International Journal of Foundations of Computer Science

OPTIMAL LAYOUT OF TRIVALENT CAYLEY INTERCONNECTION NETWORKS

TIZIANA CALAMONERI*

and

ROSSELLA PETRESCHI
Department of Computer Science
University of Rome "La Sapienza" - Italy
e-mail address:{ *calamo,petreschi* }*@dsi.uniroma1.it*

Received (received date)
Revised (revised date)
Communicated by Name of Editor

ABSTRACT

In this paper we deal with the layout of Trivalent Cayley Interconnection Networks. Namely, we prove that a lower bound on their layout area is $\Omega(2^{n-1} \times 2^{n-1})$ and we exhibit some methods to lay these networks out in $O(2^{n-1} \times 2^{n-1})$.
We point out that these layout methods work for other networks decomposable into equal length cycles connected by following fixed rules.

Keywords: Trivalent Cayley Networks, Layout, Cubic Graphs, Interconnection Network.

1. Introduction

One of the problems in designing a network of processors on a commercial basis is to find an underlying structure which can be defined 'good'. There are several parameters to analyze in order to measure the 'goodness' of networks. The most common among these include fixed degree, regularity, logarithmic diameter and maximal fault tolerance. A further parameter – as important as the previous ones – is whether the network can be laid out compactly in the orthogonal grid. The importance of this criterion is born out from the applications both in the study of the VLSI layout problem [1] and in the mapping problem for parallel architectures [2, 3].

The optimal layout problem has been solved for a lot of particular networks, such as Butterfly, CCC, Sorting, Shuffle-Exchange, and deBruijn networks [4, 5, 6, 7, 8].

In this paper we deal with Trivalent Cayley Interconnection Networks (TCINs), that are represented by regular degree 3 graphs based on permutation groups. This

*Supported by the Italian Research Council (CNR).

model has been introduced in 1995 [9], and, from that time has been extensively studied [10, 11, 12], since it is an example of 'good' network respect to the parameters listed above. In this paper we show its 'goodness' also respect to optimal layout. Namely, for the first time we study the layout problem on $TCIN$s proving that the lower bound of their layout area is $\Omega(2^{n-1} \times 2^{n-1})$ and exhibiting a couple of methods to suitably lay $TCIN$s out in $O(2^{n-1} \times 2^{n-1})$ area. Both methods reach the same result in terms of order of magnitude, but the second one improves the first one of a constant factor. We show both methods for emphatiseing that even an improvement of a modest constant factor may render a layout more appealing than another one for real implementation.

Finally, we point out that our layout methods work for other networks decomposable into equal length cycles connected by following fixed rules.

2. Definitions and Preliminaries

Interconnection networks can be conveniently modeled by using tools from graph theory since their processors may be put in correspondence with the vertices of a graph and the communication links between processors may be seen as the edges connecting the vertices. Henceforth, now on we will use the terms 'networks' and 'graphs' interchangeably; the same holds for the terms 'communication links' and 'edges', though usually the term 'node' will be preferred to 'vertex'.

Definition 1 *[9] An n-dimensional Trivalent Cayley interconnection network ($TCIN_n$) is a network whose topology is a symmetric undirected 3-regular graph $G = (V, E)$. Each node corresponds to a circular permutation in lexicographic order of n symbols, $a_1, a_2, \ldots, a_n$, complemented or uncomplemented. Each edge is of the type $(v, \delta(v))$, where $\delta \in \{g, f, f^{-1}\}$, defined in the following way:*

- $f(a^*_k a^*_{k+1} \ldots a^*_n a^*_1 \ldots a^*_{k-1}) = a^*_{k+1} \ldots a^*_{k-1} \overline{a^*_k}$ *(f-edge)*
- $f^{-1}(a^*_k \ldots a^*_n a^*_1 \ldots a^*_{k-2} a^*_{k-1}) = \overline{a^*_{k-1}} a^*_k \ldots a^*_{k-2}$ *(f^{-1}-edge)*
- $g(a^*_k \ldots a^*_n a^*_1 \ldots a^*_{k-1}) = a^*_k \ldots \overline{a^*_{k-1}}$ *(g-edge)*

where $a^*_k \ldots a^*_n a^*_1 \ldots a^*_{k-1}$ denotes the label of an arbitrary node and a^*_i denotes either a_i or $\overline{a_i}$. It is to notice that $g = g^{-1}$.

In Fig. 1 a $TCIN$ of dimension 3 is depicted.

The cardinality of V is $n2^n$ since for n distinct symbols there are exactly n different cyclic permutations and each symbol can be present either in complemented or uncomplemented form.

We call *f-cycle* a cycle consisting of only f-edges (equivalently f^{-1}-edges). All the $n2^n$ nodes are partitioned into 2^{n-1} disjoint f-cycles of length $2n$. Nodes v and $\overline{v}$ are in the same f-cycle.

Each f-cycle has a unique node starting with a_1 that is numbered 1 and all the other nodes are consecutively numbered from 1 to $2n$, following f-edges.

Following function g, each f-cycle is connected to exactly n different f-cycles.

Now, the connections between different f-cycles are analyzed. For the symmetry of the structure, it is not restrictive to consider only nodes from 2 to $n+1$ on each

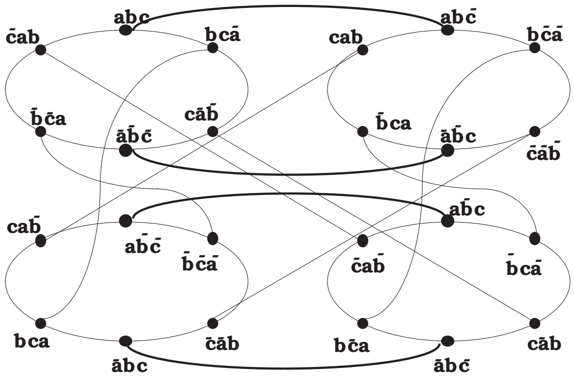

Figure 1: A $TCIN$ of dimension 3.

f-cycle, since each node $n+k$ has the same behavior than node k.

In the following we associate to each label starting with a_1 a binary string: the b_{n-k}-th bit is equal to 1 or to 0 if symbol $a_k, k = 2, \ldots, n$ is uncomplemented or complemented, respectively. The resulting string identifies the cycle the node belongs to. In the following we number cycles in increasing order according to the associated binary strings and we call 'stratificated' structure the representation of $TCIN$s as f-cycles put one on top of the other, sorted according to their numbering.

Fact 1 *[10] Given a f-cycle $b_{n-2}\ldots b_0$, each node k $(2 \leq k \leq n+1)$ on it is connected by function g to node v. If $k \neq 2$, node v is the k-th node on the cycle $b_{n-2}\ldots\overline{b_{n-k+1}}\ldots b_0$. In other words, we can say that node k skips $2^{n-k+1}-1$ cycles to find its adjacent node v in the 'stratificated' structure. Namely, nodes 3, 4, $\ldots n, n+1$ skip $2^{n-2}, 2^{n-3}, \ldots 1, 0$ cycles, respectively. The only node that has a behavior different from the other ones is node 2 $(n+2)$ in view of the fact that function g complements a_1. Then, node 2 of a fixed cycle $b_{n-2}\ldots b_0$ is connected to the $(n+2)$-th node (2-nd node) of the cycle $\overline{b_{n-2}\ldots b_0}$.*

In Fig. 2 a $TCIN_3$ is shown as the stratification of f-cycles joint by vertical and slanting connections.

Fact 2 *[9] Shrinking each f-cycle of a $TCIN_n$ in a node, the reduced graph $RTCIN_n$ contains a $(n-1)$-dimensional hypercube network. Namely, $RTCIN_n$ is a graph given by a $(n-1)$-dimensional hypercube having each edge duplicated and all its diagonals.*

In Fig. 3 an $RTCIN_4$ is shown.

3. Upper Bound on the Layout Area of a *TCIN*

In this section we will show a couple of methods to lay a $TCIN_n$ out in $O(2^{n-1} \times 2^{n-1})$ area, while in the next section we will prove that this is an optimum value for area.

The stratificated structure of Fig.2 induces to understand that the layout strongly depends on the mutual position of f-cycles and on the layout of g-edges. Indeed,

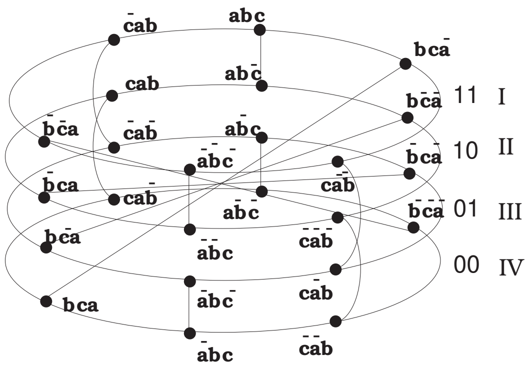

Figure 2: Stratification of f-cycles of the $TCIN_3$.

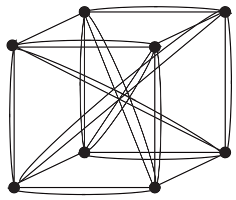

Figure 3: Reduced $TCIN_4$ network, $RTCIN_4$.

the layout of f- and f^{-1}-edges consists of the representation on the grid of the cycles they belong to.

In the first method we describe in this section, we use the standard representation of disjoint cycles by means of rectangles of height 1 [13], while in the second one we map disjoint cycles into concentric rectangles.

These two methods differ for a constant factor, and both prove the following result:

Theorem 1 $O(2^{n-1} \times 2^{n-1})$ *area is sufficient to orthogonally lay a* $TCIN_n$ *out.*

3.1. The First Method to Lay a $TCIN$ Out

f-cycles are mapped onto rectangles of height 1 and all vertices lie on the same side of each rectangle; then they will be represented as flattened on a segment (see Fig. 4). Moreover, for the symmetry of $TCIN$s, we may consider only half of the nodes (numbered from 2 to $n+1$).

Definition 2 *Given a* $TCIN_n$, *we call* k-th block, $1 \leq k \leq 2n, k \neq 2, n+2$, *the set of all g-edges connecting two k-th nodes on different f-cycles. The set of g-edges connecting a 2-nd node on a f-cycle with a $(n+2)$-th node on another f-cycle is called* extra-block.

From the previous considerations about cycles and from Fact 1, it follows that it is not difficult to put nodes on cycles they belong to at horizontal coordinates such that, for any fixed $k \neq 2$, all g-edges belonging to the k-th block are vertical segments

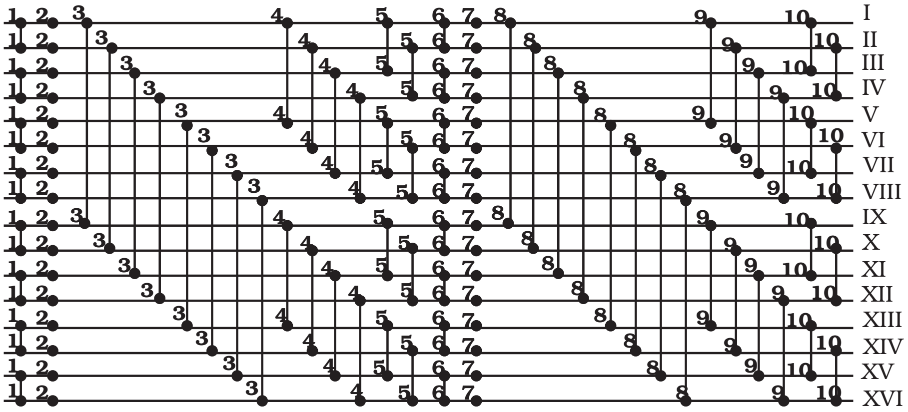

Figure 4: Rectilinear scheme for a $TCIN_5$.

of length 2^{n-k+1} and the global width of the k-th block is $2^{n-k+1}, 3 \leq k \leq n+1$ (see Fig. 4). If $k = 2$, the length of g-edges belonging to the extra-block is not fixed yet. Indeed, the fact that node 2 is not connected to one of its homonyms implies that all these edges cannot be laid out as straight-lines in the scheme of Fig. 4.

In order to obtain an orthogonal layout and to optimize the area, it is profitable to invert the order of the nodes on the cycles numbered from $2^{n-2}+1$ to 2^{n-1}. This choice implies that all the k-th blocks with $4 \leq k \leq n+1$ (i.e. all the k-th blocks whose g-edges have length no greater than 2^{n-2}) are divided into two symmetrical sub-blocks of width 2^{n-k+1}, $4 \leq k \leq n+1$. As consequence, the area for representing g-edges with end-point k, $4 \leq k \leq n+1$, will be duplicated.

It is to notice that 4-th sub-blocks can be fit into the area allocated for the 3-rd block. The general layout for a $TCIN_n$ is shown in Fig. 5 and its area is

$$(2 \cdot 2^{n-1}) \times 3 \cdot 2^{n-1}.$$

It is obtained by multiplying:

- height$=2 \cdot 2^{n-1}$:
  each of 2^{n-1} cycles is laid out as a rectangle on two rows.
- width$=2(2 + 2 \cdot 2^{n-2} + 2 + 2\sum_{k=5}^{n} 2^{n-k+1}) = 3 \cdot 2^{n-1}$:
  - (i) each sub-block of the $(n+1)$-th block (and the 1-st block) is laid out on one column;
  - (ii) 3-rd block and extra-block are laid out in 2^{n-2} columns, each;
  - (iii) also the 4-th block needs 2^{n-2} columns, globally; but since it shares $2^{n-2}-2$ columns with the 3-rd block, it contributes with only 2 columns to the computation of the total width;
  - (iv) each of the remaining k-th blocks needs $2 \cdot 2^{n-k+1}$ columns, globally, $5 \leq k \leq n$.

  The sum of all these contributions gives half of the total width.

For what concerns the maximum wire length of this layout, it is $2 \cdot 2^{n-1}$, that is the length of a g-edge of the extra-block.

The result of this subsection is summarized in the following lemma:

Lemma 1 *A $TCIN_n$ can be orthogonally laid out in area $(2 \cdot 2^{n-1}) \times (3 \cdot 2^{n-1})$ with maximum wire length $O(2^{n-1})$.*

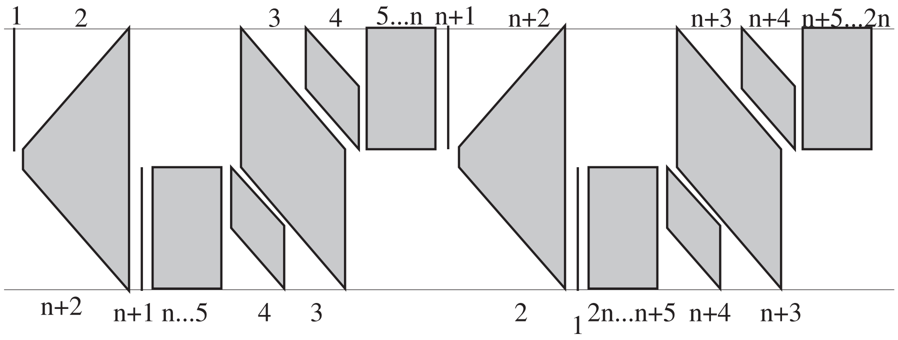

Figure 5: General scheme of a $TCIN_n$ on flattened rectangles.

3.2. The Second Method to Lay a $TCIN_n$ Out

The scheme of the layout in Fig.5 shows that some space is wasted. In order to reduce the area, we transform the layout of cycles from a 'stratificated' model to a 'concentric' one, as shown in Fig.6. The area of this new scheme is

$$(2 \cdot 2^{n-1}) \times (2 \cdot 2^{n-1} + 2^{n-3}).$$

Indeed:

- height$=2 \cdot 2^{n-1}$:
  no additional rows are needed besides the rows used by the cycles (i.e. the most interior cycle has height equal to 1) because the k-th blocks, $5 \leq k \leq n$, require $\sum_{k=5}^{n} 2^{n-k+1} = 2^{n-3} - 2$ rows among 2^{n-2} available. Therefore the height is the same than in the flattened model.
- width$=2 \cdot 2^{n-1} + 2^{n-3}$:
  it differs from height just for the width of the most interior cycle, that is equal to the width globally required by the k-th blocks, $5 \leq k \leq n$ (i.e. $2^{n-3} + 2$) plus the contribution of 1 due to the $(n+1)$-th block and the contribution of 1 due to the 4-th block (see Fig. 6, upper side of the most interior cycle).

For what concerns the maximum wire length, it is easy to see that it is of the same order of magnitude as the previous case.

As consequence of the previous arguments, we have:

Lemma 2 *A $TCIN_n$ can be orthogonally laid out in area $(2 \cdot 2^{n-1}) \times (\frac{9}{4} \cdot 2^{n-1})$ with $O(2^{n-1})$ maximum wire length.*

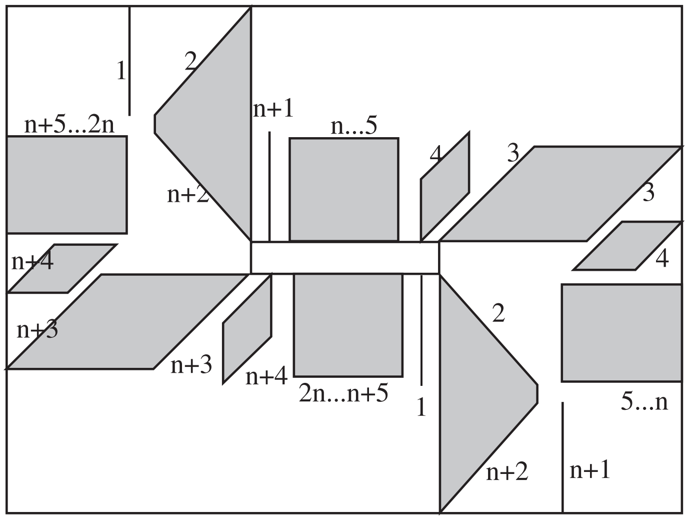

Figure 6: General scheme of a $TCIN$ on concentric rectangles.

4. Lower Bound on the Layout Area of a *TCIN*

In this section we will prove that the result of Theorem 1 is optimum because it cannot differ from the lower bound more than a constant factor, as the following theorem states:

Theorem 2 $\Omega(2^{n-1} \times 2^{n-1})$ *area is necessary to orthogonally lay a* $TCIN_n$ *out.*

Before detailing the proof of Theorem 2 we introduce some preliminary definitions and lemmas; furthermore we recall the standard strategy used to prove the lower bound on the layout area of networks introduced in [1] for what concerns the bisection width and in [14] for what concerns the congestion arguments, and refined in [15].

Definition 3 *[13] An* embedding *of a guest graph G into a host graph H is defined by an injective correspondence α from nodes of G into nodes of H and a correspondence ρ which associates to each edge (u, v) of G a path connecting $\alpha(u)$ and $\alpha(v)$ in H. The* congestion *of an embedding is the maximum, over all edges e of H, of the number of edges of G whose ρ-paths contain edge e.*

Definition 4 *[13] The* minimum bisection width *of a graph G, MBW(G), is the smallest number of edges whose removal partitions G into two disjoint subgraphs, each containing half of the nodes.*

Lemma 3 *[1] For any graph G, the area of the smallest grid in which G can be laid out is greater or equal to $(MBW(G) - 1)^2$.*

Lemma 4 *[14] Let ϵ be an embedding of a graph G into a graph H which has congestion C, then it holds:*

$$MBW(H) \geq \frac{1}{C} MBW(G).$$

Definition 5 *[16] The r-dimensional Cube Connected Cycles network CCC_r is constructed from the r-dimensional hypercube by replacing each node of the hypercube with a cycle of r nodes in CCC_r. The i-th dimension edge incident to a node of the hypercube is then connected to the i-th node of the corresponding cycle in CCC_r. CCC_r has $r2^r$ nodes each with degree 3.*

In the following, we will call reduced CCC_r, $RCCC_r$, the r-dimensional hypercube underlying CCC_r.

Fact 3 *[13] The MBW of a CCC_r is $\Theta(2^r)$.*

Basic Strategy: *As consequence of Lemmas 3 and 4, the lower bound on the area of a network H can be computed through an embedding ϵ into H of a graph G whose MBW is known. Moreover, we need to know the congestion C of ϵ. In such a way, we will have that:*

$$lower\, bound\, on\, area\, of\, H \geq (MBW(H)-1)^2 \geq (\frac{1}{C}MBW(G)-1)^2.$$

Now we are ready to detail the proof of Theorem 2.

Proof. We choose as guest graph G of the embedding ϵ the CCC_{n-1} network and as host graph H the $TCIN_n$.

Then, to compute embedding ϵ, we work on the reduced networks $RCCC_{n-1}$ and $RTCIN_n$ and define the natural correspondence ϕ from vertices of $RCCC_{n-1}$ to vertices of $RTCIN_n$, and from edges of $RCCC_{n-1}$ to edges of the hypercube contained into $RTCIN_n$ (i.e. one edge from each double edge in $RTCIN_n$ except edges derived from the edges with endpoint 2 in $TCIN_n$).

To define α we re-expand vertices of $RCCC_{n-1}$ and $RTCIN_n$ in cycles of CCC_{n-1} and $TCIN_n$, respectively: each vertex v in $RCCC_{n-1}$ is a cycle individuated by vertices $v_1, v_2, \ldots, v_{n-1}$ in CCC_{n-1}; v's correspondent $\phi(v)$ in $RTCIN_n$ is a cycle individuated by vertices $\phi(v)_1, \phi(v)_2, \ldots, \phi(v)_{2n}$ in $TCIN_n$. Then:

- $\alpha(v_1) = \phi(v)_1$;
- $\alpha(v_i) = \phi(v)_{i+1}$, for $i = 2, \ldots n-1$.

For each edge e in CCC_{n-1} we define the following ρ correspondence:

(i) if $e = (v_i, v_{i+1})$, i.e. e connects two vertices in the same "reduced" node, then:
$\rho((v_1, v_2)) = path(\phi(v)_1, \phi(v)_2, \phi(v)_3)$;
$\rho((v_i, v_{i+1})) = (\phi(v)_{i+1}, \phi(v)_{i+2})$ for $i = 2, \ldots n-2$;
$\rho((v_{n-1}, v_1)) = path(\phi(v)_n, \phi(v)_{n+1}, \ldots, \phi(v)_{2n}, \phi(v)_1)$.

(ii) if $e = (v_i, w_i)$, i.e. e coincides with an edge in $RCCC_{n-1}$, then:
$\rho(e) = \phi(e)$.

It is easy to see that the congestion C of such an embedding is 1 because the consecutive numbering of vertices on each cycle of CCC_{n-1} (as Fact 1 for the $TCIN_n$) guarantees both that ρ considers each edge just once and that item (ii) is well defined. Applying to $TCIN_n$ the basic strategy described to compute the lower bound on area of a network, using the embedding just described and recalling Fact 3, the thesis follows. □

5. Generalizations

The bases of the three-dimensional 'stratificated' model of $TCIN$s (Fig.2) lie in the existence of:

a. a partition of nodes in cycles of equal length (f-cycles) and

b. fixed rules to connect different cycles (Fact 1).

Then, we can think to model each interconnection network having both a. and b. properties as stratification of cycles. As an example, in Figs. 7.a and 7.b we show the stratificated structure of a CCC and a Wrapped Butterfly [13].

It is natural to wonder if the methods used to optimally lay the $TCIN$ out starting from the stratificated model can be extended to other stratificated networks.

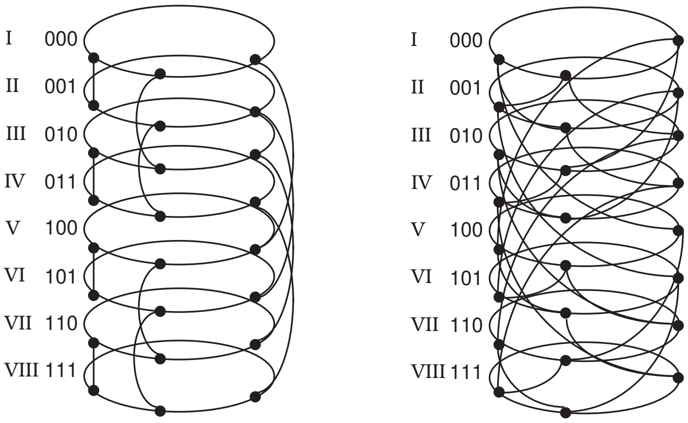

Figure 7: Stratificated structure of CCC_3 and WB_3.

For what concerns CCC network, it is possible to see that a variant of the general scheme shown in Fig. 5 brings to the layout of CCC shown in Fig. 8. This variant is obtained in the following way:

a. eliminate the modification introduced because of the extra-blocks of $TCIN$, absent in CCC and

b. recursively invert the order of the cycles in the latter half of the layout.

Notice that this variant represents an algorithmic method to optimally lay a CCC out, as shown in [16] as 'slight improvement'.

The idea of recursively inverting the order of the cycles in the latter half allows one to lay a $TCIN$ out in $(2 \cdot 2^{n-1}) \times (3 \cdot 2^{n-1} + \Theta(n))$ area; therefore it does not bring any further improvement to the upper bound on the layout area of $TCIN$ network.

For what concerns the Wrapped Butterfly, the degree 4 seems to give problems, so its layout appears more tricky.

It remains an open problem to understand for which stratificated networks our method can be used.

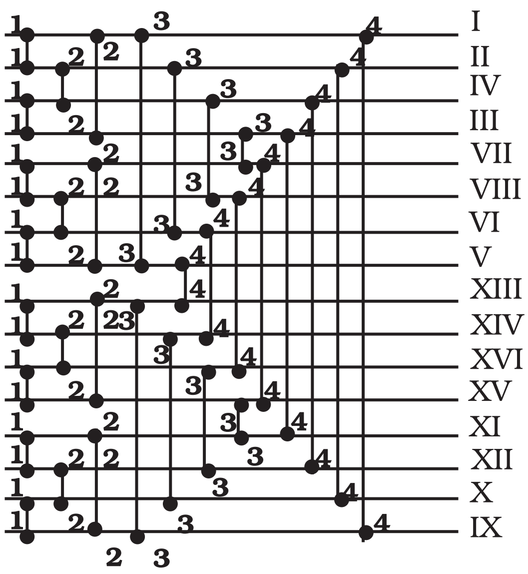

Figure 8: Classical layout of CCC_4.

References

1. C.D. Thompson, "A complexity theory for VLSI", Ph.D. thesis, Carnegie-Mellon Univ. Pittsburgh, 1980.
2. F. Berman and L. Snyder, "On mapping parallel algorithms into parallel architectures", *J. Parallel Distr. Comput.* **4** (1987) 439–458.
3. S.H. Bokari, "On the mapping problem", *IEEE Trans. Comp.* **C-30** (1981) 207-214.
4. A. Avior, T. Calamoneri, S. Even, A. Litman and A.L. Rosenberg, "A Tight Layout of the Butterfly Network", *Theory of Computing Systems (Math. Syst. Theory)* **31** (1998) 475–487. Preliminary version in *Proc. 8-th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA '96)*, ACM Press Ed. (1996) 170–175.
5. G. Chen and C.M. Lau, "Tight Layout of the Cube-Connected Cycles", *Proc. 4th Intern. Conf. of High Performance Computing (HiPC '97)*, (1987) 422–427.
6. S. Even, "Layout Area of Sorting Networks", submitted to *Networks* (1998).
7. D. Kleitman, F.T. Leighton, M. Lepley and G.L. Miller, "New Layouts for the Shuffle-Exchange Graph", *13h Annual ACM Symposium on Theory of Computings (STOC '81)* (1981) 278–292.
8. D.K. Pradhan and M.R. Samatham, "The deBruijn multiprocessor network: a versatile parallel processing and sorting network for VLSI", *IEEE Trans. Comput.* **38** (1989) 567–581.
9. P. Vadapalliand P.K. Srimani, "Trivalent Cayley Graphs for Interconnection Networks", *Informat.Process.Letters* **54** (1995) 329–335.
10. T. Calamoneri and R. Petreschi, "A new 3D representation of Trivalent Cayley Networks", *Inform. Proc. Letters* **61** (1997) 247–252.
11. P. Vadapalli and P.K. Srimani, "Shortest Routing in Trivalent Cayley Graph Networks", *Informat.Process.Letters* **57(4)** (1996) 183–188.

12. M.D. Wagh and J. Mo, "Hamiltonian cycles in Trivalent Cayley Graphs", *Inform. Process. Letters* **60** (1996) 177–181.

13. F.T. Leighton, *Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes* (Morgan Kaufmann Publishers, Inc.,1992).

14. F.T. Leighton, *Complexity Issues in VLSI: Optimal Layouts for the Shuffle-Exchange Graph and Other Networks.* MIT Press, Cambridge Mass, ,1983.

15. A.L. Rosenberg and L.S. Heath, *Graph Separators, with Applications*, in preparation.

16. F.P. Preparata and J. Vuillemin, "The Cube-Connected Cycles: A Versatile Network for Parallel Computation", *Communications of ACM* **24(5)** (1981) 300–309.