

Optimising designs by transposition

Wayne Luk*

The purpose of this paper is fourfold: first, to describe some observations about how array-based designs can be optimised by transposition – a method of rearranging components and their interconnections; second, to provide concise parametric representations of such designs; third, to present simple equations that correspond to correctness-preserving transformations of these parametric representations; and finally, to suggest quantitative measures of design trade-offs involved in this kind of transformation.

Motivation

Consider the two designs in Figure 1. Notice that despite the structural differences of the circuits, the relative positions of the external connections are the same. This is achieved by the interleaving wires at the top and bottom of Design B. Transposition is the name for this kind of interleaving and also for the subsequent structure of components, since the rearrangement is similar to transposing matrices. It is clear that the functional behaviour (but not timing characteristics) of the two designs is indistinguishable; we shall show later how this situation can be expressed mathematically as a correctness-preserving transformation.

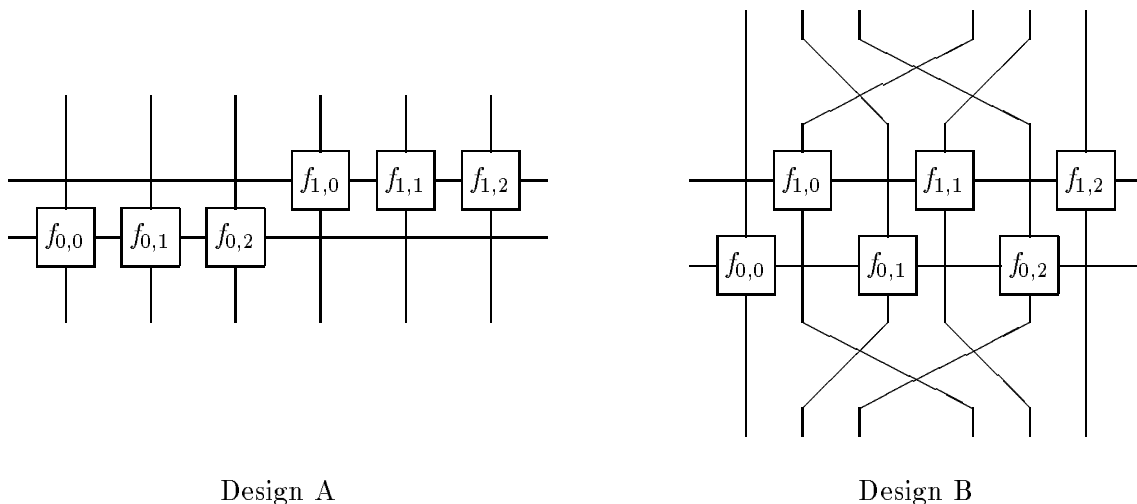


Figure 1: An example of transposing designs.

*Programming Research Group, Oxford University Computing Laboratory, 11 Keble Road, Oxford, England OX1 3QD; wayne@prg.oxford.ac.uk

Why are these designs interesting? In our experience Design A is a generic pattern that arises frequently in the development of circuits. For instance in developing an arithmetic and logic unit, it is more convenient to specify separately the arithmetic and logic operations: $f_{0,0}, f_{0,1}$ and $f_{0,2}$ may represent full adders that process interleaved representations of integers, and $f_{1,0}, f_{1,1}$ and $f_{1,2}$ may perform a specific logic operation, depending on the horizontal input, on their vertical inputs. Figure 2(a) shows a typical configuration in which the outputs of the full adders and the logic circuits are interleaved – transposed – before feeding into an array of multiplexers that select either output. It is, however, more advantageous to perform the transposition at the outset using the transformation described in Figure 1, since we would then be able to form a single array of composite cells each consisting of a full adder, a logic circuit and a multiplexer as shown in Figure 2(b). It is the aim of the circuit designer to use as few types of cell as possible in order to reduce the design and validation effort, because attention can be focused on optimising and testing the basic units which can then be replicated with confidence.

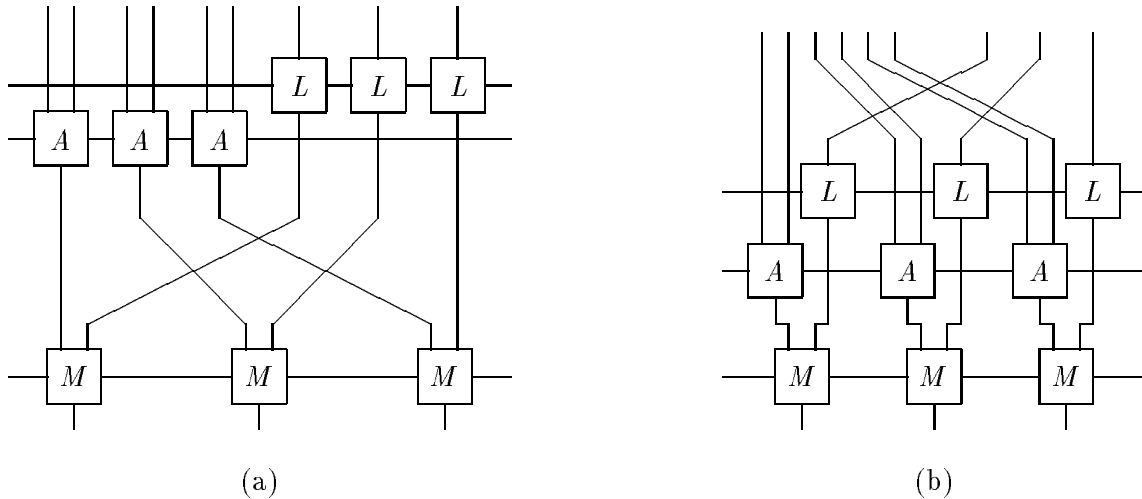


Figure 2: Two implementations of an ALU. A , L and M correspond respectively to a full adder, a logic circuit, and a multiplexer.

Another situation in which Design B in Figure 1 offers an improvement over Design A is as follows. Suppose that every component in Design A is latched. However, instead of depending on the propagation delay of the components, the clock speed of the system may be restricted by the propagation delay associated with the two long horizontal wires. In Design B the components are more evenly distributed and the problem of propagation delay is transferred to the wires implementing the transpositions. We shall see later the circumstances in which these peripheral transpositions can be eliminated to produce an efficient layout.

A further benefit of transposing components can be derived from altering the aspect ratio of a design. Given that one can overlay a wire on a component and that the connections of a component can be shifted as long as their direction and order are maintained, the configuration in Figure 3(a) can be implemented as described in Figure 3(b) or Figure 3(c). Hence if every component in Figure 1 has the same height h and the same width w , and

ignoring the contribution of the peripheral interleaving wires, then Design B may have an aspect ratio of either $2h \times 3w$ or $h \times 6w$. Design A is not as flexible; the same assumptions will result in an aspect ratio of $h \times 6w$.

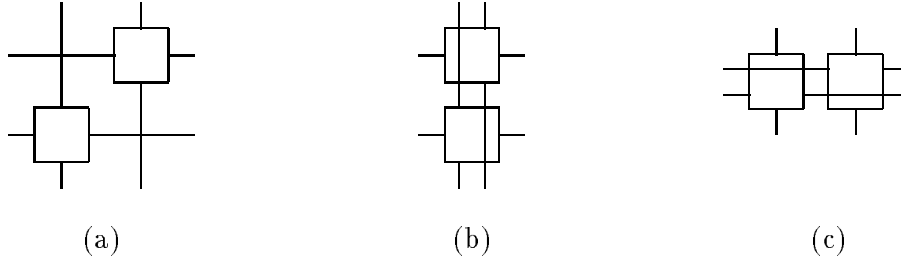


Figure 3: Implementing transposition.

So much for motivation. The rest of the paper will describe a framework for expressing transposition parametrically, and will present a catalogue of such transformations together with quantitative measures of the trade-offs involved.

Notation

The formalism we use is based on Sheeran’s relational framework and the author’s heterogeneous combinators. The background and the details of this approach have been described elsewhere (see [2], [3]), and only the definitions and concepts relevant to our discussion will be introduced here.

A design will be described by a binary relation of the form $x f y \stackrel{\text{def}}{=} \mathcal{P}(x, y)$ where x, y represent the interface signals and belong to $\text{dom}(f)$ and $\text{rng}(f)$ (domain and range of f) respectively, and \mathcal{P} is a predicate describing the intended behaviour. For example, an inverter can be specified as $x \text{ inv } y \stackrel{\text{def}}{=} x = -y$.

The converse f^{-1} of a relation f is defined by $x (f^{-1}) y \stackrel{\text{def}}{=} y f x$, and the identity relation is given by $x \text{ id } y \stackrel{\text{def}}{=} x = y$. If f is a function, then $f x$ represents the value of f for the argument x .

Objects in our notation are either atoms (such as numbers or relations) or tuples of objects: for instance the object $\langle 0, \langle 1, 2 \rangle \rangle$ is a 2-tuple containing the number 0 and the tuple $\langle 1, 2 \rangle$. A tuple is an ordered collection of elements, with the empty tuple denoted by $\langle \rangle$. Tuples are concatenated by ‘^’ (pronounced ‘append’), so that

$$\langle a \rangle \wedge \langle b, c, d \rangle = \langle a, b, c \rangle \wedge \langle d \rangle = \langle a, b, c, d \rangle.$$

Given that x is a tuple, $\#x$ represents the number of elements in it, and x_i (where $0 \leq i < \#x$) is its i -th element. Let us call a tuple of tuples a two-level tuple, a tuple of tuples of tuples a three-level tuple and so on. If x is a tuple with two or more levels, then $x_{i,j} \stackrel{\text{def}}{=} (x_i)_j$. Some operations on tuples are given below:

$$\begin{aligned} x \pi_{n+1} z &\stackrel{\text{def}}{=} z = x_n && (\text{select } n\text{-th } x, 0 \leq n < \#x), \\ \langle x, y \rangle \text{ swap } z &\stackrel{\text{def}}{=} z = \langle y, x \rangle, && (\text{swap components of 2-tuple}), \end{aligned}$$

$$x \text{ tran } z \stackrel{\text{def}}{=} \forall i, j : 0 \leq i < \#x, 0 \leq j < \#z. \quad (\text{transpose } x, \text{ given that } \\ (\#x_i = \#z) \wedge (z_{j,i} = x_{i,j}) \quad x \text{ is a tuple of tuples}).$$

For example, $\text{tran} \langle \langle 1, 2, 3 \rangle, \langle 4, 5, 6 \rangle \rangle = \langle \langle 1, 4 \rangle, \langle 2, 5 \rangle, \langle 3, 6 \rangle \rangle$. Notice that tran is only defined for rectangular tuples – tuples in which all sub-tuples have the same length.

A rectangular circuit with connections on every side is modelled by a relation that relates 2-tuples, with the components in the domain corresponding to signals for the west and north side and those in the range corresponding to signals for the south and east side. In general, composite signals are represented as tuples with the position of a particular signal corresponding to its relative position, and with its structure – the grouping of signals – reflecting the logical organisation of adjacent signals.

Given a circuit f with connections on all four sides, we can use the generic reverse function recrev which recursively reverses a tuple and all its component tuples

$$\begin{aligned} \text{recrev } x &\stackrel{\text{def}}{=} x \text{ if } x = \langle \rangle \text{ or if } x \text{ is an atom,} \\ \text{recrev } (\langle x \rangle \hat{ } xs) &\stackrel{\text{def}}{=} (\text{recrev } xs) \hat{ } (\text{recrev } x) \text{ otherwise,} \end{aligned}$$

to define $f^\mathcal{V}$ and $f^\mathcal{H}$ which denote respectively the reflection of f in a vertical and in a horizontal axis, and $f^{\mathcal{V}\mathcal{H}}$ which denotes the reflection of f both vertically and horizontally:

$$\begin{aligned} \langle x, y \rangle f^\mathcal{V} \langle x', y' \rangle &\stackrel{\text{def}}{=} \langle y', \text{recrev } y \rangle f \langle \text{recrev } x', x \rangle, \\ \langle x, y \rangle f^\mathcal{H} \langle x', y' \rangle &\stackrel{\text{def}}{=} \langle \text{recrev } x, x' \rangle f \langle y, \text{recrev } y' \rangle, \\ \langle x, y \rangle f^{\mathcal{V}\mathcal{H}} \langle x', y' \rangle &\stackrel{\text{def}}{=} \langle \text{recrev } y', \text{recrev } x' \rangle f \langle \text{recrev } y, \text{recrev } x \rangle. \end{aligned}$$

To deal with sequential circuits an expression is considered as relating a stream (an infinite tuple of data) in its domain to a stream in its range. In most cases, such as in the absence of conditionals, the same algebraic theorems can be applied to expressions representing both combinational and sequential systems [1].

Combinators

Combinators are higher-order functions that capture common patterns of computation as parametrised expressions. These patterns can represent behaviour, in which case the behaviour of a composite device is expressed in terms of the behaviour of its components; or they can represent the spatial organisation of a circuit, in which case they describe the connection of components to form the composite device. The following binary combinators are adapted from Ruby [2] to define various ways that two components can be connected together:

$$\begin{aligned} x (f; g) z &\stackrel{\text{def}}{=} \exists y. (x f y) \wedge (y g z) && \text{(relational composition),} \\ \langle x, y \rangle (f \parallel g) \langle x', y' \rangle &\stackrel{\text{def}}{=} (x f x') \wedge (y g y') && \text{(parallel composition),} \\ \langle a, \langle b, c \rangle \rangle (f \# g) \langle \langle p, q \rangle, r \rangle &\stackrel{\text{def}}{=} \exists s. \langle a, b \rangle f \langle p, s \rangle \wedge \langle s, c \rangle g \langle q, r \rangle && \text{(beside),} \\ \langle \langle a, b \rangle, c \rangle (f \ddagger g) \langle p, \langle q, r \rangle \rangle &\stackrel{\text{def}}{=} \exists s. \langle a, s \rangle f \langle p, q \rangle \wedge \langle b, c \rangle g \langle s, r \rangle && \text{(below).} \end{aligned}$$

Since $(f \ddagger g)^{-1} = f^{-1} \# g^{-1}$, we only need to work out the properties of either *beside* or *below*.

We shall use the abbreviations $\text{fst } f \stackrel{\text{def}}{=} f \parallel id$, $\text{snd } f \stackrel{\text{def}}{=} id \parallel f$, $\text{fsth } f \stackrel{\text{def}}{=} f \# swap$, $\text{sndh } f \stackrel{\text{def}}{=} swap \# f$, $\text{fstv } f \stackrel{\text{def}}{=} f \ddagger swap$, $\text{sndv } f \stackrel{\text{def}}{=} swap \ddagger f$, $f \setminus g \stackrel{\text{def}}{=} g^{-1}; f; g$ and $f \setminus\setminus g \stackrel{\text{def}}{=} f \setminus (g \setminus f)$.

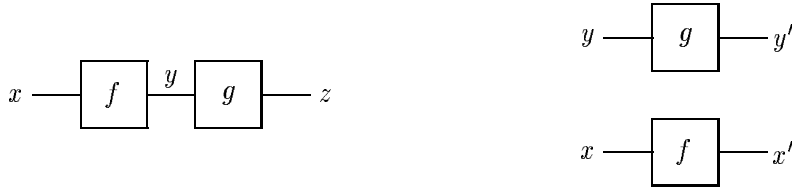


Figure 4: Relational and parallel composition.

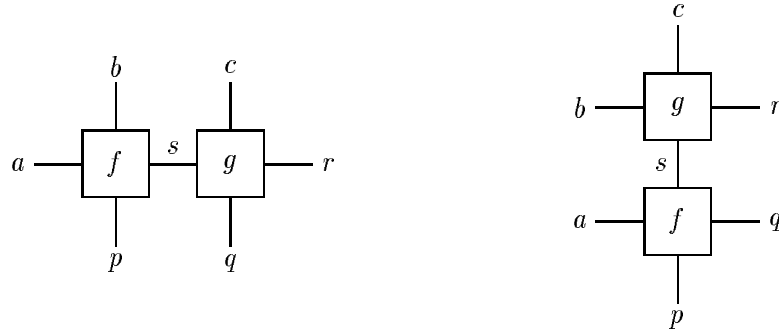


Figure 5: Beside and below.

$g^{-1} \backslash \text{swap}; f; g$. These abbreviations have interesting properties like $\text{fst}f; \text{fst}g = \text{fst}(f; g)$, $\text{fst}h \backslash \text{snd}v f = \text{snd}v(\text{fst}h f)$, $(f \parallel g) \parallel h = f \parallel (g; h)$, $\text{snd}f = (\text{fst}f) \backslash \text{swap}$, and $\text{snd}f^{-1}; g; \text{fst}f = g \parallel (\text{fst}f)$. Note that prefix combinators have a higher precedence than infix ones, and that relational composition has the lowest precedence. The combinators \backslash, \parallel have a lower precedence than all other binary combinators except relational composition.

One can check that in general none of the above binary combinators is commutative, and that only relational composition is associative. For instance, since $\langle \langle x, y \rangle, z \rangle$ and $\langle x, \langle y, z \rangle \rangle$ are distinct tuples, $(f \parallel g) \parallel h \neq f \parallel (g \parallel h)$. We now introduce a class of prefix combinators, called heterogeneous combinators, each of which takes a tuple of components and returns a binary relation corresponding to the composite circuit; and the components that are wired together can be different from one another. For instance, the parallel composition of three distinct components can be expressed as $(\parallel \langle f, g, h \rangle)$ which relates signals of the form $\langle x, y, z \rangle$. Given that $\#f = \#x = \#y = N$, the four common flavours of heterogeneous combinators are described below:

$$a \left(\begin{array}{c} \circ \\ \text{; } f \end{array} \right) b \equiv \exists s. (s_0 = a) \wedge (s_N = b) \wedge \forall i: 0 \leq i < N. s_i f_i s_{i+1} \quad (\text{het. chain}),$$

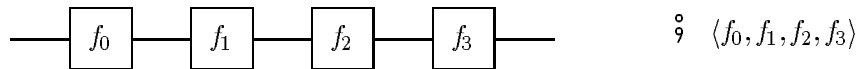


Figure 6: Heterogeneous chain.

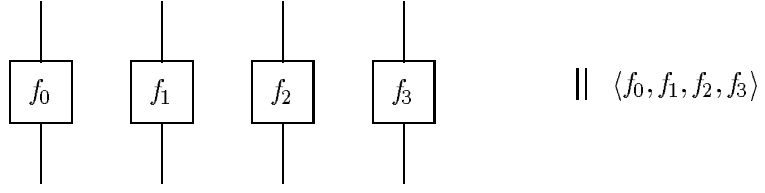


Figure 7: Heterogeneous map.

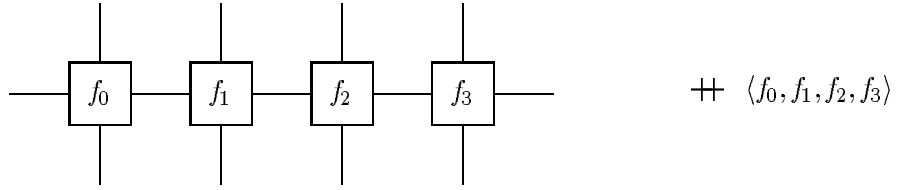


Figure 8: Heterogeneous row.

$$x \left(\parallel f \right) y \equiv \forall i : 0 \leq i < N. x_i f_i y_i \quad (\text{het. map}),$$

$$\langle a, x \rangle \left(\dashv\vdash f \right) \langle y, b \rangle \equiv \exists s. (s_0 = a) \wedge (s_N = b) \wedge \forall i : 0 \leq i < N. \langle s_i, x_i \rangle f_i \langle y_i, s_{i+1} \rangle \quad (\text{het. row}),$$

$$\langle x, a \rangle \left(\ddagger f \right) \langle b, y \rangle \equiv \exists s. (s_0 = b) \wedge (s_N = a) \wedge \forall i : 0 \leq i < N. \langle x_i, s_{i+1} \rangle f_i \langle s_i, y_i \rangle \quad (\text{het. column}).$$

Given $\Psi \in \{ \ddagger, \parallel, \dashv\vdash, \ddagger \}$ and $0 \leq N \leq \#f$, we shall adopt the abbreviation

$$\Psi_{i < N} f_i \stackrel{\text{def}}{=} \Psi \langle f_i \mid 0 \leq i < N \rangle.$$

The following combinator will be used to describe a rectangular grid of heterogeneous components:

$$\ddagger_{i,j < M,N} f_{i,j} \stackrel{\text{def}}{=} \ddagger_{i < M} \left(\dashv\vdash_{j < N} f_{i,j} \right).$$

For many applications the generality of heterogeneous combinators is not required. If f does not depend on i , then the following homogeneous combinators [2] can be used to

describe arrays of identical components.

$$f^N \stackrel{\text{def}}{=} \underset{i < N}{\overset{\circ}{\parallel}} f \quad (\text{chain}), \quad (1)$$

$$\alpha_N f \stackrel{\text{def}}{=} \underset{i < N}{\parallel} f \quad (\text{map}), \quad (2)$$

$$\text{row}_N f \stackrel{\text{def}}{=} \underset{i < N}{\text{++}} f \quad (\text{row}), \quad (3)$$

$$\text{col}_N f \stackrel{\text{def}}{=} \underset{i < N}{\text{+}} f \quad (\text{column}), \quad (4)$$

$$\text{grid}_{M,N} f \stackrel{\text{def}}{=} \text{col}_M (\text{row}_N f) \quad (\text{rectangular grid}). \quad (5)$$

The subscripts associated with these combinators correspond to the size of the arrays, often omitted when they can be deduced from context.

Properties of transposition

Let us look at some properties of *tran* that will prove useful. If we write $\langle x_i | 0 \leq i < \#x \rangle$ as $\langle x_i | i \in I \rangle$ where I is the set of subscripts $\{i | 0 \leq i < \#x\}$, then the previous definition of *tran* can be rewritten as

$$\text{tran} \langle \langle x_{i,j} | j \in J \rangle | i \in I \rangle = \langle \langle x_{i,j} | i \in I \rangle | j \in J \rangle. \quad (6)$$

This shows that transposing a tuple with two or more levels corresponds to permuting the positions of the two subscript generators $i \in I$ and $j \in J$.

From Equation 6 we can conclude that transposing an expression twice leaves it unchanged:

$$\text{tran}^2 = \text{id}_2, \quad (7)$$

since permuting the positions of the subscript generators $i \in I$ and $j \in J$ twice will return the original expression. Here id_n is the identity relation on n -level rectangular tuples. Equation 7 also shows that $\text{tran}^{-1} = \text{tran}$.

Now consider the effect of the function $(\text{tran}; \alpha \text{tran})^3$ on tuples with three or more levels:

$$\begin{aligned} & ((\text{tran}; \alpha \text{tran})^3) \langle \langle \langle x_{i,j,k} | k \in K \rangle | j \in J \rangle | i \in I \rangle \\ &= (\alpha \text{tran}; (\text{tran}; \alpha \text{tran})^2) \langle \langle \langle x_{i,j,k} | k \in K \rangle | i \in I \rangle | j \in J \rangle \\ &= ((\text{tran}; \alpha \text{tran})^2) \langle \langle \langle x_{i,j,k} | i \in I \rangle | k \in K \rangle | j \in J \rangle \\ &= (\alpha \text{tran}; (\text{tran}; \alpha \text{tran})) \langle \langle \langle x_{i,j,k} | i \in I \rangle | j \in J \rangle | k \in K \rangle \\ &= ((\text{tran}; \alpha \text{tran})) \langle \langle \langle x_{i,j,k} | j \in J \rangle | i \in I \rangle | k \in K \rangle \\ &= (\alpha \text{tran}) \langle \langle \langle x_{i,j,k} | j \in J \rangle | k \in K \rangle | i \in I \rangle \\ &= \langle \langle \langle x_{i,j,k} | k \in K \rangle | j \in J \rangle | i \in I \rangle, \end{aligned}$$

so we obtain

$$(\text{tran}; \alpha \text{tran})^3 = \text{id}_3. \quad (8)$$

Notice that at each step of the above derivation, the positions of two of the subscript generators are permuted. One can generalise this result for an M -level tuple where $M > N$ by applying a cyclic permutation to the first $N + 1$ subscript generators, so that

$$(tran; \alpha tran; \alpha^2 tran; \dots; \alpha^{N-1} tran)^{N+1} = id_{N+1}.$$

In other words,

$$\left(\underset{i < N}{\overset{\circ}{\circ}} \alpha^i tran \right)^{N+1} = id_{N+1}.$$

With $N = 1$ we obtain Equation 7, and with $N = 2$ we obtain Equation 8.

Next, we shall present three ways of eliminating $tran$ from an expression. The purpose is to remove unnecessary transpositions which are wasteful of area, or to make the description more uniform so that further optimisations can be applied.

Theorems with preconditions

The first method of eliminating $tran$ involves using theorems with preconditions. For example, with the precondition $h; g = id$, it can be shown that

$$\begin{aligned} \underset{i < N}{\overset{\circ}{\circ}} (g; f_i; h) &= g; \left(\underset{i < N}{\overset{\circ}{\circ}} f_i \right); h, \\ \underset{i < N}{\text{++}} (\text{fst } g; f_i; \text{snd } h) &= \text{fst } g; \left(\underset{i < N}{\text{++}} f_i \right); \text{snd } h, \\ \underset{i, j < M, N}{\text{##}} (g \parallel g; f_{i, j}; h \parallel h) &= \alpha g \parallel \alpha g; \underset{i, j < M, N}{\text{##}} f_{i, j}; \alpha h \parallel \alpha h, \end{aligned}$$

since the g 's and h 's within the array cancel one another. Instances of these theorems include

$$\begin{aligned} (g; f; h)^N &= g; f^N; h, \\ \text{row}(\text{fst } g; f; \text{snd } h) &= \text{fst } g; \text{row } f; \text{snd } h, \\ \text{grid}(g \parallel g; f; h \parallel h) &= \alpha g \parallel \alpha g; \text{grid } f; \alpha h \parallel \alpha h. \end{aligned}$$

Now we can use the result obtained in the last section: the above theorems will be applicable for $g = tran$ and $h = tran$, or $g = (tran; \alpha tran)$ and $h = (tran; \alpha tran)^2$, and so on. In this way transpositions in the wiring between adjacent cells can be eliminated.

Array of wiring cells

The second method of eliminating $tran$ involves developing an array of wiring cells that corresponds to an expression involving $tran$. For instance, given that

$$\begin{aligned} x \text{ fork } y &\stackrel{\text{def}}{=} y = \langle x, x \rangle && \text{(duplication),} \\ \langle x, xs \rangle \text{ apl } y &\stackrel{\text{def}}{=} y = \langle x \rangle \hat{ } xs && \text{(append left),} \\ \langle xs, x \rangle \text{ apr } y &\stackrel{\text{def}}{=} y = xs \hat{ } \langle x \rangle && \text{(append right),} \\ \langle \langle x \rangle \hat{ } xs, y \rangle \text{ shl } u &\stackrel{\text{def}}{=} u = \langle x, xs \hat{ } \langle y \rangle \rangle && \text{(shift left),} \\ \langle x, xs \hat{ } \langle y \rangle \rangle \text{ shr } u &\stackrel{\text{def}}{=} u = \langle \langle x \rangle \hat{ } xs, y \rangle && \text{(shift right),} \end{aligned}$$

it can be shown that

$$\mathit{fork}; (\mathit{apr}^{-1}; \pi_1) \parallel (\mathit{apl}^{-1}; \pi_2); \mathit{tran} = \mathit{apl}^{-1}; \mathbf{row}(\mathit{snd fork}; \mathit{shr}); \pi_1,$$

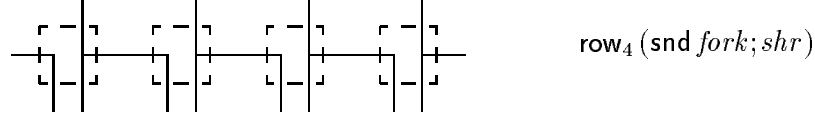


Figure 9: A row of wiring cells.

The right-hand side of the above equation is in the form of a row (Figure 9) and can be optimised further, such as combining its components with those of adjacent arrays. Other examples include

$$\begin{aligned} (f \parallel g \setminus \mathit{tran}) \parallel \mathit{snd swap} &= (\mathit{sndv} f) \# (\mathit{fstv} g), \\ f \parallel g \setminus \mathit{tran} &= (\mathit{fstv} f) \# (\mathit{sndv} g). \end{aligned}$$

Duplication and truncation

A further possibility of eliminating tran arises when signals are being duplicated or discarded. For instance, given the function dup_N which duplicates its argument N times,

$$x \mathit{dup}_N y \stackrel{\text{def}}{=} (\#y = N) \wedge (\forall i : 0 \leq i < N. y_i = x),$$

it can be shown that

$$\mathit{dup}_N; \mathit{tran} = \propto \mathit{dup}_N.$$

Similarly, for rectangular tuples of length less than or equal to i , we have

$$\mathit{tran}; \pi_i = \propto \pi_i.$$

Transposition theorems

Transposition theorems belong to one of several classes of theorems useful for circuit optimisation [4]. They relate circuits interleaved in different ways. In this section we give a number of transposing theorems that have been found useful in developing designs. Theorems involving columns can be derived from the corresponding theorems for rows, since $\mathit{col} f = (\mathbf{row}(f^{-1}))^{-1}$.

Linear structures will be discussed first. The theorems obtained will then be extended to cover rectangular structures.

Linear arrays

First, let us consider a linear array that can be described as a map of maps. The components of the array can be transposed if the domain and the range signals are transposed accordingly

(Figure 10):

$$\parallel_{i < M} \left(\parallel_{j < N} f_{i,j} \right) = \parallel_{j < N} \left(\parallel_{i < M} f_{i,j} \right) \setminus tran. \quad (9)$$

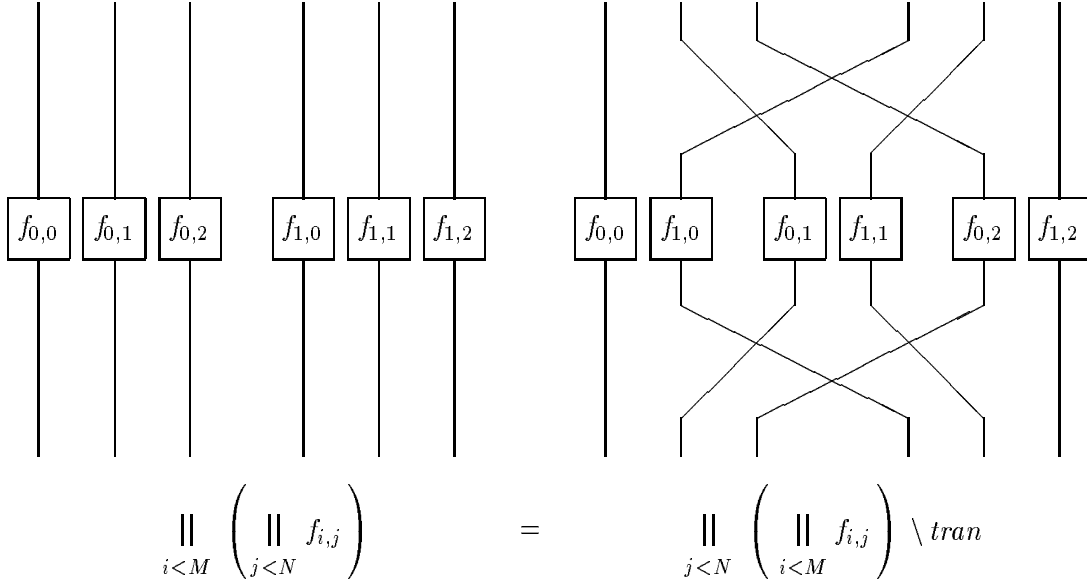


Figure 10: A theorem for transposing a map of maps ($M = 2$, $N = 3$).

A common special case of this theorem is

$$\alpha(f \parallel g) = (\alpha f \parallel \alpha g) \setminus tran.$$

Next, we consider a collection of rows displaced horizontally and vertically from each other. They can be combined into a single row with the displacement transferred to the components; this usually results in a more compact layout as described in the introductory section:

$$\parallel_{i < M} \left(\# f_{i,j} \right) \setminus tran = \# \left(\parallel_{j < N} \left(\parallel_{i < M} f_{i,j} \setminus tran \right) \right) \parallel \mathbf{fst} \, tran. \quad (10)$$

An example of the circuits represented by this equation, with $M = 2$ and $N = 3$, is given in Figure 1.

Special cases of Equation 10 include

$$\begin{aligned} \alpha(f \# g) \setminus tran &= (\alpha f \setminus tran) \# (\alpha g \setminus tran) \parallel \mathbf{fst} \, tran, \\ (\mathbf{row} f \parallel \mathbf{row} g) \setminus tran &= \mathbf{row}(f \parallel g \setminus tran) \parallel \mathbf{fst} \, tran. \end{aligned}$$

Rectangular arrays

The theorem for transposing linear arrays can be extended to deal with rectangular arrays:

$$\parallel_{k < K} \left(\#_{i,j < M,N} f_{k,i,j} \right) \setminus tran = \#_{i,j < M,N} \left(\parallel_{k < K} f_{k,i,j} \setminus tran \right) \setminus (tran \parallel tran). \quad (11)$$

Special cases of this theorem include

$$\begin{aligned} \alpha(\mathbf{grid} f) \setminus tran &= \mathbf{grid}(\alpha f \setminus tran) \setminus (tran \parallel tran), \\ (\mathbf{grid} f \parallel \mathbf{grid} g) \setminus tran &= \mathbf{grid}(f \parallel g \setminus tran) \setminus (tran \parallel tran). \end{aligned}$$

Distributing components in arrays with bends

In this section we present some theorems for reasoning about arrays with bends in them. The idea is to distribute the components as evenly as possible throughout the network. The circuit *bend* is a piece of wire that bends backwards:

$$\langle x, y \rangle \mathit{bend} z \stackrel{\text{def}}{=} x = y.$$

Note that x and y can themselves be tuples. In such cases cross-overs are needed to maintain the order of the signals (Figure 11). z is a dummy signal that corresponds to an unused connection.

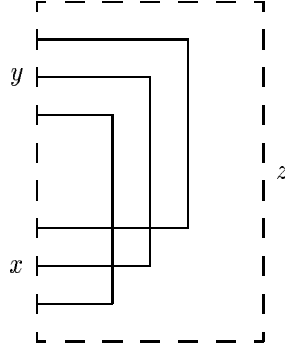


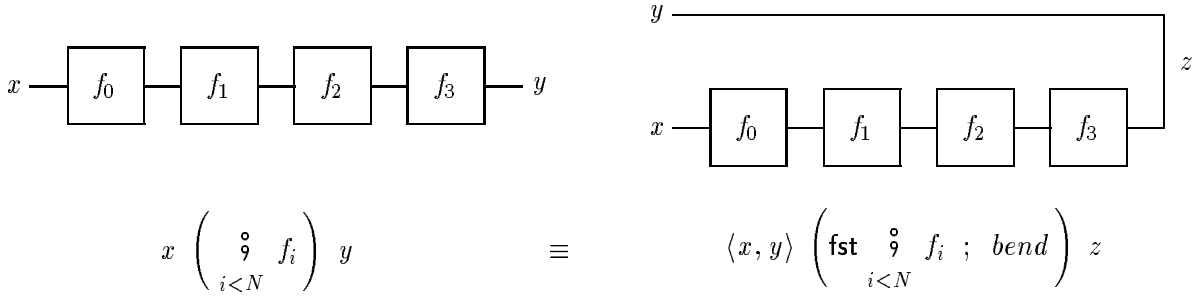
Figure 11: A vertical instance of the circuit *bend*.

We shall start by looking at how components can be distributed in linear arrays with a single bend. The results will then be extended to include linear and rectangular arrays with multiple bends.

Linear arrays with a single bend

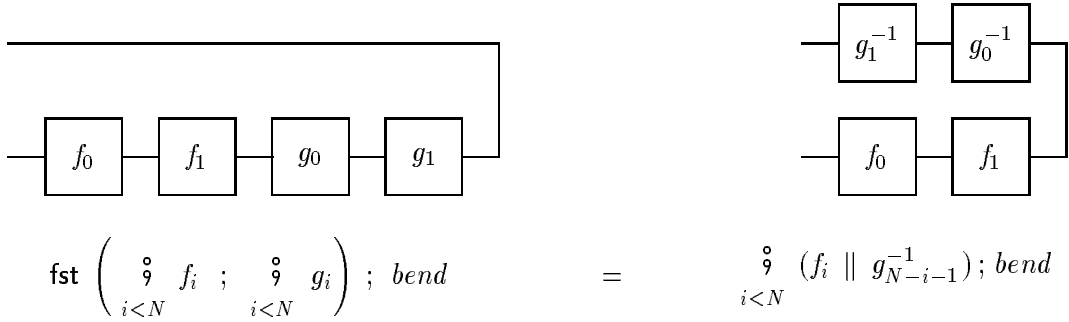
A bend can be added to a chain of N components (Figure 12):

$$x \left(\begin{array}{c} \circ \\ \mathfrak{g} \\ f_i \\ i < N \end{array} \right) y \equiv \langle x, y \rangle \left(\mathbf{fst} \begin{array}{c} \circ \\ \mathfrak{g} \\ f_i \\ i < N \end{array} ; \mathit{bend} \right) z.$$

Figure 12: Adding a vertical bend to a chain ($N = 4$).

We can now move some of the components from the lower branch to the upper branch (Figure 13):

$$\begin{aligned} \text{fst} \left(\text{fst}_{i<N} f_i ; \text{fst}_{i<N} g_i \right) ; \text{bend} &= \left(\text{fst}_{i<N} f_i \right) \parallel \left(\text{fst}_{i<N} g_{N-i-1}^{-1} \right) ; \text{bend} \\ &= \text{fst}_{i<N} (f_i \parallel g_{N-i-1}^{-1}) ; \text{bend}. \end{aligned} \quad (12)$$

Figure 13: Transforming a chain with a vertical bend ($N = 2$).

For homogeneous arrays Equation 12 becomes

$$\text{fst} (f^N ; g^N) ; \text{bend} = (f \parallel g^{-1})^N ; \text{bend}.$$

Similarly, a bend can be added to a component with connections on every side:

$$\langle x, y \rangle f \langle u, v \rangle \equiv \langle \langle x, v \rangle, y \rangle (\text{fstv } f ; \text{snd bend} ; \pi_1) u.$$

One can move the component to the upper branch provided that it is reflected vertically,

$$\text{fstv } f ; \text{snd bend} = (\text{sndv } f^\vee) \parallel (\text{fst } \text{recrv}) ; \text{snd bend}.$$

From this, it is easy to derive an equation which introduces transposition to a circuit with two components:

$$\mathbf{fstv}(f \# g) ; \mathbf{snd bend} = (f \parallel g^\vee \setminus \mathit{tran}) \parallel \mathbf{fst}(\mathbf{snd} \mathit{recrev}) ; \mathbf{snd bend}.$$

If both components are themselves rows of N components, then we can move half of the components from the lower branch to the upper branch and transpose the resulting circuit (Figure 14):

$$\begin{aligned} & \mathbf{fstv} \left(\begin{array}{c} \# f_i \# \\ \# g_i \# \end{array} \right) ; \mathbf{snd bend} \\ &= \left(\left(\begin{array}{c} \# f_i \parallel \\ \# g_{N-i-1}^\vee \end{array} \right) \setminus \mathit{tran} \right) \parallel \mathbf{fst}(\mathbf{snd} \mathit{recrev}) ; \mathbf{snd bend} \\ &= \#_{i < N} (f_i \parallel g_{N-i-1}^\vee \setminus \mathit{tran}) \parallel \mathbf{fst} \mathit{transrev} ; \mathbf{snd bend}, \end{aligned} \quad (13)$$

where

$$\mathit{transrev} \stackrel{\text{def}}{=} \mathit{tran} ; \mathbf{snd} \mathit{recrev}.$$

If the two arrays are homogeneous then Equation 13 becomes

$$\mathbf{fstv}(\mathbf{row} f \# \mathbf{row} g) ; \mathbf{snd bend} = \mathbf{row}(f \parallel g^\vee \setminus \mathit{tran}) \parallel \mathbf{fst} \mathit{transrev} ; \mathbf{snd bend}.$$

Linear arrays with multiple bends

We now explore how multiple bends can be introduced in linear arrays. First, observe that a bend can be twisted to form a zig-zag piece of wire containing three bends:

$$\mathit{bend} = \mathbf{snd}(\pi_2^{-1} ; \mathit{shl} \parallel (\mathbf{snd} \mathit{bend}) ; \pi_1) ; \mathit{bend}$$

This process can be continued so that a single bend can be twisted to form a piece of wire containing an odd number of bends:

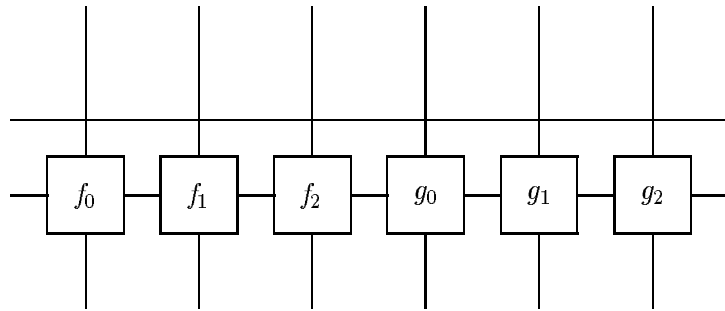
$$\mathit{bend} = \mathbf{snd}(\pi_2^{-1} ; (\mathbf{col}_N \mathit{shl}) \parallel \mathbf{snd}(\alpha_N \mathit{bend}) ; \pi_1) ; \mathit{bend}.$$

In the same vein one can add $2N - 1$ bends to a chain of $2N$ components by the following theorem:

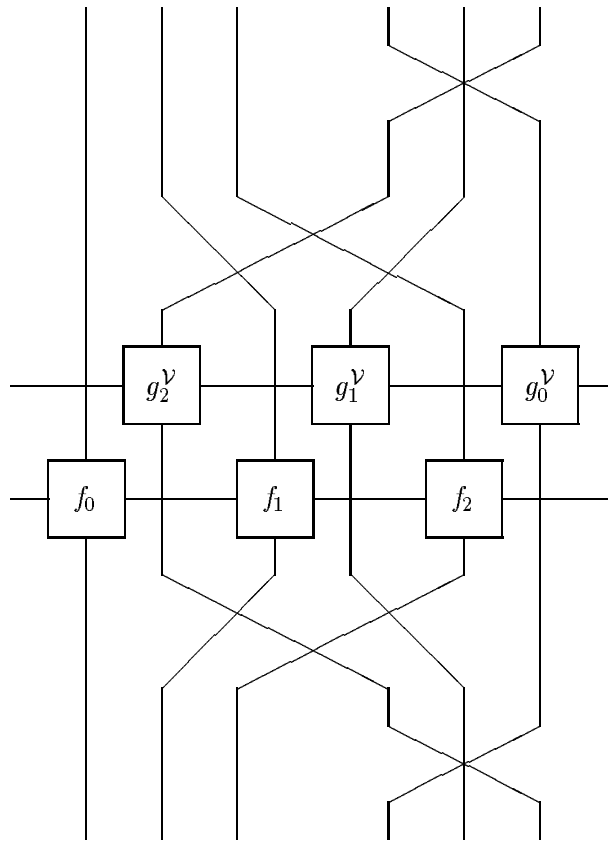
$$\mathbf{fst} \left(\begin{array}{c} \circ \\ \circ \\ \circ \end{array} f_i \right) ; \mathit{bend} = \mathit{sndbends}_N ; \mathbf{fst} \left(\mathbf{fst} \begin{array}{c} \circ \\ \circ \\ \circ \end{array} f_i \right) \setminus \mathit{apl} ; \alpha_N \mathit{bend}$$

where

$$\mathit{sndbends}_N \stackrel{\text{def}}{=} \mathbf{snd}(\pi_2^{-1} ; \mathbf{fst}(\alpha_{N-1} \mathit{bend}^{-1}) ; \mathbf{col}_{N-1} \mathit{shl}) ; \mathit{shr} ; \mathit{apl}.$$



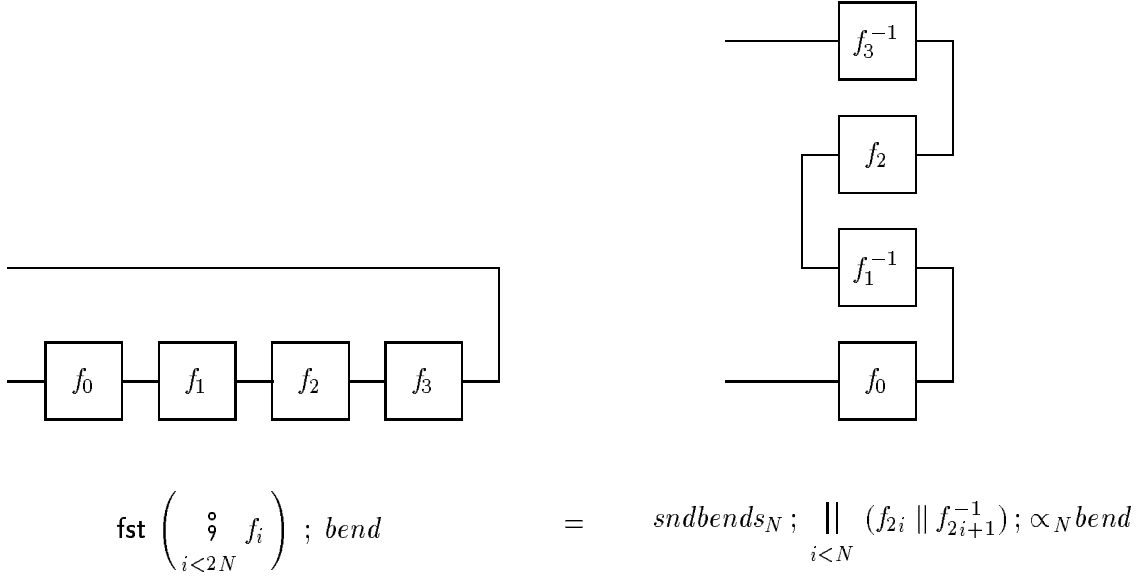
$$\text{fstv} \left(\begin{array}{cc} \text{++ } f_i \text{ ++} & \text{++ } g_i \text{ ++} \\ i < N & i < N \end{array} \right) ; \text{snd bend}$$



$$= \text{++ } (f_i \parallel g_{N-i-1}^v \setminus \text{tran}) \parallel \text{fst transrev} ; \text{snd bend}$$

$$i < N$$

Figure 14: Transforming a row with a vertical bend ($N = 3$).

Figure 15: Adding multiple bends to a chain ($N = 2$).

We then move some components from the bottom branch onto the other branches (Figure 15):

$$\mathbf{fst} \left(\overset{\circ}{\parallel}_{i < 2N} f_i \right); \mathit{bend} = \mathit{sndbends}_N; \overset{\parallel}{\parallel}_{i < N} (f_{2i} \parallel f_{2i+1}^{-1}); \propto_N \mathit{bend}. \quad (14)$$

The same procedure can be applied to a row of components (Figure 16):

$$\mathbf{fstv} \left(\overset{\ddagger}{\parallel}_{i < N} (f_{2i} \ddagger f_{2i+1}) \right); \mathit{snd} \mathit{bend} = \mathbf{fst} \mathit{sndbends}_N; \Phi_N; \mathit{snd} \propto_N \mathit{bend}$$

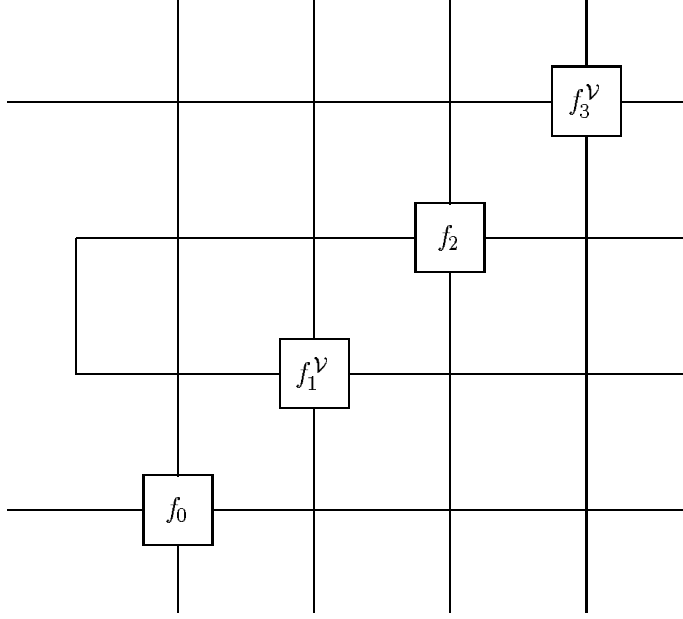
where

$$\Phi_N \stackrel{\text{def}}{=} \left(\overset{\parallel}{\parallel}_{i < N} (f_{2i} \parallel f_{2i+1}^{\vee} \setminus \mathit{tran}) \setminus \mathit{tran} \right) \parallel \mathbf{fst} \propto(\mathit{sndrecrev}). \quad (15)$$

The above two theorems can be specialised to deal with homogeneous arrays:

$$\begin{aligned} & \mathbf{fst} f^{2N}; \mathit{bend} \\ &= \mathit{sndbends}_N; \propto_N (f \parallel f^{-1}); \propto_N \mathit{bend}, \end{aligned}$$

$$\begin{aligned} & \mathbf{fstv} (\mathit{row}_N (f \ddagger f)); \mathit{snd} \mathit{bend} \\ &= \mathbf{fst} \mathit{sndbends}_N; (\propto_N (f \parallel f^{\vee} \setminus \mathit{tran}) \setminus \mathit{tran}) \parallel \mathbf{fst} \propto(\mathit{sndrecrev}); \mathit{snd} \propto \mathit{bend}_N. \end{aligned}$$



$$\mathbf{fst} \mathit{sndbends}_N ; \left(\parallel_{i < N} (f_{2i} \parallel f_{2i+1}^v \setminus \mathit{tran}) \setminus \mathit{tran} \right) \parallel \mathbf{fst} \alpha(\mathit{sndrecrev}) ; \mathbf{snd} \alpha_N \mathit{bend}$$

Figure 16: Adding multiple bends to a row of components ($N = 2$).

Rectangular arrays with multiple bends

In this section we extend the results for linear arrays to cover rectangular arrays. First, let us state two theorems for rearranging vertical bends in rectangular arrays:

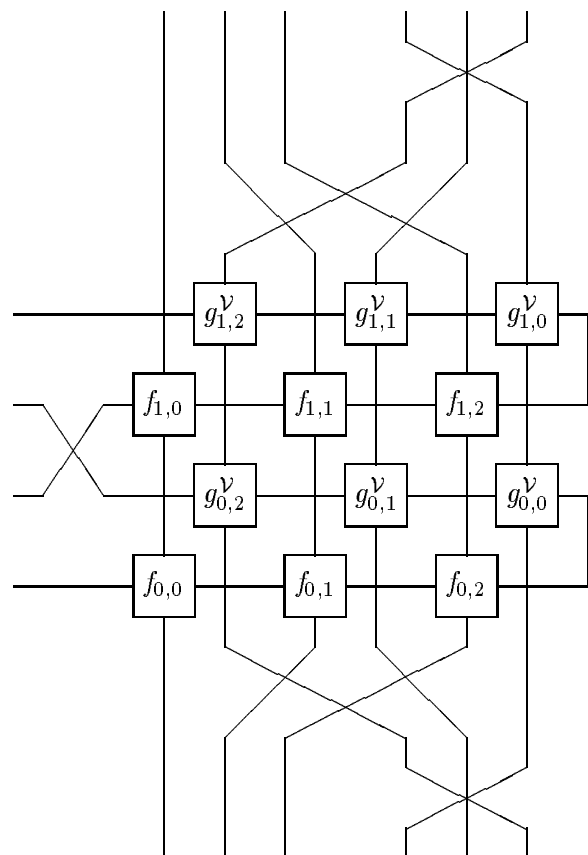
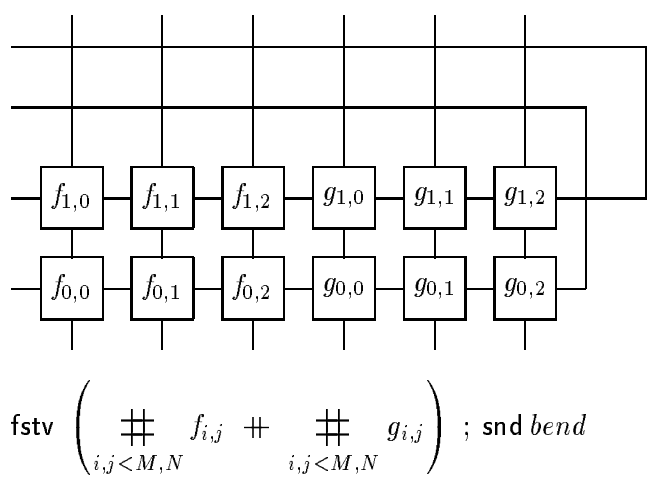
$$\begin{aligned} & \mathbf{fstv} \left(\#_{i,j < M,N} f_{i,j} \# \#_{i,j < M,N} g_{i,j} \right) ; \mathbf{snd} \mathit{bend} \\ &= \mathbf{fst} \mathit{tran} ; \#_{i,j < M,N} (\mathbf{fstv} f_{i,j}) \# \#_{i,j < M,N} (\mathbf{fstv} g_{i,j}) ; \mathbf{snd} (\alpha \mathit{bend}) \end{aligned} \quad (16)$$

$$= \mathbf{fst} \mathit{tran} ; \#_{i,j < M,N} (f_{i,j} \parallel g_{i,N-j-1}^v \setminus \mathit{tran}) \parallel \mathbf{fst} \mathit{transrev} ; \mathbf{snd} (\alpha \mathit{bend}) \quad (17)$$

where $\mathit{transrev} \stackrel{\text{def}}{=} \mathit{tran} ; \mathbf{snd} \mathit{recrev}$ as before. Equation 16 corresponds to implementing the feedback paths as internal wiring cells, and Equation 17 corresponds to a transposition forming composite cells each consisting of an $f_{i,j}$ and a reflected version of $g_{i,N-j-1}$ (Figure 17).

Again these can be specialised for homogeneous grids, for example:

$$\begin{aligned} & \mathbf{fstv} (\mathit{grid}f \# \mathit{grid}g) ; \mathbf{snd} \mathit{bend} \\ &= \mathbf{fst} \mathit{tran} ; \mathit{grid} (f \parallel g^v \setminus \mathit{tran}) \parallel \mathbf{fst} \mathit{transrev} ; \mathbf{snd} (\alpha \mathit{bend}). \end{aligned}$$



$$= \mathbf{fst} \mathit{tran} ; \#_{i,j < M,N} (f_{i,j} \parallel g_{i,N-j-1}^v \setminus \mathit{tran}) \parallel \mathbf{fst} \mathit{transrev} ; \mathbf{snd} (\times \mathit{bend})$$

Figure 17: Rectangular arrays with multiple vertical bends ($M = 2, N = 3$).

Next, consider a grid consisting of four heterogeneous arrays:

$$grid4 \langle p, q, r, s \rangle \stackrel{\text{def}}{=} \left(\#_{i,j < M,N} p_{i,j} \# \#_{i,j < M,N} q_{i,j} \right) \# \left(\#_{i,j < M,N} r_{i,j} \# \#_{i,j < M,N} s_{i,j} \right).$$

A rectangular array with vertical and horizontal bends (the top diagram of Figure 19) can be described by

$$grid4bends \langle p, q, r, s \rangle \stackrel{\text{def}}{=} \mathbf{fsth} (\mathbf{fstv} (grid4 \langle p, q, r, s \rangle)); \mathit{bend} \parallel \mathit{bend}.$$

The first optimisation, as before, is to implement the feedback paths as wiring cells within the array. This is captured by the following theorem:

$$grid4bends \langle p, q, r, s \rangle = \mathit{map2tran} ; grid4 \langle \Omega p, \Omega q, \Omega r, \Omega s \rangle ; \mathit{map2bend} \quad (18)$$

where

$$\begin{aligned} \Omega f &\stackrel{\text{def}}{=} \#_{i,j < M,N} (\mathbf{fsth} (\mathbf{fstv} f_{i,j})), \\ \mathit{map2tran} &\stackrel{\text{def}}{=} (\mathit{tran} ; \times \mathit{tran}) \parallel (\mathit{tran} ; \times \mathit{tran}), \\ \mathit{map2bend} &\stackrel{\text{def}}{=} (\times \mathit{bend} \parallel \times \mathit{bend}) \parallel (\times \mathit{bend} \parallel \times \mathit{bend}). \end{aligned}$$

However, there are still long combinational paths in the circuit which can be eliminated by a further rearrangement of components. Before doing that, let us define the combinator representing the wiring pattern that will be required (Figure 18):

$$\mathit{transwap} \langle a, b, c, d \rangle \stackrel{\text{def}}{=} ((\mathbf{sndh} a \parallel \mathbf{sndh} b) \setminus \mathit{tran}) \# ((\mathbf{fsth} c \parallel \mathbf{fsth} d) \setminus \mathit{tran}).$$

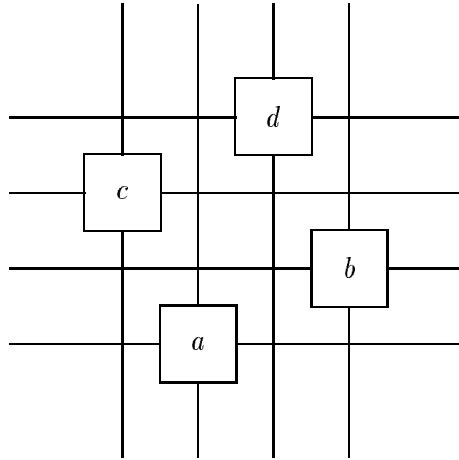
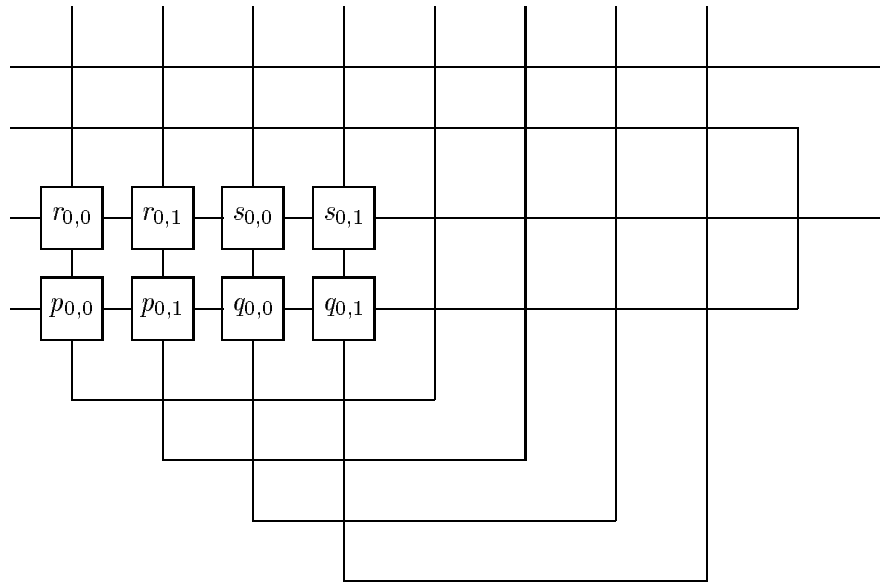
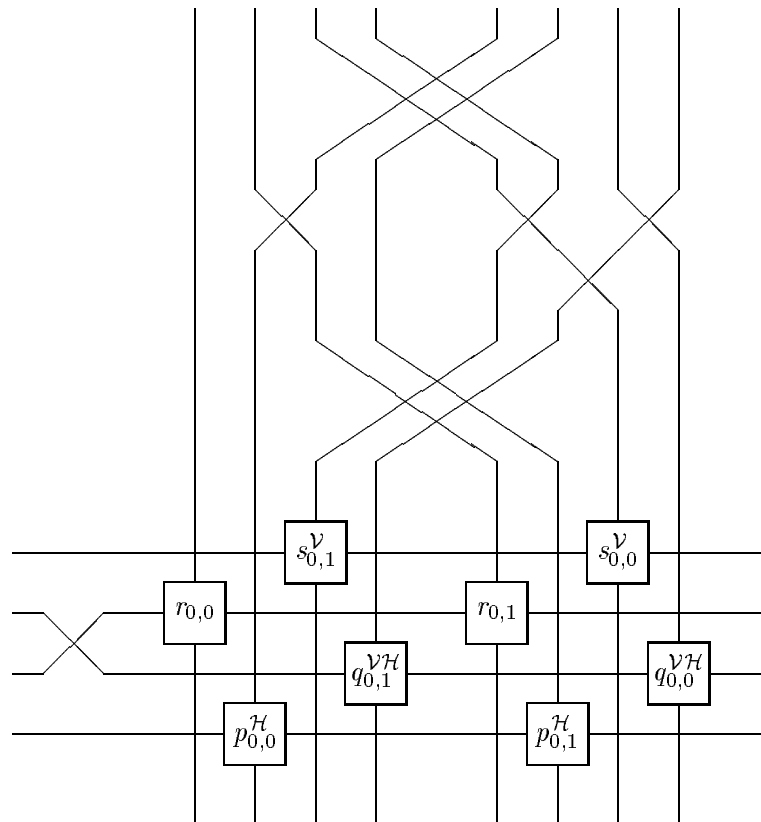


Figure 18: $\mathit{transwap} \langle a, b, c, d \rangle$.



$$\text{fsth}(\text{fstv}(\text{grid4}(p, q, r, s))); \text{bend} || \text{bend}$$



$$= \text{maprevs} ; \#_{i,j < M,N} \left(\text{transwap} \langle p_{M-i-1,j}^H, q_{M-i-1,N-j-1}^{VH}, r_{i,j}, s_{i,N-j-1}^V \rangle \right) ; \text{mapbend2}$$

Figure 19: Rectangular arrays with vertical and horizontal bends ($M = 1, N = 2$).

A useful result concerning *transwap* is

$$\mathbf{fstv} (\mathbf{fsth} \# \langle \langle a, b \rangle, \langle c, d \rangle \rangle); \mathbf{bend} \parallel \mathbf{bend} = \mathbf{revs}; \mathbf{transwap} \langle a^{\mathcal{H}}, b^{\mathcal{VH}}, c, d^{\mathcal{V}} \rangle; \mathbf{parbend2},$$

where

$$\begin{aligned} \mathbf{revs} &\stackrel{\text{def}}{=} (\mathbf{tran}; \mathbf{fst} (\mathbf{recrev} \parallel \mathbf{recrev})) \parallel (\mathbf{tran}; \mathbf{snd} (\mathbf{recrev} \parallel \mathbf{recrev})), \\ \mathbf{parbend2} &\stackrel{\text{def}}{=} (\mathbf{bend} \parallel \mathbf{bend}) \parallel (\mathbf{bend} \parallel \mathbf{bend}). \end{aligned}$$

Using this we can obtain a theorem that produces a more even distribution of components for *grid4bends* $\langle p, q, r, s \rangle$, an instance of which is depicted in Figure 19:

$$\begin{aligned} \mathbf{grid4bends} \langle p, q, r, s \rangle \\ = \mathbf{maprevs}; \#_{i,j < M,N} \left(\mathbf{transwap} \langle p_{M-i-1,j}^{\mathcal{H}}, q_{M-i-1,N-j-1}^{\mathcal{VH}}, r_{i,j}, s_{i,N-j-1}^{\mathcal{V}} \rangle \right); \\ \mathbf{mapbend2} \end{aligned} \tag{19}$$

where

$$\begin{aligned} \mathbf{maprevs} &\stackrel{\text{def}}{=} (\mathbf{tran}; \mathbf{fst} \times \mathbf{recrev}; \times \mathbf{tran}; \mathbf{tran}) \parallel (\mathbf{tran}; \mathbf{snd} \times \mathbf{recrev}; \times \mathbf{tran}; \mathbf{tran}), \\ \mathbf{mapbend2} &\stackrel{\text{def}}{=} \times (\mathbf{bend} \parallel \mathbf{bend}) \parallel \times (\mathbf{bend} \parallel \mathbf{bend}). \end{aligned}$$

These theorems can be specialised for homogeneous networks as well; for instance one can show that

$$\begin{aligned} \mathbf{fstv} (\mathbf{fsth} (\mathbf{grid}_{2,2} (\mathbf{grid}_{M,N} f))); \mathbf{bend} \parallel \mathbf{bend} \\ = \mathbf{maprevs}; \mathbf{grid}_{M,N} (\mathbf{transwap} \langle f^{\mathcal{H}}, f^{\mathcal{VH}}, f, f^{\mathcal{V}} \rangle); \mathbf{mapbend2}. \end{aligned}$$

Trade-off analysis

This section suggests some quantitative measures of the trade-offs involved in the transformations discussed in the preceding sections. As explained in the introductory section, transposition usually results in a more uniform distribution of components in the array at the expense of increasing the complexity of the interface. In the following estimation of size and performance the contribution of the interface will not be included since it is usually application- or implementation-dependent. Hence the reader must check for the particular situation whether the improvement brought by the transformation is offset by the increased complexity of the interface.

Table 1 summarises the effects of various transposition strategies introduced earlier. The calculation is based on the simplified case in which all components have the same aspect ratio $h \times w$. We also assume, as explained in the introductory section, that wires can be overlaid on components in assessing the alteration of aspect ratio, and that all components are latched in estimating the reduction of long combinational paths.

Table 1: Summary of trade-offs in transposition strategies, showing the features of the designs represented by the left- and the right-hand side of a given equation in the text.

Equation	LHS aspect ratio	RHS aspect ratio	LHS longest path	RHS longest path
9	$h \times MNw$	$h \times MNw$ or $Mh \times Nw$	0	h or 0
10	$h \times MNw$	$h \times MNw$ or $Mh \times Nw$	h or Nw	h or w
11	$KMh \times KNw$	$KMh \times Nw$ or $Mh \times KNw$	$(K - 1)Mh$ or $(K - 1)Nw$	$(K - 1)h$ or $(K - 1)w$
14	$h \times 2Nw$	$2Nh \times w$	$2Nw$	0
15	$h \times 2Nw$	$2Nh \times w$	$2Nw$	$(2N - 1)h$
16	$Mh \times 2Nw$	$Mh \times 2Nw$	$(M - 1)h + 2Nw$	$2Nw$
17	$Mh \times 2Nw$	$Mh \times 2Nw$ or $2Mh \times Nw$	$(M - 1)h + 2Nw$	h or w
18	$2Mh \times 2Nw$	$2Mh \times 2Nw$	$2(M - 1)h + 2Nw$ or $2Mh + 2(N - 1)w$	$2Mh$ or $2Nw$
19	$2Mh \times 2Nw$	$4Mh \times Nw$ or $Mh \times 4Nw$	$2(M - 1)h + 2Nw$ or $2Mh + 2(N - 1)w$	$4h$ or $4w$

Concluding remarks

We have developed a number of techniques for optimising array-based designs by component transposition. The major innovations include the description of properties of *tran* that are useful for design optimisation and the collection of theorems for transforming expressions with *tran* and *bend*. Let us summarise the main features of our work and examine its design implications.

Summary

A principal challenge in providing a theory for transforming designs is to find useful design abstractions. We have identified the parametrised building blocks *tran* and *bend* and showed how they correspond to common ways of interconnecting components. These generic building blocks allow concise descriptions of complex wiring patterns.

The use of *tran* and *bend* also enables us to follow a simple equational style of presenting transformations. Several schemes for transposing components of regular array circuits have been discussed. Although the more refined versions require a higher degree of sophistication in interface arrangements, they can still be captured succinctly using the appropriate combinatorics and primitives. These transformations form a valuable part of a coherent framework for developing and optimising designs (see [2], [3], [4], [6], [7]).

Design implications

The increase in performance brought by transposing components comes mainly from localising interconnections by wiring cells. The elimination of long wires reduces the area and power required for a circuit. We have also indicated how further improvements can be obtained by wiring over the computational blocks, a technique which is eased by the adoption of multi-layer interconnections in some CMOS technologies. An advantage of our approach is that the transformations preserve the regularity of the architecture. This simplifies both the implementation of regular array circuits and the comparison of alternative designs.

Our work, however, is not restricted to custom integrated circuit designs; for instance, transposition can also be applied to vary the array dimensions of programmable cellular structures in order to maximise cell utilisation.

The introduction of bends provides a further opportunity for meeting design constraints, such as satisfying requirements for an array to conform to a given aspect ratio or for the input and output ports to be arranged in a specific manner. In particular, our work provides a possible optimisation for systolic implementations involving long feedback paths [5]. The optimised design can be achieved with little additional effort as the transformations only involve the reflection of components. A more complicated interface may be required, however, although this may not be an issue if the data come in the right format or if the resulting increase in performance justifies the complication.

Further work is needed in three areas. First, a deeper understanding of the scope and the value of transposition requires the application of this method to circuits of greater complexity than those discussed in this paper. Second, it is important to study how transposition fits in with other techniques – such as pipelining – in order to provide an overall strategy for optimising designs. Third, the implementation of transposition should be facilitated by appropriate computer-based tools which are compatible with other circuit design aids and cell libraries.

Acknowledgements. I thank Geraint Jones for contributing the bottom array in Figure 19 and for providing useful suggestions. I also thank Michael Jampel, Mary Sheeran and the referees of this paper for their comments. The support of the UK Alvey Programme (Project ARCH 013), the Croucher Foundation and Rank Xerox (UK) Limited is gratefully acknowledged.

References

- [1] G. Jones and M. Sheeran, *Timeless truths about sequential circuits*, in S. K. Tewksbury, B. W. Dickinson and S. C. Schwartz (eds.), *Concurrent computations: algorithms, architectures and technology*, pp. 245–259, Plenum Press, 1988.
- [2] G. Jones and M. Sheeran, *Circuit design in Ruby*, in J. Staunstrup (ed.), *Formal methods for VLSI design*, North-Holland, 1990.
- [3] W. Luk, *Specifying and developing regular heterogeneous designs*, in L. Claesen (ed.), *Formal VLSI specification and synthesis*, pp. 391–409, North-Holland, 1990.
- [4] W. Luk and G. Jones, *The derivation of regular synchronous circuits*, in K. Bromley, S. Y. Kung and E. Swartzlander (eds.), *Proceedings of International Conference on systolic arrays*, pp. 305–314, IEEE Computer Society Press, 1988.
- [5] J. Moreno and T. Lang, *On partitioning the Faddeev algorithm*, in K. Bromley, S. Y. Kung and E. Swartzlander (eds.), *Proceedings of International Conference on systolic arrays*, pp. 125–134, IEEE Computer Society Press, 1988.
- [6] L. Rossen, *HOL formalisation of Ruby*, Technical Report ID-TR 1989–61, Department of Computer Science, Technical University of Denmark, 1989.
- [7] S. Singh, *An application of non-standard interpretation: testability*, in L. Claesen (ed.), *Formal VLSI correctness verification*, pp. 235–244, North-Holland, 1990.