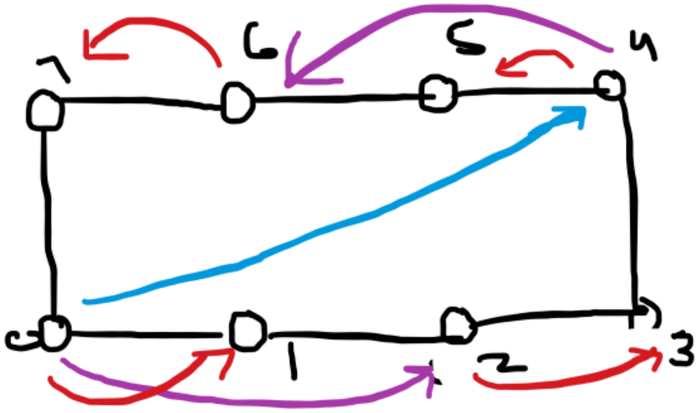
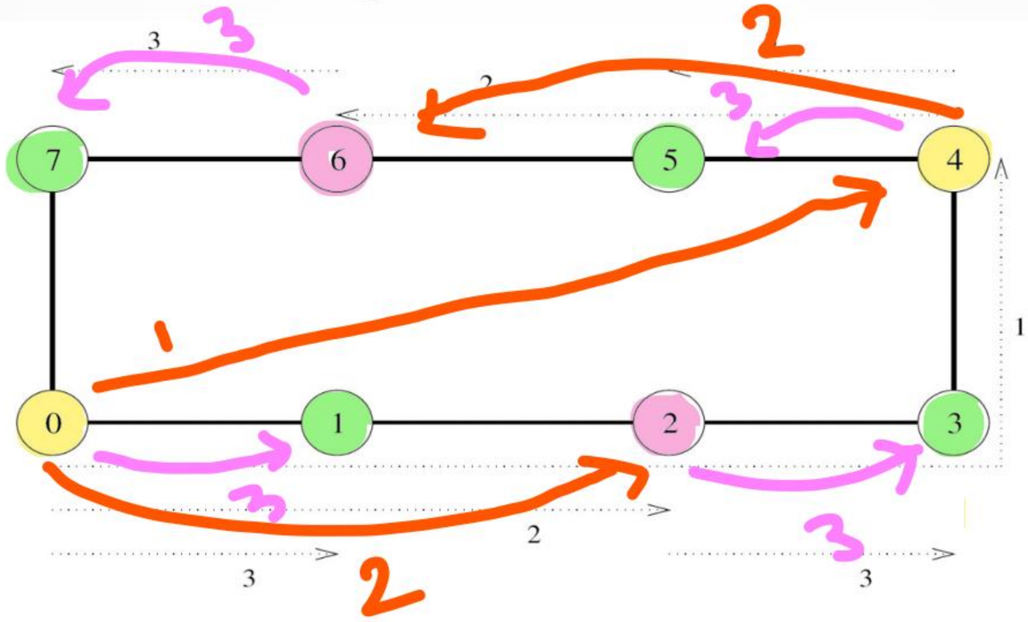
One to All Broadcast:



1st: 1 → 8/2=4

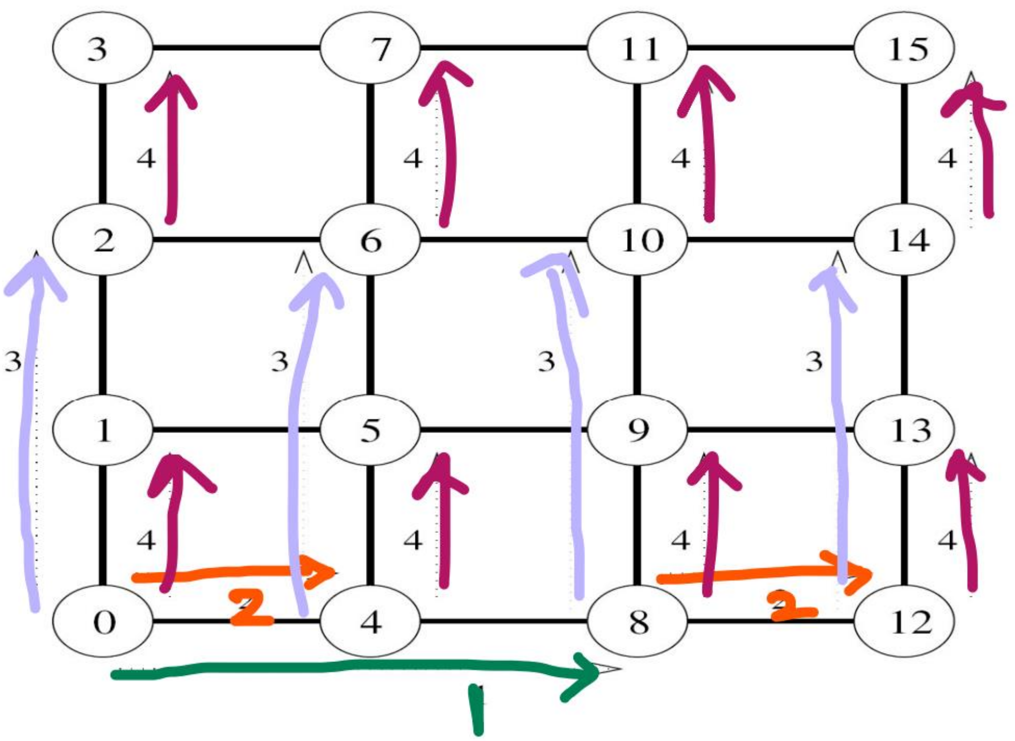
2nd: 1 → P/4 :
5 → P/4+P/2   2+4:6

Recursive Doubling Broadcast

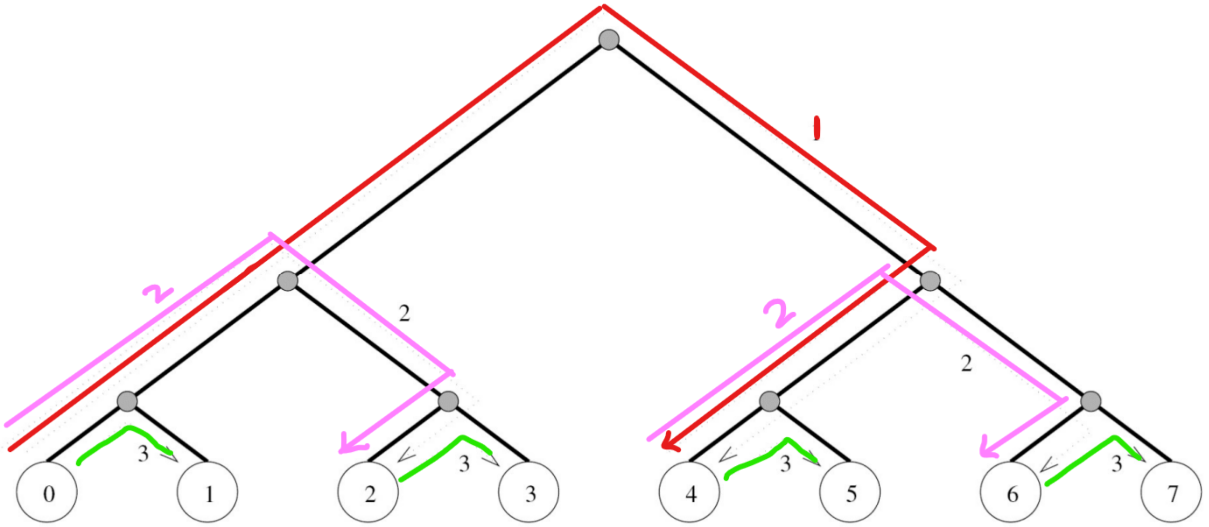


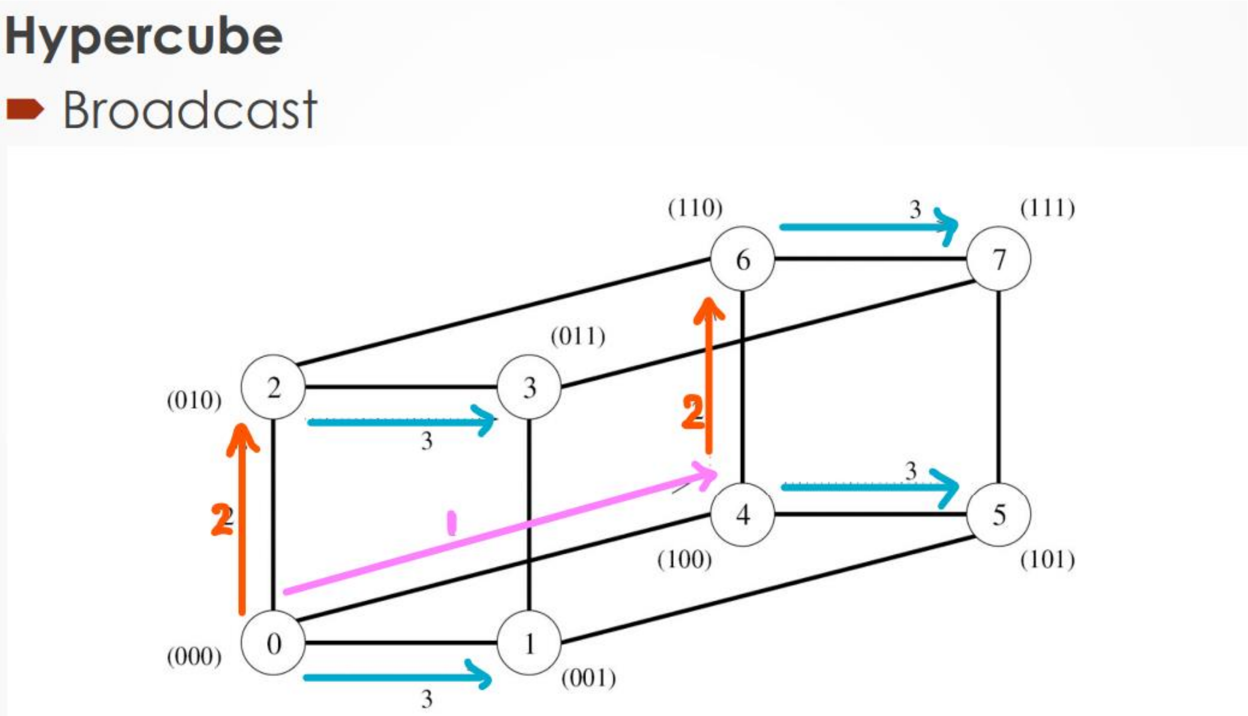
Cost: logp(ts+mtw)

Mesh (Broadcast and Reduction)



- Broadcast

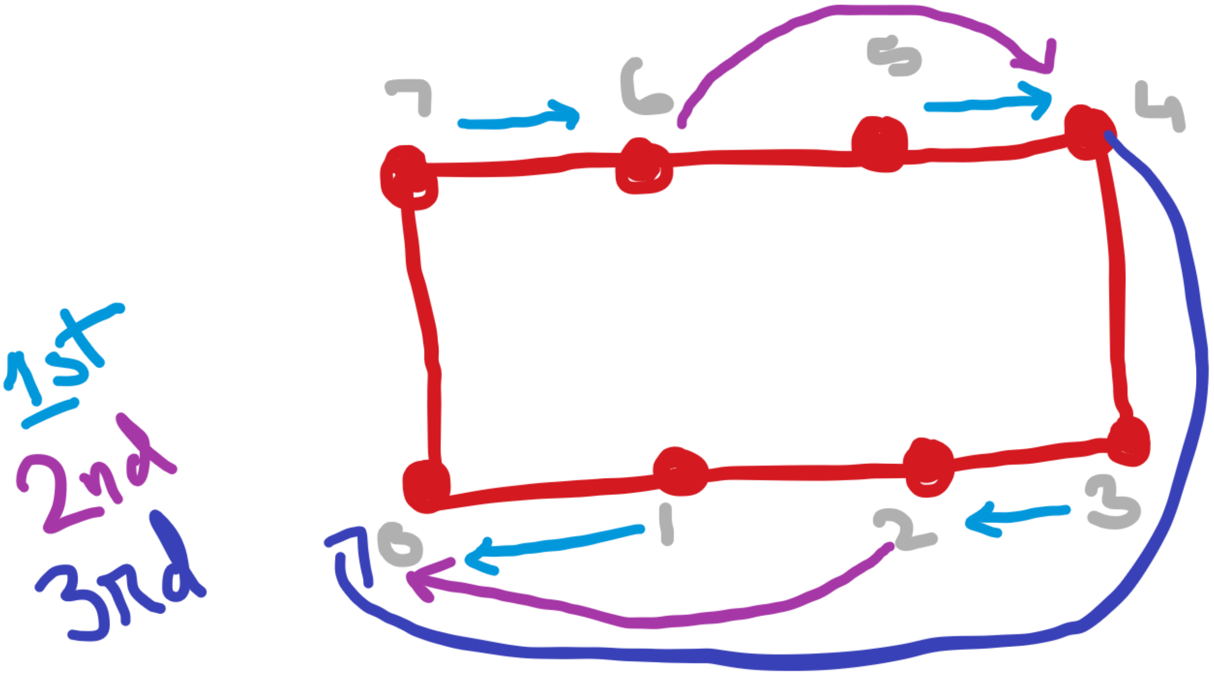


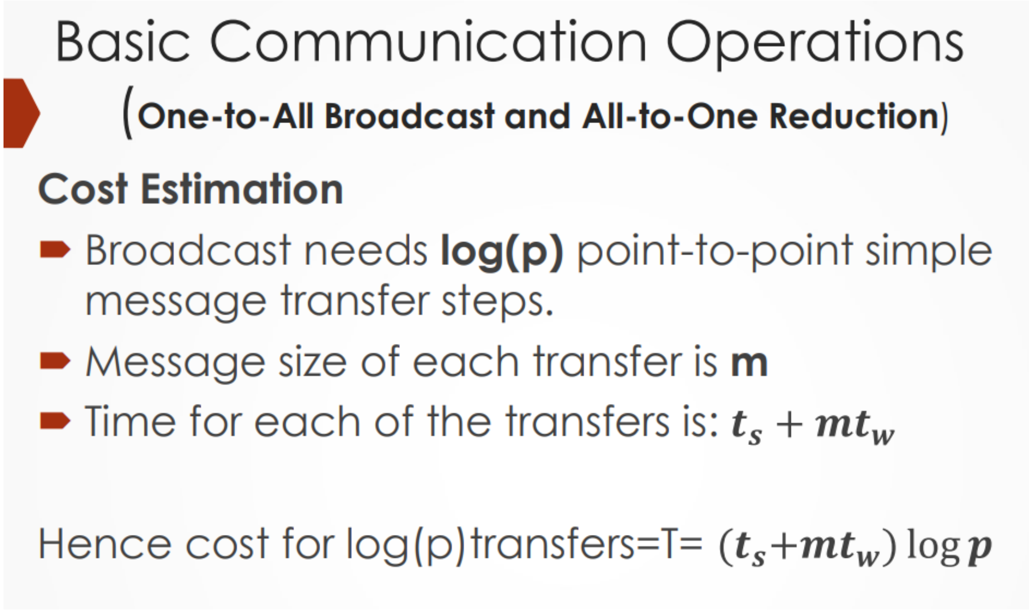


Hypercube

- Broadcast

All to One Reduction:





Basic Communication Operations

(One-to-All Broadcast and All-to-One Reduction)

Cost Estimation

- Broadcast needs **log(p)** point-to-point simple message transfer steps.
- Message size of each transfer is **m**
- Time for each of the transfers is: $t_s + mt_w$

Hence cost for log(p)transfers=T= $(t_s + mt_w)\log p$

All to All Broadcast:

2D Mesh:

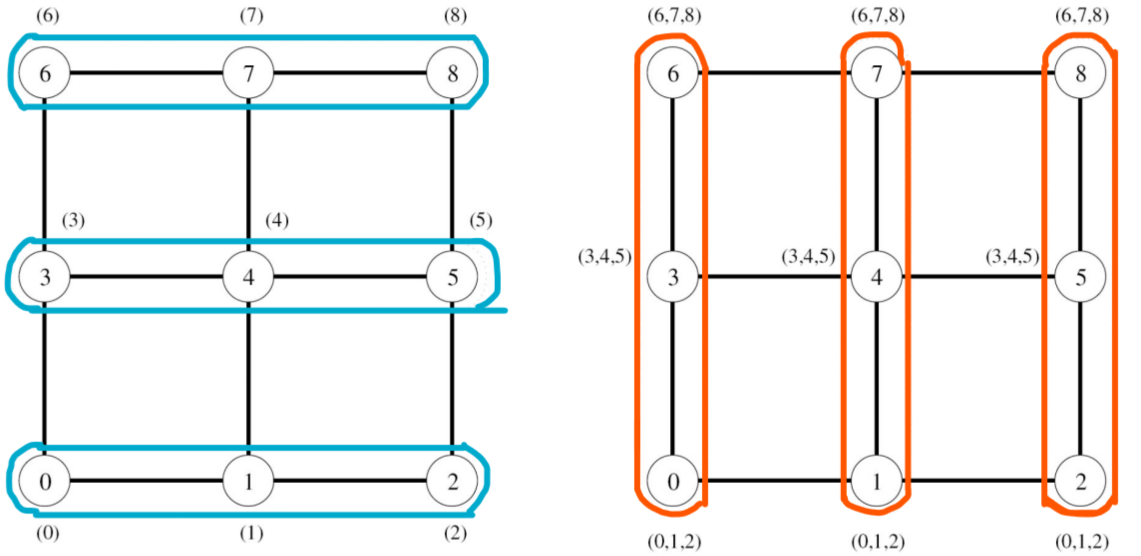
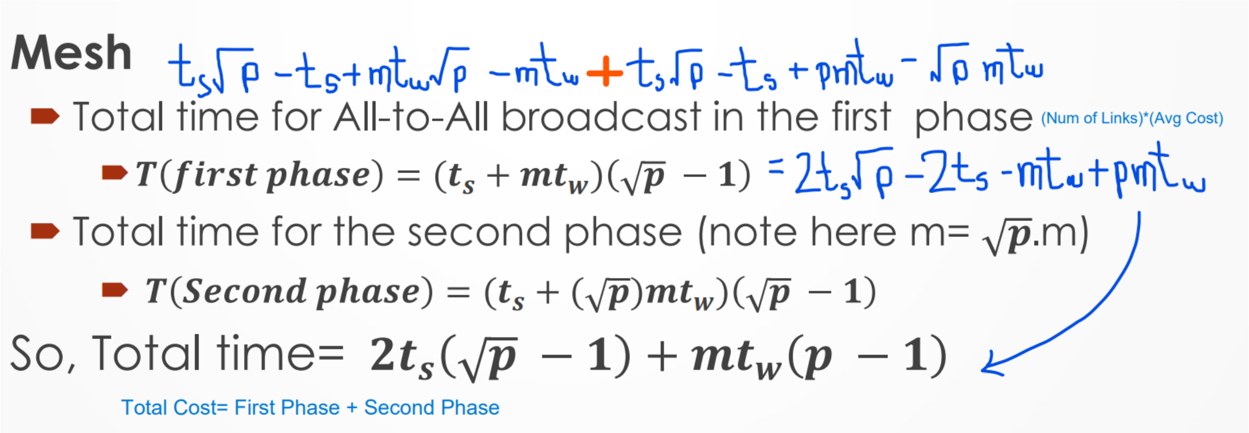


Figure 4.10 All-to-all broadcast on a 3 × 3 mesh. The groups of nodes communicating with each other in each phase are enclosed by dotted boundaries. By the end of the second phase, all nodes get (0,1,2,3,4,5,6,7) (that is, a message from each node).



Mesh

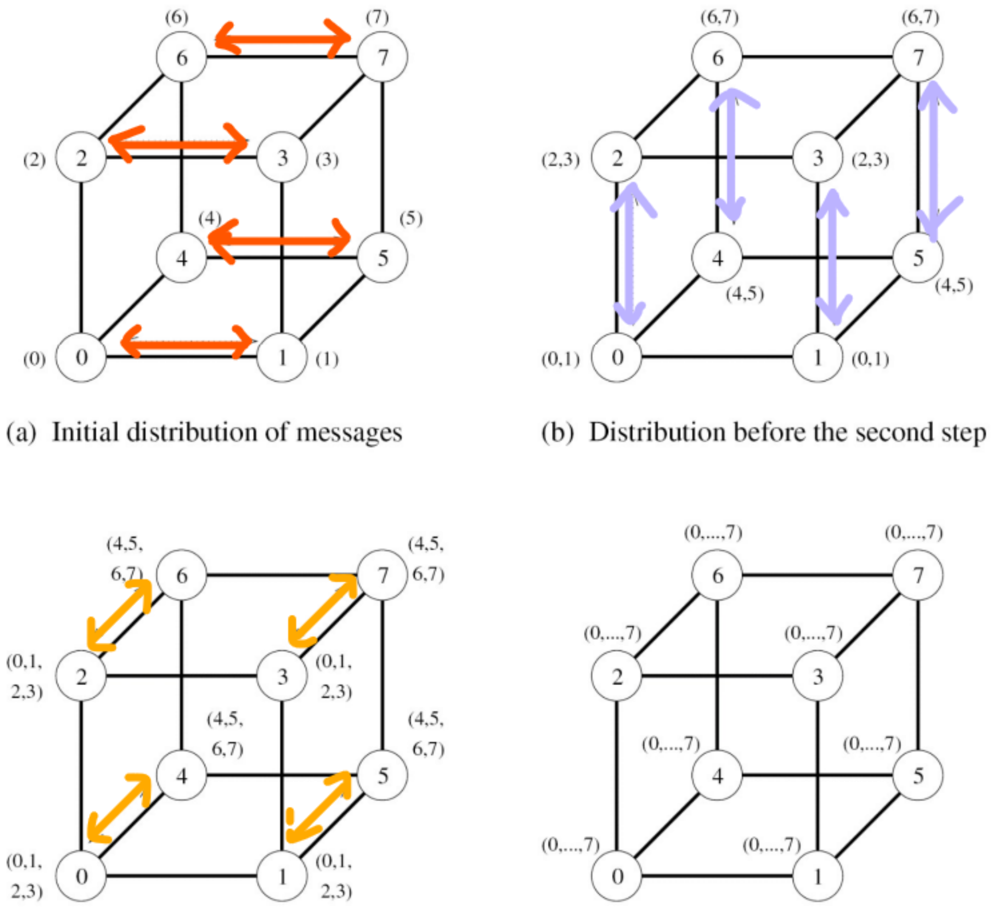
$t_s\sqrt{p} - t_s + mt_w\sqrt{p} - mt_w + t_s\sqrt{p} - t_s + pmt_w - \sqrt{p}\,mt_w$

- Total time for All-to-All broadcast in the first phase (Num of Links)*(Avg Cost)
  - $T(first\ phase) = (t_s + mt_w)(\sqrt{p} - 1)$ $= 2t_s\sqrt{p} - 2t_s - mt_w + pmt_w$
- Total time for the second phase (note here m= $\sqrt{p}$.m)
  - $T(Second\ phase) = (t_s + (\sqrt{p})mt_w)(\sqrt{p} - 1)$

So, Total time= $2t_s(\sqrt{p} - 1) + mt_w(p - 1)$

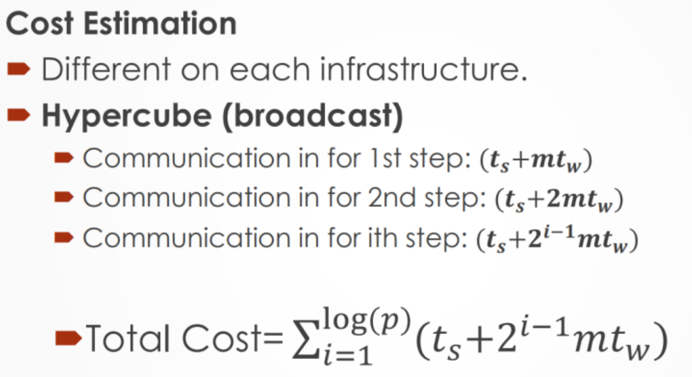
Total Cost= First Phase + Second Phase

Note: m=underroot(p)m cuz now the message length has changed and the msg len is equal to the num of rows.



(a) Initial distribution of messages

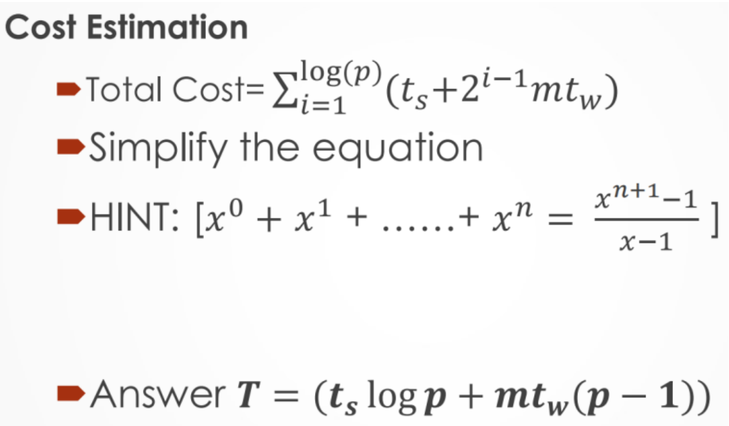
(b) Distribution before the second step



Cost Estimation

- Different on each infrastructure.
- **Hypercube (broadcast)**
  - Communication in for 1st step: $(t_s + mt_w)$
  - Communication in for 2nd step: $(t_s + 2mt_w)$
  - Communication in for ith step: $(t_s + 2^{i-1}mt_w)$

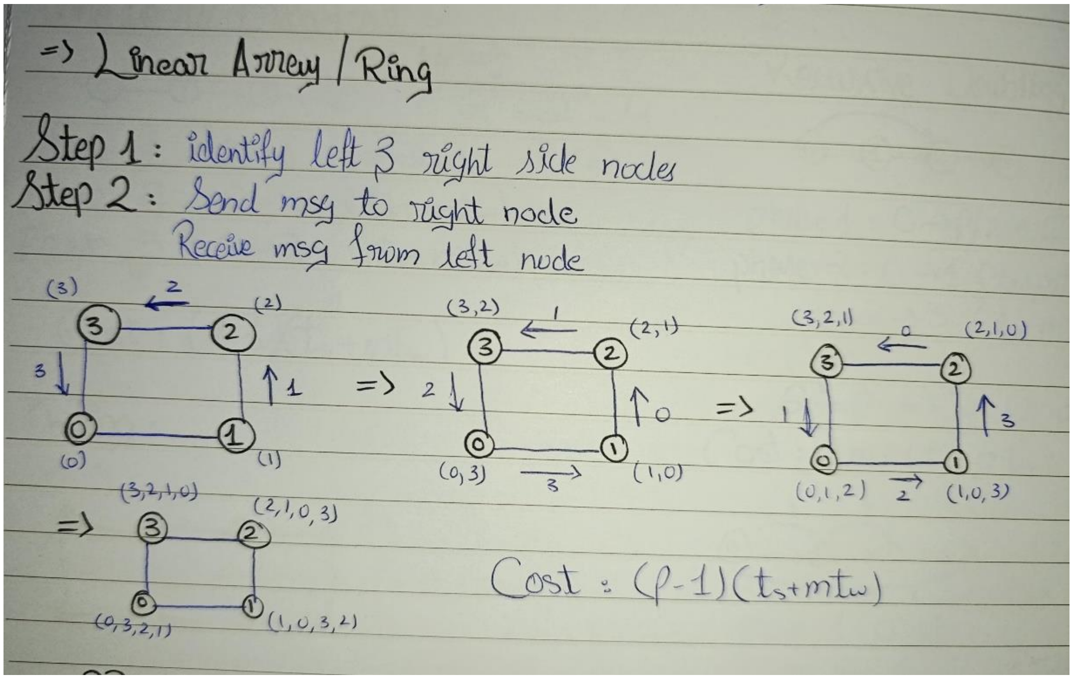
- Total Cost= $\sum_{i=1}^{\log(p)}(t_s + 2^{i-1}mt_w)$



Cost Estimation

- Total Cost= $\sum_{i=1}^{\log(p)}(t_s + 2^{i-1}mt_w)$
- Simplify the equation
- HINT: $[x^0 + x^1 + \ldots\ldots + x^n = \frac{x^{n+1} - 1}{x-1}]$

- Answer $T = (t_s \log p + mt_w(p-1))$



=> Linear Array / Ring

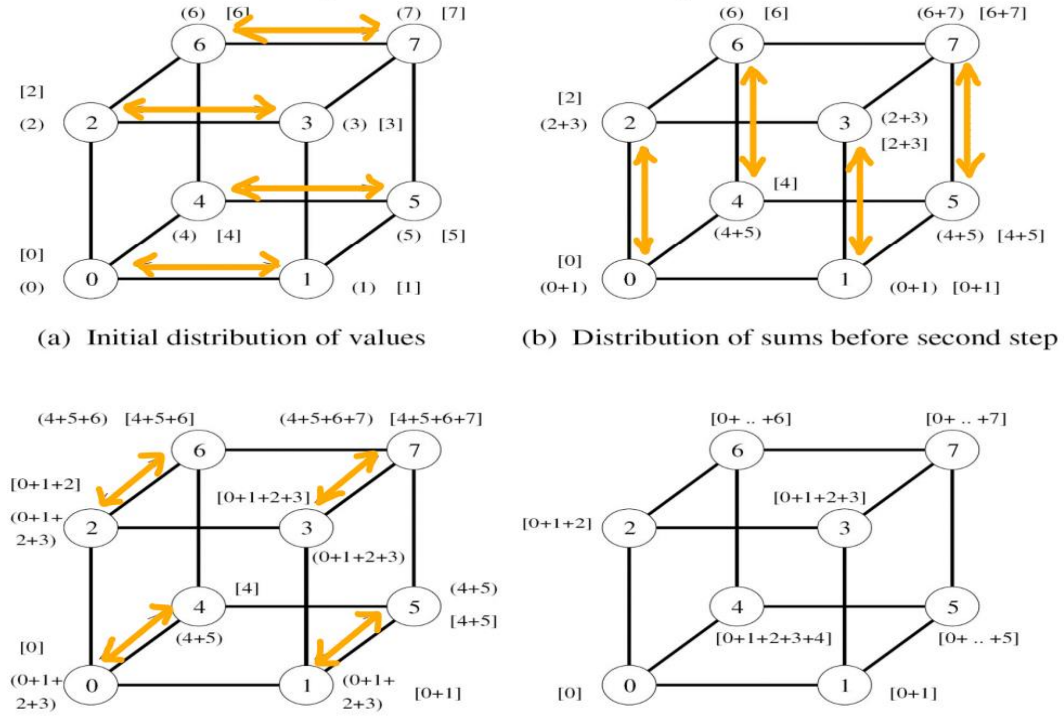
Step 1 : identify left & right side nodes
Step 2 : Send msg to right node
Receive msg from left node

Cost : $(p-1)(t_s+mt_w)$

- **Linear Ring**
  - $T = (t_s + mt_w)\,(p-1)$

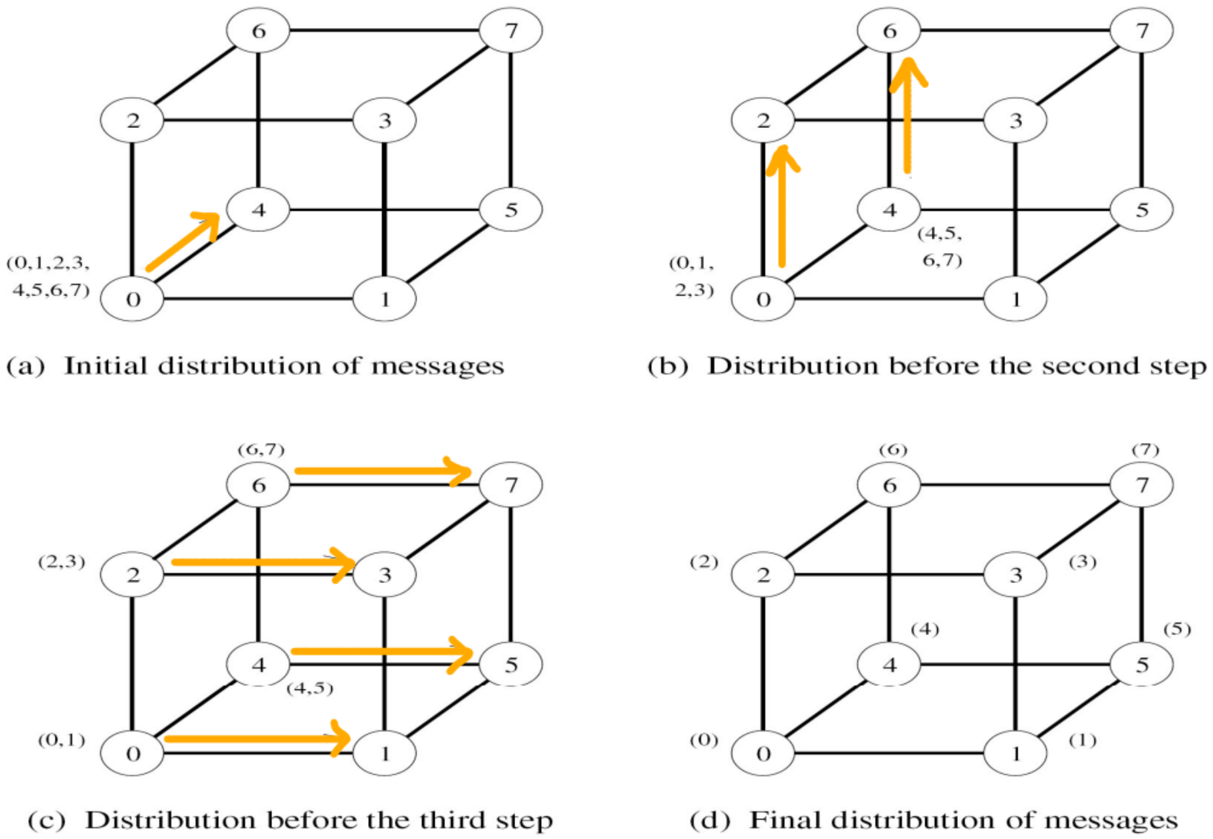
Prefix Sums:

(Prefix-Sums)



(a) Initial distribution of values

(b) Distribution of sums before second step



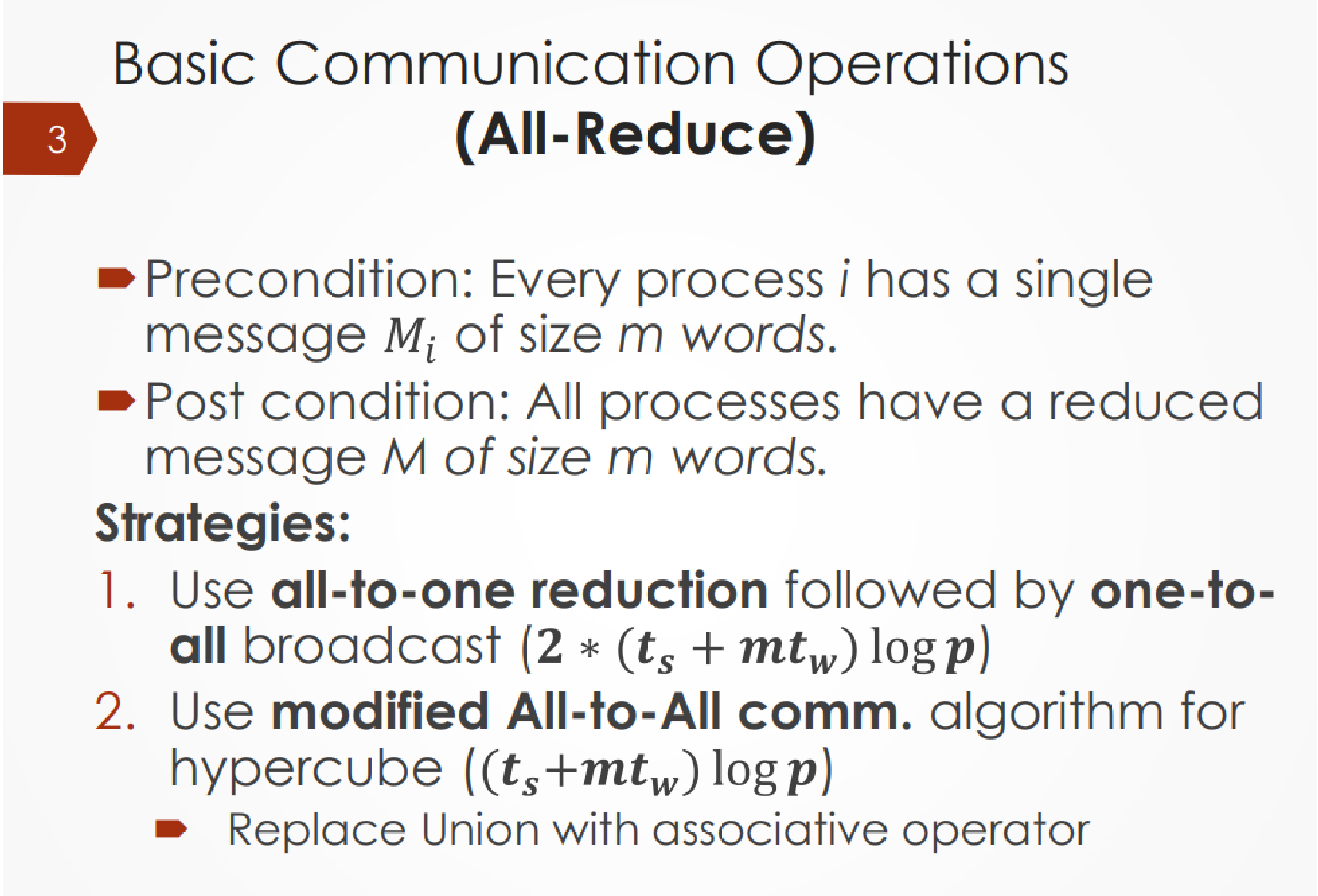
(a) Initial distribution of messages

(b) Distribution before the second step

(c) Distribution before the third step

(d) Final distribution of messages

Figure 4.15 The scatter operation on an eight-node hypercube.



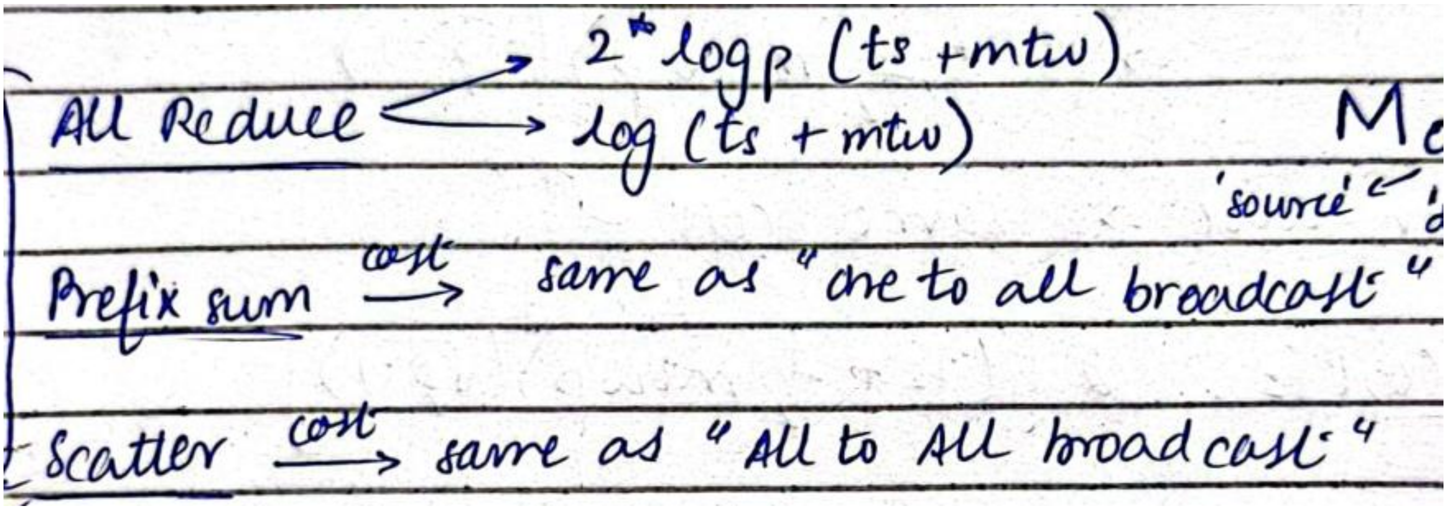
Basic Communication Operations (All-Reduce)

3

- Precondition: Every process *i* has a single message M_i of size *m words.*
- Post condition: All processes have a reduced message *M of size m words.*

Strategies:

1. Use **all-to-one reduction** followed by **one-to-all** broadcast ($\boldsymbol{2 * (t_s + mt_w) \log p}$)
2. Use **modified All-to-All comm.** algorithm for hypercube ($\boldsymbol{(t_s + mt_w) \log p}$)
   - Replace Union with associative operator



All Reduce → $2^{*} \log p \,(ts + mtw)$

All Reduce → $\log (ts + mtw)$

M_c

'source' ←

Prefix sum $\xrightarrow{cost}$ same as "one to all broadcast"

Scatter $\xrightarrow{cost}$ same as "All to All broadcast"