}^{n/2}(n+w+1-2i)+2\sum_{i=1}^{w/2}(n+w+1-2i) = 2nw + \frac{n^2+w^2}{2}$. Moreover, if a net is routed along a shortest path then its routing belongs to a single layer. Hence the number of such nets is at most k since no two nets can be routed on a common single layer for simple topological reasons. Now $n+w > 3 \geq k$ implies $N > 2nw + \frac{n^2+w^2}{2}$. On the other hand, $N \leq knw$ since each layer has nw vertices. From this we have $k > 2 + \frac{n^2+w^2}{2nw} \geq 3$, a contradiction.

□

The above bounds show that the minimum number of layers in the worst case for the routing of a switchbox problem can be lower bounded as a function of the ratio $m = \max(\frac{n}{w}, \frac{w}{n})$. It is a natural question whether there is also an upper bound on the necessary number of layers as a function of m? The following theorem answers this question in the affirmative.

Theorem 12 (E. Boros, A. Recski and F. Wettl, 1995) [5] *Any switchbox routing problem can be solved in linear time on at most 18 layers if $m \leq 2$ and on at most $2m + 14$ layers if $m > 2$ in the unconstrained model.*

The authors conjecture that actually $\lceil m \rceil + 3$ layers also suffice; this, however, is still open.

Section 4 will be dedicated completely to switchbox routing in the Manhattan model. Among other results we will prove an improvement on the above theorem: we will show that $2\lceil m \rceil + 4$ layers suffice, even in the Manhattan model.

3.5 Gamma Routing

The special case of the switchbox routing problem in which all the terminals of each net are situated on two adjacent boundaries of the grid (say, they are all northern or western) is called *gamma routing*. The objective in a gamma routing problem is again to minimize the number of layers needed (since both the length n and the width w are fixed).

Although the complexity of the gamma routing problem is not known, it seems to be substantially easier than general switchbox routing (or even channel routing) and certain special cases are known to be polynomially tractable. For example, the following observation is trivial: if each net has at least one terminal both on the northern and on the western boundaries then the problem instance can be solved on two layers in the Manhattan model.

The problem of gamma routing seems to be completely solved in the two-layer Manhattan model if each net contains two terminals only: S. A. Wu and J. JáJá [57] gave an algorithm which decides the solvability of such a problem instance and constructs a solution if possible. The running time of their algorithm is linear in the number of nets and, furthermore, it uses the minimum number of vias. The core of the algorithm is an extension of Gallai's method (Theorem 1).

E. Boros, A. Recski, T. Szkaliczki and F. Wettl [4] provide solutions under partly more general conditions: they allow each net to have more than one terminals either on the northern or the western boundary (but not both). On the other hand, the routing model is restricted to the two-layer, dogleg free Manhattan model, where a dogleg free solution means that the routing of a net with k terminals uses $k-1$ vias. They give an algorithm which decides solvability and, if possible, constructs a solution for such problems. The running time of their algorithm is $O(nwt)$, where n and w are the length and the width of the board and t is the number of nets. Their method is simplified and slightly extended in [54].

We mention that the minimum number of layers needed for a gamma routing problem in the Manhattan model can be approximated in linear time with an additive constant of 3 in a trivial way; see the last paragraph of Section 4.2.

3.6 Edge-Disjoint Routing

In Definition 2 we defined the solution of a detailed routing problem as a set of pairwise vertex-disjoint, connected subgraphs. However, as we have already mentioned in Section 2.1, the variant of the above definition, in which the subgraphs are required to be edge-disjoint only, is also of interest. This is called *edge-disjoint routing* and it is considered mostly in the single layer (planar) case.

Apparently, a planar edge-disjoint solution of a routing problem has no direct application. However, it can be regarded as a first step to obtain a vertex-disjoint routing. Assume, for example, that a (vertex-disjoint) solution of a routing problem is given in the 2-layer Manhattan model. Then the projection of the two layers on a single layer gives an edge-disjoint solution of the same problem. Moreover, wires of the projection can only intersect in a + shape, they never share a corner as in Figure 9. (The situation shown in Figure 9 is called a *knock-knee*, and planar edge-disjoint routing is sometimes also referred to as *knock-knee routing*.) Conversely, if no knock-knee occurs in a planar edge-disjoint routing, then it can easily be converted into a two-layer Manhattan routing.

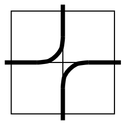

Figure 9

In general, in the *layer assignment problem* we aim at transforming an edge-disjoint routing into a vertex-disjoint one on as few layers as possible. That is, a layer is assigned to each edge of the edge-disjoint routing, such that the following conditions are fulfilled:

1. wire edges sharing a vertex are placed on different layers, if they belong to different nets;

2. wire edges sharing a vertex v that belong to the same net are

   (a) either placed on the same layer,

   (b) or they are placed on two different layers such that the vertices corresponding to v on the layers between the two assigned layers are not occupied by any other wire (so that the necessary vias can be inserted).

It is fairly easy to decide whether a planar edge-disjoint routing can be transformed into a vertex-disjoint one on two layers. However, the range of such routings is very limited. The problem is much more complicated with three layers.

Theorem 13 (W. Lipski, Jr., 1984) [32] *It is NP-complete to decide, whether a planar edge-disjoint routing can be transformed into a vertex disjoint one on three layers.*

In light of the above theorem the following result can be surprising.

Theorem 14 (M.L. Brady and D. J. Brown, 1984) [6] *Every planar edge-disjoint routing can be transformed into a vertex disjoint one on four layers.*

The research of edge-disjoint routing started with the pioneering paper of A. Frank [13], in which a necessary and sufficient condition for the solvability of a routing problem is given, provided that each net consists of two terminals only. This theorem adds a new item into the toolbox for proving unsolvability of routing problems: besides the usual arguments involving congestion, the notion of parity also plays an important role.

Before we state the theorem we mention that, in case of edge-disjoint routing, the usual formulation of the routing problem differs from the one given in Section 3.1 in two details: terminals can appear on the corners of the board and wires can use grid edges of the boundaries of the board. Furthermore, in Frank's theorem it is also permitted to place at most two terminals on the corners of the board. With this extension it is easy to see that the setting of the problem is equivalent to the original one: if a planar routing problem is defined in accordance with Section 3.1 and the terminals are moved one unit to the direction of the opposite boundary then we get an equivalent routing problem in which the wires can use the boundaries.

The notion of congestion can be defined for edge-disjoint routing problems as before: if e is a vertical or a horizontal line that cuts the grid into two then $c(e)$ denotes the number of nets separated by e. If w and n denote the number of rows and columns of the grid, respectively then $c(e) \leq w$ for each vertical line e and $c(e) \leq n$ for each horizontal line e are straightforward necessary conditions for solvability. However, these conditions are not sufficient as shown by the example of Figure 10.

In the remaining part of this section we restrict ourselves to planar, edge-disjoint routing problems in which each net has two terminals only. Therefore each wire in any solution can be assumed to be a path. Let X be a subset of the vertices of the grid. Denote by $e(X)$ the number of grid edges leaving X and denote by $n(X)$ the number of nets having exactly one terminal in

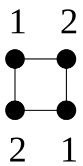

Figure 10

X. Call X an *odd set* if $e(X)+n(X)$ is odd. The following simple lemma is going to be the main tool for obtaining a stronger necessary condition.

Lemma 15 (A. Frank, 1982) [13] *If X is an odd set then no solution of a given routing problem can use all grid edges leaving X.*

Proof: Denote by E_X the set of edges leaving X. If a net N has exactly one terminal in X then its wire uses an odd number of edges from E_X. Similarly, if a net N has no terminal in X or it has both terminals in X then its wire uses an even number of edges from E_X. Therefore the number of edges in E_X which are used by an arbitrary routing has the same parity as $n(X)$. Since X is odd, $n(X)$ and $e(X)$ have different parities, which proves the lemma. □

Call a horizontal line e a *saturated line* if $c(e) = n$ holds; similarly, a vertical line e is saturated if $c(e) = w$. Denote the number of horizontal saturated lines by h and let e be any vertical line. Then the part of the board left from e is cut into $h+1$ rectangular areas by the saturated horizontal lines; the number of odd sets among the subsets of vertices corresponding to these $h+1$ rectangular areas is called the *parity congestion* of e. The parity congestion, which is denoted by $p(e)$, is defined in an analogous way for horizontal lines e.

Theorem 16 (A. Frank, 1982) [13] *If each net contains two terminals only then a planar, edge-disjoint routing problem is solvable if and only if $c(e)+p(e) \leq w$ holds for each vertical line e and $c(e)+p(e) \leq n$ holds for each horizontal line e.*

Proof of necessity: Assume that a solution is given. Let e be any vertical line. Denote the odd sets in the definition of $p(e)$ by $X_1, \ldots, X_{p(e)}$. For all

$1 \leq i \leq p(e)$ there is an edge e_i leaving X_i which is not used by the routing by Lemma 15. Obviously, e_i cannot intersect a saturated line, thus it is horizontal. Therefore e is intersected by at least $p(e)$ edges not used by the routing and at least $c(e)$ further edges which are used by the routing. This proves $c(e) + p(e) \leq w$. The proof is analogous for horizontal lines e. □

Figure 11 (which is taken from [13]) shows an example for the above theorem. The parity congestion of the line e in the left hand side routing problem is $p(e) = 4$, since the number of saturated horizontal lines (shown with dotted lines) is 3 and all 4 rectangular areas left from e (shown with dashed lines) correspond to odd sets. The congestion of e is $c(e) = 4$, thus the problem is not solvable by $4+4 > 6$. The right hand side routing problem together with its solution is shown only to illustrate that a minor change in the specification can result in a solvable problem.

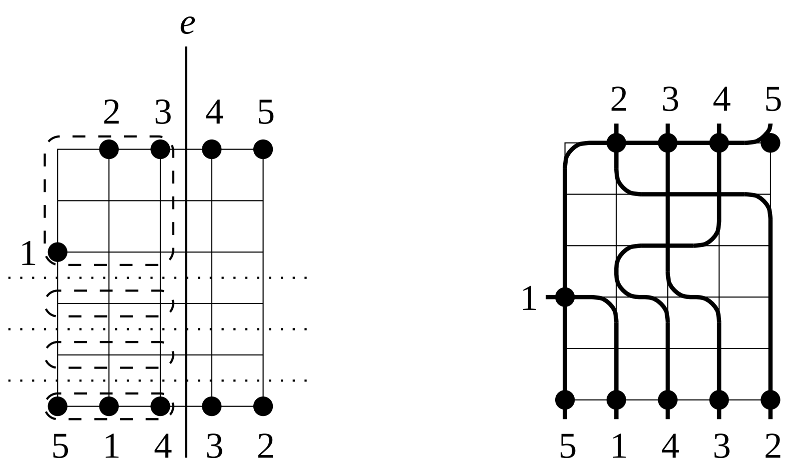

Figure 11

The proof of sufficiency of Theorem 16 also involves an algorithm which runs is linear time in the area of the obtained routing. A sublinear algorithm in the area was also found by K. Mehlhorn and F. P. Preparata [36]. Many papers have considered the general case (in which the nets can have arbitrarily many terminals), see [38] for a survey.

4 Switchbox Routing in the Multilayer Manhattan Model

In this section we consider the switchbox routing problem again. We present a linear time algorithm that solves any such problem in the Manhattan model. The number of layers used will not be too far from the optimum; see the details below. As a corollary, we will be able to improve the result of Theorem 12.

4.1 The Square-Shaped Switchbox

In this section we consider the square-shaped switchbox first, that is, for which $n = w$ (and thus the value of m defined in Section 3.4 is 1). We have seen that the necessary number of layers to solve such a problem is at least 4 in the worst case, even in the unconstrained model (see Proposition 11). Furthermore, it is verified in [19] that the number of layers needed is at least 6 in the Manhattan model for the 2×2 problem of Figure 12. Although no better lower bound than 4 is known in the $n \times n$ case, the following theorem is not far from being optimal.

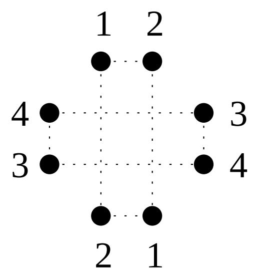

Figure 12

Theorem 17 (Sz. D., 1997) [51] *Every square-shaped switchbox can be solved on 6 layers in the Manhattan model.*

Proof: We illustrate the proof on the switchbox routing problem of Figure 13.

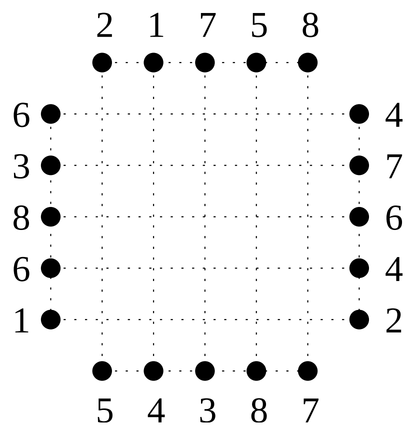

Figure 13

Obviously, one can assume without loss of generality that every terminal belongs to some net and that every net has at least two elements.

In order to simplify reference to them, we classify the nets according to the boundaries of the switchbox on which they occur and the number of occurrences on the boundaries. We say, for example, that a net is SE type if it has terminals on the southern and eastern boundaries, but not on the other two ones; a net is N type if it has only northern terminals; a net is NSW type if it has terminals on the northern, southern and western boundaries, but not on the eastern one, etc. Furthermore we say, for example, that a net is $\mathsf{S}_1\mathsf{W}_1$ type if it has exactly two terminals, one of which is on the southern boundary and the other is on the western one; a net is $\mathsf{N}_1\mathsf{W}$ type if it has one terminal on the north, some (maybe one) on the west and none elsewhere; a net is $\mathsf{S}_{\geq 2}\mathsf{W}_1$ type if it has at least two terminals on the south, exactly one on the west and none elsewhere, etc.

Getting on to the proof at last, we may obviously assume without loss of generality that out of NE, NW, SE, SW type nets NE nets are (one of) the greatest in number. (In the example of Figure 13 there is exactly one of all the four types, so the condition holds.) Having this in mind, the sketch of the construction is shown in Figure 14. The term S-comb means, for example, that on that layer a wire segment leads from each southern terminal to (for the sake of simplicity) the opposite boundary (unnecessary wire ends can be removed later on). The 2nd layer will contain horizontal wire segments while the 5th will contain vertical ones, thus the construction will indeed be in the Manhattan model.

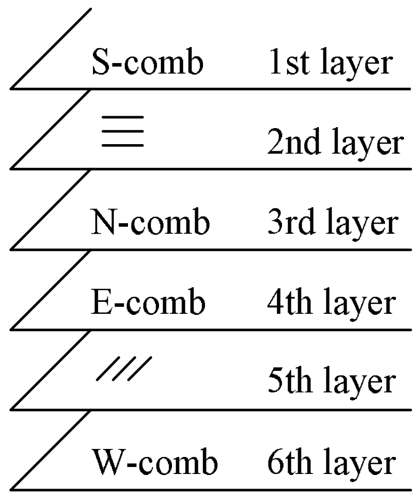

Figure 14

We associate tracks of the 2nd layer and columns of the 5th layer with the nets: these tracks and columns may only contain wire segments belonging to the net with which they are associated. The following nets are given a track on the 2nd layer (each listed net is given one track):

- nets having altogether at least two terminals on the northern and southern boundaries, but which are not NE type;
- $\mathsf{N}_1\mathsf{W}$ type nets;
- half of the $\mathsf{S}_1\mathsf{W}_1$ type nets (if there is an odd number of these then we mean the upper integer part).

Similarly, the following nets are given a column on the 5th layer:

- nets having altogether at least two terminals on the eastern and western boundaries, but which are not NE type;
- SE_1 type nets;

- the remaining half of the S_1W_1 type nets.

Consequently, in the example of Figure 13 nets 1,3,5,7 and 8 are given a track on the 2nd layer, nets 4 and 6 are given a column on the 5th layer. (The rest of the columns of the 5th layer are not associated with any net, these columns remain empty.)

Let us accept for a while that there is indeed a sufficient number of tracks and columns on the 2nd and 5th layers to accomodate the listed nets; we will come back to this at the end of the proof. Instead we make rules for how the nets that are given a track (or column) on the 2nd (or 5th) layer should share these. If a net is given a track on the 2nd layer and there is a terminal belonging to this net on the **western** boundary then this net must be given a track next to (one of) its western occurrence(s). (This obviously can always be managed.) Such nets will further on be called *fixed to the west.* If a net does not have a terminal on the western boundary then there is no restriction as to which track it should be given. Similarly, if a net is given a column on the 5th layer and there is a terminal belonging to this net on the **southern** boundary then this net must be given a track above (one of) its southern occurrence(s) (and such nets are called *fixed to the south*). A possible distribution of the tracks and columns on the 2nd and 5th layers of the illustrative example is shown in Figure 15. The tracks and columns are marked with the number of the net with which they are associated. Tracks and columns assigned to nets that are fixed to the west or to the south are extended to the respective boundary.

After all this preparation the desired routing can at last be given. Vertical wire segments on the 1st and 3rd layers coming from the northern and southern members of a net N_i can be interconnected by means of a single wire segment with the necessary vias in the track of the 2nd layer associated with the net N_i. If N_i is fixed to the west then the western terminal left to the track assigned to N_i also belongs to N_i, so the wire segment in the track should be extended to reach this terminal as well. Similarly, wire segments (on the 4th and 6th layers) coming from the western and eastern members of a net can be interconnected in the columns of the 5th layer. Wire segments belonging to nets that are fixed to the south are likewise extended to

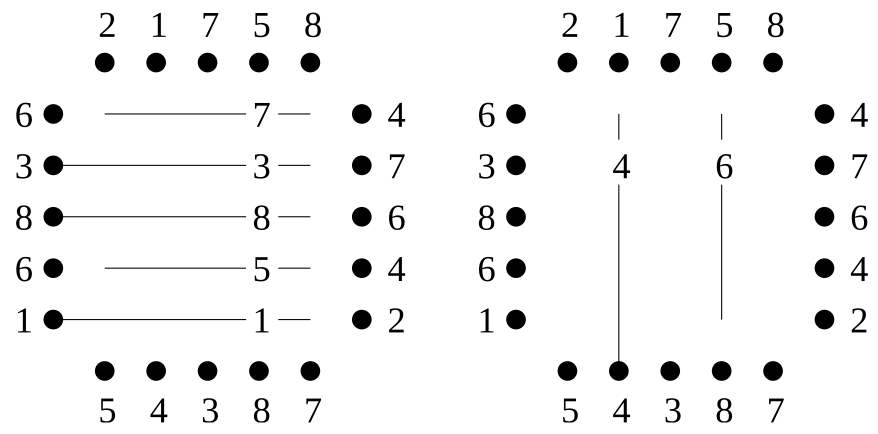

Figure 15

reach the southern terminals. Finally, we introduce extra vias between the 3rd and 4th layers suitably: if a northern and an eastern terminal belong to the same net then the vertical wire segment coming from the northern terminal on the 3rd layer and the horizontal wire segment coming from the eastern terminal on the 4th layer can be connected through an appropriate via. (Evidently, it is not necessary to take advantage of all the possibilities to introduce the above mentioned vias. It is trivial that for each net the number of vias needed between the 3rd and 4th layers is at most the total number northern and eastern terminals minus 1.)

The final routing of our illustrative example is shown in Figure 16 (unnecessary wire ends have also been removed from the 'comb layers'). Vias going up and down from a layer are denoted by squares and circles, respectively.

We claim that the above described routing is good. This must be verified by distinguishing cases. If a net is of N, S or NS type then it was given a track on the 2nd layer (every net has at least two members) so its members became interconnected in the above described way. The situation is the same with E, W and EW type nets which were given a column on the 5th layer. If a net is NW type then it was surely given a track on the 2nd layer, through which northern members of the net became interconnected. In addition, since this net is fixed to the west, northern members also became connected with one of the western terminals. If, on the other hand, the net has more than one

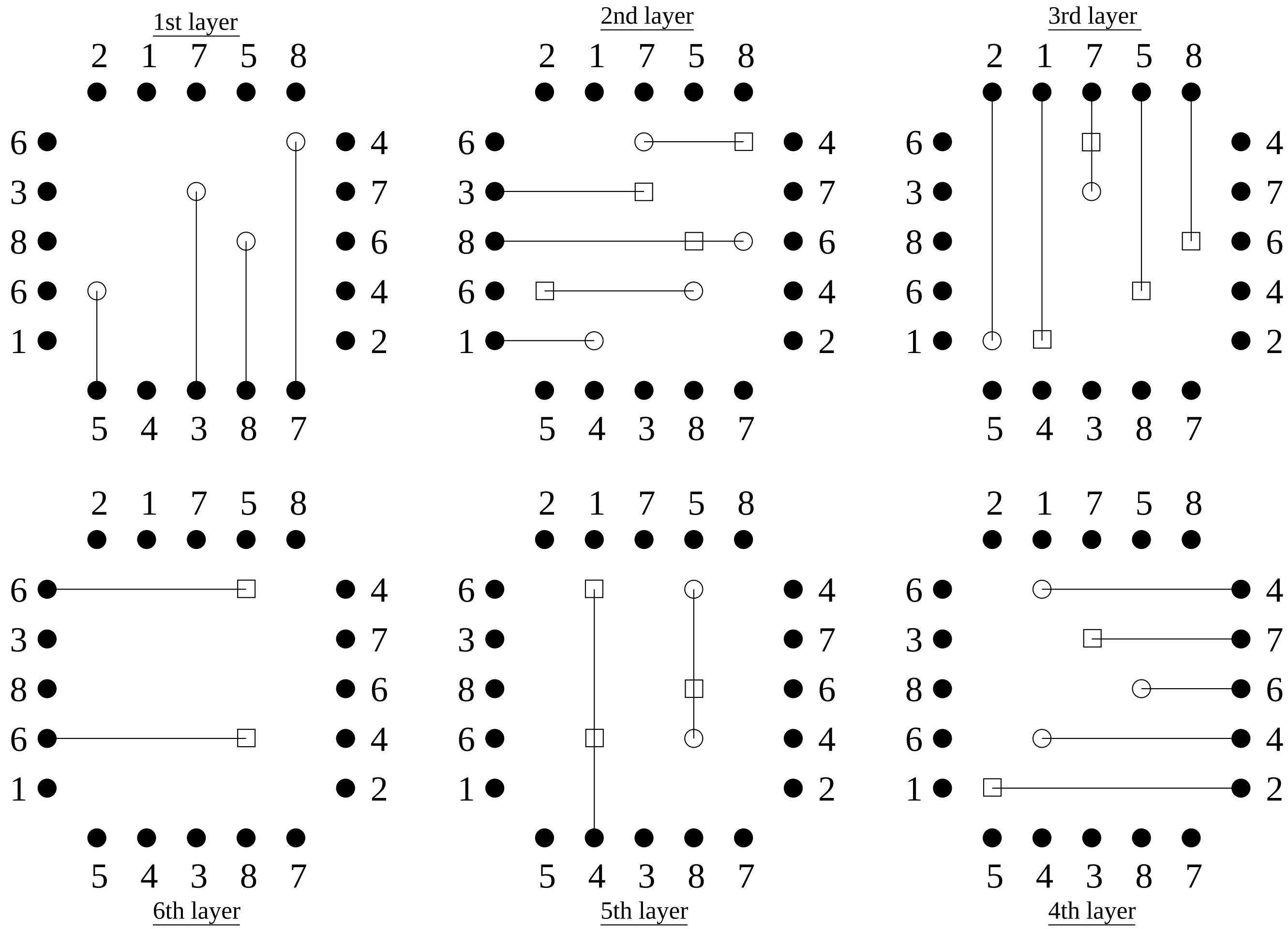

Figure 16

western terminals then it was also given a column on the 5th layer through which western members of the net also became interconnected. Analogously, SE type nets were surely given a column on the 5th layer, these nets are fixed to the south and if one of them has more than one southern terminals then it was also given a track on the 2nd layer. The wiring of NE type nets is solved simply by the extra vias between the 3rd and 4th layers. The most difficult is the case of SW type nets. $\mathsf{S}_{\geq 2}\mathsf{W}_{\geq 2}$ nets were both given a track on the 2nd and a column on the 5th layer, thus both the southern and the western members of these nets became interconnected within themselves. Moreover, since such a net is fixed both to the west and to the south, the two groups also became (doubly) connected with each other. $\mathsf{S}_{\geq 2}\mathsf{W}_1$ nets were only given a track on the 2nd layer, but such a net is fixed to the west, thus the southern terminals are connected with each other and the single western terminal. Wiring of the $\mathsf{S}_1\mathsf{W}_{\geq 2}$ nets is analogous. Finally, whether a $\mathsf{S}_1\mathsf{W}_1$ type net was given a track on the 2nd or a column on the 5th layer, its wiring is solved since the net is fixed either to the west or to the south.

We can start to deal with nets appearing on at least 3 boundaries. NSW nets were surely given a track on the 2nd layer and they are fixed to the west, which solves the interconnection of the northern and southern members as well as the connection of all these to (one of) the western one(s). If there are more than one western members then the net was also given a column on the 5th layer, thus the wiring is complete. In case of NSE nets the only difference is that such a net is not fixed to the west, but this is compensated by the extra vias between the 3rd and 4th layers which interconnect the northern and eastern terminals. Wiring of the SEW and NEW type nets is analogous to the above. Finally, NSWE type nets were both given a track on the 2nd and a column on the 5th layer, such nets are fixed both to the west and the south, thus the wiring of these nets is more than complete.

We still owe the proof that there is a sufficient number of tracks and columns on the 2nd and 5th layers, respectively, to be enough for the listed nets. We prove this only for the 2nd layer, the proof for the 5th layer is similar. Let n denote the length (and at the same time width) of the switchbox; let x denote the number of nets having altogether at least 2 terminals on the northern and southern boundaries but which are not NE type; let y denote the number of $\mathsf{N_1W}$ type nets; let z denote the number of $\mathsf{S_1W_1}$ nets; finally, let t denote the number of NE nets. Using these notations, $x + y + \lceil \frac{z}{2} \rceil \leq n$ is to be proved. Since there are altogether $2n$ terminals on the northern and southern boundaries, obviously $2x + y + z + t \leq 2n$. Because of the assumption made at the beginning of the proof $y \leq t$ holds, thus $2x + 2y + z \leq 2n$ is also true. Divided by 2 we get $x + y + \frac{z}{2} \leq n$ and since x, y and n are integers, the proof is complete. □

Proposition 18 *The construction of the above proof can be realized by a linear time algorithm.*

Proof: The length of the input is obviously $4n$ (since four sequences of terminals form the input.)

The algorithm should consider the nets one by one and decide whether the examined net should be given a track on the 2nd layer and/or a column on the 5th one (this requires nothing but scanning the terminals belonging to the net). It can at the same time assign the possibly awarded tracks and/or columns to the nets and place the necessary wire segments and the vias in the tracks/columns. For all this, it is enough to scan the terminals of the nets a second time. Meanwhile, the algorithm must make sure to give appropriate tracks and columns to the nets that are fixed to the west or to the south. If this meets with difficulties because a necessary track or column was previously given to a formerly examined net then it should simply remove the wire segment in way to a still empty track or column (preserving the position of the wire segment and the vias in the track or column). During the repeated scanning of the nets, the algorithm deals twice with each terminal ($4n$ in number), so the number of operations is linear.

In the next phase the extra vias are placed between the 3rd and 4th layers. We mentioned before that the necessary number of vias for each net is at most the number of northern and eastern terminals (minus 1). Hence the total number of vias needed between the 3rd and 4th layers is not more than the number of northern and eastern terminals together, that is, $2n$.

The ‘combs’ of the 1st, 3rd, 4th and 6th layers are made in the last phase. If we want to avoid unnecessary wire ends, there is no need to draw the wires to the opposite boundary, it is enough to draw them to the last via (which we already know). $\square$

4.2 The Switchbox of Arbitrary Shape

In this section we consider the general switchbox routing problem. The following theorem, which is a new result of this dissertation, shows that the minimum number of layers needed for a switchbox routing problem in the Manhattan model can be approximated with an additive constant of 5 in linear time.

Theorem 19 *Assume that a switchbox routing problem instance with length n and width w is given such that $n \geq w$. Denote by d the maximum congestion of all vertical lines that cut the grid into two. Then the minimum number of layers needed for a solution in the Manhattan model is between $2\lceil \frac{d}{w} \rceil - 1$ and $2\lceil \frac{d}{w} \rceil + 4$. A solution on $2\lceil \frac{d}{w} \rceil + 4$ layers can be found in linear time.*

Proof: If e is a vertical line with congestion d then d wires must intersect e. At most w wires can intersect e on each layer reserved for horizontal wire segments. Thus the number of such layers is at least $\lceil \frac{d}{w} \rceil$ which proves that $2\lceil \frac{d}{w} \rceil - 1$ is a lower bound on the total number of layers.

We suitably modify the proof of Theorem 17 to obtain the upper bound. There is nothing to change on layers 4, 5 and 6 (see Figure 14). A similar calculation as that in the last paragraph of the proof of Theorem 17 shows that the number of nets that require a column on the 5th layer (see page 40) is at most w which shows that the same construction works by $w \leq n$.

The main modification compared to the proof of Theorem 17 will be that, instead of awarding complete tracks of the (original) 2nd layer to nets, we will use the same track to accomodate horizontal wire segments for more than one nets. (The idea is going to be very similar to that of Proposition 9.) For this, we associate a horizontal interval to each net that received a track on the 2nd layer in the proof of Theorem 17 (see page 40); denote the set of these nets by $\mathcal{N}'$. An interval corresponding to a net $N \in \mathcal{N}'$ stretches from the column of the leftmost terminal to the column of the rightmost terminal belonging to N. (Here we consider all four boundaries; if, for example a net has a western terminal then the left endpoint of its interval is on the western boundary.)

Denote the density of the routing problem defined by the nets of $\mathcal{N}'$ by d'. Using Gallai's algorithm (Theorem 1) we can pack the intervals assigned to nets of $\mathcal{N}'$ into d' colour classes. If $d' \leq w$ then the original construction of Theorem 17 works with a straightforward modification: sets of horizontal wire segments corresponding to the colour classes are placed in a common track of the 2nd layer. Each colour class contains at most one net that is fixed to the west (since the intervals of any two such nets intersect on the western boundary), thus wire segments belonging to these nets can be extended to reach the terminal on the western boundary in the same way. If, on the other hand, $d' > w$ holds then we add further layers to the construction of Figure 14: a new layer is placed above the 1st one to hold horizontal wire segments, a N-comb comes on top of this one, again a new layer for horizontal wire segments, then a S-comb, etc. In other words, new layers are sandwiched between N-combs and S-combs. The horizontal tracks on these new layers are going to serve as the tracks of the 2nd layer did in Theorem 17.

Since d' colour classes have to be packed into tracks of the "sandwiched" layers (including the original 2nd layer), $\lceil \frac{d'}{w} \rceil$ such layers are needed. Therefore, as it can be seen from the construction, $2\lceil \frac{d'}{w} \rceil + 4$ layers are used altogether. This proves the statement of the theorem by $d' \leq d$.

It is obvious that the algorithm realizing the above construction is also linear, as both Gallai's algorithm (Theorem 1) and the algorithm of Proposition 18 were linear. □

We have seen in Section 3.4 that at least $2\lceil m \rceil + 1$ layers are needed in the worst case to solve a switchbox routing problem in the Manhattan model. (Recall that $m = \max(\frac{n}{w}, \frac{w}{n})$, where n and w are the length and the width of the switchbox, respectively.) An upper bound of $\max\{18, 2m + 14\}$ was provided by Theorem 12, but only in the unconstrained model. As a corollary of the above theorem we can improve this bound, even in the

Manhattan model.

Corollary 20 (Sz. D., 1997) [51] *Every switchbox can be solved in linear time on* $2\lceil m\rceil + 4$ *layers in the Manhattan model.*

Proof: $n \geq w$ can be assumed. We show that the construction of the proof of Theorem 19 proves the present theorem as well; we adopt the terminology of that proof. A similar calculation as that of the last paragraph in the proof of Theorem 17 shows that the number of nets in $\mathcal{N}'$ is at most n, hence $d' \leq n$ holds. Therefore the number of layers used in the construction is $2\lceil \frac{d'}{w} \rceil + 4 \leq 2\lceil \frac{n}{w} \rceil + 4 = 2\lceil m\rceil + 4$. □

We mention two further corollaries of Theorem 19. Assume that one of the four boundaries of a switchbox routing problem contains no terminal; this special case is sometimes referred to as *C-shaped routing problem*. This means that the lowest layer in the construction of Theorem 19 (see Figure 14) contains no wire segment, thus it can be omitted. (This corresponds to the assumption in the proof of Theorem 17 that the number of NE nets is at least the number of NW, SE and SW nets.) Hence the upper bound can be improved to $2\lceil \frac{d}{w} \rceil + 3$ in case of a C-shaped routing problem (while the lower bound remains unchanged). In particular, every C-shaped routing problem can be solved on 5 layers in the Manhattan model if $n = w$ [41, 42].

Furthermore, if we restrict ourselves to the gamma routing problem (see Section 3.5) then the topmost layer of the construction of Theorem 19 can also be removed. Assume now that the western and southern boundaries are empty. In this case, instead of adding N-combs and S-combs alternatingly, we only need N-combs (for every second layer). Hence the topmost layer, which would be a N-comb, is not needed, since the intervals on any "sandwiched layer" are connected to the northern terminals on the layer below. Therefore the upper bound becomes $2\lceil \frac{d}{w} \rceil + 2$ (while the lower bound remains $2\lceil \frac{d}{w} \rceil - 1$). We mention, however, that in this case there are no fixed nets any more, so the construction becomes trivial; it can be regarded as a double application of Gallai's algorithm (Theorem 1).

5 Single Active Layer Routing

Although the solution of a switchbox problem can require arbitrarily many layers, switchbox routing can still be regarded as a 2-dimensional problem: the input consists of four sequences (the terminals on the four boundaries) and the output consists of a fixed number of planar layers (provided that the value of m defined in Section 3.4 is fixed).

Due to the quick improvement of routing technology, research has recently turned towards 'real' 3-dimensional routing. There are plenty of deep results in this area, see [1, 2, 10, 12, 15, 27, 28, 29, 48, 50], for example. Most of them embed certain 'universal-purpose' graphs (like n-permuters, n-rearrangeable permutation networks, shuffle-exchange graphs) into the 3-dimensional grid, ensuring that *pairs* of terminals can be connected, moreover, in some papers along *edge-disjoint* paths. In the results of this section we allow *multiterminal nets* as well, and ensure *vertex disjoint* paths (or Steiner-trees) for the interconnections of the terminals within each net.

Throughout, we restrict ourselves to the single active layer routing problem, where terminals to be interconnected are situated on a rectangular planar grid of size $n \times w$ and the routing should be realized in a cubic grid of height h above the original grid that contains the terminals. Evidently, the height h is to be optimized. Henceforth we will use the term 'vertical direction' to refer to the direction of h (that is, the direction perpendicular to the $n \times w$ rectangle) and not for the direction of w.

One can easily see even in small instances like 4×1 or 2×2 (see Figure 17) that a routing is usually impossible unless either the length n or the width w may be extended by introducing extra rows or columns between rows and columns of the original grid. It seems that the nature of the problem depends fundamentally on whether only one of the dimensions n and w is extended by a constant factor or both of them.

In the next section we define the single active layer routing problem and we prove two straightforward bounds. Then we consider the case in which only the width w can be extended by a contstant factor (and n is extended by at most 1). In the last section we deal with the case in which both n and

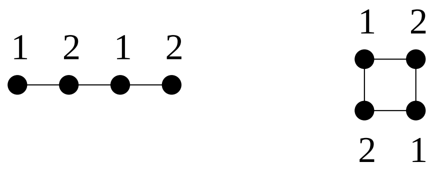

Figure 17

w are extended.

5.1 Definitions and Straightforward Bounds

The single active layer routing problem is a special case of the detailed routing problem defined in Definitions 1 and 2. However, for the sake of convenience, we now give an independent definition for this problem.

Definition 7 *The vertices of a given (planar) grid of size $w \times n$ (consisting of w rows and n columns) are called* terminals. *A* net *N is a set of terminals. A single active layer routing problem (or* SALRP *for short) is a set $\mathcal{N} = \{N_1, N_2, \ldots, N_t\}$ of pairwise disjoint nets. n and w are the* length *and the* width *of the routing problem, respectively.*

Definition 8 *By a* spacing of s_w in direction w *we are going to mean that we introduce $s_w - 1$ pieces of extra columns between every two consecutive columns (and also to the right hand side of the rightmost column) of the original grid. This way the width of the grid is extended to $w' = s_w \cdot w$. A* spacing of s_n in direction n *is defined analogously.*

Definition 9 *A* solution with a given spacing s_w and s_n *of a routing problem $\mathcal{N} = \{N_1, N_2, \ldots, N_t\}$ is a set $\mathcal{H} = \{H_1, H_2, \ldots, H_t\}$ of pairwise vertex-disjoint, connected subgraphs in the cubic grid of size $(w \cdot s_w) \times (n \cdot s_n) \times h$ (above the original planar grid containing the terminals) such that $N_i \subset V(H_i)$, that is, H_i connects the terminals of N_i. The subgraphs H_i are again called* wires. *h is called the* height *of the routing and a cross-section of the cubic grid perpendicular to the height (and therefore of size $(w \cdot s_w) \times (n \cdot s_n)$) is called a* layer.

Again, the wires can evidently be chosen to be Steiner-trees.

In order to simplify the description of the routings, we are going to use the term *w-wire segment* to refer to a wire segment that is parallel with the width w of the grid. The meaning of the terms *n-wire segment* and *h-wire segment* is analogous.

Lemma 21 *For any given n and s_w there exists a routing problem that cannot be solved with height h smaller than $\frac{n}{2s_w}$.*

Proof: Let, for simplicity, the width and the length be even, let $w = 2a$ and $n = 2b$. Consider the following example (the idea is very similar to that in Theorem 10). Suppose that each net consists of two terminals in central-symmetric position as shown in Figure 18.

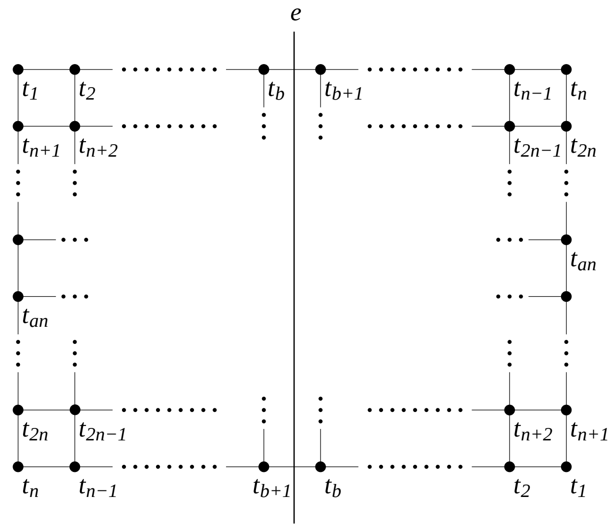

Figure 18

The number of nets is an. Since each net is cut into two by the central vertical line e, any routing with width $w' = s_w \cdot w$ and height h must satisfy $w'h \geq an$. Therefore $h \geq (w/2w')n$, hence $h \geq \frac{n}{2s_w}$. □

Lemma 22 *If $s_w \geq 2$ and $s_n \geq 2$ then every routing problem can be solved with height $h \leq \frac{wn}{2}$.*

Proof: We assign a separate layer to each net. For every terminal we introduce an h-wire segment to connect the terminal with the layer of its net. The interconnection of the terminals of each net can now be performed trivially on its layer using the extra rows and columns guaranteed by the spacing in both directions.

Since 1-terminal nets can be disregarded, the number of nets is at most $\frac{1}{2}nw$ thus $h \leq \frac{wn}{2}$ follows immediately. □

5.2 The $s_n = 1$ Case

An alternative interpretation of Lemma 22 is that if we fix w then there is a routing of height $h = O(n)$, provided that $s_w, s_n \geq 2$. However, the truth of the same statement is not at all obvious in the $s_n = 1$ case. In this section we aim at proving that if w is fixed and $s_w \geq 8$ holds then a routing of height $h = O(n)$ can be found even in the $s_n = 1$ case (and in $t = O(n)$ time).

Theorem 23 (Recski A. and Sz. D., 2001) [44] *If $s_w \geq 8$ then for any fixed value of w and for any n a single active layer routing problem can always be solved in time $t = O(n)$ and with height $h = O(n)$ such that the length n is preserved or increased by at most one. Both linear bounds are best possible.*

Our algorithm gives $t = O(w^3 n)$ and $h = O(wD)$, where D is defined below. (The straightforward lower bound for the time is the length of the input, that is, $t = \Omega(wn)$.)

Proof: Since w is fixed, one can also imagine the input as w rows of terminals (each of length n) or as a set of $\binom{w}{2}$ channel routing problems, each with length n and with a given density. Let D be the maximum of these densities. Clearly, $D \leq n$. (We mention that in the example of Lemma 21 $D = n$ holds, therefore the lower bound $h = \Omega(D)$ is also true.)

5.2.1 Two Simple Steps

1. Suppose at first that $w = 1$. Then what we have is essentially a single row routing problem with density d. Each net determines an interval of length at most n and these intervals can be packed in a vertex-disjoint way into d parallel lines (or tracks), using Gallai's algorithm (Theorem 1). Using the 2-layer Manhattan model, we can arrange the tracks in a horizontal plane, as shown in the top of Figure 19, thus realizing a routing with $w' = d$ and $h = 2$. However, alternatively these tracks can occupy either a vertical plane, leading to $w' = 2$ and $h = d$, or two vertical planes, leading to $w' = 3$ and $h = \lceil d/2 \rceil$, see the middle and the bottom drawing of Figure 19, respectively. (Theoretically one can pack the tracks to more vertical planes and thus ensure $h = \lceil 3d/(2w') \rceil$ for larger values of w' as well but it does not seem to be interesting.) Throughout in Figures 19, 20 and 23 continuous lines denote wires while dotted lines are for the indication of coplanarity only.

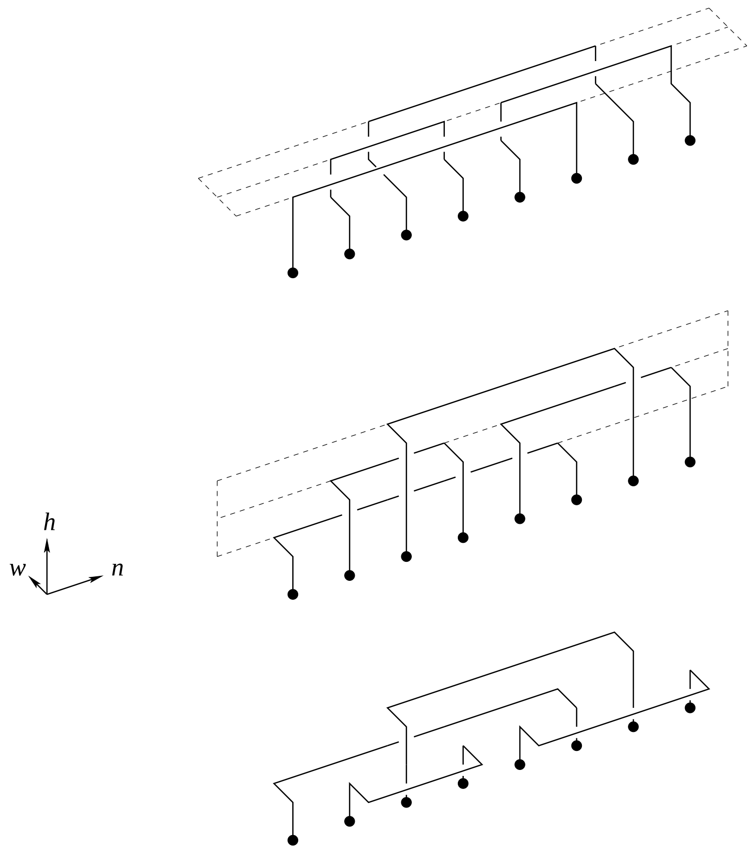

Figure 19

Similarly, if $w = 2$ then we have a channel routing problem with density d and using the same linear time algorithm we can always realize a routing with $w' = d + 1$ and $h = 3$ or with $w' = 3$ and $h = d + 1$, see Figure 20.

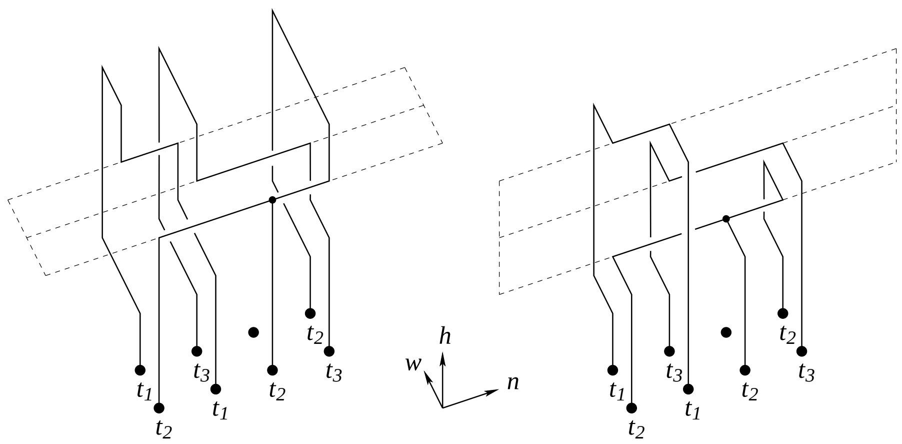

Figure 20

The right hand side of Figure 20 shows the essential idea of our algorithm: immediately at the level of the terminals (when leaving the 'single active layer') we increase the width from w to $w' = s_w w$ in order to make enough space 'between the rows of the terminals' for the vertical plane(-s) containing the interconnecting wires. The same process is illustrated for $w = 6$ in Figure 22 below.

Later we shall need the following observation: if a larger horizontal area is available, we can decrease the height to $h = 2\lceil d/(w' - 2)\rceil + 2$. For this we can arrange the tracks in $\lceil d/(w' - 2)\rceil$ parallel horizontal planes - however, we need an 'empty' plane between two consecutive planes of tracks to ensure that the endpoints of the intervals can always reach the terminals, even if they are 'in the wrong side', as terminals t_2 and t_3 in Figure 20. Therefore the basic quantity hw' can be upper bounded by essentially $3d/2$ for $w = 1$ but only by $2d$ for $w = 2$.

2. Let us turn now to the general problem with width w. Since the terminals occupy certain gridpoints of an $n \times w$ rectangle, we consider them as a collection of w parallel rows, each of length n. We wish to solve $\binom{w}{2}$ channel routing problems one after the other. Figure 21 illustrates this for $w = 6$. At first (going from the bottom of the figure to the top) we solve those $w - 1$ channel routings where the rows are adjacent (first 'floor'), then those $w - 2$ ones where the distance of two rows is two (second 'floor') etc.

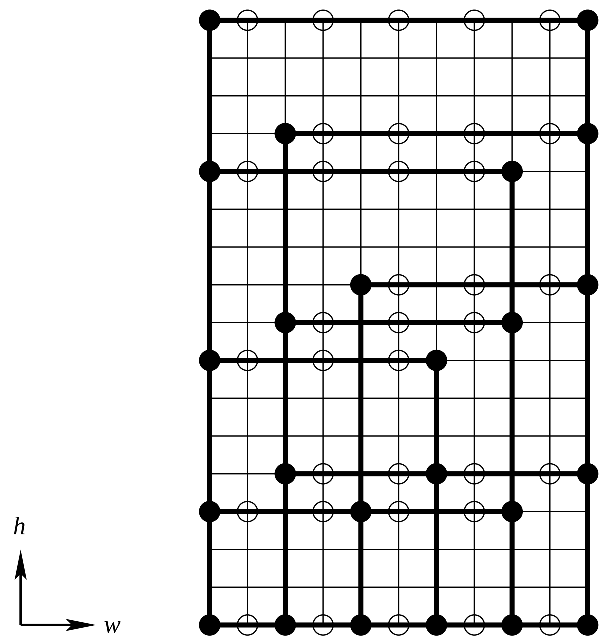

Figure 21

The $w - 1$ channel routings at the first floor (actually the $w - 1$ vertical planes containing the wire segments of these channel routings) do not interfere with one another (these are illustrated by the $w - 1$ empty dots in the bottom horizontal line of the figure).

On the other hand, the $w - 2$ channel routings at the second floor do not have this property hence this floor will have two levels, one for the row pairs 1 and 3, 3 and 5, etc and one for the row pairs 2 and 4, 4 and 6 etc. In general, floor f contains l_f levels where $l_f = f$ if $1 \leq f \leq \lfloor w/2 \rfloor$ and $l_f = w - f$ if $\lfloor w/2 \rfloor < f \leq w - 1$. Solid dots illustrate the rows, the vertical lines in the figure show that the terminals within a row may appear at different floors. Empty dots indicate the areas where these two rows can be interconnected. Hence such an empty dot may indicate a contribution of at most D to the final height (compare with the right hand side of Figure 20).

Observe that there are two empty dots between two solid ones in the second floor, three empty dots between two solid ones in the third floor etc. Hence the total height requirement is *not* $D \times \sum_{f=1}^{w-1} l_f = O(w^2 D)$ but only $2D \times \sum_{f=1}^{w-1} \frac{1}{f} l_f = O(wD)$. The extra constant 2 is due to the necessary empty planes between the consecutive planes of tracks, as explained in the last remark in Step **1** above, concerning the empty plane between the consecutive planes of tracks.

At the beginning of Section 5 we mentioned that the width of the input must be extended to $w' = s_w w$. Figure 21 might give the wrong impression that $s_w = 2$ suffices. However, as we shall see in the next section, the realization of the 'crossings' in the figure requires much more space, leading to $s_w = 8$.

5.2.2 The Real Routing

For future reference we are going to introduce the following terminology. By a *w-plane* we are going to mean a plane that is perpendicular to the width of the routing, that is, to the 'vector' w of Figures 19, 20 and 23. Analogously, *h-planes* and *n-planes* are planes perpendicular to the height and the length of the routing (or to the vectors h and n), respectively. Recall that a *w-wire segment* is a wire segment parallel to the width of the routing or to the vector w. *h-wire segments* and *n-wire segments* are defined in the same way. (Note that for example an h-plane is a horizontal plane, while an h-wire segment is a vertical wire segment.)

In Section 5.2.1 Figure 21 illustrated the order how the $\binom{w}{2}$ channel routing problems are routed one above the other. (Of course it is possible that the terminals in row i and those in row k do not share any net and therefore a whole level within a floor is missing.)

However, Figure 21 may alternatively be considered as a 'cross-section' of the routing by an n-plane of size $w' \times h$ (and then there are n copies of these cross-sections, one behind the other). In this sense an empty dot in a particular level indicates a whole w-plane of length n and of height at most $2\lceil D/f \rceil$ containing several wires one above the other for that channel routing. Hence a horizontal line in Figure 21 between a solid and an empty dot indicates a wire segment going towards this plane - but it may go for a wire segment running in this plane or it may wish to avoid it and go for a wire segment in one of the other parallel planes (that is, towards one of the further empty dots).

It is very important to realize, therefore, that there are two types of 'crossings' which have to be avoided if we wish to realize the final 3-dimensional routing along the n-planes one after the other:

- Type 1 – A vertical line, connecting two solid dots, and a horizontal line, connecting two empty dots, may cross each other in Figure 21.
- Type 2 – A horizontal line which passes through an empty dot in Figure 21 may, in fact, not use that particular n-wire (which is perpendicular to the actual n-plane).

The basic point is that crossings of Type 2 can be avoided by a detour within the actual n-plane (increasing the height by one and the width by two) but crossings of Type 1 can be avoided outside the n-plane only. Therefore we must use the adjacent n-plane as well. But what happens if in this latter n-plane there is another vertical line interconnecting two solid dots (or in the real routing: a h-wire segment coming from a terminal) blocking the detour?

We avoid this problem in the following way: since $w' > w$, we must, in any case, start the routing in each n-plane by 'expanding' the w terminals into larger distances. This increases the height by $w/2$ and can be performed like in either of the ways shown in Figure 22. Now, if we use the two kinds of expansions alternatively then the terminals within a single row will form a zigzag pattern and hence two h-wires that 'should' block each other in two consecutive n-planes, will actually be shifted by two units away.

This way we have ensured that a Type 1 detour will not be blocked by an h-wire segment coming from a terminal. However, one difficulty remains to be solved: a Type 1 detour can still cross a w-wire segment that goes in the same h-plane one unit behind (connecting two w-planes within a level).

In order to handle this, let us number the n-planes from 'front' to 'back' with the numbers $1, 2, \ldots, n$ and the h-planes form 'bottom' to 'top' with

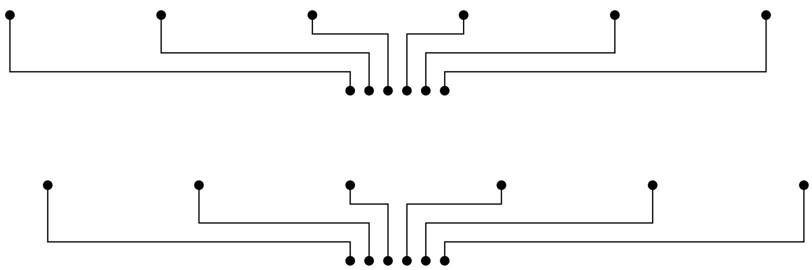

Figure 22

the numbers $1, 2, \ldots, h$. Now let us declare the following rule: if a w-wire segment goes in an n-plane numbered with an even number, then it must go in an h-plane also numbered with an even number; similarly, if a w-wire segment goes in an n-plane numbered with an odd number, then it must go in an h-plane also numbered with an odd number. Since the height of a level is always the double of the number of tracks in it, the above rule can obviously be fulfilled. (This way some of the Type 2 detours will become unnecessary: if a w-wire segment reaches a w-plane (containing some of the tracks within a level) between two consecutive tracks, then the w-wire segment can cross the w-plane without meeting the tracks, there is no need for a Type 2 detour.)

Figure 23 illustrates most of these situations in a single drawing. Recall that continuous lines are wires, dotted lines are for the indication of coplanarity only. Observe that $t_1, t_2, \ldots$ are terminals within a single row (illustrating the aforementioned zigzag pattern), while the terminals $t'_1, t'_2, \ldots$ form the next row. It might be instructive to recall that the detour between A and B is of Type 1 while that between C and D is of Type 2.

Figure 24 shows a part of Figure 23 again in order to explain why s_w needs to be as large as 8.

An algorithm realizing the above construction has to solve $\binom{w}{2}$ channel routing problems, as shown by Figure 21. For each one of these channel routing instances the intervals can be packed into the corresponding w-planes in $O(n)$ time (this is essentially Gallai's algorithm, see Theorem 1). However, for each terminal, the necessary Type 1 and Type 2 detours have to be included to reach the corresponding n-wire segment. This takes at most $O(w)$

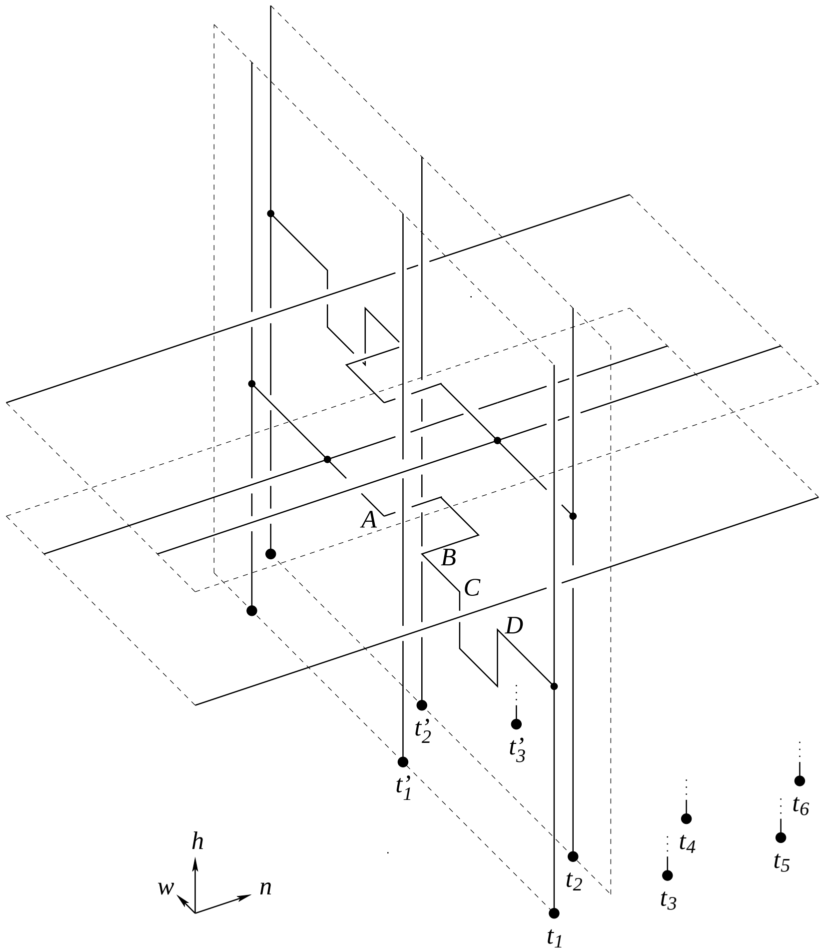

Figure 23

time for each terminal, resulting in a time consumption of $O(wn)$ for each one of the $\binom{w}{2}$ channel instances. Hence the total running time is $O(w^3n)$. $\square$

5.3 The $s_n, s_w \geq 2$ Case

We have seen in Lemma 22 that every SALRP problem can trivially be solved with height $h = \frac{wn}{2}$ if $s_w, s_n \geq 2$. This provides an upper bound of $h = O(n^2)$ in the $n = w$ case. However, in 2000 Aggarwal et al. [2] proved the following theorem using elaborate probabilistic methods.

Theorem 24 (A. Aggarwal, J. Kleinberg and D. P. Williamson, 2000) [2]*If*

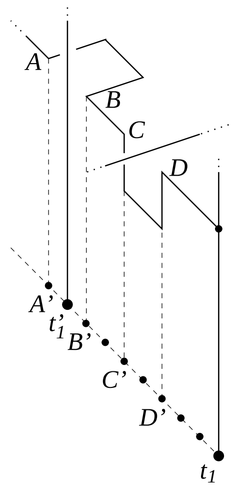

Figure 24

each net consists of two terminals only then the nets of an $n \times n$ SALRP can be partitioned into $O(n \log^2 n)$ classes such that each class of nets can be routed on a copy of the grid (of size $n \times n$).

An easy corollary of this theorem is that if $s_w = s_n = 2$ and each net consists of two terminals only than every SALRP can be solved with height $h = O(n \log^2 n)$; see Proposition 26 (with $h_0 = 1$) for a proof of this corollary from the above theorem.

The main aim of this section is to prove that $h = O(n)$ also suffices, even if the nets can have an arbitrary number of terminals.

5.3.1 2-terminal Nets

In this section we restrict ourselves to the special case in which every net contains two terminals only.

Theorem 25 (Recski A. and Sz. D.) [45] *If each net consists of two terminals only then the nets of a SALRP can be partitioned into $\lfloor \frac{3}{2} n \rfloor$ classes such that each class of nets can be routed as a separate SALRP with $s_n = \lceil \frac{w}{2n} \rceil$ and height $h = 2$.*

Proof: Assume that a single active layer routing problem with w rows and n columns is given. Let the set of rows be $R = \{r_1, r_2, \ldots, r_w\}$. We define a graph G with vertex set R where the edges correspond to the nets: if the two terminals of a net N are situated in the rows r_i and r_j then we add the edge $e_N = \{r_i, r_j\}$ to the edge set E of G. (Note that parallel edges and loops are possible in G.)

The definition of G implies that the degree of every vertex is at most n. Therefore, by Shannon's classic theorem [49], the edges can be coloured with $\lfloor \frac{3}{2}n \rfloor$ colours (such that no two adjacent edges share the same colour). Since the edges of G correspond to the nets, this colouring induces a partition on the set of nets. We claim that this partition fulfils the requirements of the theorem, that is, each partition class can be routed with $s_n = \lceil \frac{w}{2n} \rceil$ and height $h = 2$

Figure 25 shows an example for the part of such a routing (with $s_n = 1$). We have two layers of size $w \times (n \cdot s_n)$, the terminals are situated in, say, the bottom layer and the top layer is for the wires only. In Figure 25 wires of the bottom and top layer are represented by continuous and dashed lines, respectively; empty dots denote the location of the vias (that is, 1-unit h-wire segments between the two layers); terminal pairs marked with the same number form the nets; $r_1, r_2, \ldots$ and $c_1, c_2, \ldots$ denote the rows and columns, respectively. Figure 26 shows the corresponding part of the graph G associated with the routing problem of Figure 25; for every edge, the number of the corresponding net is shown in brackets.

Assume that the nets $\mathcal{N}' = \{N_{i_1}, N_{i_2}, \ldots, N_{i_k}\}$ form one of the partition classes. Call a net *trivial* if its two terminals are in the same row. It follows from the construction of G that every row contains either at most one terminal belonging to a nontrivial net in $\mathcal{N}'$, or two terminals that form a trivial net of $\mathcal{N}'$. We route each trivial net by a single n-wire segment on the bottom layer connecting the two terminals (see net 4 in Figure 25). We assign a column of the top layer to every nontrivial net arbitrarily (each column to at most one net). Since the number of nontrivial nets in $\mathcal{N}'$ is at most $\frac{w}{2}$ and the number of columns is $ns_n \geq n\frac{w}{2n} = \frac{w}{2}$, such an assignment exists. The routing of a nontrivial net N_{i_j} consists of two n-wire segments on

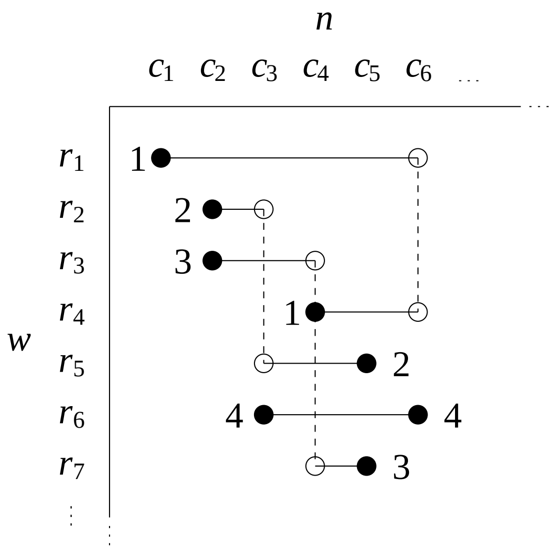

Figure 25

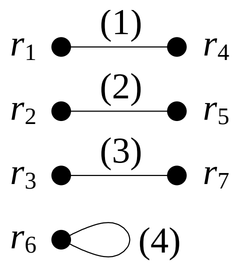

Figure 26

the bottom layer that connect the two terminals with the column assigned to N_{i_j}, a w-wire segment on the top layer running in the column assigned to N_{i_j} that connects the two rows of the two terminals, and at the endpoints of this w-segment two additional 1-unit h-wire segments (these are denoted by empty dots in Figure 25). It is easy to verify that the above described routing is good. □

Proposition 26 (Recski A. and Sz. D.) [45] *Assume that the nets of a SALRP can be partitioned into h_1 classes such that each class of nets can be routed as a separate SALRP with a given $s_w = s'$, $s_n = s''$ and height h_0.*

Then the original SALRP can be solved with $s_w = s' + 1$, $s_n = s'' + 1$ *and height* $h = h_1 \cdot h_0$.

Proof: By a *primary row* we are going to mean one of the 'original' w rows that contain the terminals. The remaining $(s_w - 1)w = s'w$ rows are going to be called *secondary rows*. We are also going to use the terms *primary column* and *secondary column* with an analogous meaning.

We introduce a 'long' h-wire segment starting from every terminal. We route each class of nets on h_0 consecutive layers (thus resulting in a total height of $h_1 \cdot h_0$). For a given class we temporarily introduce a virtual terminal for the terminals belonging to the nets of this class: a virtual terminal is always moved one unit to the right and one unit back from its original position and is placed on the lowest of the h_0 layers belonging to this class. We apply the routing guaranteed by the assumption of the theorem to connect the virtual terminals using the secondary rows and columns only. Finally, we connect every virtual terminal with the h-wire segment coming from its original version by a 1-unit w-wire segment and a 1-unit n-wire segment. (It is easy to verify that these two 1-unit segments cannot cross the routing of any other net. However, one of them can be unnecessary depending on the routing of the virtual terminals.)

Figure 27 illustrates the proof. The nets of the SALRP with $w = n = 3$ on the left hand side are first partitioned into $h_1 = 2$ classes and each class of nets is routed with height $h_0 = 1$ (and $s_w = s_n = 1$). Then the two routings are combined using the above construction to give a solution of the original problem with $s_w = s_n = 2$ and height $h = 2$. Primary and secondary rows are denoted by dashed and dotted lines, respectively; bigger dots denote the terminals, while smaller dots denote the virtual terminals; for each net N, virtual terminals corresponding to the terminals of N are denoted by N'. $\square$

Corollary 27 (Recski A. and Sz. D.) [45] *If each net consists of two terminals only then a SALRP can be solved with* $s_n = \lceil \frac{w}{2n} \rceil + 1$, $s_w = 2$ *and height* $h = 3n$.

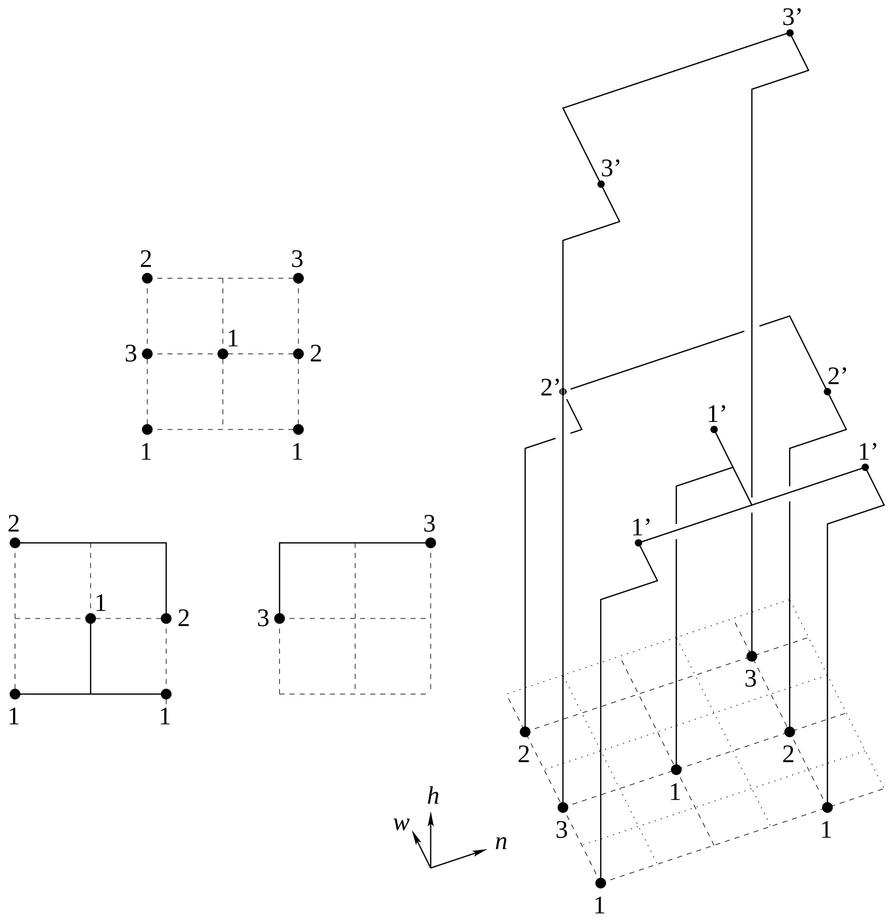

Figure 27

Proof: Apply Theorem 25 and Proposition 26. □

Corollary 28 (Recski A. and Sz. D.) [45] *If each net consists of two terminals only then a SALRP can be solved with* $s_w = s_n = 2$ *and height* $h = 3\max(n, w)$.

Proof: Since the role of w and n is symmetric, $n \geq w$ can be assumed without loss of generality. Now the statement follows form Corollary 27. □

If w is much larger than n then the statement of Corollary 27 offers two options: either the height is relatively small ($h = 3n$) and the spacing is large ($s_n = \lceil \frac{w}{2n} \rceil + 1$), or the height is large ($h = 3w$) and then the spacing is only 2. The following theorem claims that a trade-off between these two extremities can be found.

Theorem 29 (Recski A. and Sz. D.) [45] *If each net consists of two terminals only then a SALRP can be solved with $s_n \geq \lceil \frac{w}{4n} \rceil + 1$, $s_w = 2$ and height $h = \lfloor \frac{9}{2} n \rfloor$.*

Proof: Consider the partition of the nets defined in the proof of Theorem 25. We claim that the same partition can also be used to prove the statement of this theorem. Namely we show that each class can be routed as a separate 'nearly SALRP' with $s_n \geq \lceil \frac{w}{4n} \rceil$ and height $h = 3$. By a 'nearly SALRP' we mean that the terminals are situated on the middle one of the three layers, and not on the bottom one. However, this modification will not affect significantly the proof of Proposition 26 (the only change is that virtual terminals have to be placed on the middle layer rather than on the lowest one) and thus will prove the theorem.

The routing of the classes (each class on 3 layers) is again going to be similar to the one described in the proof of Theorem 25 (and in Figure 25). The only difference is that now we assign columns of the top and bottom layers to nontrivial nets arbitrarily. Since the number of nontrivial nets in each class is again at most $\frac{w}{2}$, and the number of (secondary) columns on the top and bottom layers altogether is $2n\lceil \frac{w}{4n} \rceil \geq \frac{w}{2}$, such an assignment exists. The routing of nontrivial nets is then modified in an obvious way: the two n-wire segments are placed on the middle layer, the single w-wire segment is placed either on the top or on the bottom layer (depending on where the column of the net is) and the 1-unit h-wire segments connect the w-segment with the two n-segments. □

We mention that in case of all the routings presented in Section 5.3.1 the path connecting the two terminals of any net uses at most 8 bends.

5.3.2 Multiterminal Nets

In this section we consider the general case in which the nets may have an arbitrary number of terminals. We modify the proof of Corollary 27 to get a similar result.

Theorem 30 (Recski A. and Sz. D.) [45] *Any SALRP can be solved with* $s_n \geq \lceil \frac{w}{2n} \rceil + 1$, $s_w = 2$ *and height* $h = 6n$.

Proof: Assume that a single active layer routing problem with w rows and n columns is given. We define the notion of *subnet* for further use. For each net N we fix an arbitrary order of its terminals: $N = \{t_1, t_2, \ldots, t_u\}$. Now the terminal pairs $\{t_i, t_{i+1}\}$ ($i = 1, 2, \ldots, u-1$) are going to be called *subnets* of N. (Thus each net has one less subnet than the number of its terminals.)

In what follows, we are going to route each subnet of every net such that the routings of two subnets can intersect only if they belong to the same net. Thus we obtain a routing of the whole problem. We proceed similarly as in the proof of Theorem 25.

Let again the set of rows be $R = \{r_1, r_2, \ldots, r_w\}$. We define a graph $G(R, E)$ on R with the edges corresponding to the subnets: if the two terminals of a subnet S are situated in the rows r_i and r_j then we add the edge $e_S = \{r_i, r_j\}$ to the edge set E of G. (Again, parallel edges and loops are possible in G.)

Now the degree of every vertex of G is at most $2n$ (since each terminal in a row belongs to at most two subnets). By Shannon's theorem [49], the edges can be coloured with $3n$ colours. The edges of G correspond to the subnets, so this colouring induces a partition on the set of subnets.

Again, we introduce a 'long' h-wire segment starting from every terminal. We assign two consecutive layers to each partition class and route the subnets belonging to this class on these two layers. (Thus the total height of the routing will be $6n$.) Given a partition class, we place virtual terminals for every terminal belonging to any of the subnets of this class on the lowest of the two layers assigned to the class; the position of the virtual terminals is the same as in Proposition 26: it is one unit to the right and one unit behind its original position. Now we route the virtual terminals using the secondary rows and columns of the two layers. The routing is identical to the one described in Theorem 25 (and Figure 25). To complete the routing, we connect every virtual terminal with the h-wire segment coming from its original version in the same way as in Proposition 26: through a 1-unit w-wire segment and a 1-unit n-wire segment .

It is easy to verify that the obtained routing is good. The routings of any two subnets cannot intersect in the secondary rows and columns of the layers. Since the virtual terminals are connected to the h-wire segments coming form the original terminal, subnets sharing a common terminal become interconnected and thus each net is completely routed. □

Corollary 31 (Recski A. and Sz. D.) [45] *Any SALRP can be solved with* $s_w = s_n = 2$ *and height* $h = 6\max(n, w)$.

Proof: Assume $n \geq w$ without loss of generality and apply Theorem 30. □

Similarly as it was the case with Theorem 29, the statement and the proof of Theorem 30 can again be modified to achieve a smaller spacing in return for an increase in the height.

Theorem 32 (Recski A. and Sz. D.) [45] *Any SALRP can be solved with* $s_n \geq \lceil \frac{w}{4n} \rceil + 1$, $s_w = 2$ *and height* $h = 9n$.

Proof: We proceed in a very similar way as in the proof of Theorem 29. We again route the subnets defined in the proof of Theorem 30 in such a way that two routings can intersect only if they belong to the same net. For the routing we use the same partition of the subnets as in the proof of Theorem 30. However, now we assign three consecutive layers to each partition class and we place virtual terminals in the middle one of these three layers and use the top and bottom layers to accomodate the w-wire segments. Hence $s_n \geq \lceil \frac{w}{4n} \rceil + 1$ suffices, since this way the number of secondary columns in the top and bottom layers is altogether $2n\lceil \frac{w}{4n} \rceil \geq \frac{w}{2}$. Since the number of partition classes is at most $3n$ and the routing of each class requires 3 layers, we get a total height of $9n$. □

5.3.3 Algorithmic Aspects

In order to actually generate the routings provided by the above proofs, the only real task is to edge-colour the graph assigned to the routing problems. We omit all the other details of the algorithms, we restrict ourselves to mentioning that these can be performed in linear time in the area $A = w \cdot n$ of the (planar) grid, that is, in the size of the input.

The folklore proof of Shannon's theorem using Kempe-chains (see eg. [39]) involves an algorithm to edge-colour a multigraph with v vertices, m edges and maximum degree Δ in $O(m \cdot (\Delta + v))$ time (using at most $\lfloor \frac{3}{2}\Delta \rfloor$ colours). This implies that, if t denotes the number of nets, each of the above constructions can be performed in $O(t \cdot (w + n))$ time. Since the size of the input is $A = w \cdot n$, the presented algorithms work in $O(A^{\frac{3}{2}})$ time, provided that $w = \Theta(n)$ holds.

However, it is an alternative to use the trivial greedy method for edge-colouring. In this case the number of colours increases to at most $2\Delta - 1$ and therefore the height of the routing also increases (in case of Theorem 30, for example, from $h = 6n$ to $h = 8n$), but remains to be linear in $\max(n, w)$. In return for the increase in the height, the time complexity of the routing algorithm also becomes linear in the size A of the input.

5.3.4 Comparison of the Bounds

All the results of Section 5.3 show that if $s_w, s_n \geq 2$ then the 3-dimensional single active layer routing problem can be solved with linear height in $\max(n, w)$. However, the exact height of the routings provided by the various constructions depend on different conditions: whether we restrict ourselves to 2-terminal nets or we allow multiterminal nets, the width w and length n of the routing problem and the spacing s_w and s_n.

An obvious measure to compare various solutions of the same routing problem is the volume of the routing, that is, the product $(w \cdot s_w) \cdot (n \cdot s_n) \cdot h$. Assume $w \geq n$ and denote the ratio $\frac{w}{n}$ by m. The following table shows the volume of the routings provided by the constructions of Section 5.3 (in each case, the solution with the smaller volume is shown in bold letters).

	2-terminal nets		multiterminal nets	
	Corollary 27	Theorem 29	Theorem 30	Theorem 32
$m \leq 2$	$\mathbf{12wn^2}$	$18wn^2$	$\mathbf{24wn^2}$	$36wn^2$
$2 < m \leq 4$	$\mathbf{18wn^2}$	$\mathbf{18wn^2}$	$\mathbf{36wn^2}$	$\mathbf{36wn^2}$
$4 < m \leq 6$	$\mathbf{24wn^2}$	$27wn^2$	$\mathbf{48wn^2}$	$54wn^2$
$6 < m \leq 8$	$30wn^2$	$\mathbf{27wn^2}$	$60wn^2$	$\mathbf{54wn^2}$
$8 < m \leq 10$	$\mathbf{36wn^2}$	$\mathbf{36wn^2}$	$\mathbf{72wn^2}$	$\mathbf{72wn^2}$
$10 < m \leq 12$	$42wn^2$	$\mathbf{36wn^2}$	$84wn^2$	$\mathbf{72wn^2}$
$12 < m \leq 14$	$48wn^2$	$\mathbf{45wn^2}$	$96wn^2$	$\mathbf{90wn^2}$
$14 < m \leq 16$	$54wn^2$	$\mathbf{45wn^2}$	$108wn^2$	$\mathbf{90wn^2}$

We mention, however, that the routings provided by Corollaries 27, 28 and 31 and Theorem 30 belong to the Manhattan model (see Definition 5).

On the other hand, this does not hold for the routings of Theorems 29 and 32, because these involve consecutive layers containing w-segments.

5.3.5 Final Remark

The results of this section have very recently been improved by Reiss A. and Szeszlér D. [46]. We considered "3-dimensional channel routing", that is, the terminals are placed on two parallel layers (rather than on a single one) and the routing is to be realized in between these two layers with a minimum height. Using the results of the present section we proved that every such problem can be solved with height at most $h = 15\max(n, w)$ if $s_w, s_n \geq 2$ holds.

Index

area, 6

C-shaped routing problem, 48
channel routing problem, 15, 16, 18, 20, 26, 28, 52, 54
  bipartite channel routing problem, 20
  dense channel routing problem, 20
circuit partitioning, 7
column, 11, 50
comb layer, 39, 42, 46, 47
congestion, **13**, 17, 35, 46
cost function, 6

density, **13**, 17, 26, 52, 54
detailed routing, 4, **9**, 50
dogleg free, **26**, 27, 33
  dogleg free routing, 26

edge-disjoint routing, 12, 33

floorplanning, 8
flux, 18, **19**, 26

Gallai's algorithm, **13**, 16, 24, 27, 28, 33, 47, 48, 53, 58
gamma routing problem, **32**, 33, 48
global routing, 9

h-plane, 56
h-wire segment, 51
height, 50, 69

interval graph, 13, 15

knock-knee, 34
  knock-knee routing, 34

layer, 11, 50
layer assignment problem, 34
layout, 10
length, 11, 50

Manhattan model, **12**, 13, 16, 20, 26, 28, 30, 32, 33, 38, 46–48, 53, 69

n-plane, 56
n-wire segment, 51
net, 6, 9, 50
  net fixed to the west/south, **41**, 42, 45
  trivial net, 18, 20, 61

odd set, 36

parity congestion, 36
placement, 8
primary row/column, 63

rectangular grid graph, 11
routing, 10
routing graph, 9
routing problem, 6
row, 50

saturated line, 36

secondary row/column, 63
Shannon's theorem, 61, 66, 68
shift-right-1 problem, 18, 26
single active layer routing problem (SALRP), 49, **50**, 52, 59, 60, 62–66, 68, 69
single row routing problem, **13**, 15, 27, 52
spacing, 50
Steiner-tree, 9, 10, 51
subnet, 66, 68
switchbox routing problem, **12**, **29**, 30, 32, 38, 46–48

terminal, 10, 11, 50
  virtual terminal, 63, 66
track, 11

unconstrained model, **12**, 15, 27, 30, 32

vertical constraint graph, 21
VLSI, 4
volume, 69

w-plane, 56
w-wire segment, 51
width, 11, 50
wire, 10, 50

References

[1] Aggarwal, A., M. Klawe, D. Lichtenstein, N. Linial and A. Wigderson, A lower bound on the area of permutation layouts, *Algorithmica* (1991) **6**, 241-255.

[2] Aggarwal, A., J. Kleinberg and D. P. Williamson, Node-disjoint paths on the mesh and a new trade-off in VLSI layout, *SIAM J. Computing* (2000) **29**, 1321-1333.

[3] Baker, S. B., S. N. Bhatt and F. T. Leighton, An approximation algorithm for Manhattan routing, *Advances in Computer Research, Volume 2: VLSI Theory* (F. P. Preparata, ed.), JAI Press, Reading, MA (1984), 205-229.

[4] Boros E., A. Recski, T. Szkaliczki and F. Wettl, Polynomial time Manhattan routing without doglegs — a generalization of Gallai's algorithm, *Computers and Artificial Intelligence* (1999) **18 (4)**, 403-413.

[5] Boros E., A. Recski and F. Wettl, Unconstrained multilayer switchbox routing, *Annals of Operations Research* (1995) **58**, 481-491.

[6] Brady, M. L. and D. J Brown, VLSI routing: Four layers suffice, *Advances in Computer Research, Volume 2: VLSI Theory* (F. P. Preparata, ed.), JAI Press, Reading, MA (1984), 245-257.

[7] Brown, D. J. and R. R. Rivest, New lower bounds for channel width, *VLSI Systems and Computations* (H. T. Kung, B. Sproull and G. Steele, eds.) Computer Science Press, Rockville, MD (1981), 178-185.

[8] Bui, T. N., S. Chauduri, F. T. Leighton and M. Sipser, Graph bisection algorithms with good average case behaviour, *Combinatorica* (1987) **7(2)**, 171-191.

[9] Cohoon, J.P. and P. L. Heck, BEAVER: a computational-geometry-based tool for switchbox routing, *IEEE Trans. Computer-Aided Design of Integrated Circ. Syst.* (1988), **7(6)**, 684-697.

[10] Cutler, M. and Y. Shiloach, Permutation layout, *Networks* (1978) **8**, 253-278.

[11] Deutsch, D, A dogleg channel router, *Proc. 13rd Design Automation Conf.* (1976), 425-433.

[12] Enbody, R.J., G. Lynn and K. H. Tan, Routing the 3-D chip, *Proc. 28th ACM/IEEE Design Automation Conf.* (1991), 132-137.

[13] Frank A., Disjoint paths in a rectilinear grid, *Combinatorica* (1982) **2**, 361-371.

[14] Gallai T., His unpublished results were announced in A. Hajnal and J. Surányi Über die Auflösung von Graphen in vollständige Teilgraphen, *Annales Univ. Sci. Budapest. Eötvös Sect. Math.* (1958) **1**, 115-123.

[15] Games, R. A., Optimal book embeddings of the FFT, Benes, and barrel shifter networks, *Algorithmica* (1986) **1**, 233-250.

[16] Gao, S. and M. Kaufmann, Channel routing of multiterminal nets, *Proc. 28th FOCS Symp.* (1987), 316-325.

[17] Gao, S. and M. Kaufmann, Channel routing of multiterminal nets, *J. Assoc. Comput. Mach.* (1994) **41 (4)**, 791-818.

[18] Garey, M.R. and D. S. Johnson, *Computers and Intractability: a Guide to the Theory of NP-completeness*, W. H. Freeman and Co., San Francisco, CA, 1979.

[19] Golda, B., B. Laczay, Z. Á. Mann, Cs. Megyeri and A. Recski, Implementation of VLSI Routing Algorithms, *Proc. IEEE 6th Internat. Conf. on Intelligent Engineering Systems, Croatia, 2002*, 383-388.

[20] Greenberg, R., J. JáJá and S. Krishnamurthy, On the difficulty of Manhattan channel routing, *Inform. Process. Lett.*, (1992) **44 (5)**, 281-284.

[21] Hamachi, G. T. and J. K. Ousterhout, A switchbox router with obstacle avoidance, *DAC '84: Proc. 21st Conf. on Design Automation* (1984), IEEE Press, 173-179.

[22] Hambrusch, S. E., Channel routing in overlap models, *IEEE Trans. Computer-Aided Design of Integrated Circ. Syst.* (1985) **CAD-4**, 23-30.

[23] Hashimoto, A. and J. Stevens, Wire routing by optimizing channel assignment, *Proc. 8th Design Automation Conf.* (1971), 214-224.

[24] Johnson, D. S., The NP-completeness column: An ongoing guide, *J. Algorithms* (1984) **5**, 147-160.

[25] Joobbani, R. and D. P. Siewiorek, WEAVER: a knowledge-based routing expert, *DAC '85: Proc. 22nd ACM/IEEE Conf. on Design Automation* (1985), ACM Press, 266-272.

[26] LaPaugh, A. S., A polynomial time algorithm for optimal routing around a rectangle, *Proc. 21st FOCS Symp.* (1980), 282-293.

[27] Leighton, T., *Complexity issues in VLSI: Optimal layouts for the shuffle-exchange graph and other networks*, The MIT Press, Cambridge, MA., 1983.

[28] Leighton, T. and A. L. Rosenberg, Three-dimensional circuit layouts, *SIAM J. Computing* (1986) **15**, 793-813.

[29] Leighton, T., S. Rao and A. Srinivasan, New algorithmic aspects of the Local Lemma with applications to routing and partitioning, *Proc. 10th Annual ACM-SIAM Symp. on Discrete Algorithms* ACM/SIAM, New York and Philadelphia (1999), 643-652.

[30] Lengauer, T., *Combinatorial algorithms for integrated circuit layout*, John Wiley & Sons Ltd., Chichester, 1990.

[31] Lienig, J. and K. Thulasiraman, A Genetic Algorithm for Channel Routing in VLSI Circuits, *Evolutionary Computation* (1993) **1 (4)**, 293-311.

[32] Lipski, W., Jr., On the structure of three-layer wireable layouts, *Advances in Computer Research, Volume 2: VLSI Theory* (F. P. Preparata, ed.), JAI Press, Reading, MA (1984), 231-243.

[33] Luk, W. K., A greedy switchbox router, *Integration, the VLSI Journal* (1985) **3**, 129-149.

[34] Kramer, M. R. and J. van Leeuwen, The complexity of wire routing and finding minimum area layouts for arbitrary VLSI circuits, *Advances in Computing Research, Volume 2: VLSI Theory* (F. P. Preparata, ed.), JAI Press, Reading, MA (1984), 129-146.

[35] Marek-Sadowska, M. and E. Kuh, General channel-routing algorithm, *Proc. IEE (GB)* (1983) **130**, 83-88.

[36] Mehlhorn, K. and F. P. Preparata, Routing through a rectangle, *J. Assoc. Comput. Mach.* **33** (1986), 60-85.

[37] Middendorf, M., Manhattan channel routing is NP-complete under truly restricted settings, *Chicago J. Theoret. Comput. Sci.*, Article 6, (1996), approx. 19 pp. (electronic).

[38] Möhring, R. H., D. Wagner and F. Wagner, VLSI network design, in: *Handbook in OR & MS: Network Routing*, North-Holland, Amsterdam (1995), 625-712.

[39] Nakano, S., T. Nishizeki and X. Zhou, Edge-coloring algorithms, *Computer science today*, Lecture Notes in Computer Science **1000**, 172-183, Springer, Berlin (1995).

[40] Recski A., Channel routing in the dogleg-free multilayer Manhattan model, *Proc. ECCTD'97* (1997), 39-43.

[41] Recski A., G. Salamon and D. Szeszlér, Improving Size Bounds for Subcases of Square-Shaped Switchbox Routing, *Annales Univ. Sci. Budapest. Eötvös Sect. Math.*, submitted.

[42] Recski A., G. Salamon and D. Szeszlér, Improving Size Bounds for Subcases of Square-Shaped Switchbox Routing, *Periodica Polytechnica Ser. El. Eng.* (2004) **48**, 55-60.

[43] Recski A. and F. Strzyzewski, Vertex-disjoint channel routing on two layers, *Integer programming and combinatorial optimization* (Ravi Kannan and W.R. Pulleyblank, eds.), University of Waterloo Press (1990), 397-405.

[44] Recski, A. and D. Szeszlér, 3-dimensional single active layer routing, *Discrete and Computational Geometry*, Lecture Notes in Computer Science **2098**, 318-329, Springer, Berlin (2001).

[45] Recski A. and D. Szeszlér, Routing Vertex Disjoint Steiner-Trees in a Cubic Grid – an Application in VLSI, *submitted*.

[46] Reiss A. and D. Szeszlér, *Proc. 4th Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications* (2005), 409-415.

[47] Rivest, R. L. and Fiduccia, C. M., A greedy channel router, *Proc. 19th IEEE Design Automation Conf.* (1982), 418-424.

[48] Rosenberg, A. L., Three-dimensional VLSI: a case study, *J. Assoc. Comput. Mach.* (1983) **30**, 397-416.

[49] Shannon, C., A theorem on coloring the lines of a network, *J. Math. Phys.* (1949) **28**, 148-151.

[50] Shirakawa, I., Some comments on permutation layout, *Networks* (1980) **10**, 179-182.

[51] Szeszlér D., Switchbox routing in the multilayer Manhattan model, *Annales Univ. Sci. Budapest. Eötvös Sect. Math.* (1997) **40**, 155-164.

[52] Szeszlér D., A New Algorithm for 2-layer, Manhattan Channel Routing, *Proc. 3rd Hungarian-Japanese Symposium on Discrete Mathematics and Its Applications* (2003), 179-185.

[53] SZKALICZKI T., Optimal routing on narrow channels, *Periodica Polytechnica Ser. El. Eng.* (1994) **38**, 191-196.

[54] SZKALICZKI T., Polynomial Algorithms in VLSI Routing, PhD. dissertation, BME (1996).

[55] SZYMANSKI, T. G., Dogleg channel routing is NP-complete, *IEEE Trans. Computer-Aided Design of Integrated Circ. Syst.* (1985) **CAD-4**, 31-41.

[56] WIENERS-LUMMER, C., Manhattan routing with good theoretical and practical performance, *Proc. 1st Annual ACM-SIAM Symp. on Discrete Algorithms* (1990), 465-475.

[57] WU, S. A. AND J. JÁJÁ, Optimal algorithms for adjacent side routing, *Algorithmica* (1991) **6 (4)**, 565-578.