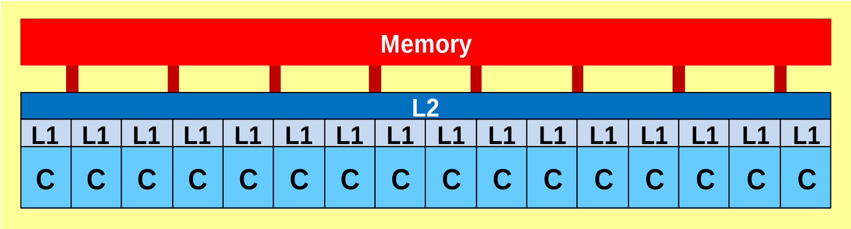
Parallel MGCG Solvers for PDE’s with Sparse Matrices

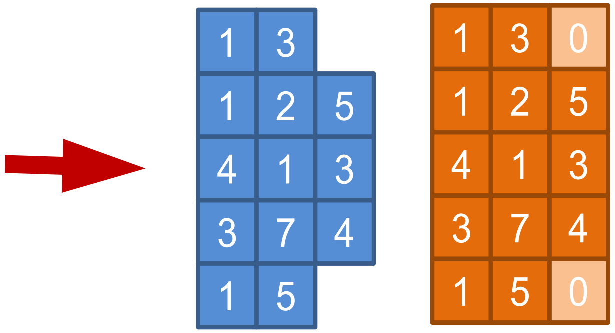
- Multigrid
  - Scalable, one of the choices for post-peta/exascale HPC
  - HPCG
- Serial Communications
  - Data Transfer through Memory Hierarchy: Sparse Matrices
- Parallel Communications
  - Message Passing through Network
- Fujitsu FX10 (Oakleaf-FX)
  - up to 4,096 nodes (~1PF)
  - Weak/Strong Scaling
  - Hybrid/Flat-MPI



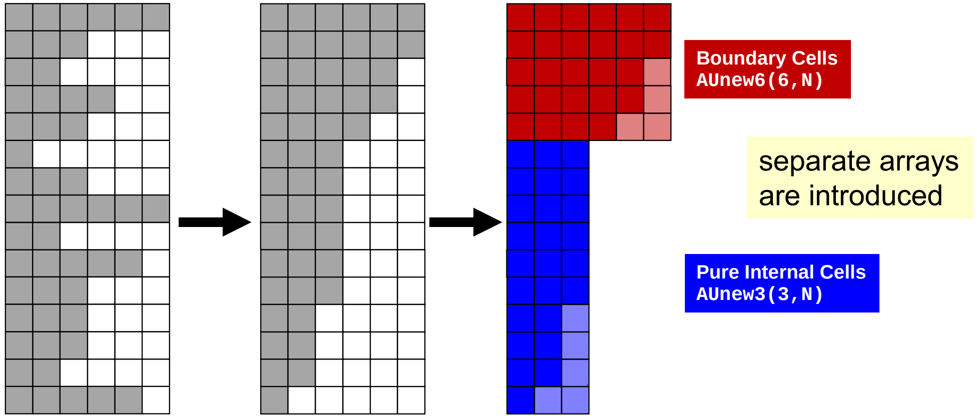
Optimization of Serial Comm.

ELL (Ellpack-Itpack), Sliced-ELL for Matrix Storage

$$\begin{bmatrix} 1 & 3 & 0 & 0 & 0 \\ 1 & 2 & 5 & 0 & 0 \\ 4 & 1 & 3 & 0 & 0 \\ 0 & 3 & 7 & 4 & 0 \\ 1 & 0 & 0 & 0 & 5 \end{bmatrix}$$



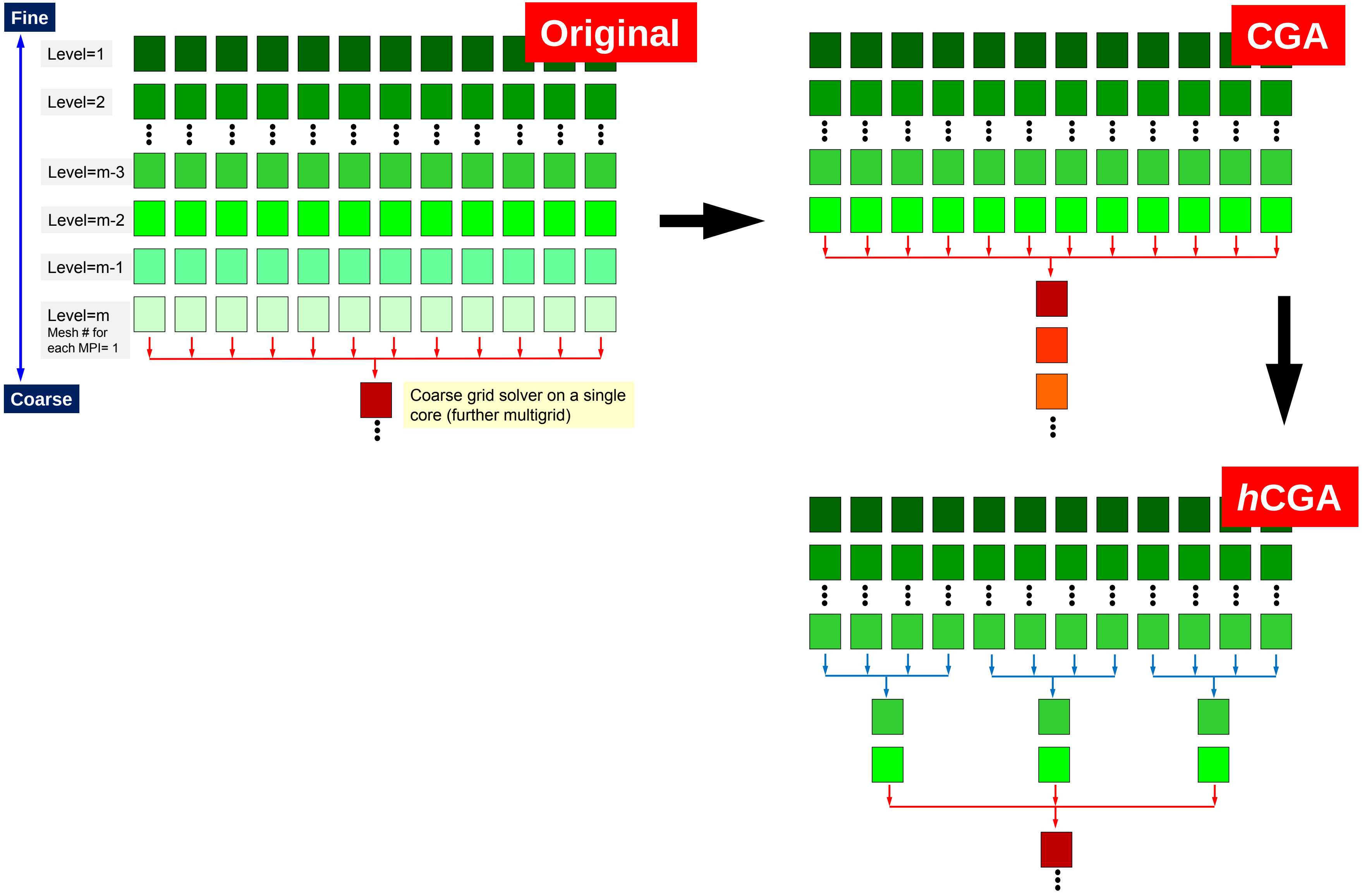
(a) CRS (b) ELL



- **A lot of X-ELL-Y-Z’s !**
  - **focusing on SpMV**
- SELL-C-σ
  - M. Kreutzer et al
- Recently, X-ELL-Y-Z’s are applied to forward/backward substitutions with data dependency
  - HPCG: SC14 BoF
  - Gauss-Seidel: Easy
- ILU
  - Much more difficult than GS

Optimization of Parallel Comm.

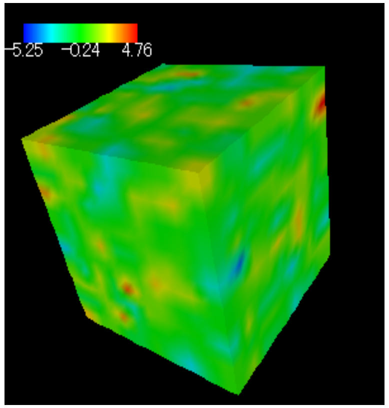
CGA(Coarse Grid Aggregation) -> *h*CGA

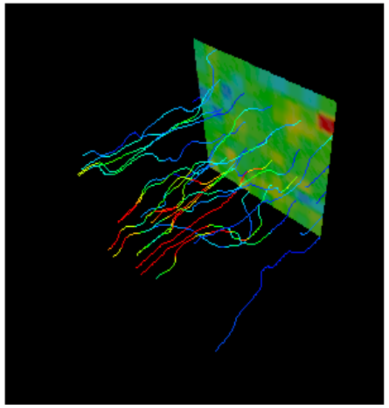


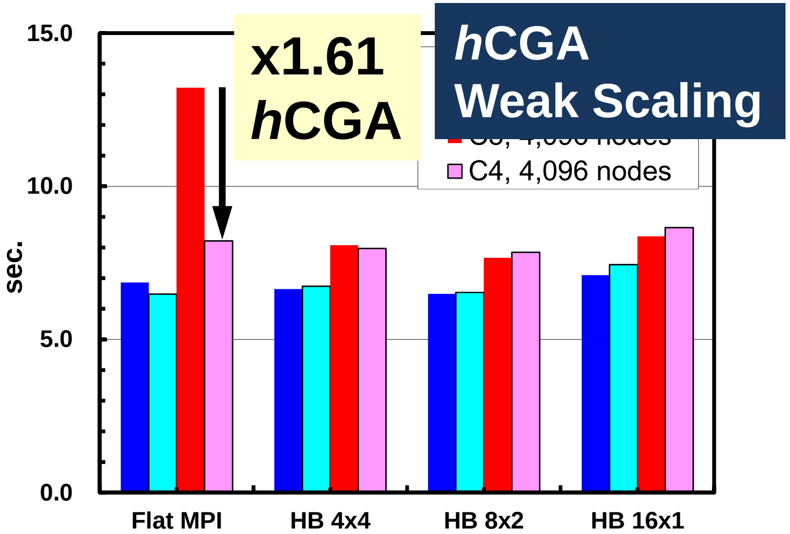
Coarse grid solver on a single core (further multigrid)

GW Flow Simulation with up to 4,096 nodes on Fujitsu FX10 (GMG-CG)

up to 17,179,869,184 meshes (64^3 meshes/core)

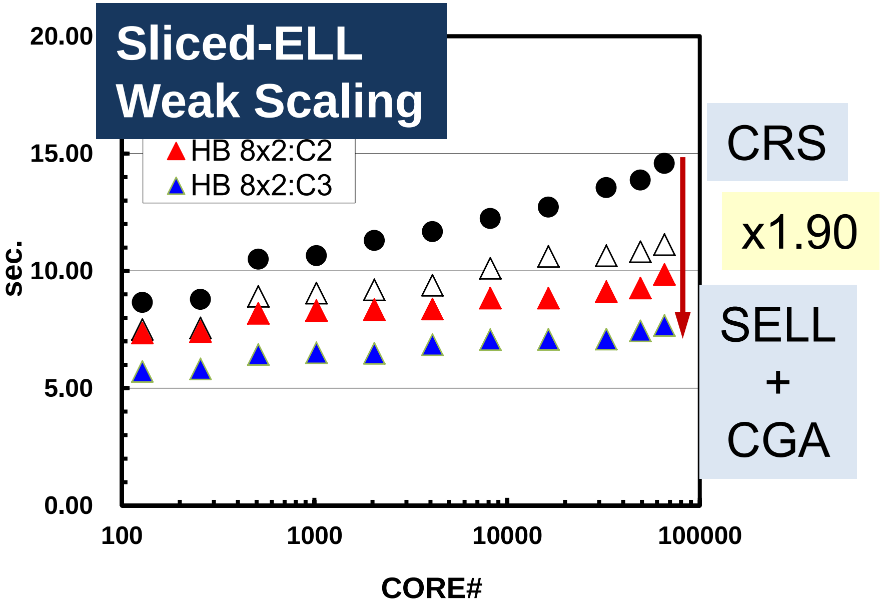






x1.61 *h*CGA

***h*CGA Weak Scaling**

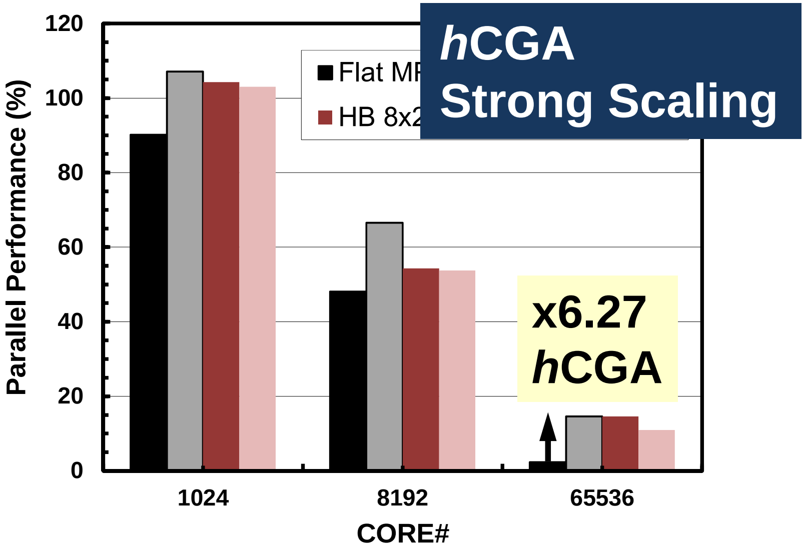


Sliced-ELL Weak Scaling

CRS

x1.90

SELL + CGA



***h*CGA Strong Scaling**

x6.27 *h*CGA

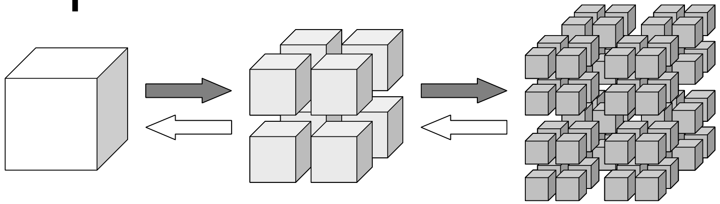
Summary

- **CRS-> Sliced ELL + CGA: x1.90**
- **CGA->*h*CGA is good for Flat-MPI**
  - **x1.61 for weak scaling, x6.27 for strong scaling at 4,096 nodes of Fujitsu FX10**
    - **Performance of coarse grid solver has been much improved**
  - ***h*CGA will be effective for HB 16x1 with more than 2.50x10^5 nodes (= 4.00x10^6 cores) of FX10 (=60 PFLOPS)**
    - Computational amount of coarse grid solver for each core of *flat MPI* is 256 (=16×16) times as large as that of HB 16×1.
    - Therefore, *h*CGA is expected to be really effective for HB 16×1 with more than 2.50×10^5 nodes (4.00×10^6 cores) of Fujitsu FX10, where the peak performance is more than 60 PFLOPS.
    - Comp. time of coarse grid solver is significant for Flat MPI with >10^3 nodes

HPC+BDA

Co-Development & Sharing Common Libraries

- Automatic Tuning using Machine Learning
  - Multigrid: So many parameters: performance, convergence
  - Xeon Phi: Combinatorial explosion for selection of optimum prefetching distance
  - Already some examples AT research with machine learning
- Visualization
  - Simplified Parallel Visualization using Background Voxel developed in ppOpen-HPC
    - http://ppopenhpc.cc.u-tokyo.ac.jp/
  - Octree-based AMR
  - Feature Detection & Simplification: 10^9 -> 10^4 meshes



Compiler Options for MIC

	AC-2		BC-2	
	sec.	GFLOPS	sec.	GFLOPS
-O3 -openmp -mmic -align array64byte (base)	2.654	10.53	2.603	10.74
-opt-streaming-stores always	3.165	8.832	3.159	8.849
-opt-streaming-cache-evict=0	2.625	10.65	2.600	10.75
-opt-streaming-cache-evict=1	2.639	10.59	2.605	10.73
-opt-streaming-stores always -opt-streaming-cache-evict=0	2.486	11.24	2.539	11.01
-opt-streaming-stores always -opt-streaming-cache-evict=1	2.477	11.29	2.556	10.94
-opt-streaming-stores always -opt-streaming-cache-evict=0 -opt-prefetch-distance=a,b	2.385 (2,0)	11.72	2.477 (8,1)	11.29
-opt-streaming-stores always -opt-streaming-cache-evict=1 -opt-prefetch-distance=c,d	2.404 (2,0)	11.63	2.487 (16,1)	11.24