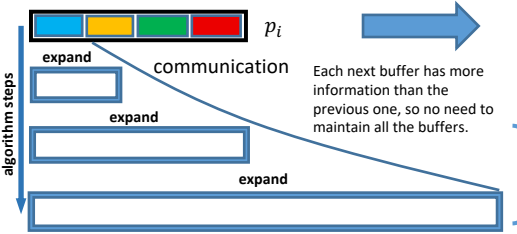
Buffers Reuse:

Our implementation decreases the total amount of memory used and the total number of buffers allocated. All the buffers are allocated just once and are then being reused throughout the application.

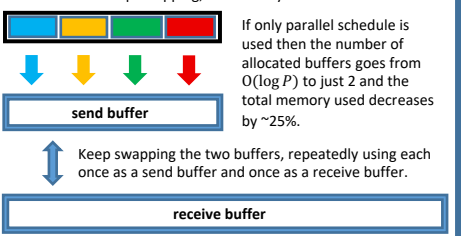
Before: each sequential subproblem allocates a new list of buffers



Each next buffer has more information than the previous one, so no need to maintain all the buffers.

Evolution of a buffer holding local data of a single matrix

Now: All subproblems reuse the same buffers. Moreover, send and receive buffers keep swapping, so that only 2 buffers suffice.



If only parallel schedule is used then the number of allocated buffers goes from $O(\log P)$ to just 2 and the total memory used decreases by ~25%.

Keep swapping the two buffers, repeatedly using each once as a send buffer and once as a receive buffer.