Merging and Sorting Networks with the Topology of the Omega Network

Gianfranco Bilardi*

87-816

March 1987

Department of Computer Science
Cornell University
Ithaca, New York 14853-7501

*This research has been supported in part by the National Science Foundation grant DCI-8602256. A preliminary version of some of the results was presented at the 23rd Annual Allerton Conference on Communication, Control and Computing, Allerton, Illinois, October 1985.

MERGING AND SORTING NETWORKS WITH THE TOPOLOGY OF THE OMEGA NETWORK

Gianfranco Bilardi

Department of Computer Science

Cornell University, Ithaca, NY 14853

Abstract

We consider a class of comparator networks obtained from the *omega* permutation network by replacing each switch with a comparator exchanger of arbitrary direction. These networks are all isomorphic to each other, have merging capabilities, and can be used as building blocks of sorting networks in ways different from the standard merge-sort scheme. It is shown that the *bitonic* merger and the *balanced* merger are members of the class. These two networks were not previously known to be isomorphic.

This research has been supported in part by the National Science Foundation grant DCI-8602256. A preliminary version of some of these results was presented at the 23rd Annual Allerton Conference on Communication, Control and Computing, Allerton, Illinois, October 1985.

1. Introduction

Networks of switches and networks of comparators have been widely investigated, especially for their relevance to parallel computing. In this paper, we study the class of comparator networks that have the same topology as the *omega* permutation network [La 75]. We develop a framework that leads both to a unified treatment of previously known networks and to the discovery of new ones.

The original motivation for this work was provided by the desire to understand the exact relationship between the classical *bitonic merger* [Ba68], and the more recently proposed *balanced merger* [DPRS83]. The similarity between the two networks is striking, but difficult to characterize precisely.

To illustrate the nature of the difficulty, let us recall that, following [Kn73], a network of comparators is typically viewed as a set of lines (which can be imagined to run parallel, horizontally) connected as a set of comparators (which can be imagined as vertical connections between pairs of lines). This view naturally suggests that two networks that differ only by a permutation of the lines are equivalent.

Indeed, in [DPRS83] it is shown that no permutation of the lines transforms the balanced merger into the bitonic one. The proof is indirect and elaborate, but very interesting in itself. It consists in showing that the cascade of $\log N$ balanced mergers of N lines sorts every input. Thus, a similar construction using the balanced merger with the lines permuted maps every input into a fixed permutation of the sorted output. It is easily shown that this property does not hold for the bitonic merger.

In spite of this fact, we shall see in this paper that the balanced merger is essentially the same network as the bitonic merger. The first step towards this result lies in the realization that permutation of lines is *not* the most general transformation that preserves the topology of the network. We define a network as a directed graph whose nodes are comparators and whose arcs are connections from an output of a comparator to an input of another comparator.

Although the proof is not straightforward, the balanced merger and the bitonic merger turn out to be isomorphic graphs. One of the consequences of this result is that the properties derived in [DPRS83] and also in [R85] for the balanced network hold, suitably reinterpreted, for the bitonic network.

Rather than concentrating on the two networks in question, we have found it fruitful to develop a general approach whereby networks of comparators are investigated in connection with their underling network of switches, obtained by replacing each comparator with a switch. With this approach, we have derived isomorphism results for a class of exponentially many networks, including the bitonic and the balanced one.

The switching network underlying these comparator networks is the well-known *omega* network [La75]. Our analysis exploits systematically some algebraic properties of the permutations passable by the omega network, obtained in [St83]. Besides establishing the isomorphism of the networks considered, the analysis sheds further light on their merging capabilities. For example, we show that there are exponentially many ways of feeding two sorted sequences to the bitonic merger and obtain a sorted output.

Another result, extending that of [DPRS83] mentioned above, is that from any of the merging networks considered, one can construct a sorting network by cascading $\log N$ copies of it, provided that the outputs of a copy are connected to the inputs of the next one according to a suitable permutation.

The remainder of this paper is organized as follows. In Section 2, we introduce the basic definitions and notations. Particular attention is devoted to various notions of equivalence among networks. In Section 3, the omega network and related concepts are discussed. A class of comparator networks is defined in Section 4. It is shown that all the members of this class are isomorphic to each other, that they have merging capabilities, and that the bitonic and the balanced mergers are among them. Sorting properties are investigated in Section 5. Conclusive remarks are in Section 6.

2. Networks of switches and comparators

A *comparator-exchanger* (CE) and a *switch* (SW) are both devices with two input terminals, receiving values a and b, and two output terminals, producing values c and d. In a CE, $c = \min(a,b)$ and $d = \max(a,b)$. In a SW, there is a binary state s that can be set independently of the inputs: if $s=0$, then $(c,d)=(a,b)$, else $(c,d)=(b,a)$. A CE can be viewed as a SW whose state is a function of the inputs ($s=0$ *iff* $a<b$).

A network T of size N is a directed acyclic graph such that :

(i) There are N nodes, called input terminals, with in-degree zero and out-degree one.

(ii) There are N nodes, called output terminals, with out-degree one and in-degree zero.

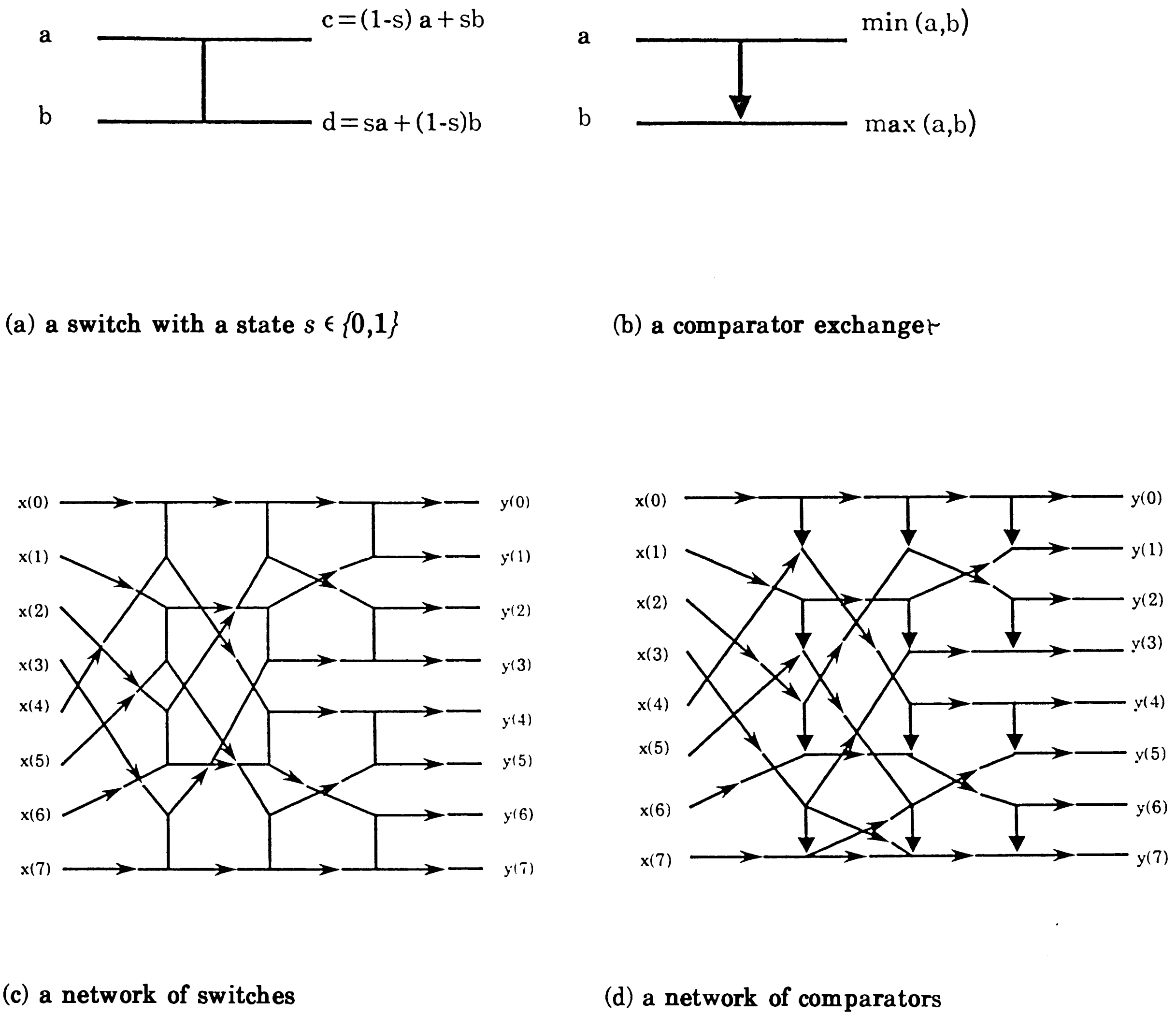
(iii) All the remaining nodes, representing devices, have in degree and out-degree two.

Depending on the nature of the devices, we have networks of switches and networks of comparators. Examples of networks are given in Figure 1 where a graphical notation for SWs and CEs is also introduced.

Given a network T (of either CEs or SWs) of size N, we assume that both the input terminals and the output terminals are numbered from 0 to $N-1$. We say that a sequence $x=(x(0),...,x(N-1))$ is the input of T when $x(i)$ is input at terminal i. Similarly, $y=(y(0),...,y(N-1))$ is the output sequence if $y(i)$ is output at terminal i. We denote by xT the output of network T corresponding to input x.

The output y of a network T of either CEs or SWs is a permutation of the input x, that is, $y=xT=(x(\tau(0)),...,x(\tau(N-1)))$, where $\tau=(\tau(0),...,\tau(N-1))$ is a permutation of $(0,...,N-1)$. We write $y=x\tau$, where the product represents composition of functions, and we say that T *realizes* τ (on x). Permutation τ is a function of x for a network of CEs, and is a function of the state of the switches for a network of SWs.

Without loss of generality, in the sequel we shall only consider input sequences that are permutations of $(0,...,N-1)$. Of particular interest will be those permutations that are



(a) a switch with a state $s \in \{0,1\}$

(b) a comparator exchanger

(c) a network of switches

(d) a network of comparators

Figure 1. Graphical notation for devices and networks.

mapped to the identity by a given network, for which we adopt the following terminology.

If T is a network of CEs, we say that permutation π is *sorted* by T if πT is the identity. If T is a network of SWs, we say that permutation π is *admissible* or *passable* by T if there is a setting of the switches such that πT is the identity. Note that, in either case, the permutation *realized* by T is $\tau = \pi^{-1}$.

It is convenient to define the *series* and the *parallel* connections of two networks T_1 and T_2 of size N_1 and N_2, respectively. The series of T_1 and T_2 is obtained by connecting the *i-th* output terminal of T_1, to the *i-th* input terminal of T_2 and is defined only when $N_1 = N_2$. We denote it by T_1T_2. The parallel $T_1 \bigcirc T_2$, is the union of T_1 and T_2, with terminal i of T_1 becoming terminal i of $T_1 \bigcirc T_2$ $(i=0,...,N_1-1)$, and terminal i of T_2 becoming terminal $i+N_1$ of $T_1 \bigcirc T_2$ $(i=0,...,N_2-1)$.

If T_r realizes permutation τ_r $(r=1,2)$, then T_1T_2 realizes the product $\tau_1\tau_2$, and $T_1 \bigcirc T_2$ realizes the parallel $\tau_1 \bigcirc \tau_2$ defined as $(\tau_1 \bigcirc \tau_2)(i) = \tau_1(i)$ for $i=0,...,N_1-1$, and $(\tau_1 \bigcirc \tau_2)(i+N_1) = \tau_2(i)+N_1$ for $i=0,...,N_2-1$.

Any network T can be conveniently viewed as a series of *stages*, each containing at most $N/2$ parallel devices. More specifically, the *k-th* stage $(k=0,1,...,d(T) \triangleq$ depth of $T)$ contains the nodes of T of depth k. The *depth* of a node u is the length of the longest path in T, from an input node to u. The maximum depth of a node in T is also called the depth of T. For example, the networks of Figure 1 consist of five stages, one of inputs, three of devices, and one of outputs.

In a network T of size N, the arcs can be partioned in N paths, each joining an input node to an output node. It is easy to see that if there are $c(T)$ devices in T, then there are $2^{c(T)}$ partitions of T in N *arc-disjoint paths.* Indeed, in a network of SWs, each partition corresponds to a setting of the SWs, the paths being those connecting the input nodes to the output nodes for that particular setting.

Following [Kn 73], it is customary to represent a network as an ordered set of N lines (drawn horizontally) connected by a set of devices (drawn as vertical connections between lines). The N lines are obviously a set of arc-disjoint paths. In fact any set of N arc-disjoint paths in a network T yields a *line-representation* of T. As an example, in Figure 2, two different line representations are given for the network of CEs of Figure 1d. As Figure 2 illustrates, two representations of the same network may look rather different from each other.

This fact motivates a close examination of some notions of equivalence between networks.

Two networks T_1 and T_2 are said to be *functionally equivalent* $(T_1 \approx T_2)$ if, for a suitable ordering of their respective input and output terminals, they realize the same function.

Two networks T_1 and T_2 are said to be *structurally equivalent* $(T_1 \equiv T_2)$ if they are *isomorphic* as graphs.

Two line representations of networks are said to be *line equivalent* if they can be obtained from one another by a reordering of the lines.

Line equivalence implies structural equivalence, which in turn implies functional equivalence. The various notions of equivalence can be reformulated with the help of a very simple kind of network called a *router*.

A *router* R^τ, with $\tau = (\tau(0),\ldots,\tau(N-1))$ a permutation of $(0,\ldots,N-1)$, is a network of size N, with no device nodes, in which the $\tau(i)$-th input node is directly connected to the i-th output node. Thus, the output corresponding to input x is $y = xR^\tau = x\tau$.

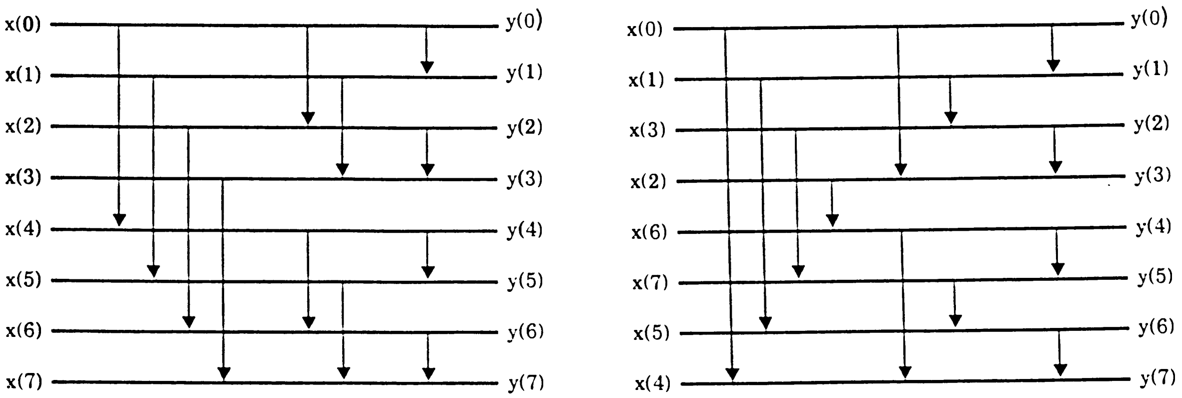


Figure 2. Two line representations for the network of CE in Figure 1d. Corresponding terminals have been labeled in the same way.

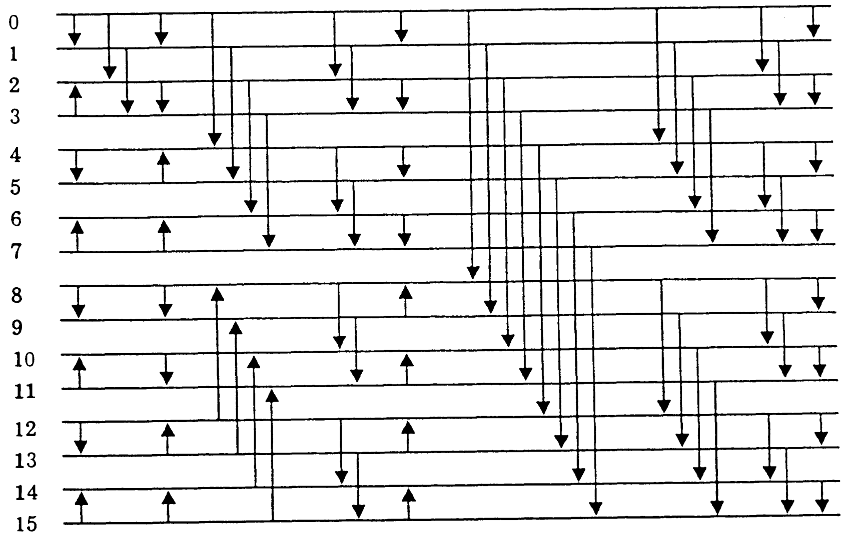
We say that T_1 and T_2 are *equal* ($T_1 = T_2$) if they are structurally equivalent and their terminals are numbered so that, for all x, $xT_1 = xT_2$. Then, T_1 and T_2 are functionally equivalent *iff* there exist permutations α and β such that , for all x, $xT_1 = xR^{\alpha}T_2R^{\beta}$. T_1 and T_2 are structurally equivalent *iff* there exist permutations α and β such that $T_1 = R^{\alpha}T_2R^{\beta}$. T_1 and T_2 are line equivalent if there exists a permutation α such that $T_1 = R^{\alpha}T_2R^{\alpha^{-1}}$ (line i of T_2 corresponds to line $\alpha(i)$ of T_1).

It is interesting to distinguish between properties of networks and properties of their representations. For example , in [Kn 73, p. 236], a network of CEs is defined to be *standard* if all the comparators are directed from the lowest numbered to the highest numbered of the two lines they connect. Clearly, it is not the network itself to be standard, but a given line representation of it, with a specified ordering of the lines. Some line representations can not be standardized regardless of the ordering of the lines, for example, because there are two opposite CEs between the same pair of lines. However, any network of CEs T can be represented in standard form, and in more than one way. Indeed, it is sufficient to choose as lines the paths traversed by the elements of an arbitrary input, ordered according to increasing values of the elements. (The input elements must be distinct.) Figure 3a shows the typical non-standard line representation of a 16-line bitonic sorter [Kn 73, p. 237]. Figure 3b shows a standard representation of the same network, where the lines have been chosen as the paths traversed in the network of Figure 3a when element i is input at line i ($i = 0,...,N-1$).

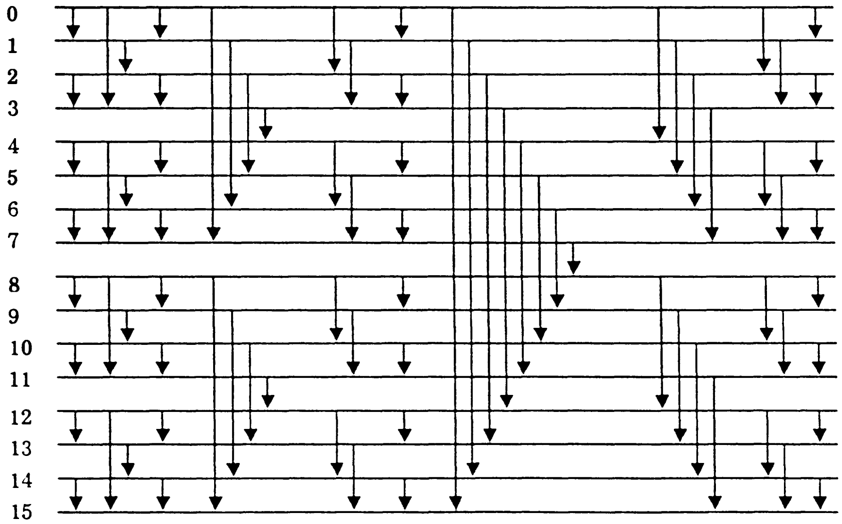
In the following sections, when defining and studying various networks, it will be convenient to use line representations. The observations of this section should help avoiding confusion between properties of the networks and properties of their representations.

3. The omega network

We are ready to define the omega network OM_n on $N = 2^n$ lines, and to discuss some of its important properties. (For alternative definitions and their equivalence see [WF 81].)



(a) the "classical", nonstandard form



(b) a standard form

Figure 3. Two representations of the bitonic sorter.

Definition 1. Let $D_n (n \geq 1)$ be a one-stage network on $N=2^n$ lines with a SW between the pair $(i, i+N/2)$, for $i=0,...,N/2-1$. Then, OM_n is recursively defined as $OM_n = D_n(OM_{n-1} \bigcirc OM_{n-1})$, for $n>1$, and $OM_1 = D_1$. (See Figure 4.)

Definition 2. Let $\Delta_n (n \geq 1)$ be the set of permutation admissible by D_n. Then, Ω_n is recursively defined as

$$\Omega_n = \{(\pi_1 \bigcirc \pi_2)\delta \mid \pi_1, \pi_2 \in \Omega_{n-1}, \delta \in \Delta_n\}$$

for $n > 1$, with Ω_1 containing both permutations of (0,1).

Ω_n is the class of permutations admissible by OM_n. The class Ω_n was introduced by [La 75] who gave a characterization of the admissible permutations, later expressed more formally in [Pe 77]. Another characterization is given in [Pr 83], where the results are actually formulated for the class of permutations sorted by the bitonic merger. How we shall see below, the latter class equals Ω_n.

It is convenient to give a name to some frequently used permutations.

Definition 3. Let $N = 2^n$. (a) We let $\iota_n = (\iota_n(0),...,\iota_n(N-1)) = (0,...,N-1)$ be the *identity* permutation. (b) We let $\sigma_n = (\sigma_n(0),...,\sigma_n(N-1)) = (N/2,...,N-1,\ 0,...,N/2-1)$ be the *swap* permutation, which is realized by D_n when all the SWs are in state 1. (c) We let $\rho_n = (\rho_n(0),...,\rho_n(N-1))$ be the *bit-reversal* permutation such that the binary representation of $\rho_n(i)$ is $i_0 i_1 ... i_{n-1}$, where $i_{n-1} ... i_1 i_0$ is the binary representation of i.

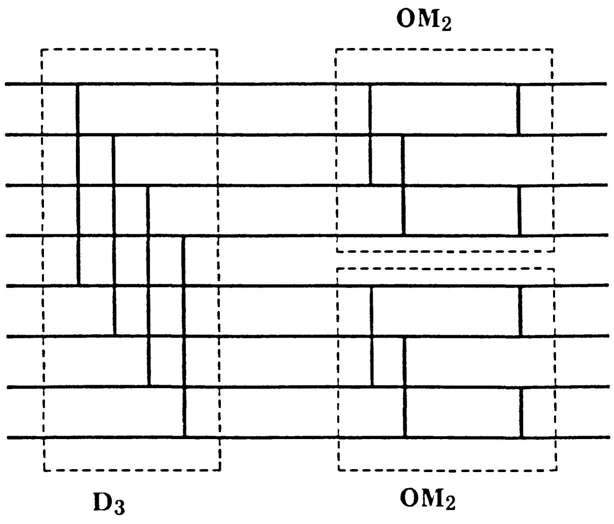


Figure 4. Recursive definition of OM_3.

The following two classes of permutations play a central role in the study of merging networks of the next section.

Definition 4. The class Λ_n of the admissible *lower triangular* permutations is recursively defined as

$$\Lambda_n = \{(\alpha \bigcirc \alpha)\delta \,|\, \alpha \in \Lambda_{n-1},\, \delta \in \Delta_n\}$$

for $n > 1$, with Λ_1 containing both permutations of (0,1).

Thus, a permutation is in Λ_n if and only if it can be passed by OM_n in a configuration such that: (i) the SW's of the first stage are set arbitrarily; (ii) the two copies of OM_{n-1} in which OM_n is recursively decomposed have the same configuration and pass a permutation in Λ_{n-1}. (See Figure 5.)

Definition 5. The class Υ_n of the admissible *upper triangular* permutations is recursively

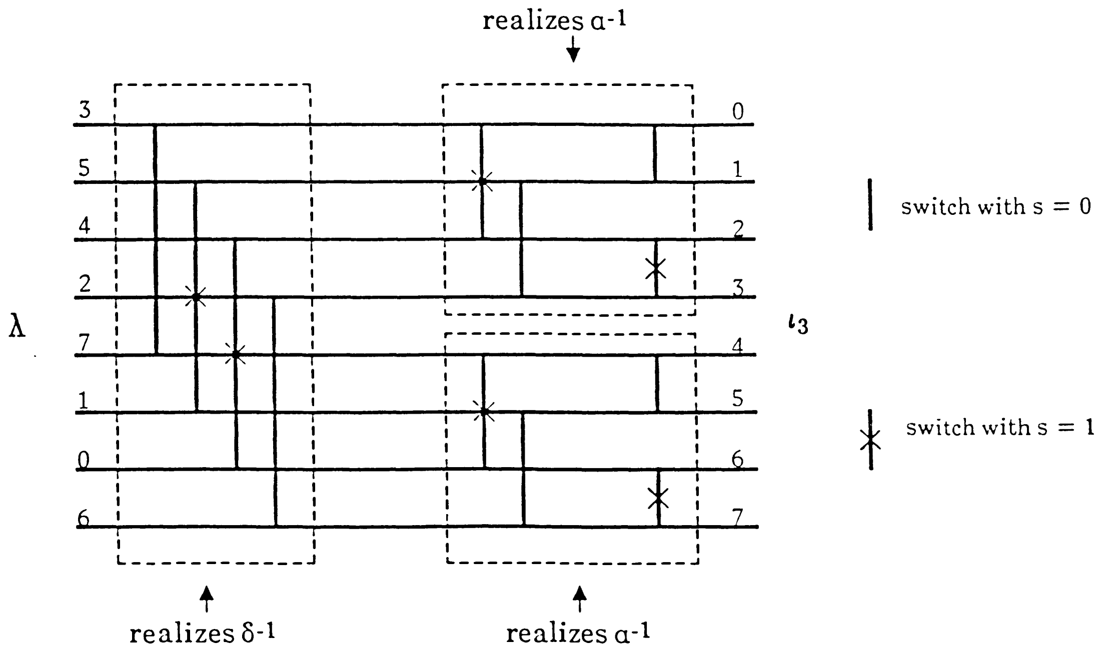


Figure 5. An admissible lower-triangular permutation $\lambda = (3,5,4,2,7,1,0,6) \in \Lambda_3$.

defined as

$$\Upsilon_n = \{(\alpha \bigcirc \beta)\delta \mid \alpha, \beta \in \Upsilon_{n-1}, \delta \in \{\iota_n, \sigma_n\}\},$$

for $n > 1$, with Υ_1 containing both permutations of (0,1).

Thus, a permutation is in Υ_n if and only if it can be passed by OM_n in a configuration such that: (i) the SWs of the first stage are all set in the same state; (ii) each of the two copies of OM_{n-1} in which OM_n is recursively decomposed passes a (possibly different) permutation in Υ_{n-1}. (See Figure 6.)

The classes Λ_n and Υ_n, originally introduced in [Pe 77], were shown to contain respectively the right and the left invariant permutations of Ω_n by [St 83]. The recursive definitions we have adopted for Λ_n and Υ_n are actually different from the algebraic ones used by [Pe 77] and [St 83]. Below, we state and proof the invariant properties of Λ_n and Υ_n according to our definitions. This indirectly proves the equivalence between our definitions and those of Pease and Steinberg. We begin with a lemma, whose proof is simple and hence omitted.

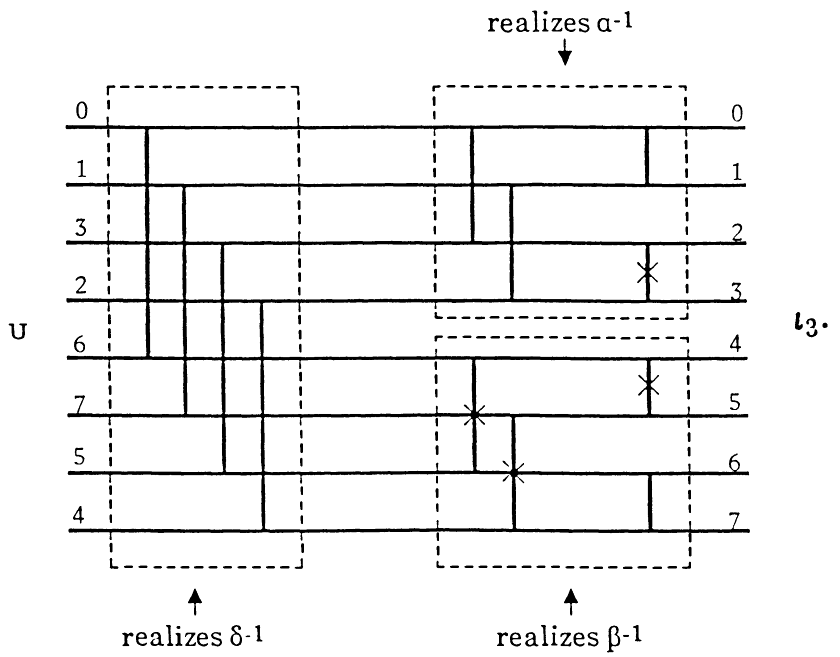


Figure 6. An admissible upper-triangular permutation $\upsilon = (0,1,3,2,6,7,5,4) \in \Upsilon_3$.

Lemma 1. Let α be a permutation of $(0,...,N/2-1)$ and let $\delta \in \Delta_n$. Then, $\delta(\alpha \bigcirc \alpha) = (\alpha \bigcirc \alpha)\delta'$, where $\delta' \in \Delta_n$ is defined as $\delta'(i) = i+\delta(\alpha(i))-\alpha(i)$, and $\delta'(i+N/2) = (\delta'(i)+N/2) \ mod \ (N/2)$, for $i=0,1,...,N/2-1$.

Proposition 1. $\lambda \in \Lambda_n$ if and only if, for all $\pi \in \Omega_n$, $\pi\lambda \in \Omega_n$, that is, λ is a right invariant of Ω_n.

Proof. Proceeding inductively, we can easily verify the proposition for $n=1$, and assume it true for $n-1$.

We prove first that if $\lambda \in \Lambda_n$ and $\pi \in \Omega_n$, then $\pi\lambda \in \Omega_n$. From Definition 4, $\lambda = (\alpha \bigcirc \alpha)\delta$, with $\alpha \in \Lambda_{n-1}$ and $\delta \in \Delta_n$. From Definition 2, $\pi = (\pi_1 \bigcirc \pi_2)\beta$, with $\pi_1, \pi_2 \in \Omega_{n-1}$ and $\beta \in \Delta_n$. Thus,

$$\begin{aligned} \pi\lambda &= (\pi_1 \bigcirc \pi_2)\beta(\alpha \bigcirc \alpha)\delta \\ &= (\pi_1 \bigcirc \pi_2)(\alpha \bigcirc \alpha)\beta'\delta \\ &= (\pi_1\alpha \bigcirc \pi_2\alpha)(\beta'\delta) \end{aligned}$$

where, in the second line, we have used Lemma 1, so that $\beta' \in \Delta_n$. By the inductive hypothesis, $\pi_1\alpha$ and $\pi_2\alpha$ are in Ω_{n-1}. The product $\beta'\delta$ is trivially in Δ_n. Therefore, by Definition 2, $\pi\lambda \in \Omega_n$.

We now prove that if $\pi\lambda \in \Omega_n$ for all $\pi \in \Omega_n$, then $\lambda \in \Lambda_n$. Considering $\pi = \iota_n$, we see that $\lambda \in \Omega_n$, hence, $\lambda = (\alpha_1 \bigcirc \alpha_2)\delta$ with $\alpha_1, \alpha_2 \in \Omega_{n-1}$ and $\delta \in \Delta_n$. Considering $\pi = \pi_1 \bigcirc \pi_2$, we also see that $\alpha_1, \alpha_2 \in \Lambda_{n-1}$. It remains to show that $\alpha_1 = \alpha_2$ to conclude, from Definition 4, that $\lambda \in \Lambda_n$. If $\alpha_1 \neq \alpha_2$, let j be such that $\alpha_1(j) \neq \alpha_2(j)$, and let π be the transposition that exchanges $\alpha_2(j)$ and $\alpha_2(j) + N/2$. Consider now the product $\beta = \pi\lambda = \pi(\alpha_1 \bigcirc \alpha_2)\delta$. If $\delta(j) = j$, then $\beta(j) = \alpha_1(j)$ and $\beta(j + N/2) = \alpha_2(j)$, else $(\delta(j) = j+N/2)$ $\beta(j) = \alpha_2(j)$ and $\beta(j+N/2) = \alpha_1(j)$. In either case $0 \leq \beta(j), \beta(j+N/2) < N/2$, which is easily seen to be impossible for $\beta \in \Omega_n$. □

Proposition 2. $\upsilon \in \Upsilon_n$ if and only if, for all $\pi \in \Omega_n$, $\upsilon\pi \in \Omega_n$, that is, υ is a left invariant of Ω_n.

Proof. Proceeding inductively, we can easily verify the proposition for $n=1$, and assume it true for $n-1$.

We prove first that if $\upsilon \in \Upsilon_n$ and $\pi \in \Omega_n$, then $\upsilon\pi \in \Omega_n$. From Definition 5, $\upsilon = (\alpha_1 \bigcirc \alpha_2)\delta$, with $\alpha_1, \alpha_2 \in \Upsilon_{n-1}$ and $\delta \in \{\iota_n, \sigma_n\}$. From Definition 2, $\pi = (\pi_1 \bigcirc \pi_2)\beta$, with $\pi_1, \pi_2 \in \Omega_{n-1}$ and $\beta \in \Delta_n$. If $\delta = \iota_n$, then $\upsilon\pi = (\alpha_1 \bigcirc \alpha_2)(\pi_1 \bigcirc \pi_2)\beta = (\alpha_1\pi_1 \;\; \alpha_2\pi_2)\beta$. If $\delta = \sigma_n$, then $\upsilon\pi = (\alpha_1 \bigcirc \alpha_2)\sigma_n(\pi_1 \bigcirc \pi_2)\beta = (\alpha_1 \bigcirc \alpha_2)(\pi_2 \bigcirc \pi_1)\sigma_n\beta = (\alpha_1\pi_2 \bigcirc \alpha_2\pi_1)(\sigma_n\beta)$. In either case, by the inductive hypothesis and Definition 2, $\upsilon\pi \in \Omega_n$.

We now prove that if $\upsilon\pi \in \Omega_n$ for all $\pi \in \Omega_n$, then $\upsilon \in \Upsilon_n$. Considering $\pi = \iota_n$, we see that $\upsilon \in \Omega_n$, hence $\upsilon = (\alpha_1 \bigcirc \alpha_2)\delta$ with $\alpha_1, \alpha_2 \in \Omega_{n-1}$ and $\delta \in \Delta_n$. If $\delta \notin \{\iota_n, \sigma_n\}$, then there is an $i<N/2-1$ such that either $\delta(i) = i$ and $\delta(i+1) = i+1+N/2$ or $\delta(i) = i+N/2$ and $\delta(i+1) = i+1$. Assume, for example, that the first is the case. Let $\pi = (\pi_1 \bigcirc \pi_2)\beta$ with $\beta = \iota_n$, $\pi_1 = \iota_{n-1}$, and $\pi_2(j) = (j+1)mod(N/2)$ (which can be easily shown to be in Ω_{n-1} [Le 75]), and consider the product $\gamma = \upsilon\pi = (\alpha_1 \bigcirc \alpha_2)\delta(\iota_{n-1} \bigcirc \pi_2)$. Then, $\gamma(i) = \alpha_1(i)$ and $\gamma(i+N/2) = \alpha_1(i+1)$, so that $0 \le \gamma(i), \gamma(i+N/2) < N/2$, showing that $\gamma \notin \Omega_n$. We conclude that $\delta \in \{i_n, \sigma_n\}$. It is then easy to see that α_1 and α_2 are in Υ_{n-1}, and therefore that $\upsilon \in \Upsilon_n$. $\square$

Simple arguments yield the following result on the algebraic structure of Λ_n and Υ_n.

Proposition 3. Λ_n and Υ_n are groups of order $2^{N-1}(N=2^n)$.

4. Merging Networks

The networks of comparators considered in this section are obtained from the line representation of the omega network (according to Definition 1) by replacing each SW with a CE of suitable direction. Since OM_n contains $(N \log N)/2$ SWs, each of which can be replaced by a CE in two ways, there are $N^{N/2}$ such networks of CEs. For example, if all the

SWs of OM_3 (Figure 4) are replaced by CEs with the same orientation, one obtains the network of CEs of Figure 2a, which is the bitonic merger on 8 lines [Ba 68]. Another choice of comparator directions yields the network of Figure 7, which turns out to be line equivalent to the network of Figure 2b. The latter is the balanced merger on 8 lines [DPRS 83].

Below, we show that a large family of comparator networks obtained from OM_n have merging capabilities. While in general different members of this family are not line equivalent, they are all structurally equivalent. This family includes both the bitonic merger [Ba 68] and the balanced merger [DPRS 83], which are therefore isomorphic to each other.

4.1. Isomorphism properties

We begin with a result that allows to parametrize the comparator networks under consideration by permutations in Ω_n.

Proposition 4. For each $\omega \in \Omega_n$, there is a *unique* network B_n^ω obtained by replacing each SW of OM_n with a CE of suitable direction, such that $\iota_n B_n^\omega = \omega^{-1}$.

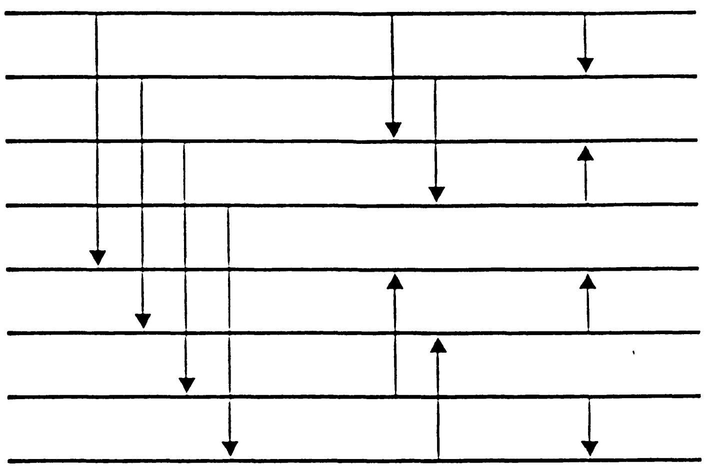


Figure 7. A Network of CEs obtained from OM_3 by replacing each SW with a CE. When redrawn with the lines ordered as (0,1,3,2,6,7,5,4) this network becomes the same as the one in Figure 2b.

Proof. Given $\omega \in \Omega_n$, there is a setting of the switches of OM_n that passes ω (equivalently, realizes ω^{-1}), connecting input line j to output line $\omega^{-1}(j)$. Clearly, with such setting, $\iota_n OM_n = \omega^{-1}$. Let B_n^ω be constructed by replacing each SW by a CE directed so as to have the same output as the SW, when OM_n is set to realize ω^{-1} and the input is ι_n. By construction, B_n^ω satisfies $\iota_n B_n^\omega = \omega^{-1}$.

To prove the uniqueness of B_n^ω, we show that if T is obtained by replacing the SWs of OM_n with CEs, then the direction of the CEs is uniquely determined by $\omega^{-1} = \iota_n T$, and $\omega \in \Omega_n$. We observe that in OM_n, and hence in T, there is a unique path from a given input to a given output. Therefore, the permutation ω^{-1} realized by T on input ι_n uniquely determines the path traversed by each of the input elements. Thus, for each CE in T, the inputs and the outputs are uniquely determined, and so must be the direction. Obviously, ω^{-1} is realizable by OM_n and hence $\omega \in \Omega_n$. □

An important network in the class $\{B_n^\omega : \omega \in \Omega_n\}$ is $B_n^{\iota_n}$, in which all the CEs are directed from the lower-numbered to the higher-numbered line. Indeed, $B_n^{\iota_n}$ is the well-known Batcher's *bitonic merger* [Ba 68]. Whenever convenient, we shall drop the superscript ι_n and denote the bitonic merger by B_n.

In general, for $\omega \in \Omega_n$, B_n^ω is neither structurally nor functionally equivalent to B_n. However, as stated in the following theorem, if ω is admissible upper triangular, then B_n^ω is structurally equivalent to B_n. Typically, even for $\omega \in \Upsilon_n$, B_n^ω is not line- equivalent to B_n, and hence the structural equivalence is not apparent from drawings of the networks. Indeed, the proof of equivalence is not trivial.

Theorem 1. If $\upsilon \in \Upsilon_n$, then $B_n^\upsilon = B_n R^{\upsilon^{-1}}$.

Proof. Proceeding inductively, we can easily verify the theorem for $n=1$, and assume it true for $n-1$. By Definition 5, $\upsilon = (\upsilon_1 \bigcirc \upsilon_2)\delta$, with $\upsilon_1, \upsilon_2 \in \Upsilon_n$, and $\delta \in \{\iota_n, \sigma_n\}$. By the definition of B_n^υ (Proposition 4) it is not difficult to see that $B_n^\upsilon = S^\delta(B_{n-1}^{\upsilon_1} \bigcirc B_{n-1}^{\upsilon_2})$ where, for

$i=0,1,\ldots,N/2-1$,

S^{δ} contains a CE directed from i to $i+N/2$ when $\delta = \iota_n$, and from $i+N/2$ when $\delta=\sigma_n$.

If $\delta = \iota_n$, making use of the inductive assumption, we have that

$$\begin{aligned} B_n^{\upsilon} &= S^{\iota_n} (B_{n-1}R^{\upsilon_1^{-1}} \bigcirc B_{n-1}R^{\upsilon_2^{-1}}) \\ &= S^{\iota_n} (B_{n-1} \bigcirc B_{n-1}) (R^{\upsilon_1^{-1}} \bigcirc R^{\upsilon_2^{-1}}) \\ &= B_n R^{\upsilon^{-1}}. \end{aligned}$$

If $\delta = \sigma_n$, we can write

$$\begin{aligned} B_n^{\upsilon} &= S^{\sigma_n} (B_{n-1}^{\upsilon_1} \bigcirc B_{n-1}^{\upsilon_2}) \\ &= S^{\iota_n} (B_{n-1}^{\upsilon_2} \bigcirc B_{n-1}^{\upsilon_1}) R^{\sigma_n} \ , \end{aligned}$$

and then make use of the inductive assumption to obtain the following relations:

$$\begin{aligned} B_n^{\upsilon} &= S^{\iota_n} (B_{n-1}R^{\upsilon_2^{-1}} \bigcirc B_{n-1}R^{\upsilon_1^{-1}}) R^{\sigma_n} \\ &= S^{\iota_n} (B_{n-1} \bigcirc B_{n-1})(R^{\upsilon_2^{-1}} \bigcirc R^{\upsilon_1^{-1}}) R^{\sigma_n} \\ &= B_n R^{(\upsilon_2^{-1} \bigcirc \upsilon_1^{-1})\sigma_n} \\ &= B_n R^{\upsilon^{-1}}, \end{aligned}$$

where we have exploited the fact that, since $\upsilon=(\upsilon_1 \bigcirc \upsilon_2)\sigma_n$, $n^{-1} = \sigma_n^{-1}(\upsilon_1^{-1} \bigcirc \upsilon_2^{-1}) = \sigma_n(\upsilon_1^{-1} \bigcirc \upsilon_2^{-1}) = (\upsilon_2^{-1} \bigcirc \upsilon_1^{-1})\sigma_n$. $\Box$

Example. Let $\upsilon=(0,1,3,2,6,7,5,4) \in \Upsilon_3$ (see Figure 6). Then, B_3^{υ} is the network shown in Figure 7. According to Theorem 1, this network is equivalent of the series of B_3 (shown in Figure 2a) and the router $R^{\upsilon^{-1}}$, where $\upsilon^{-1}=(0,1,3,2,7,6,4,5)$.

Since Theorem 1 shows that , for $\upsilon \in \Upsilon_n$, B_n^{υ} is closely related to B_n, it is worthwhile to study some properties of B_n in detail. We begin by considering the class of permutations sorted by the bitonic merger, a characterization of which has been given in [Pr 83], and we show that these permutations are exactly those admissible by OM_n.

Proposition 5. A permutation π is sorted by B_n if and only if $\pi \in \Omega_n$.

Proof. Proceeding inductively, we can easily verify the proposition for $n=1$, and assume it true for $n-1$. If $\pi \in \Omega_n$ then, by Definition 2, $\pi = (\pi_1 \bigcirc \pi_2)\delta$, with $\pi_1, \pi_2 \in \Omega_{n-1}$, and $\delta \in \Delta_n$. With the notation used in the proof of Theorem 1, $B_n = S^{\iota_n}\,(B_{n-1} \bigcirc B_{n-1})$. It is easy to see that S^{ι_n} sorts any permutation in Δ_n ($\delta S^{\iota_n} = \iota_n$, for $\delta \in \Delta_n$), hence

$$\begin{aligned}
\pi B_n &= (\pi_1 \bigcirc \pi_2)\ \delta S^{\iota_n}\,(B_{n-1} \bigcirc B_{n-1}) \\
&= (\pi_1 \bigcirc \pi_2)(B_{n-1} \bigcirc B_{n-1}) \\
&= (\pi_1 B_{n-1} \bigcirc \pi_2 B_{n-1}) \\
&= (\iota_{n-1} \bigcirc \iota_{n-1}) \\
&= \iota_n,
\end{aligned}$$

where we have made use of the inductive assumption. Thus, if $\pi \in \Omega_n$, then π is sorted by B_n.

A simple counting argument shows that a network with c comparators sorts at most 2^c different permutations. Since B_n has $(NlogN)/2$ comparators ($N=2^n$), and $|\Omega_n| = 2^{(NlogN)/2}$, then π is sorted by B_n only if $\pi \in \Omega_n$. $\square$

As a simple corollary of Theorem 1 and Proposition 5, we have the following result.

Proposition 6. If $\upsilon \in \Upsilon_n$ and $\pi \in \Omega_n$, then $\pi B_n^{\upsilon} = \upsilon^{-1}$.

4.2. Merging properties

Having established some basic properties of the networks B_n^{υ} , $\upsilon \in \Upsilon_n$, we focus now on their merging capabilities. We begin by recalling the notion of bitonic permutation [Ba 68].

Definition 6. A permutation π is *bitonic* if there exists a cyclic shift of π consisting of an increasing sequence followed by a decreasing one.

Proposition 7. Bitonic permutations belong to Ω_n.

Proof. It follows from Proposition 4 and Batcher's proof that all bitonic permutations are sorted by B_n. □

If a network sorts all bitonic permutations, it can be used as a merger. In fact, it is sufficient to form the input by concatenating, in opposite order, the two sorted sequences to be merged. We are then ready to state our main result on merging.

Theorem 2. Let us define the class of comparator networks

$$F_n \triangleq \{R^{\lambda} B_n^{\upsilon} R^{\upsilon} \mid \lambda \in \Lambda_n, \upsilon \in \Upsilon_n\}.$$

Then, each member of F_n sorts all bitonic permutations, and hence is a merger.

Proof. Let π be bitonic. From Proposition 7, $\pi \in \Omega_n$. Let $T = R^{\lambda} B_n^{\upsilon} R^{\upsilon} \in F_n$. Then, $\pi T = \pi R^{\lambda} B_n^{\upsilon} R^{\upsilon} = (\pi\lambda) B_n^{\upsilon} R^{\upsilon}$. Since $\lambda \in \Lambda_n$, by Proposition 1, $(\pi\lambda) \in \Omega_n$ so that, by Proposition 6, $(\pi\lambda) B_n^{\upsilon} R^{\upsilon} = \upsilon^{-1} R^{\upsilon} = \iota_n$. In conclusion, $\pi T = \iota_n$, that is, π is sorted by T. □

It is interesting to consider the consequences of Theorem 2 for a fixed υ, say, $\upsilon = \iota_n$. We see that, for $\lambda \in \Lambda_n$, $R^{\lambda} B_n$ is a merger. For $\lambda = \iota_n$, this is the classical result of [Ba 68]. But, for arbitrary λ, the theorem says something new, namely that there are many ways, precisely $|\Lambda_n| = 2^{N-1}$ (Proposition 3), to input a bitonic sequence in a bitonic merger and obtain a sorted output.

The definition of the network B_n^{ω} implicit in Proposition 4 suggests a line representation similar to that of OM_n (shown in Figure 4 for $n=3$). This representation is non-standard, except when ω is the identity or the reversal permutation ($\omega(i) = N-1-i$), since the CEs do not have all the same direction. However for $\upsilon \in \Upsilon_n$, B_n^{υ} can be always redrawn in standard form by changing the order of the lines, as shown by the following result.

Proposition 8. If $\upsilon \in \Upsilon_n$, then $\hat{B}_n^{\upsilon} \triangleq R^{\upsilon^{-1}} B_n^{\upsilon} R^{\upsilon}$ is in standard form.

Proof. Since Υ_n is a group, $\upsilon^{-1} \in \Upsilon_n$ and therefore $\upsilon^{-1} \in \Omega_n$. From Proposition 6, $\upsilon^{-1} B_n^{\upsilon} = \upsilon^{-1}$

so that B_n^υ realizes the identity permutation on input υ^{-1}. Since in OM_n, and hence in B_n^υ, there is a unique path between a given input terminal and a given output terminal, B_n^υ can realize the identity only if no CE exchanges, and element $\upsilon^{-1}(i)$ traces line i from input to output. We observe now that $\hat{B}_n^\upsilon$ is line equivalent to B_n^υ and is obtained from the latter by renaming line i as line $\upsilon^{-1}(i)$. Thus, when ι_n is input to $\hat{B}_n^\upsilon$, element i traces line i. Then, a CE between lines i and j must compare elements i and j without exchanging them, which implies that the CE is directed from $min(i,j)$ to $max(i,j)$. □

For $\upsilon \in \Upsilon_n$, B_n^υ can be readily obtained as indicated by the following proposition. The proof is simple and hence omitted.

Proposition 9. Let $\upsilon \in \Upsilon_n$, and let the SWs of OM_n be set to pass υ. Then, B_n^υ is obtained by replacing each SW of OM_n by a CE whose direction is from lower to higher line number if the SW is straight ($s=0$), and in the opposite direction otherwise.

Example. Consider the permutation $\upsilon \in \Upsilon_3$ passed by OM_3 as shown in Figure 6. Following the procedure outlined in Proposition 9, one obtains the network B_3^υ shown in Figure 7. To obtain the standard form B_3^n, one only needs to rearrange the lines according to υ^{-1} (Proposition 8). The result is the network of Figure 2b.

4.3. Balanced merger

This section is devoted to the study of the balanced merger of [DPRS 83], and the proof that this network is one of the form $B_n^{\tau_n}$, where τ_n is a suitable upper-triangular permutation. First, we need to define some permutations and to establish some of their properties.

Definition 7. Let $N=2^n$. We let $\varphi_n = (\varphi_n(0),...,\varphi_n(N-1))$ be the *reversal* permutation such that $\varphi_n(i) = N-1-i$.

The following properties of φ_n are easy to prove.

Proposition 10. $\varphi_n \in \Lambda_n \cap \Upsilon_n$. For $n>1$, $\varphi_n = (\varphi_{n-1} \bigcirc \varphi_{n-1})\sigma_n$.

Definition 8. Let $N = 2^n$. Permutation τ_n is recursively defined as $\tau_n = (\tau_{n-1} \bigcirc \tau_{n-1}\varphi_{n-1})$, for $n>1$, with $\tau_1 = \iota_1$.

Proposition 11. $\tau_n \in \Upsilon_n$.

Proof. For $n=1$, $\tau_1 \in \Upsilon_1$ is obvious. Assuming inductively that $\tau_{n-1} \in \Upsilon_{n-1}$, we have that $(\tau_{n-1}\varphi_{n-1}) \in \Upsilon_n$ since, by Proposition 10, $\varphi_{n-1} \in \Upsilon_{n-1}$, and Υ_{n-1} is a group. Then, by Definition 5, $\tau_n = (\tau_{n-1} \bigcirc \tau_{n-1}\varphi_{n-1})\iota_n \in \Upsilon_n$. □

Proposition 12. The binary representation $i_{n-1}i_{n-1}...i_0$ of i and the binary representation $j_{n-1}\, j_{n-2}..j_0$ of $j = \tau_n(i)$ are related as follows:

$$j_k = i_k \oplus i_{k+1}, \tag{1}$$

$$i_k = j_{n-1} \oplus j_{n-2} \oplus ... \oplus j_k, \tag{2}$$

for $k=0,1,...,n-1$, and $i_n \triangleq 0$.

Proof. Proceeding inductively, we can easily verify Equation (1) for $n=1$, and assume it true for $n-1$. For $k=n-1$, it is sufficient to observe that, a simple consequence of Definition 8, $\tau_n(i)<N/2$ if and only if $i<N/2$. Thus, $j_{n-1} = \iota_{n-1}$, as in Equation (1).

For $k<n-1$, we argue as follows. If $i<N/2$, then $j = \tau_n(i) = \tau_{n-1}(i)$, and (1) is satisfied because of the inductive assumption. If $i>N/2$, then $j = \tau_n(i) = \tau_{n-1}(\varphi_{n-1}(i-N/2)) + N/2 = \tau_{n-1}(N-1-i)+N/2$. Since the binary representation of $N-1-i$ is the bitwise complement of the binary representation of i, Equation (1), which is not altered if both i_k and i_{k+1} are complemented, is again satisfied because of the inductive assumption.

We conclude by observing that (2) is a trivial consequence of (1). □

Definition 9. The network $\hat{B}_n^{\tau_n} \triangleq R^{\tau_n^{-1}} B_n^{\tau_n} R^{\tau_n}$ is called *balanced merger.*

The above definition of balanced merger is equivalent to the original definition given in [DPRS 83], as shown by the next result.

Proposition 13. Let $B_n^{\tau_n}$ be decomposed into the cascade of n stages as $B_n^{\tau_n} = H_{n-1}H_{n-2}...H_0$. Then, in H_r, there is a CE from line h to line l if and only if $h<l$ and the two following relations hold: (i) $h-(h \ mod \ 2^{r+1}) = l-(l \ mod \ 2^{r+1})$, and (ii) $h \ mod \ 2^r + l \ mod \ 2^r = 2^{r+1}-1$.

Proof. From Propositions 4 and 8, $\hat{B}_n^{\tau_n}$ can be obtained from OM_n by remaining line i as line $\tau_n^{-1}(i)$, and then replacing each SW with a CE directed from the lower-numbered to the higher-numbered line. Stage H_r is thus obtained from the corresponding stage of OM_n, in which there is a SW between line i and line C_ri, where C_ri is the integer whose binary representation differs from that of i exactly in the coefficient of 2^r.

Therefore, if lines h and l are connected in H_r, then for some i, $h = \tau_n^{-1}(i)$, and $l = \tau_n^{-1}(C_ri) = \tau_n^{-1}C_r\tau_n(h)$. Let $h_{n-1}h_{n-2}...h_0$ be the binary representation of h. Simple manipulations show that applying Equation (1) to this representation, complementing the rth bit, and then applying Equation (2) yields for l the binary representation $h_{n-1}...h_{r+1}\overline{h}_r...\overline{h}_0$, where $\overline{h}_i$ is the complement of h_i. The relation between the binary representations of h and l readily implies (i) and (ii). □

In [DPRS 83] it is shown that the balanced merger merges two sorted sequences of $N/2$ elements each, when one sequence is input on the even-numbered lines and the other on the odd-numbered lines. This input protocol could be shown to correspond to choosing $\lambda = \rho_n\tau_n\rho_n$ in Theorem 2, where ρ_n is the bit-reversal permutation (Definition 3). As remarked above, there are many other input protocols under which the balanced merger behaves as a merger.

5. Sorting Networks

Given a merging network, one can easily construct a sorting network using standard merge-sort techniques. However, the merging networks described in the previous section can be used as building blocks of sorting networks according to a scheme different from merge-sort. The first result of this type was obtained in [DPRS 83], where the series $(\hat{B}_n^{\tau_n})^n$ of n balanced mergers was shown to be a sorter, and called the *balanced sorter.*

With the techniques of the preceeding sections, this result can be readily generalized as follows.

Theorem 3. If $\upsilon \in \Upsilon_n$, then the network $(R^{\tau_n^{-1}\upsilon}\,\hat{B}_n^{\upsilon})^n$ is a sorter.

Proof. From [DPRS 83], we have that $(\hat{B}^{\tau_n})^n$ is a sorter. $\hat{B}^{\tau_n}$ can be rewritten as follows:

$$\begin{aligned}
\hat{B}^{\tau_n} &= R^{\tau_n^{-1}}\, B_n^{\tau_n}\, R^{\tau_n} \\
&= R^{\tau_n^{-1}}\, B_n \\
&= R^{\tau_n^{-1}}\, B_n^{\upsilon}\, R^{\upsilon} \\
&= R^{\tau_n^{-1}\upsilon}\, R^{\upsilon-1}\, B_n^{\upsilon}\, R^{\upsilon} \\
&= R^{\tau_n^{-1}\upsilon}\, \hat{B}_n^{\upsilon},
\end{aligned}$$

where we have used Definition 9 in the first line, Theorem 1 in the second and in the third line, and the definition of B_n^{υ} (given in Proposition 8) in the last line. $\square$

We observe that, for $\upsilon = \tau_n$, Theorem 3 specializes to the result of [DPRS 83], and that, for $\upsilon = \iota_n$, it yields the following property of the bitonic merger.

Proposition 14. The network $(R^{\tau_n^{-1}}\, B_n)^n$ is a sorter.

6. Conclusions

We have considered the class $\{B_n^{\omega} : \omega \in \Omega_n\}$ of the networks of comparators that result from replacing each switch of the omega network with a comparator of arbitrary direction. For the subclass $\{B_n^{\omega} : \omega \in \Upsilon_n\}$ we have established isomorphism, merging, and sorting results. Among other things, these results clarify the relationship between the bitonic and the balanced merger. An interesting subject for further investigation are the properties of B_n^{ω} when ω is not upper triangular.

The approach of analyzing classes of networks that differ only in the direction of their comparators in terms of the properties of their common underlying switching network has proven fruitful. It may be worthwhile to give a general formulation to this approach and to adopt it for other problems.

References

[Ba 68] K.E. Batcher, "Sorting networks and their applications," in *Proceedings AFIPS Spring Joint Comp. Conf.*, Vol. 32. pp. 307-314, April 1968.

[DPRS 83] M. Dowd, Y. Perl, L. Rudolph, and M. Saks, "The balanced sorting network," in *Proceedings ACM Principles of Distributed. Computing*, pp. 161-172, August 1983.

[HS 82] Z. Ahong, R. Sedgewick, "Notes on merging networks," *Proceedings of the 14th Annual ACM Symposium on Theory of Computing*, pp. 296-302, San Francisco, CA, May 1982.

[Kn 73] D.E. Knuth, "The art of computer programming," *Vol. 3: Sorting and Searching*, Addison Wesley, Reading, MA, 1973.

[La 75] D. Lawrie, "Access and alignment in an array processor," *IEEE Transactions on Computers*, Vol. C-24, No. 12, pp. 1145-1155, December 1975.

[Pe 77] M.C. Pease, "The indirect binary n-cube microprocessor array," *IEEE Transaction on Computers*, Vol. C-26, No. 5, pp. 458-473, May 1977.

[Pr 83] Y. Perl, "The bitonic and odd-even networks are more than merging," *The State University of New Jersey Rutgers* DCS-TR-123, February 1983.

[R 85] L. Rudolph, "A robust sorting network," *IEEE Transactions on Computers*, Vol. C-34, No. 4, pp.326-335, April 1985.

[St 83] D. Steinberg, "Invariant properties of the shuffle-exchange and a simplified cost-effective version of the omega network," *IEEE Transactions on Computers*, Vol. C-32, No. 5, pp. 444-450, May 1983.

[WF 81] C. Wu and T. Ferg, "The universality of the shuffle-exchange network," *IEEE Transactions on Computers*, Vol. C-30, No. 5, pp. 324-332, May 1981.