Notes on Bitonic Merge Sort

Scott B. Baden
Department of Computer Science and Engineering
University of California, San Diego **October 11, 2006**

Problem: Given a list $x = (x_0, x_1, ..., x_{n+1})$ of n elements, produce a new list $x' = (x_{i_0}, x_{i_1}, ..., x_{i_{n-1}})$ such that $x_{i_j} \leq x_{i_k}$ when $j < k$ and $p_i = (i_0, i_1, ..., i_{n-1})$ is a permutation vector describing how indices of x are mapped to x'.

Definition:
A bitonic sequence is a sequence of numbers $x_0, x_1 ... x_{n-1}$ with the following properties:

1. There exists an index i where $0 \leq i \leq n-1$ such that

$$a_0 \leq a_1 \leq ... \leq a_i \;\; \textbf{and} \;\; a_i \geq a_i + 1 \geq ... \geq a_{n-1}$$

2. We may cyclically shift the a_i such that (1) is true.

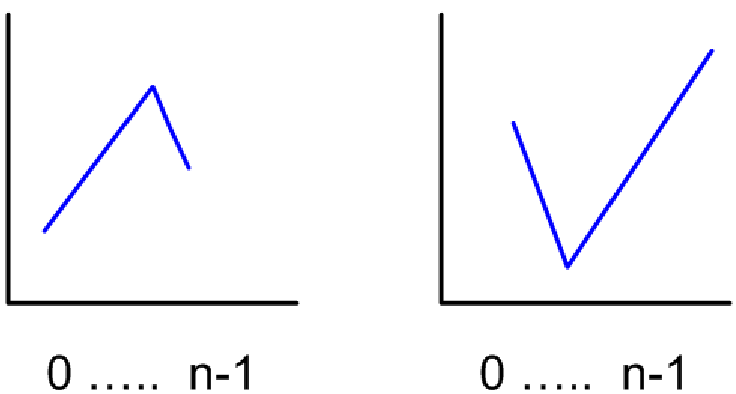


Figure 1: Examples of bitonic sequences.

Unique Crossover properties of bitonic sequences

Let $x = (x_0, x_1, ..., x_{n-1})$ be a bitonic sequence such that

$$x_0 \leq x_1 \leq ... \leq x_{\frac{n}{2}-1} \;\; AND \;\; x_{\frac{n}{2}} \geq x_{\frac{n}{2}+1} \geq ... \geq x_{n-1}$$

Let's represent x as follows:

```
                    |   x_0           /|\ x_n/2
                    |   x_1            |  x_(n/2)-1
                    |    .             |     .
                    |    .    "<="     |     .
                    |    .             |     .
                    |   x_i-1          |  x_(n/2)+i-1
           ------------------------------------- H
                    |   x_i            |  x_(n/2)+i
                    |    .             |     .
                    |    .    ">"      |     .
                    |    .             |     .
 increasing        \|/ x_(n/2)-1       |  x_n-1
    size
```

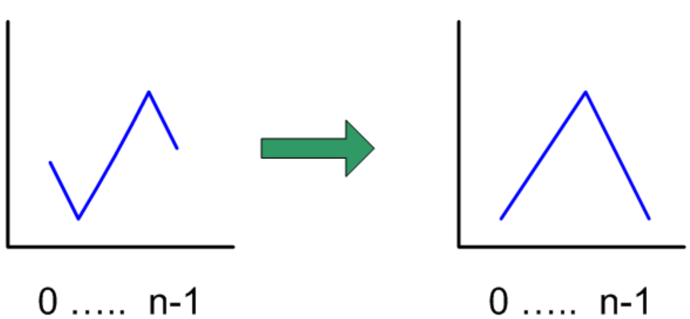
There exists a unique line H dividing the region between the arrows into two subsections labeled $'' <= ''$ and $'' > ''$ such that

1. An arbitrary pair of sub-elements in the same region–one from the left, the other from the right–satisfies the relation $'' <= ''$ or $'' > ''$ in a subsection.
2. Every element in the left column of sub-region $'' <= ''$ is less than or equal to every element in the sub-region $'' > ''$ in the left column. (And a similar property holds for the right).

A horizontal line H that induces two region with these two properties is said to have the *unique crossover property.* Consider the bitonic sequence:

$$x = (21, 20, 14, 10, -6, -4, 0, 1, 2, 18, 19, 30, 31, 25, 23, 22)$$

We cyclically rotate the original list as shown:



The result is to produce the two monotonic subsequences as shown:

```
     |   -4            25 /|\
     |    0    "<="    23  |
     |    1            22  |
     |    2            21  |
     |   18            20  |
-------------------------------
     |   19    ">"     14  |
     |   30            10  |
    \|/  31            -6  |
```

But how do we know that a bitonic sequence is unique (there is no other "lowest" H line)? Now we show that the unique crossover property holds in general.

Theorem: Any bitonic sequence satisfies the crossover property.

Proof: We will show that there always exists a horizontal cut splitting the region into two subsections that satisfy the following properties:

1. Given any x_i and x_j in the same subregion where x_i comes from the left and x_j comes from the right then $x_i \leq x_j$ if in subsection "$\leq$", else $x_i > x_j$.
2. Given any x_i and x_j in the same column C, say x_i in "$\leq$" and x_j in "$>$" then $x_i \leq x_j$ if C is the left column, otherwise $x_i > x_j$.

Let $x = (x_0, x_1..., x_{n-1})$ be an arbitrary bitonic sequence. Without any loss of generality, assume there exists an index l which divides this list into two parts.

$$x_0 \leq x_1 \leq ... \leq x_l \;\; AND \;\; x_{l+1} \geq x_{l+2} \geq ... \geq x_{n-1}$$

```
                 x_0              x_n/2
             |    .          /|\    .
             |    .           |     .
             |    .   "<="    |     .
             |    .           |     .
             |   x_l-(n/2)    |   x_l
     ------------------------------------- H   |
             |    .           |     .         \|/
     -------------------------------------
             |   x_i          |   x_i+(n/2)
             |    .           |     .
             |    .   ">"     |     .
             |    .           |     .
 increasing \|/ x_(n/2)-1     |   x_n-1
   size
```

Assume $l \geq (n/2)$

We draw H temporarily below $x_{l-\frac{n}{2}}$; Clearly the upper region is $" <= "$ because $x_0 \leq x_1 \leq x_l$. We then lower line until one of the following conditions is met:

1. we find an index i such that $x_i > x_{\frac{n}{2}+i}$ where $0 \leq i \leq (n/2) - 1$, OR
2. we determine that no such line exists of $x_{\frac{n}{2}-1} \leq x_{n-1}$.

In case (1) we leave H just above x_i and $x_{l+\frac{n}{2}}$.

In case (2) H is at the bottom below x_{n-2} and x_{n-1}. (verify that H satisfies the desired property...)

In this sense "unique" actually means that there is no H that is lower.

END PROOF.

Given a bitonic sequence x we define two operations on x:

$$L(x) = MIN(x_0, x_{\frac{n}{2}}), ..., MIN(x_i, x_{\frac{n}{2}+i})..., MIN(x_{\frac{n}{2}-1}, x_{n-1})$$

and

$$R(x) = MAX(x_0, x_{\frac{n}{2}}), ..., MAX(x_i, x_{\frac{n}{2}+i})..., MAX(x_{\frac{n}{2}-1}, x_{n-1})$$

Both $L(x)$ and $R(x)$ are bitonic and every element y of $L(x)$ is less than or equal to every element y of $R(x)$. These functions divide x into two sets.

Proof: By the unique crossover property there exists a horizontal cut immediately below x_i and $x_{\frac{n}{2}+i}$.

From the properties of the H cut

$$L(x) = (x_0, x_1, ..., x_i, x_{\frac{n}{2}+i+1}, ..., x_{n-1})$$

$$R(x) = (x_{\frac{n}{2}}, x_{\frac{n}{2}+1}, ..., x_{\frac{n}{2}+i}, x_{i+1}, x_{\frac{n}{2}-1})$$

```
                    x_0         /|\/  x_n/2       \
        /       |    .       \  | |     .         |
L(x) ---|       |    .   "<=" |  | |     .         |----R(x)
        |       |    .       |  | |     .         |
        \       |   x_i      /  | \  x_(n/2)+i    /
           ------------------------------------- H
        /       |   x_i+1    \   | /  x_(n/2)+i+1  \
R(x) ---|       |    .       |   | |     .          |
        |       |    .   ">" |   | |     .          |-----L(x)
        \       |    .       |   | |     .          |
 increasing    \|/ x_n-1     /   | \  x_n-1         /
    size
```

Note: $L(x)$ and $R(x)$ are both subsets of a bitonic sequence, and are therefore themselves bitonic.

Sorting Strategy

A) Create a bitonic sequence x from an unsorted list y using the procedure described below.
B) Sort x by splitting into two bitonic sequences L(x) and R(x), sorting these recursively and then merging.

Let B(n) be a module that constructs a bitonic sequence from an n-element sequence, and S(n) a module that sorts an n-element bitonic sequence. Thus, a bitonic sort starts by constructing n/4 4-element bitonic sequences (using B(4) modules), then n/8 8-element sequences, and so on. Each 4-element sequence is formed using 2 S(2) modules. We do this on a shuffle exchange network with (n/2) processors, where n = # of elements in the list.

```
a ---      min(a,b)
       >
b ---      max(a,b)


a ---      max(a,b)
       <
b ---      min(a,b)
```

Let $n = 2^k$. In the shuffle exchange we have partitions at each iteration:

```
n/2       2*(n/4)      4*(n/8)    etc.....
---        ---          ---
| |        | |          | |
| |        | |          | |
| |        | |          ---
| |        | |          | |
| |        | |          | |
| |        ---          ---
| |        | |          | |
| |        | |          | |
| |        | |          ---
| |        | |          | |
| |        | |          | |
---        ---          ---     until each partition is a single element.
```

The following procedure for sorting a bitonic sequence helps explain the organization of modules and the inherent parallelism

```
    procedure BSORT(i, j, >)                               // S(n)

      if |j - i + 2| >= 2 then
         return min(a_i,a_i+1) CONCAT max (a_i,a_i+1)
      else
         SHUFFLE(i, j, >)
         UNSUFFLE(i, j, >)
         paralleldo
             BSORT(i, i + (j-i+1)/2 - 1, >)                // S(n/2)
             BSORT(i + (j-i+1)/2, j, >)                    // S(n/2)

   end BSORT
```

Forming a bitonic sequence

There is a simple procedure for transforming an unsorted sequence into a bitonic sequence. Trivially, any 2-element sequence of numbers form a bitonic sequence. But most sequences are longer than 2 elements, so we inductively construct a bitonic sequence from smaller bitonic sequences.

We start by forming 4-element bitonic sequences from consecutive two element sequence. Consider four elements in sequence x_0, x_1, x_2, and x_3. We sort x_0 and x_1 into ascending order, while we sort x_2 and x_3 in descending order. We then concatenate the two pairs to form a 4 element bitonic sequence.

Next, we take two 4 element bitonic sequences, sorting one in ascending order, the other in descending order (using the sort procedure described above), and so on, until we obtain the following bitonic sequence, which we then sort:

```
           _.
    +---+  /|
    | n | /
    | - |/ sort in ascending order
    | 2 |
    +---+
    | n |
    | _ |\ sort in descending order
    | 2 | \
    +---+  \|
      n    -'
```

An example:

Input	bitonic 4	sort 4	bitonic 8
5	2	0	-7
2	5	2	0
-			
3	3	3	2
0	0	5	3
-	-	-	
9	4	9	4
4	9	8	5
-			
-7	8	4	8
8	-7	-7	9

The following procedure for generating a bitonic sort helps explain the organization of modules, and the inherent parallelism:

```
procedure bitonic (i, j)            // B(n)
  s = (j-i+1)/2
  if (j-i+1) > 2 then
    do in parallel
      begin BITONIC (i, i+s-1)      // B(n/2)
            BSORT (i, i+s-1, +)     // S(n/2)
      end
      begin BITONIC (i+s, j)        // B(n/2)
            BSORT (i+s, j, -)       // S(n/2)
      end
  end do parallel
```

The main program for bitonic sort is as follows:

```
  BITONIC (0, n-1)
  BSORT (0, n-1, +)
```

where $n = 2^j$