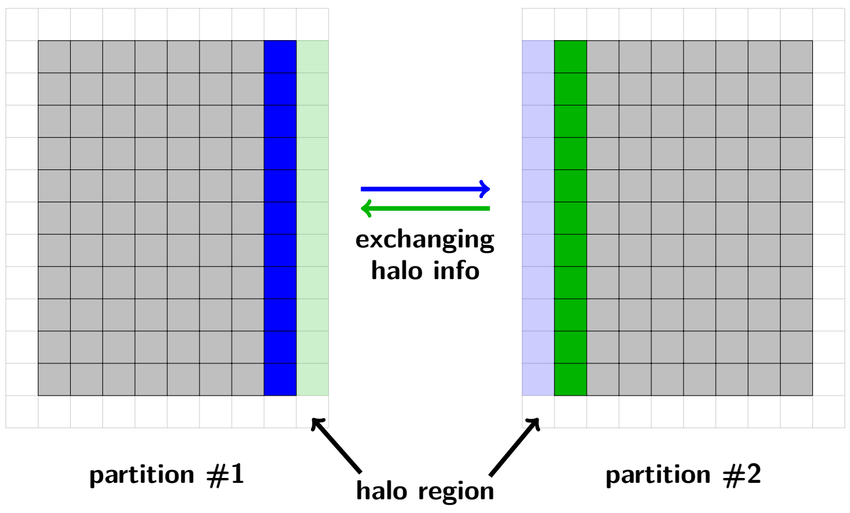
# Tausch - A Halo Exchange Library

Lukas Spies <lspies@illinois.edu>
Computer Science · University of Illinois

ILLINOIS

## Problem Statement

- Halo exchanges (both across MPI ranks and compute units) are part of most Scientific Applications:




- Halo exchanges are often the bottleneck
  $\rightarrow$ efficient handling key to performance
- re-writing halo exchange code for each project
  $\rightarrow$ bugs more likely, sophisticated performance tweaks often too cumbersome to re-implement

$\rightarrow$ **Tausch to the rescue**

## Approach

- Solution: Abstraction for halo exchanges (semi-)agnostic of underlying hardware
- inter-CPU communication: MPI
- inter-device communication: OpenCL
- dimension-independent handling of halo information
- programming language: C++ (with C API)
- header-only library $\rightarrow$ no linking necessary
- templated code $\rightarrow$ works with any data type

## Design Considerations to Boost Performance

- MPI Persistent communication decreases overhead on repeated halo exchanges
- Ability to combine messages
  $\rightarrow$ fast(er) communication
- Aligned memory buffers
- No safety checks (user responsibility)

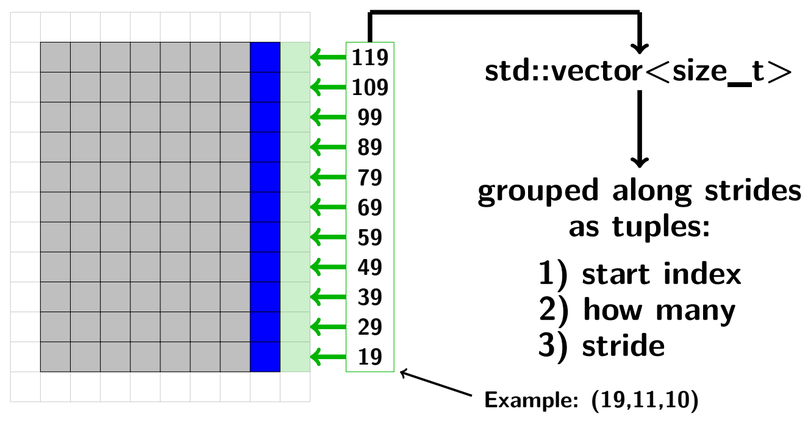
## Glossary

- **remote halo:** values needed by one partition that are computed on another partition.
- **local halo:** values computed by one partition that are needed by another partition.

## API

- Setting halo information:

```
int addLocalHaloInfo(std::vector<size_t>);
int addRemoteHaloInfo(std::vector<size_t>);
```



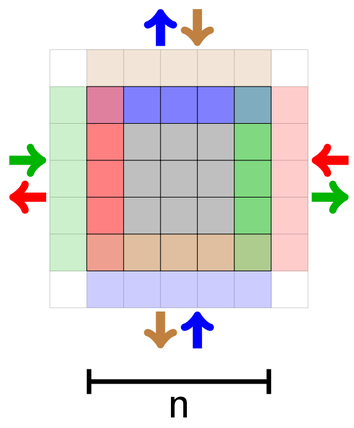

- Performing halo exchange:

```
void packSendBuffer(size_t, size_t, buf_t*);
void send(size_t, int, int);
void recv(size_t, int, buf_t*);
void unpackRecvBuffer(size_t, size_t, buf_t*);
```

- clean and straight-forward API

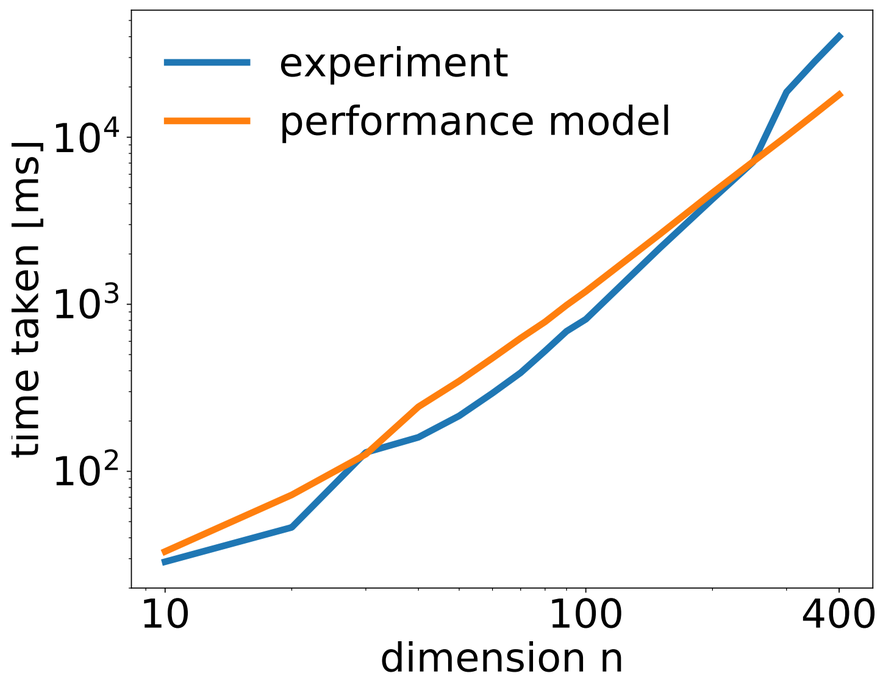
## Performance Evaluation

Performance Model in 3D:




$$T_{perf} = 6n^2\left(\frac{T_{b2h,nseq}+T_{b2h,seq}}{2} + T_{send,recv} + \frac{T_{h2b,nseq}+T_{h2b,seq}}{2}\right)$$

- 1000 iterations (one set of halo exchanges per iteration)
- test system: Blue Waters




- very simple performance model, yet it closely matches measured performance

## Acknowledgements

Supervision and/or input given by

- Luke Olson
- Andrew Reisner
- Mike Campbell

This material is based in part upon work supported by the Department of Energy, National Nuclear Security Administration, under Award Number DE-NA0002374.