An Attractor-Repeller Approach to Floorplanning*

Miguel F. Anjos †   Anthony Vannelli ‡

April 5, 2001

Abstract

The floorplanning (or facility layout) problem consists in finding the optimal positions for a given set of modules of fixed area (but perhaps varying dimensions) within a facility such that the distances between pairs of modules that have a positive connection cost are minimized. This is a hard discrete optimization problem; even the restricted version where the shapes of the modules are fixed and the optimization is taken over a fixed finite set of possible module locations is NP-hard. In this paper, we extend the concept of target distance introduced by Etawil and Vannelli and apply it to derive the AR (Attractor-Repeller) model which is designed to improve upon the NLT method of van Camp et al. This new model is designed to find a good initial point for the Stage-3 NLT solver and has the advantage that it can be solved very efficiently using a suitable optimization algorithm. Because the AR model is not a convex optimization problem, we also derive a convex version of the model and explore the generalized target distance that arises in this derivation. Numerical results demonstrating the potential of our approach are presented.

Key words: Facilities planning and design, Convex Programming, Global Optimization, VLSI Macro-Cell Layout.

*Released as Technical Report UW-E&CE#2001-05.

†Combinatorics & Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Research partially supported by a Ph.D. Research Scholarship from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (Fonds FCAR) of Québec. Email `anjos@stanfordalumni.org`

‡Electrical & Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Research partially supported by NSERC Operating Grant 15296. Email `vannelli@cheetah.vlsi.uwaterloo.ca`

1 Introduction

The floorplanning (or facility layout) problem consists in finding the optimal positions for a given set of modules of fixed area (but perhaps varying dimensions) within a facility. The objective is to minimize the distances between pairs of modules that have a nonzero connection "cost" while ensuring that no modules overlap. If the modules have varying dimensions, then finding their optimal shapes is also a part of the problem.

This is a hard problem; even the restricted version where the shapes of the modules are fixed and the optimization is taken over a fixed finite set of possible module locations is NP-hard. (This restriction is known as the Quadratic Assignment Problem, see [7] for example.) For this reason, most of the approaches in the literature are based on heuristics. For a recent survey on the facility layout problem, see [6].

The floorplanning problem has been throughly studied in the Operations Research literature. In order to put our contribution in context, we begin by reviewing two previous methods that relate closely to our new approach.

1.1 Background

1.1.1 The DISCON Method

In 1980 Drezner introduced the DISCON (DISpersion-CONcentration) method [4]. To describe the DISCON method, let us suppose that the modules are labelled $1, \ldots, N$, where N is the total number of modules, and that:

1. Each module is a circle (or can be approximated by a circle) of given radius $r_i, i = 1, \ldots, N$.
2. The distance between two modules is measured as the Euclidean distance between the centres of the circles.
3. The non-negative costs c_{ij} per unit distance between modules i and j are given.

The DISCON method uses a formulation equivalent to

$$\text{(DISCON)} \quad \begin{array}{ll} \min\limits_{(x_i,y_i)} & \sum\limits_{1 \le i < j \le N} c_{ij}\, d_{ij} \\ \text{s.t.} & \\ & d_{ij} \ge r_i + r_j, \ \forall\ 1 \le i < j \le N, \end{array}$$

where

- (x_i, y_i) denotes the centre of the i^{th} module, and
- $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.

Drezner solves the DISCON problem using a penalty algorithm and a two-phase approach:

- In Phase-1 (the dispersion phase), all the circles are dispersed far from the origin. This phase provides a starting point for the next phase.
- In Phase-2 (the concentration phase), the circles are concentrated (i.e. they are brought as close together as possible) and the resulting arrangement is the final solution.

We point out that both the objective function and the feasible set of the DISCON problem are non-convex. Furthermore, DISCON gives the user no control over the dimensions of the resulting layout of the circles (modules).

1.1.2 The NLT Method

More recently, van Camp, Carter and Vannelli [8] introduced a technique where all the modules as well as the facility are restricted to having fixed (given) areas and rectangular shapes, but the dimensions of the rectangles are optimized by the mathematical model.

Specifically, they introduce the following model, which we denote by vCCV:

$$
\text{(vCCV)} \quad
\begin{array}{ll}
\min\limits_{(x_i, y_i)} & \sum\limits_{1 \le i < j \le N} c_{ij}\, d_{ij} \\
\text{s.t.} & \\
& |x_i - x_j| - \frac{1}{2}(w_i + w_j) \ge 0 \quad \text{if} \quad |y_i - y_j| - \frac{1}{2}(h_i + h_j) < 0 \\
& |y_i - y_j| - \frac{1}{2}(h_i + h_j) \ge 0 \quad \text{if} \quad |x_i - x_j| - \frac{1}{2}(w_i + w_j) < 0 \\
& \frac{1}{2}w_T - (x_i + \frac{1}{2}w_i) \ge 0 \quad \text{for} \quad i = 1, \ldots, N \\
& \frac{1}{2}h_T - (y_i + \frac{1}{2}h_i) \ge 0 \quad \text{for} \quad i = 1, \ldots, N \\
& (x_i - \frac{1}{2}w_i) + \frac{1}{2}w_T \ge 0 \quad \text{for} \quad i = 1, \ldots, N \\
& (y_i - \frac{1}{2}h_i) + \frac{1}{2}h_T \ge 0 \quad \text{for} \quad i = 1, \ldots, N \\
& \min(w_i, h_i) - l_i^{\min} \ge 0 \quad \text{for} \quad i = 1, \ldots, N \\
& l_i^{\max} - \min(w_i, h_i) \ge 0 \quad \text{for} \quad i = 1, \ldots, N \\
& \min(w_T, h_T) - l_T^{\min} \ge 0 \\
& l_T^{\max} - \min(w_T, h_T) \ge 0,
\end{array}
$$

where (x_i, y_i) and d_{ij} are as previously defined, and

- w_i, h_i are the width and height of module i;
- $l_i^{\min}, l_i^{\max}$ are the minimum and maximum allowable lengths for the shortest side of module i;
- w_T, h_T are the width and height of the facility; and
- $l_T^{\min}, l_T^{\max}$ are the minimum and maximum allowable lengths for the shortest side of the facility.

Therefore, with the vCCV model, the user can input ranges for the shortest sides of the modules and of the resulting facility, and the model optimizes all the dimensions within the given ranges.

The NLT method is based on the above model and employs a three-stage approach:

1. Stage-1 aims to evenly distribute the centres of the modules inside the facility;
2. Stage-2 aims to reduce the overlap between modules;
3. Stage-3 determines the final solution.

Stage-3 consists of solving the complete vCCV model, whereas the problems solved at Stages 1 and 2 correspond to relaxations of the vCCV model. These relaxations approximate each module by a circle whose radius is proportional to the area of the module. The Stage-2 relaxation is presented in Section 2. We also observe that the models for all three stages are solved using a penalty-based method.

1.2 Contribution of this Paper

Our contribution is the AR (Attractor-Repeller) model which is designed to improve upon the 3-stage approach of van Camp et al. [8]. In Section 2, we motivate and derive the AR model which replaces Stages 1 and 2 by a single mathematical model that finds a "good" initial point for the solver of the Stage-3 model, that is, the vCCV model. The design of the AR model is inspired by the work of Etawil and Vannelli [5] for the VLSI placement problem. In Section 3, we isolate the source of the non-convexity in the AR model and derive a convex version. In Section 4, we observe that the AR model can be solved very efficiently using a suitable optimization algorithm. Numerical results demonstrating the potential of the AR model are presented in Section 5 and finally we discuss several possible directions for future research in Section 6.

2 Derivation of the AR Model

2.1 The Stage-2 Model of the NLT Method

In order to motivate the AR model, we begin by introducing the Stage-2 model used in the NLT method. (Recall that our objective is to improve on this Stage-2 model in terms of generating an initial point for the Stage-3 vCCV problem above.) This model is a relaxation of vCCV where each module is approximated by a circle whose radius is equal to half of the square root of the area of the module.

Given all the radii $r_i, i = 1, \ldots, N$, the Stage-2 model (St-2) is:

$$\text{(St-2)} \quad \begin{array}{ll} \displaystyle\min_{(x_i,y_i),h_T,w_T} & \displaystyle\sum_{1\le i<j\le N} c_{ij}\, d_{ij} \\ \text{subject to} & \\ & d_{ij} \ge r_i + r_j,\ \forall\ i,j \\ & \frac{1}{2}w_T \ge x_i + r_i \quad \text{for} \quad i = 1,\ldots,N \\ & \frac{1}{2}h_T \ge y_i + r_i \quad \text{for} \quad i = 1,\ldots,N \\ & \frac{1}{2}w_T \ge r_i - x_i \quad \text{for} \quad i = 1,\ldots,N \\ & \frac{1}{2}h_T \ge r_i - y_i \quad \text{for} \quad i = 1,\ldots,N \\ & l_T^{\max}\ \ge\ \min(w_T, h_T)\ \ge\ l_T^{\min}. \end{array}$$

The objective function is clear. As for the constraints, the first set of constraints in this model are those used in DISCON:

$$d_{ij} \ge r_i + r_j,\ \forall\ 1 \le i < j \le N. \tag{1}$$

These constraints require for each pair of circles that they do not overlap.

Recall that in DISCON, the user has no control of the overall dimensions of the final arrangement of circles. The model St-2 addresses this issue via the next four sets of constraints, which require that all circles remain inside the facility, and the last two constraints which ensure that the facility's shortest dimension is within the given bounds.

2.2 Target Distances Concept

A significant difficulty with the model St-2 is that the constraints (1) make the feasible set of St-2 non-convex. In our new model, we enforce these constraints drawing from the target distance paradigm introduced by Etawil and Vannelli [5].

Note that if the constraints (1) are removed from St-2, then the (global) optimal solution is simply to place all the circles with their centres at the same point (since we assume all the costs c_{ij} are non-negative). One way to understand why this happens is to interpret the objective function

$$\sum_{1 \leq i < j \leq N} c_{ij}\, d_{ij}$$

as an *attractor*, i.e. a function that seeks to make the values d_{ij} as small as possible by attracting all pairs of circles i, j to each other. To prevent this from happening, we will add to the objective function a *repeller* term that seeks to enforce the constraints (1).

For convenience we work with the *squares* of the distances between each pair of circles. It is clear that this does not change the optimal layout. Let us therefore define the variables

$$D_{ij} := {d_{ij}}^2 = (x_i - x_j)^2 + (y_i - y_j)^2$$

and the parameters

$$t_{ij} := \alpha\,(r_i + r_j)^2,\ \forall\ 1 \leq i < j \leq N,\ \text{ where } \alpha > 0.$$

The parameters t_{ij} are key to our strategy for enforcing the constraints (1), i.e. the separation of the circles representing the modules. The idea is that

the value $\sqrt{t_{ij}}$ is the *target distance* between the pair of circles i, j.

Equivalently, t_{ij} is the target value for D_{ij}, which is the square of the distance between the circles i and j with radii r_i and r_j respectively.

The parameter $\alpha > 0$ is introduced to provide some flexibility as to how tightly the user wishes to enforce the constraints (1). In theory, our algorithm will ensure that

$$\frac{D_{ij}}{t_{ij}} = 1$$

at optimality, so choosing $\alpha < 1$ sets a target value t_{ij} that allows some overlap of the areas of the respective circles, which means that a "relaxed" version of the constraints (1) on the d_{ij}'s is enforced. Similarly, $\alpha = 1$ means that there should be no overlap and the circles should intersect at exactly one point on their boundaries. This is illustrated in Figure 1. Note that the parameters t_{ij} may be interpreted as a generalization of the constant d in [5] since we need to set a different target for each pair of circles because

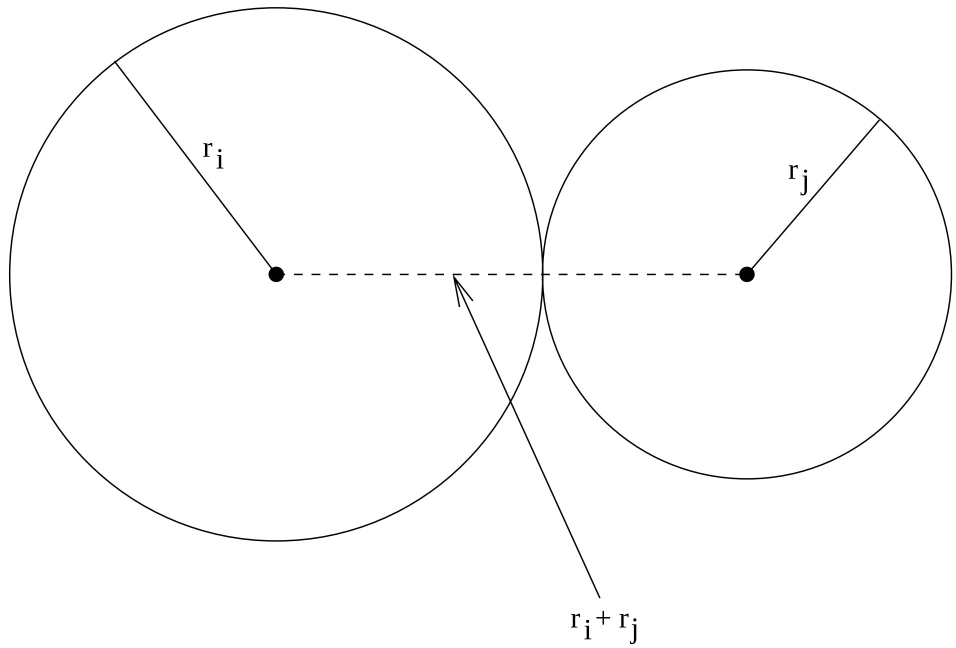


Figure 1: Motivation of the concept of target distances

of the variation among the radii.

In practice, we choose α empirically in such a way that we achieve a reasonable separation between all pairs of circles. The choice of α is explored further in Section 5. We now proceed to describing the mathematical model used to enforce the target distances between pairs of circles.

2.3 Enforcing the Target Distances

For each pair of circles i and j, if their overlap becomes too large with respect to the target distance t_{ij}, we penalize it by introducing the following term in the objective function of the AR model:

$$f\left(\frac{D_{ij}}{t_{ij}}\right)$$

where the function f is defined by

$$f(z) = \frac{1}{z} - 1, \quad z > 0.$$

The choice of $f(z)$ is inspired by the work of Etawil and Vannelli [5]. The function (of one variable) $f(z)$ is continuously differentiable and strictly decreasing on $(0, \infty)$. We exploit these properties of f for our purposes as follows. We start by placing the circles i and j in some initial configuration where $D_{ij} \gg t_{ij}$, and so $\frac{D_{ij}}{t_{ij}} > 1$. Then, as the ratio $\frac{D_{ij}}{t_{ij}}$ decreases with the effect of the "attractor" component of the objective function, the function f behaves as a monotonically increasing "repeller", and we let them interact until an equilibrium is attained. The "repeller" effect of f as the ratio $\frac{D_{ij}}{t_{ij}}$ decreases is illustrated in Figure 3, where we see that this effect is accentuated by scaling f by a (large) constant K.

By properly adjusting the value of the parameter α, we aim to attain this equilibrium at the point where $\frac{D_{ij}}{t_{ij}} \approx 1$, i.e. where the target distance is attained. In other words, the "attractor" component of our objective function makes the two circles move closer together and pulls them towards a layout where $D_{ij} = 0$, while the "repeller" component prevents the circles from overlapping.

2.4 Additional Design Features of the AR Model

Beyond the use of target distances, the AR model offers the user two other features for layout design.

Firstly, the AR model offers the possibility to distinguish between two kinds of modules: those that are mobile and those that are fixed (pads). This is motivated by our intended application of this approach to macro-cell layout problems. We present several such examples in Section 5. Another motivation is that it allows for a multiple-round strategy for attacking large problems. The idea is that in each round, the model is solved and some modules are chosen to be fixed before re-solving in the next round.

Hence, in the AR model, we let M and P denote the set of (mobile) modules and the set of (fixed) pads respectively. Note that target distances are introduced only for pairs of mobile modules. For all mixed pairs, only the attractor term is present in the objective function. That is, we permit overlap of a module and a pad, and interpret such overlap as an indication that the module in question should be placed very close to the pad.

Secondly, the ability to specify bounds on the desired dimensions of the rectangular facility, which is available in St-2, is also available in AR in the sense that it allows the user to specify bounds $w_T^{\min} \leq w_T^{\max}$ on the width of the facility, and $h_T^{\min} \leq h_T^{\max}$ on the height. In particular, if the user knows in advance that the facility should have width $\bar{w}$ and height $\bar{h}$, then these constraints can be enforced by setting

$$w_T^{\min} = w_T^{\max} = \bar{w}, \quad \text{and} \quad h_T^{\min} = h_T^{\max} = \bar{h}.$$

2.5 The AR Model

The above ideas yield the AR model:

$$\min_{\substack{(x_i,y_i),\, i \in M,\\ h_T, w_T}} \quad \sum_{i,j \in M \cup P} c_{ij} D_{ij} \quad + \quad \sum_{i,j \in M} f\left(\frac{D_{ij}}{t_{ij}}\right)$$

(AR) subject to

$$\begin{aligned}
\tfrac{1}{2} w_T \geq x_i + r_i \quad &\text{and} \quad \tfrac{1}{2} w_T \geq r_i - x_i, \text{ for all } i \in M,\\
\tfrac{1}{2} h_T \geq y_i + r_i \quad &\text{and} \quad \tfrac{1}{2} h_T \geq r_i - y_i, \text{ for all } i \in M,\\
w_T^{\max} \;\geq\; &w_T \;\geq\; w_T^{\min},\\
h_T^{\max} \;\geq\; &h_T \;\geq\; h_T^{\min}.
\end{aligned}$$

We point out that the AR model has only linear constraints (on the variables x_i, y_i, h_T, and w_T). From an optimization point of view, this is a significant advantage over St-2. In Section 4 we discuss the methodology we employed to test the efficiency on this model. However, we first derive in the next section a convex version of the AR model.

3 Convex Version of the AR Model

The AR model as stated above is not a convex problem because, although all the constraints are linear (and hence convex), the objective function is not convex. In this section we show that, under the mild assumption that $c_{ij} \neq 0$ for all pairs of modules $i, j \in M$, an appropriate modification of the objective function yields a convex problem. This convex problem can be thought of as a "convexification" of the AR model, and its derivation will naturally lead us to define certain parameters T_{ij}. These parameters can be viewed as a generalisation of the target distances t_{ij}, and also have an interesting interpretation from a practical point of view. This is further discussed in Section 3.1 below.

We begin by isolating the source of the non-convexity in the objective function of the AR model. We will use the well-known fact that

> If $f : \Re^m \mapsto \Re$ is twice continuously differentiable then f is convex over a convex set $C \subseteq \Re^m$ if its Hessian $\nabla^2 f$ is positive semidefinite for all points in C.

(This and other elementary results from convex analysis are presented in [2] for example. Recall that a symmetric $m \times m$ matrix is positive semidefinite if and only if all its eigenvalues are non-negative.)

Let us rewrite the objective function as

$$\sum_{i,j\in M} c_{ij}D_{ij} \quad + \quad \sum_{i\in M, j\in P} c_{ij}D_{ij} \quad + \quad \sum_{i,j\in P} c_{ij}D_{ij} \quad + \quad \sum_{i,j\in M} \left(\frac{t_{ij}}{D_{ij}} - 1\right),$$

where we have applied the function f in the argument of the last summation. We observe that

- The term $\sum_{i,j\in P} c_{ij}D_{ij}$ is constant, since the modules in P are fixed pads. Hence we can omit this term from the objective function without changing the optimal solution.

- The term $\sum_{i\in M, j\in P} c_{ij}D_{ij}$ is convex. Indeed, since x_j and y_j are fixed, its Hessian is a diagonal matrix whose diagonal entries are either 2 (for the rows corresponding to x_i and y_i, $i \in M$) or 0 (for all the other rows). Such a matrix is clearly positive semidefinite over the whole space and hence this term is convex.

- Bringing the remaining two summations together, we can rewrite the remainder of the objective function as

  $$\sum_{i,j\in M} c_{ij}D_{ij} + \frac{t_{ij}}{D_{ij}} - 1.$$

  This is where convexity can fail and where we now focus our attention.

We consider each term in this last summation independently, so without loss of generality, let us consider the term

$$c_{12}D_{12} + \frac{t_{12}}{D_{12}} - 1.$$

For clarity of the presentation, denote c_{12} by c, t_{12} by t and D_{12} by z. We have the following result:

Theorem 3.1 *Let $g : \Re^4 \mapsto \Re$ be given by*

$$g(x_1, x_2, y_1, y_2) = c\,z + \frac{t}{z} - 1,$$

where $c > 0$, $t > 0$ and $z > 0$, $z = (x_1 - x_2)^2 + (y_1 - y_2)^2$. The following statements hold for g:

1. *If $z \geq \sqrt{\frac{t}{c}}$, then the Hessian of g is positive semidefinite.*
2. *If $z = \sqrt{\frac{t}{c}}$, then the gradient of g is zero.*

Proof: At any given point (x_1, x_2, y_1, y_2), define the vector

$$w := \begin{bmatrix} (x_1 - x_2) \\ -(x_1 - x_2) \\ (y_1 - y_2) \\ -(y_1 - y_2) \end{bmatrix}$$

and the scalars $\rho := \left(c - \frac{t}{z^2}\right)$ and $\sigma := \frac{t}{z^3}$.

It is straightforward to check that the gradient and Hessian of g are respectively $\nabla g = 2\,\rho\, w$ and

$$\nabla^2 g = \rho \begin{bmatrix} 2 & -2 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & -2 & 2 \end{bmatrix} + 4\,\sigma\, w\, w^T,$$

where w^T denotes the transpose of w.

Now, the matrix $\begin{bmatrix} 2 & -2 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & -2 & 2 \end{bmatrix}$ has only two distinct eigenvalues, namely 0 and 4, so it is positive semidefinite. Furthermore,

$$z \geq \sqrt{\frac{t}{c}} \quad \Rightarrow \quad \rho \geq 0,$$

and so the first term of $\nabla^2 g$ is positive semidefinite. In the second term, the rank-one matrix $w\,w^T$ is positive semidefinite for any w, and the constant σ is always positive. Hence the second term is also positive semidefinite. Since the sum of positive semidefinite matrices is positive semidefinite, this proves the first claim.

The second claim is immediate:

$$z = \sqrt{\frac{t}{c}} \quad \Rightarrow \quad \rho = 0 \quad \Rightarrow \quad \nabla g = 0.$$

■

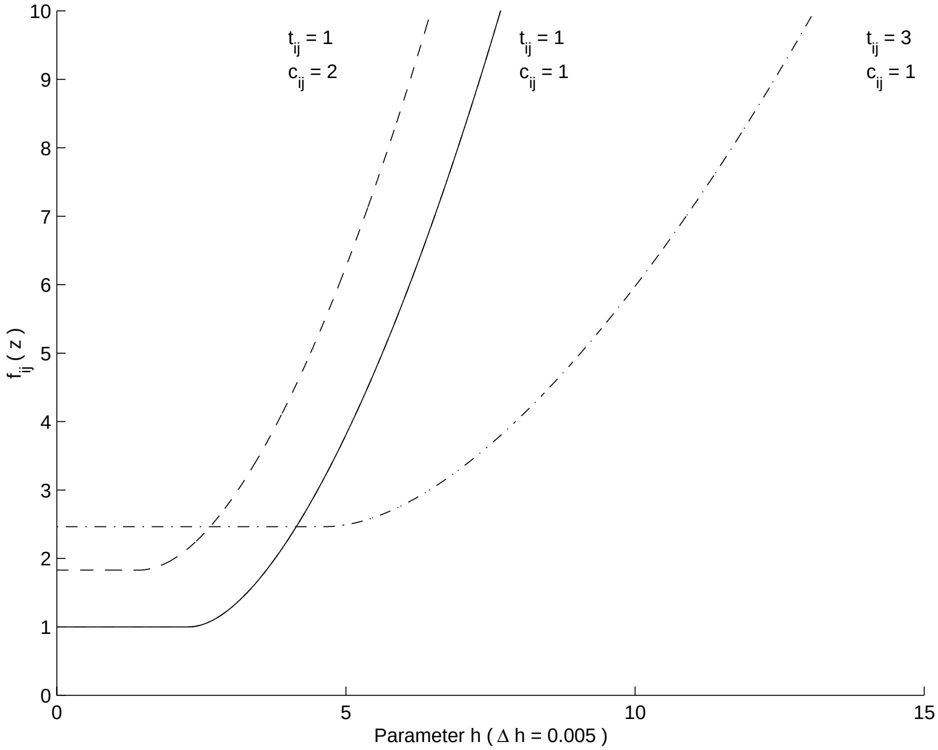


Figure 2: Graph of the convex function f_{ij} for several values of c_{ij} and t_{ij}

For $c_{ij} > 0$, $t_{ij} > 0$ and $z = (x_i - x_j)^2 + (y_i - y_j)^2$, define the piecewise function

$$f_{ij}(x_i, x_j, y_i, y_j) = \begin{cases} c_{ij}\, z + \frac{t_{ij}}{z} - 1, & z \geq \sqrt{\frac{t_{ij}}{c_{ij}}} \\ 2\sqrt{c_{ij}\, t_{ij}} - 1, & 0 \leq z < \sqrt{\frac{t_{ij}}{c_{ij}}}. \end{cases}$$

Figure 2 illustrates the function f_{ij} for various values of c_{ij} and t_{ij}. It was generated by taking random vectors x, y, Δx and Δy and using h to parametrize the change in z: for each step h, $h = 0, 0.005, 0.010, \ldots, 15$, $z = \|x - y + h(\Delta x - \Delta y)\|_2^2$.

By construction, f_{ij} is continuous on $\Re^4$ and since, by Theorem 3.1, $\nabla f_{ij}(\sqrt{\frac{t_{ij}}{c_{ij}}}) = 0$, f_{ij} is continuously differentiable on $\Re^4$. Also by Theorem 3.1 and since $\nabla^2 f_{ij} = 0$ for $z < \sqrt{\frac{t_{ij}}{c_{ij}}}$, we see that f_{ij} is convex over $\Re^4$, since its Hessian is everywhere positive semidefinite.

The "convexified" AR model, denoted CoAR, is then:

$$
\text{(CoAR)} \quad
\begin{array}{l}
\displaystyle \min_{\substack{(x_i,y_i),\, i\in M,\\ h_T,\, w_T}} \quad \sum_{i\in M, j\in P} c_{ij} D_{ij} \quad + \quad \sum_{i,j\in M} f_{ij}(x_i,x_j,y_i,y_j) \\[2ex]
\text{subject to} \\[1ex]
\qquad \begin{array}{lll}
\frac{1}{2}w_T \geq x_i + r_i & \text{and} & \frac{1}{2}w_T \geq r_i - x_i, \text{ for all } i \in M, \\
\frac{1}{2}h_T \geq y_i + r_i & \text{and} & \frac{1}{2}h_T \geq r_i - y_i, \text{ for all } i \in M, \\
 & w_T^{\max} \; \geq \; w_T \; \geq \; w_T^{\min}, & \\
 & h_T^{\max} \; \geq \; h_T \; \geq \; h_T^{\min}. &
\end{array}
\end{array}
$$

3.1 Generalized Target Distances

From the construction of f_{ij}, we observe that the minimum value of f_{ij} is attained for all the positions of modules i and j for which $D_{ij} \leq \sqrt{\frac{t_{ij}}{c_{ij}}}$. This includes the case $D_{ij} = 0$, i.e. both modules completely overlap. Since we do not want such a solution to our layout problem, the algorithm employed to solve CoAR must use a line search that knows the structure of f_{ij} and stops at a point where $D_{ij} \approx \sqrt{\frac{t_{ij}}{c_{ij}}}$, that is, close to the boundary of the flat portion of f_{ij}.

Note that if $D_{ij} \approx \sqrt{\frac{t_{ij}}{c_{ij}}}$, then the corresponding layout of the two modules has D_{ij} proportional to t_{ij}. Therefore such a solution to the model CoAR is still enforcing the target distances defined in Section 2.2.

Furthermore, if we define

$$T_{ij} := \sqrt{\frac{t_{ij}}{c_{ij}}}, \quad i, j \in M, i \neq j,$$

then we can think of the parameter T_{ij} as a generalized target distance for modules i and j which takes both t_{ij} and c_{ij} into account. Indeed, it is reasonable that the target distance be inversely proportional to c_{ij} since, from a practical point of view,

- if c_{ij} is small, then the two modules are likely to be places far apart in the layout, and correspondingly the generalized target distance should be large;

- if c_{ij} is large, then the opposite reasoning applies and the generalized target distance should be small.

4 Solution Methodology

We tested the validity and efficiency of our attractor-repeller paradigm by solving the AR model using the optimization package MINOS 5.3 of Gill et al. accessed via the modelling language GAMS (release 2.25) on a SUNSparc. The model has $2|M|+2$ variables and $4|M|+4$ inequality constraints, all of which are linear. Therefore only the objective function is non-linear and MINOS specifically exploits this structure by applying a reduced-gradient approach combined with a quasi-Newton algorithm. More details on these algorithms can be found in the GAMS User's Guide [3].

This is a significant advantage for the AR model because in general we can expect the algorithm to be *superlinearly convergent* [3]. In practice, this implies that we are able to solve the model AR quite efficiently even for large $|M|$. Some empirical evidence for the fact that the AR model can be solved more efficiently than the Stage-1/Stage-2 sequence of the NLT method is presented for Example 5.4 in Section 5.

To accentuate the effect of the repeller function $f(z) = \frac{1}{z} - 1$ for $z < 1$, we scale it by a large constant K. Indeed, even mild scaling of f can significantly increase the "barrier" effect that occurs as soon as z decreases below 1, as the graphs in Figure 3 show. Like Figure 2, Figure 3 was generated by taking random vectors x, y, Δx and Δy and using h to parametrize the change in D_{ij}. For each step h, $h = 0, 0.005, 0.010, \ldots, 20$, $z = \|x - y + h(\Delta x - \Delta y)\|_2^2$. Note that we set $t_{ij} = 1$.

Hence the problem we solve is

$$
\text{(AR)} \qquad
\begin{array}{ll}
\min\limits_{(x_i,y_i),h_T,w_T} & \sum\limits_{i,j\in M\cup P} c_{ij} D_{ij} \quad + \quad K \cdot \sum\limits_{i,j\in M} f\left(\frac{D_{ij}}{t_{ij}}\right) \\
\text{subject to} & \\
& \frac{1}{2} w_T \geq x_i + r_i \quad \text{and} \quad \frac{1}{2} w_T \geq r_i - x_i, \text{ for all } i \in M, \\
& \frac{1}{2} h_T \geq y_i + r_i \quad \text{and} \quad \frac{1}{2} h_T \geq r_i - y_i, \text{ for all } i \in M, \\
& w_T^{\max} \; \geq \; w_T \; \geq \; w_T^{\min}, \\
& h_T^{\max} \; \geq \; h_T \; \geq \; h_T^{\min}.
\end{array}
$$

For our numerical tests, we chose

$$
K = 10 \cdot \left(\sum_{i,j\in M} c_{ij} \right)
$$

so that K clearly dominates all the cost coefficients corresponding to pairs of mobile modules in the objective function of AR.

Since the algorithm is an iterative method, MINOS requires the user to supply an initial configuration. It is not clear a priori what the "best" starting configuration is, so we place the centres of the $|M|$ mobile

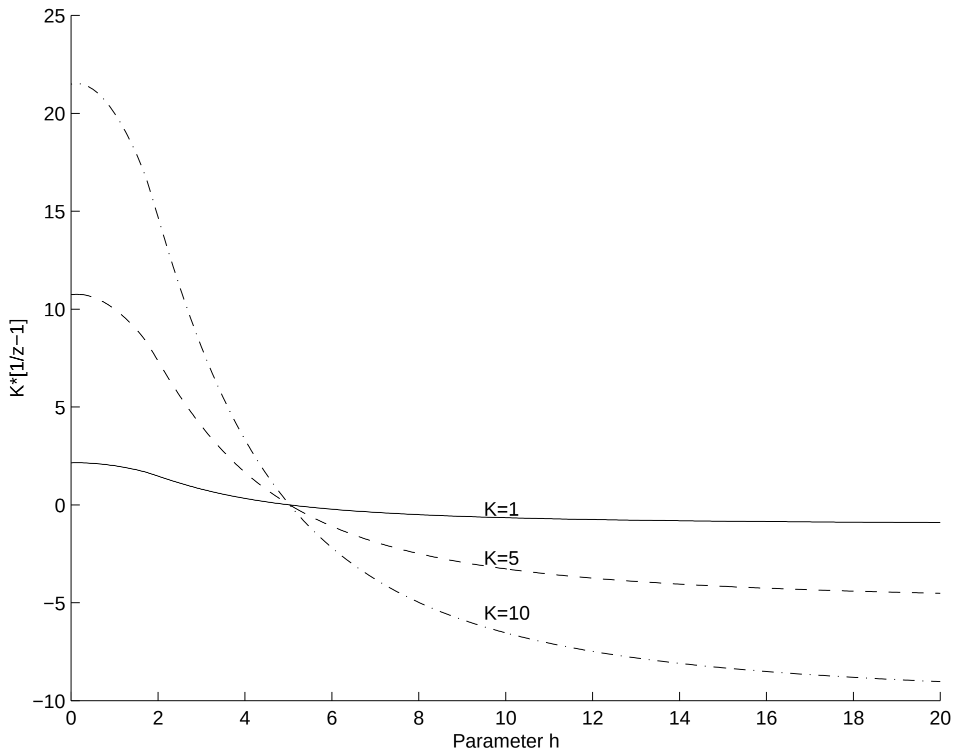


Figure 3: Effect of scaling on the function $f(z) = \frac{1}{z} - 1, z > 0$.

modules at regular intervals around a circle of radius $r = w_T^{\max} + h_T^{\max}$. Thus, letting $\theta_i = \frac{2\pi(i-1)}{|M|}$ and $r = w_T^{\max} + h_T^{\max}$, we initialize the centre (x_i, y_i) of the mobile module to

$$x_i = r\cos\theta_i, \quad y_i = r\sin\theta_i, \quad i = 1, \ldots, |M|.$$

We did not solve the CoAR model because this requires a carefully designed line search procedure, as discussed in Section 3. Furthermore, as the results in Section 5 show, the non-convex AR model with the scaling by K performs remarkably well in practice, in spite of its lack of convexity.

5 Numerical Results

We present the results of applying the AR approach to four floorplanning problems. These examples illustrate the characteristics of the AR model. In the first three examples some of the modules are fixed. Of course, our model applies equally to problems where the position of every module may vary. This is illustrated in the last example.

5.1 Macro-Cell Placement Example

We first apply our approach to a macro-cell layout example in which 8 modules out of 19 are fixed. The *optimal* layout for this problem is shown in Figure 4. The layout is optimal because any other layouts can be obtained by pairwise cell interchanges, and these clearly increase the total “wirelength”.

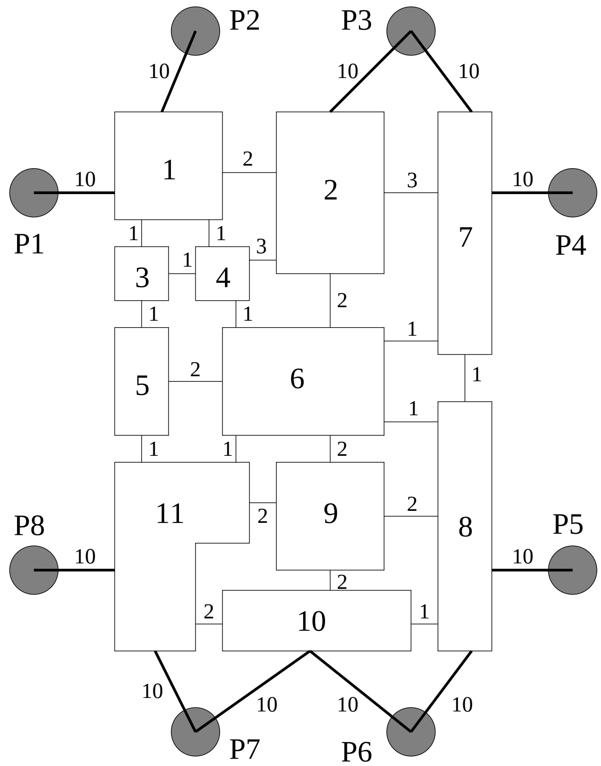


Figure 4: Optimal layout for the first macro-cell example

In Figure 4, the costs c_{ij} are indicated on the connections of the layout, and the modules P1,. . .,P8 are fixed. (They represent peripheral pads in the macro-cell circuit design.)

For our numerical tests, we used $w_T^{\min} = 8$, $w_T^{\max} = 10$, $h_T^{\min} = 10$, $h_T^{\max} = 14$ and the radii for the mobile modules M1, . . ., M11 are (respectively) 1,1.225,0.5,0.5,0.707,1.225,1,1,1,0.866,1.225, while each of the eight pads has radius equal to 0.5. The initial configuration used is shown in Figure 5.

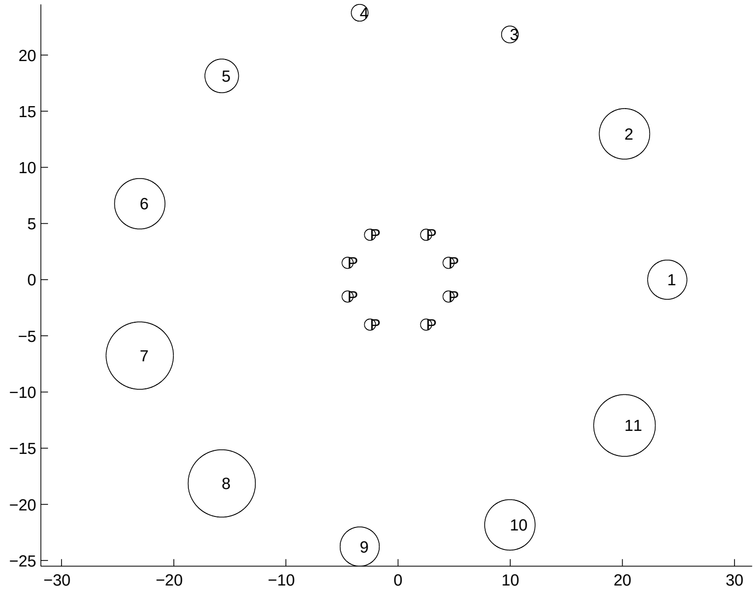
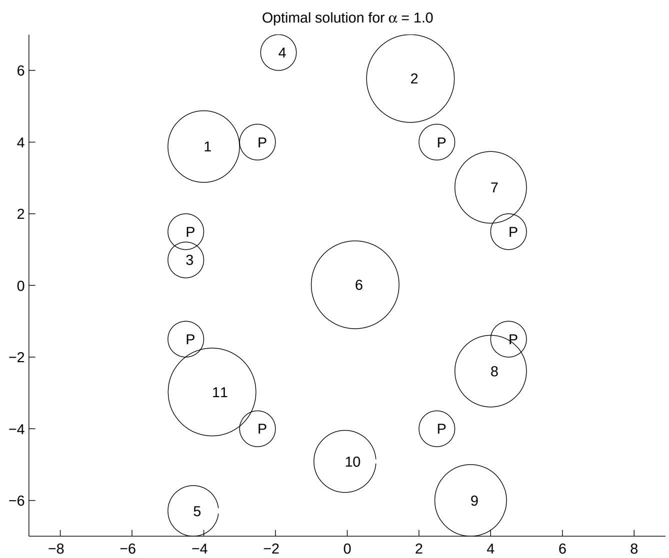
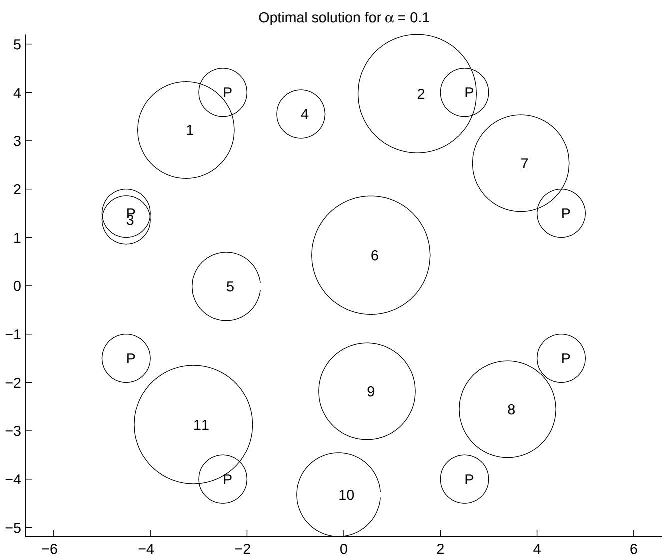
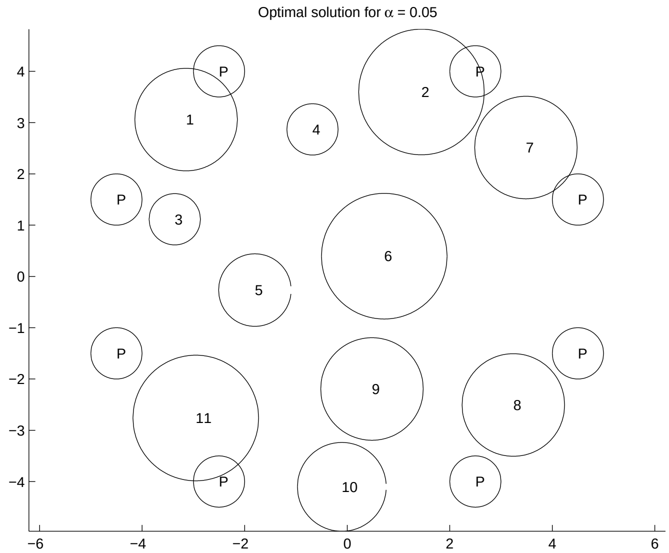


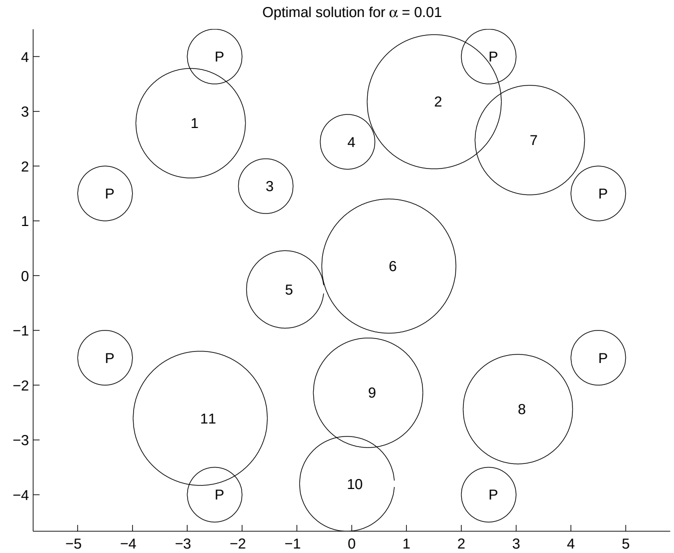
Figure 5: Starting configuration for the first macro-cell example

The results for varying values of α are as follows:

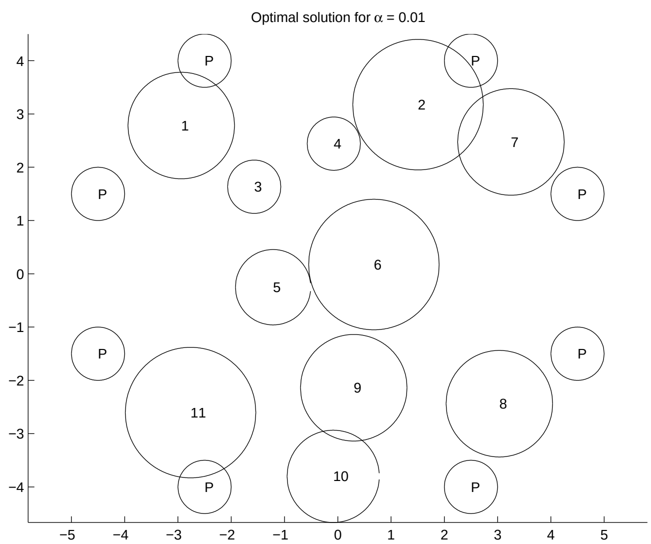


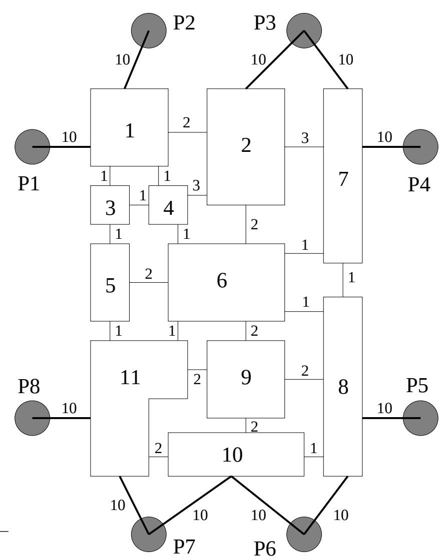






Surprisingly, we find that setting $\alpha = 1.0$ results in excessive separation of the modules, which are thrown against the boundaries of the facility. Therefore we need to reduce α in order to obtain a good placement. Furthermore, we find that the solution for $\alpha = 0.01$ yields a layout which is very similar to the (known) global optimal solution, as can be seen if we display them side by side:



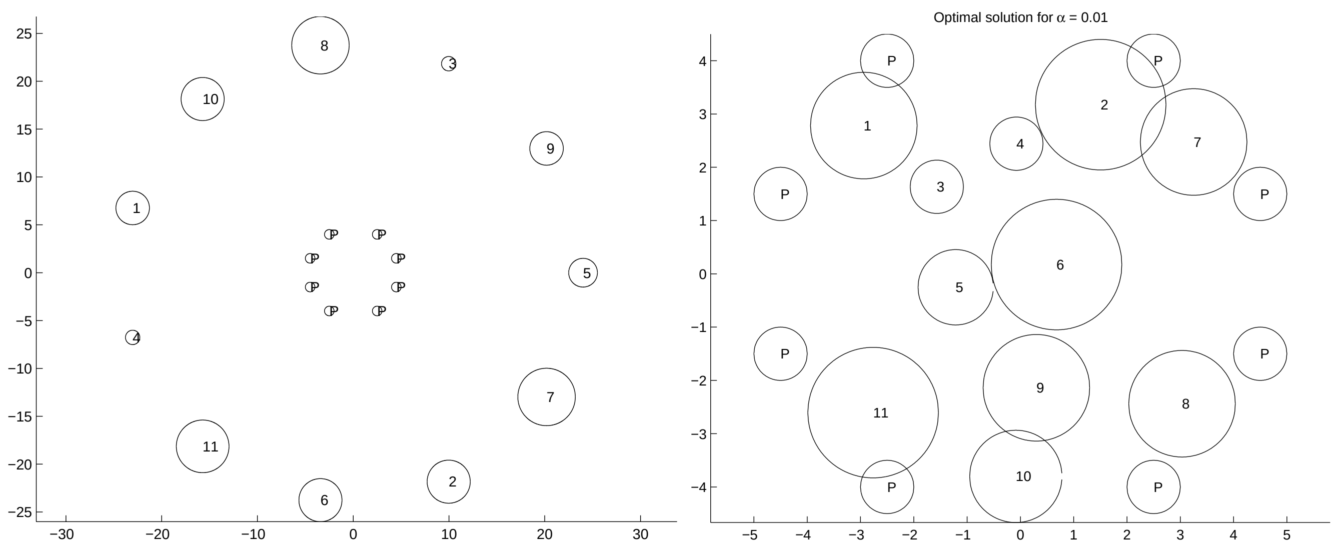


Clearly, the modules that have connections to pads are all located in the right areas, and all the modules are quite close to their (known) optimal positions.

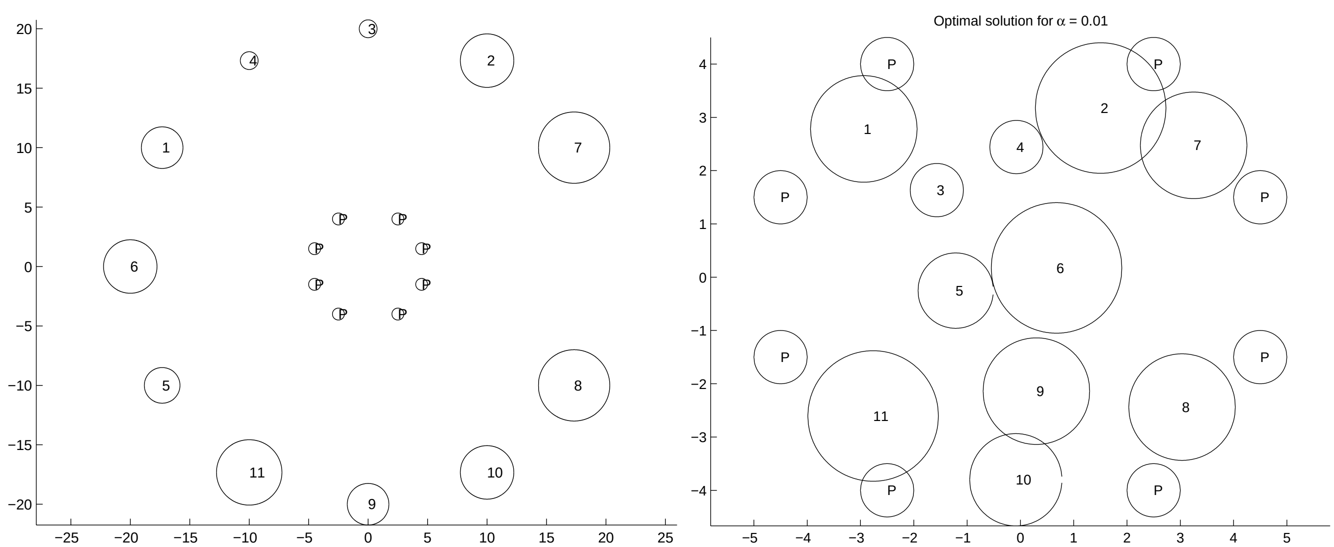
5.1.1 Alternative Initial Configurations

It is unclear at this point how much the success of the AR model depends upon our choice of initial configuration. Therefore we tested the robustness of the model for this example by trying two alternative initial configurations.

The first alternative configuration and corresponding solution are:



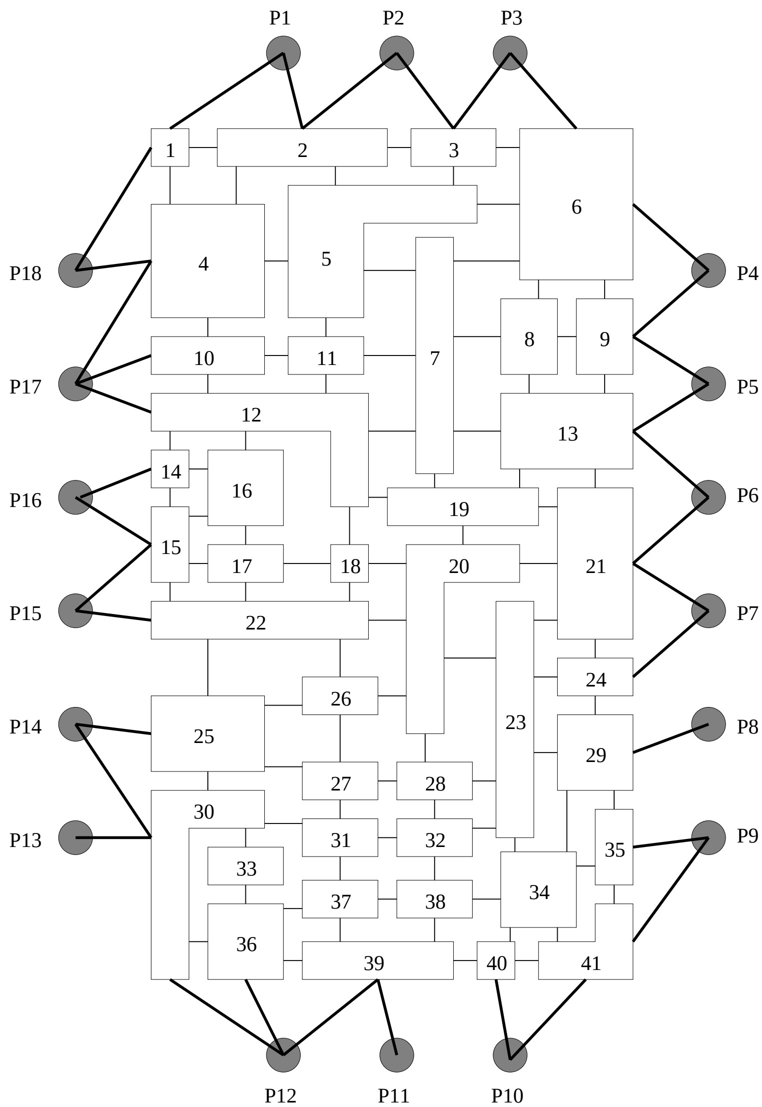
The second alternative configuration and corresponding solution are:



We see that with all three starting configurations, we always obtain the same optimal solution. This illustrates the robustness of the AR model when applied to this test problem.

5.2 Medium-Sized Macro-Cell Placement Example

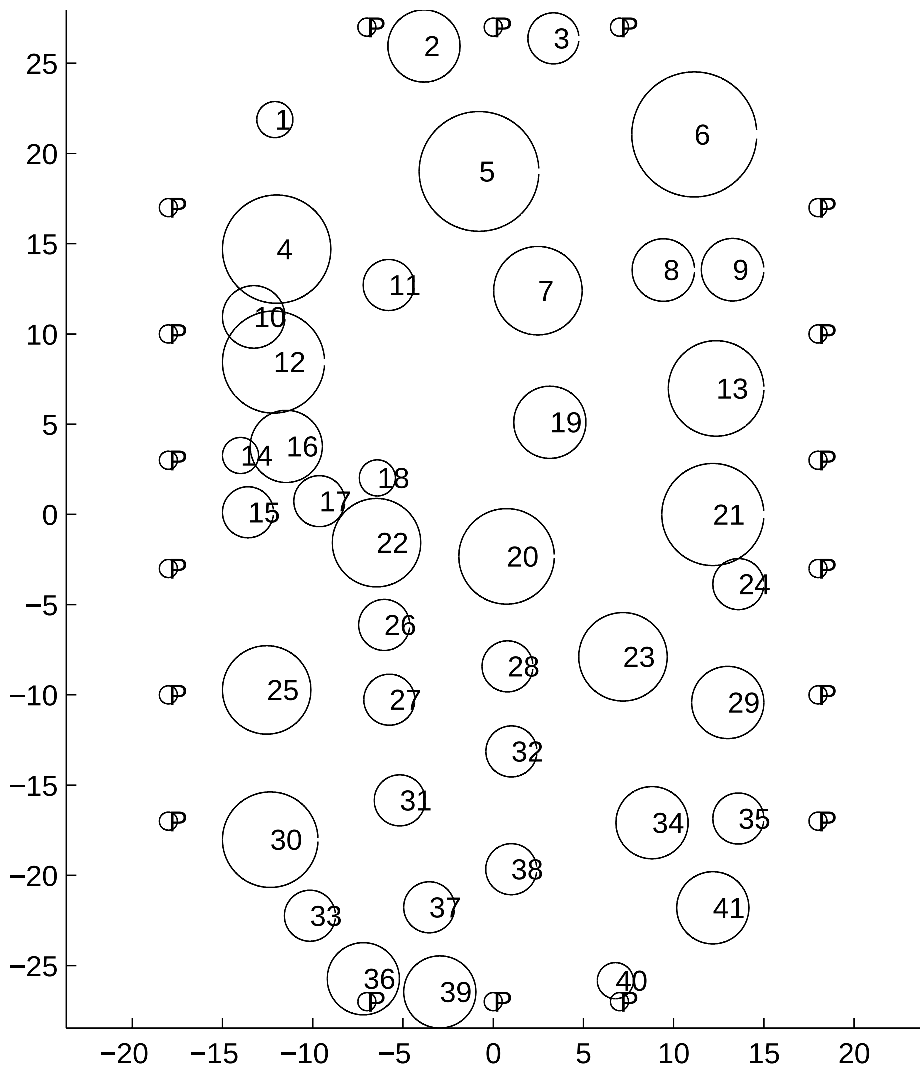
Next, we experimented with a larger test problem of macro-cell layout. Again, it has the following *optimal* layout:



where $c_{ij} = 1$ for the connections with the lighter edges (these are connections between mobile modules) and $c_{ij} = 10$ for the connections with the darker edges (these are connections between a fixed pad and a mobile module).

For our numerical tests, we used $w_T^{\min} = 25, \quad w_T^{\max} = 30, \quad h_T^{\min} = 50, \quad h_T^{\max} = 60$ and the radii for the mobile modules M1, ..., M41 are (respectively) 1,2,1.414,3,3.317,3.464,2.450,1.732,1.732,1.732,1.414,2.828, 2.646,1,1.414,2,1.414,1,2,2.646,2.828,2.450,2.450,1.414,2.450,1.414,1.414,1.414,2,2.646,1.414,1.414,1.414,2, 1.414,2,1.414,1.414,2,1,2, while each of the 18 pads has radius equal to 0.5. The initial configuration used was as described above, that is, the 41 mobile modules were placed at regular intervals on a circle of radius 200.

The result for $\alpha = 0.01$ is:

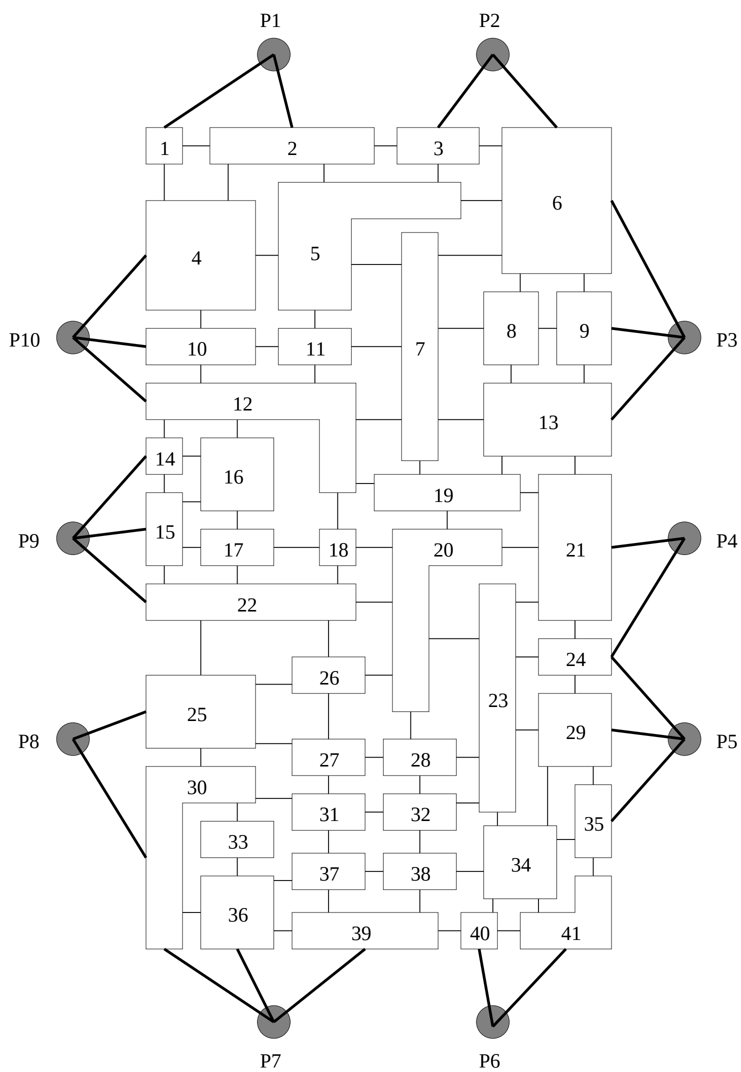


Note that, in comparison with the example in Section 5.1, we have significantly increased the number of modules in the problem, and furthermore, the ratio of fixed pads to mobile modules has decreased from

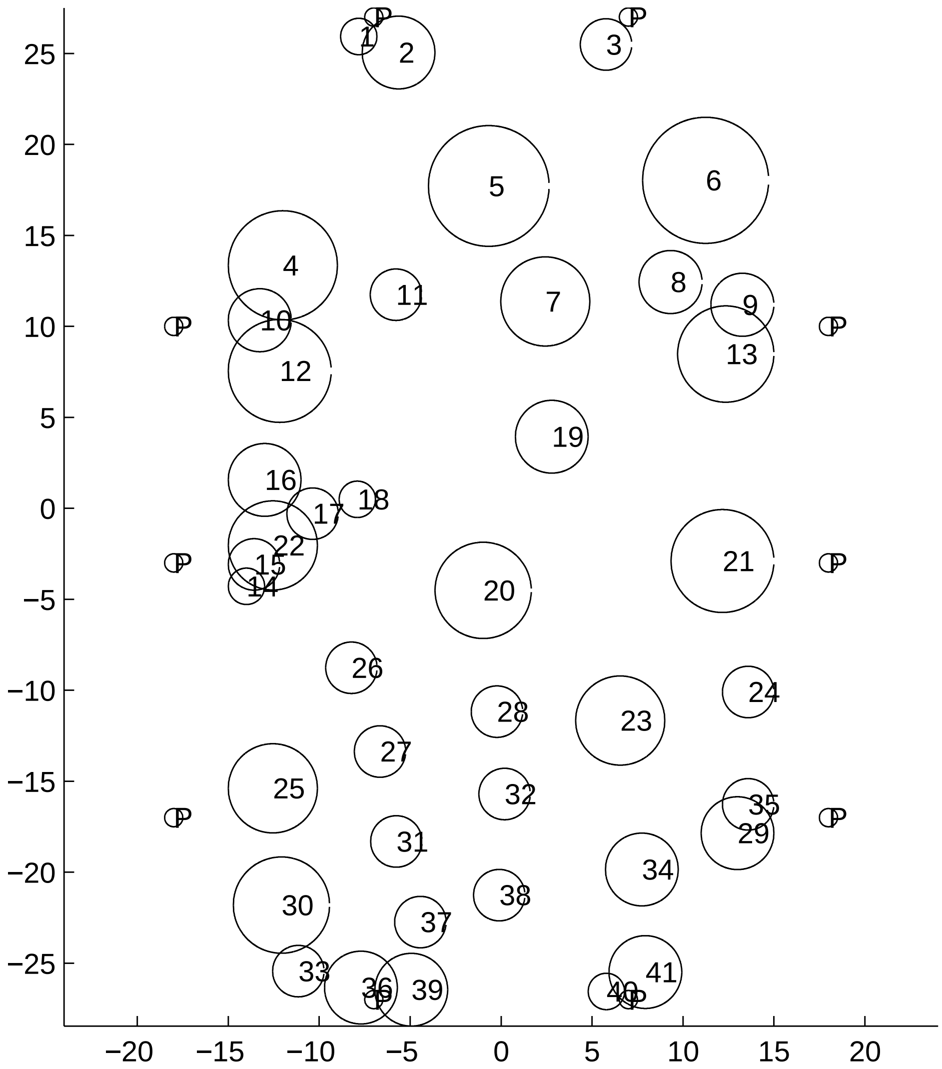
$8/11 \approx 0.73$ to $18/41 \approx 0.44$. Nonetheless, we again obtain from the AR model (with the same value of $\alpha = 0.01$) a solution whose layout is very similar to the (known) global optimal solution, i.e. the modules are all quite close to their optimal positions.

5.3 Alternative Medium-Sized Macro-Cell Example

To further test the robustness of the AR model for macro-cell applications, we chose our third example to be the same circuit as in the second example, but with even fewer fixed pads. This example has now only 10 pads and therefore a ratio of fixed pads to mobile modules of $10/41 \approx 0.24$. The dimensional bounds, radii and initial configuration were left unchanged, since our purpose is to make the proportion of fixed pads *smaller* and observe the effect on the overall layout. The optimal layout for this third example is:



And the result (again leaving $\alpha = 0.01$ unchanged) is:



We see that the obtained layout is not as clearly defined. Indeed, there are two clusters of modules, one consisting of modules M29 and M35, and the other consisting of modules M14, M15, M16, M17 and M22, whose circles overlap significantly and whose relative positions are either unclear or different from those in the optimal layout. Nonetheless, the clusters themselves are in the correct area of the layout and thus the overall result is still an excellent approximation of the optimal layout.

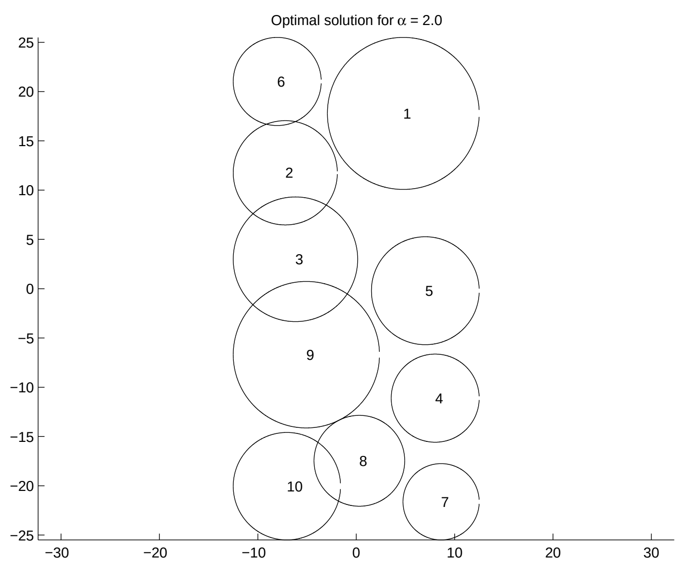
We tried varying α between 0.005 and 0.02 with similar results. The key observation is that a lower ratio of fixed pads to mobile modules has little effect on the overall optimal positions of the circles in the layout.

5.4 Facility Layout Example

Finally, we applied the AR approach to the 10-facility example from [8]. This problem comes from an existing production plant and hence the dimensions of the facility are fixed in advance. Therefore we set $w_T^{\min} = w_T^{\max} = 25, \quad h_T^{\min} = h_T^{\max} = 51$. This problem has 10 mobile modules (departments) and, in contrast with our previous examples, has *no* fixed modules. We use the same approximating circles as in [8], therefore the radii for M1, ..., M10 are set to 7.7136,5.2915,6.3246,4.4721,5.4772,4.4721,3.8730,4.6098,7.4330,5.4544. These radii were chosen so that the radius of each circle equals half of the square root of the (given) area for each department. Also as in [8], the costs c_{ij} are as follows:

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
M1	0	0	0	0	0	218	0	0	0	0
M2	0	0	0	0	0	148	0	0	296	0
M3	0	0	0	28	70	0	0	0	0	0
M4	0	0	28	0	0	28	70	140	0	0
M5	0	0	70	0	0	0	0	210	0	0
M6	218	148	0	28	0	0	0	0	0	0
M7	0	0	0	70	0	0	0	0	0	28
M8	0	0	0	140	210	0	0	0	0	888
M9	0	296	0	0	0	0	0	0	0	59.2
M10	0	0	0	0	0	0	28	888	59.2	0

Here is the best layout we obtained:



We see that (for $\alpha = 2.0$) the AR model yields a solution exhibiting a fairly good separation of the circles. The optimal layout for this problem is not known, and interestingly this solution is different from the Stage-2 solution reported in [8].

Furthermore, we can compare the AR model and the Stage-1/Stage-2 sequence of models of the NLT method in terms of the computational effort required to find a solution for the layout of the circles. We took the cumulative summary of computational data reported in [8] to solve the Stage-1 and Stage-2 problems and compared the number of iterations with the results reported by MINOS. The results are presented in Table 1 and illustrate the advantage of having a superlinearly convergent algorithm as is the case for the AR model.

	Number of iterations
NLT method (Stage-1 & Stage-2)	6607
AR model with $\alpha = 2.0$	31

Table 1: Comparison of iteration counts between NLT and AR

Finally, we remark that Etawil and Vannelli studied three possible choices of "repeller" functions:

- $z - \ln(z) - 1$;
- $z + e^{1-z} - 2$;
- $\frac{1}{z} - 1$.

Any of these choices can be used in the AR model. Our experimental results showed that the choice of $\frac{1}{z} - 1$ outperformed the other two choices for the floorplanning examples we considered.

6 Conclusions

Our numerical results show that the AR (Attractor-Repeller) model was robust on several "optimal" floorplanning problems irrespective of the number of fixed pads on the boundary. This new model *globally* places modules or cells of different sizes in optimal two-dimensional locations. This global optimality is the most important feature of this new model. When compared with the Stage-1/Stage-2 sequence of the NLT method, the new AR model finds in just one stage, and furthermore much more quickly and efficiently, a good initial point for the Stage-3 model of the NLT method.

Our results also show that the choice of the parameter α in the AR model has a significant influence on the optimal solution. Hence the question of how to choose α appropriately deserves further study.

We presented a convex version of the AR model, CoAR, but also observed that solving it requires a carefully designed line search algorithm. Further work is required to implement such an algorithm and compare the results with those of the (non-convex) AR model.

Finally, we believe that the ideas presented in Sections 2 and 3 are applicable to other layout problems. For example, they may be quite effective for solving another important layout problem in circuit design, namely the standard-cell problem [5]. In view of our computational results here, this seems a promising avenue of research.

References

[1] *GAMS – The Solver Manuals.* GAMS Development Corporation, Washington, D.C., 1996.

[2] D.P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, MA, 1995.

[3] A. Brooke, D. Kendrick, and A. Meeraus. *GAMS – A User's Guide, Release 2.25.* The Scientific Press, South San Francisco, CA, 1992.

[4] Z. Drezner. DISCON: A new method for the layout problem. *Operations Research*, 28(6):1375–1384, 1980.

[5] H.A.Y. Etawil and A. Vannelli. Target distance models for VLSI placement problem. To appear in *Journal of VLSI.*

[6] R.D. Meller and K.-Y. Gau. The facility layout problem: Recent and emerging trends and perspectives. *Journal of Manufacturing Systems*, 15(5):351–366, 1996.

[7] P. Pardalos and H. Wolkowicz, editors. *Quadratic assignment and related problems.* American Mathematical Society, Providence, RI, 1994. Papers from the workshop held at Rutgers University, New Brunswick, New Jersey, May 20–21, 1993.

[8] D.J. van Camp, M.W. Carter, and A. Vannelli. A nonlinear optimization approach for solving facility layout problems. *European Journal of Operational Research*, 57:174–189, 1991.