Reducing Data Movement using Cache-Oblivious Layouts

Peter Lindstrom

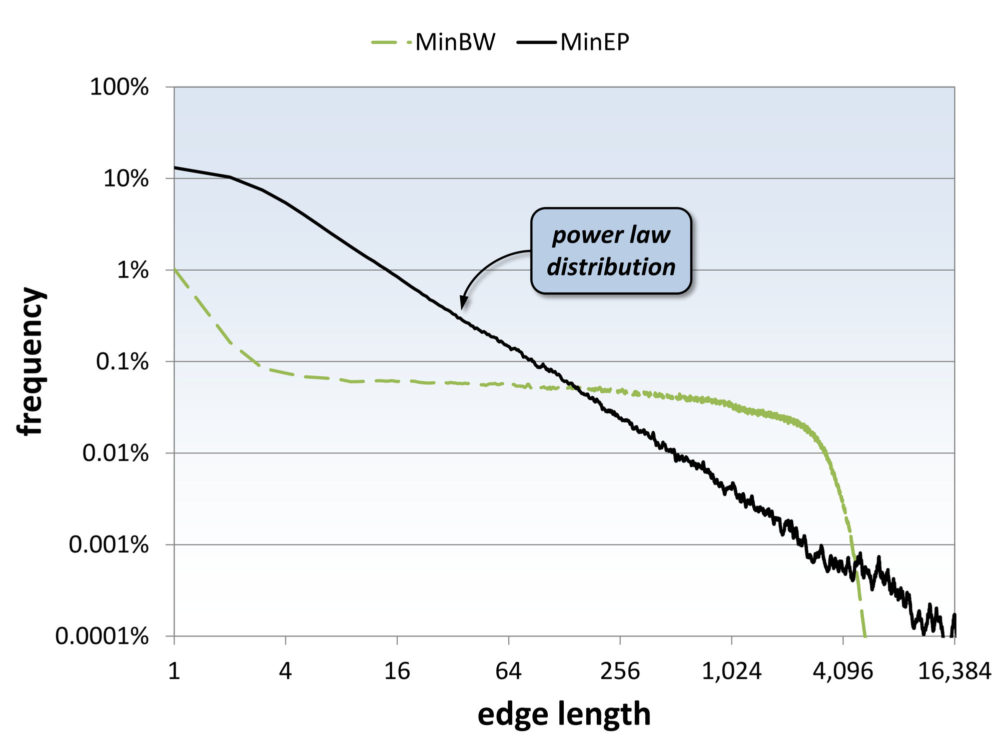
Lawrence Livermore National Laboratory

One of the key challenges to achieving high performance in extreme-scale computing is how to limit data movement between memory banks, distributed compute nodes, and storage media. Via caching, the need to move data can be reduced, as long as the access pattern and data organization exhibit locality. To this end, space-filling curves are often used to linearly order multi-dimensional grids, but are ineffective for unstructured or non-geometric data. To address this problem, we have developed a simple measure of spatial locality for data that can be modeled as an affinity graph, whose edges encode which data elements to store close together. Our measure is designed to be cache-oblivious, to account for the heterogeneity of most memory hierarchies. We show that data layouts optimized for our measure generalize space-filling curves to unstructured graphs, and that our layouts are effective in reducing data movement in a variety of applications.

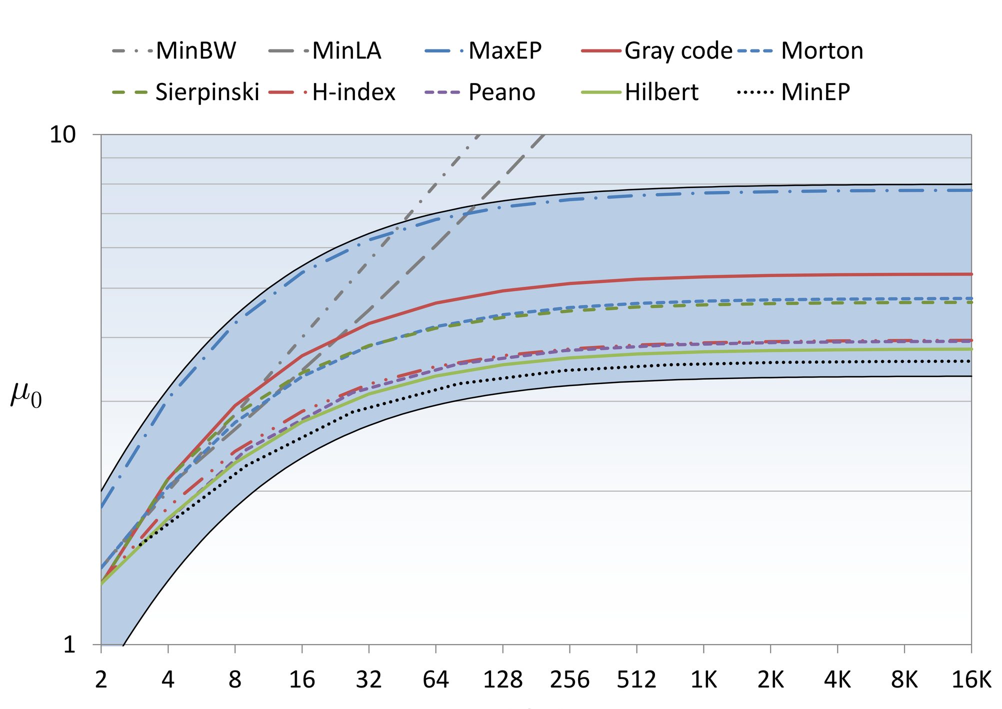
$$\mu_0 = \left(\prod_{ij\in E} |\varphi(i) - \varphi(j)|\right)^{1/|E|}$$

$$= \exp\left(\frac{1}{|E|}\sum_{ij\in E} \log|\varphi(i) - \varphi(j)|\right)$$

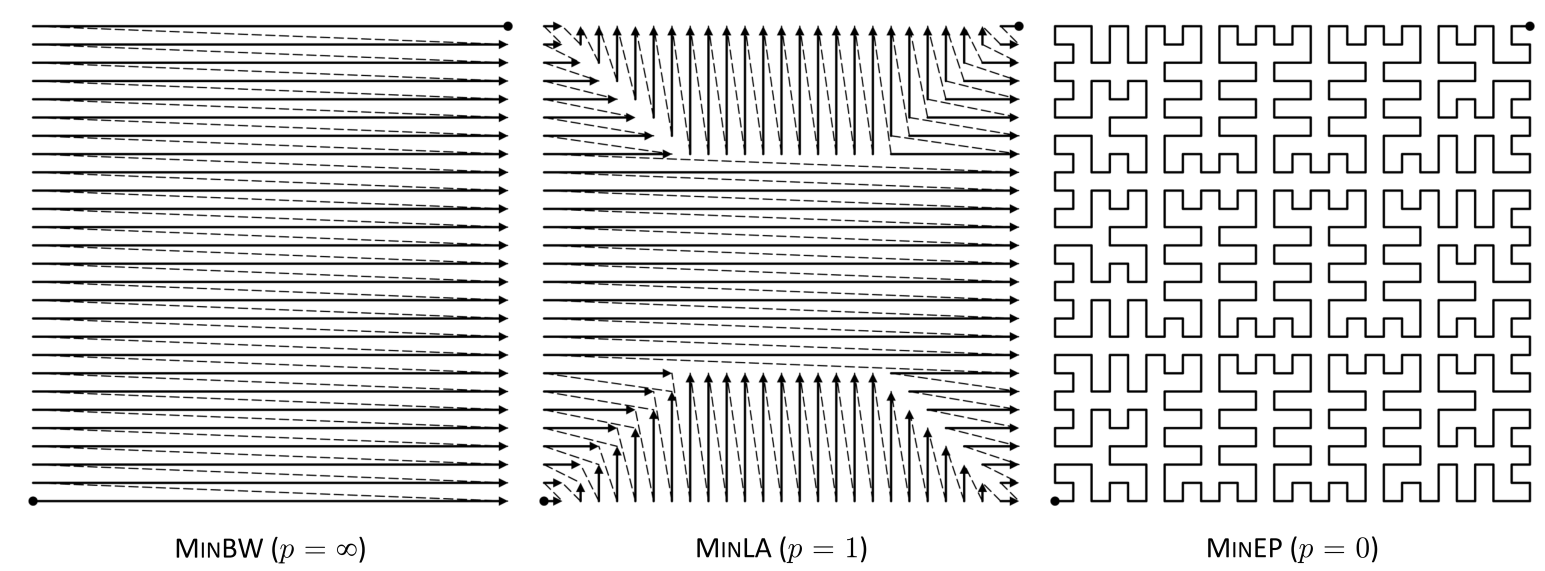
1. Given a graph $G(V, E)$, the ***minimum edge product*** (MinEP) problem is to find a linear ordering $\varphi : V \to \{1, ..., |V|\}$ of the nodes V that minimizes the ***geometric mean*** edge length μ_0.



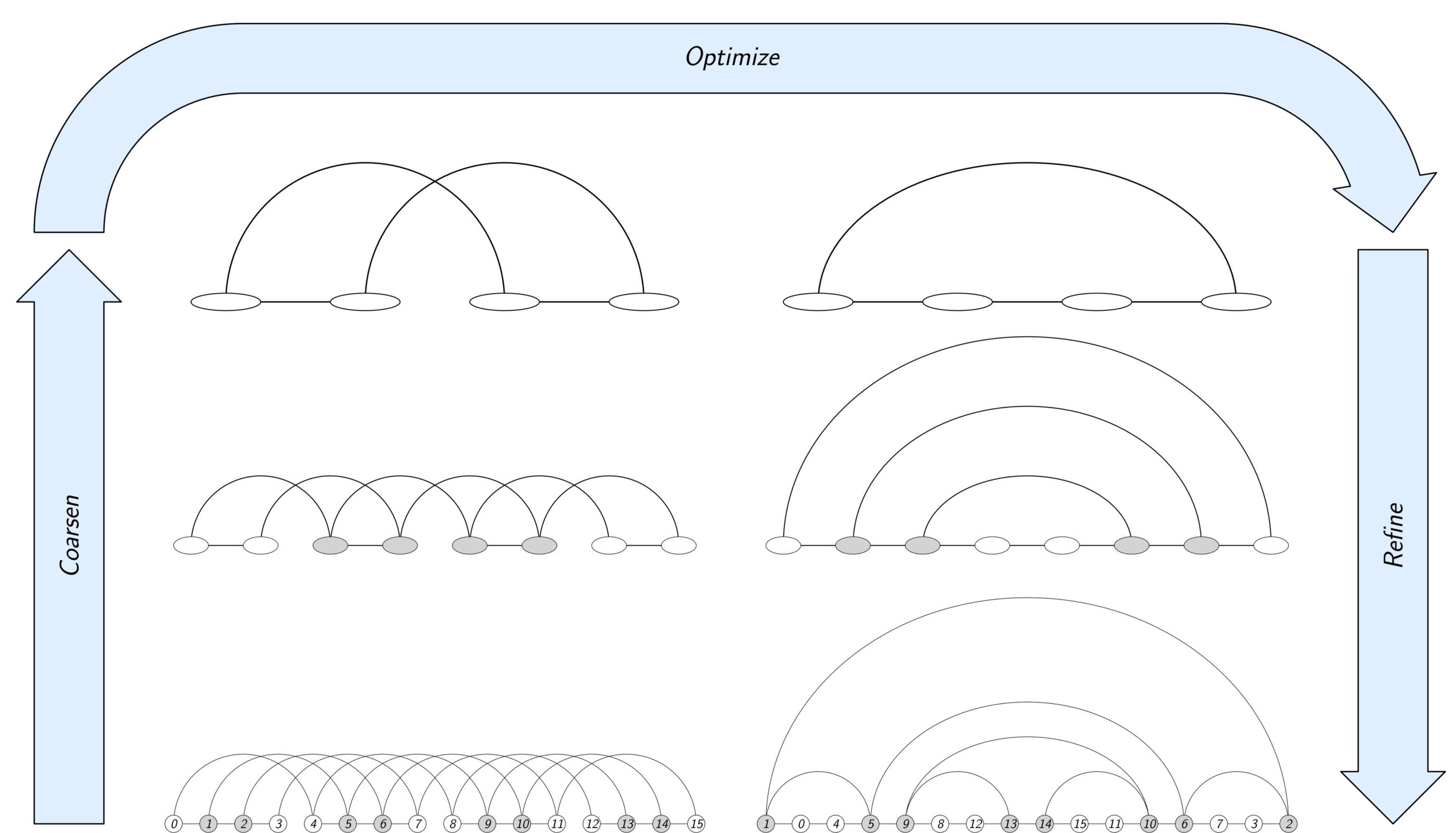
2. MinEP edge length distributions tend to be ***scale-free***. By allowing occasional long edges, many short edges can be formed that "fit" in cache, which provides good multiscale locality. We call such linear orderings ***cache-oblivious layouts***.



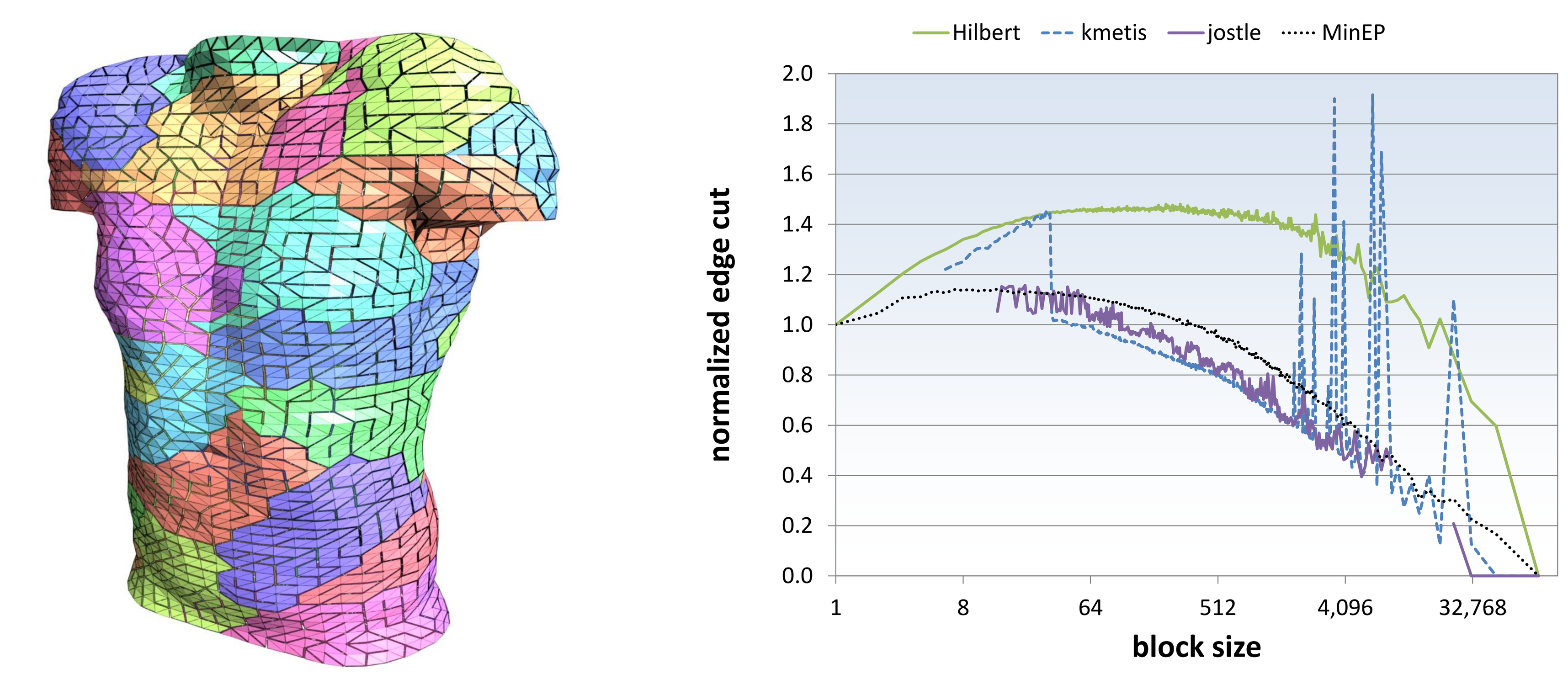
3. For regular grids ***recursively partitioned*** into tiles of contiguous nodes, our μ_0 measure is ***bounded***, regardless of grid size and how nodes within each tile are ordered.



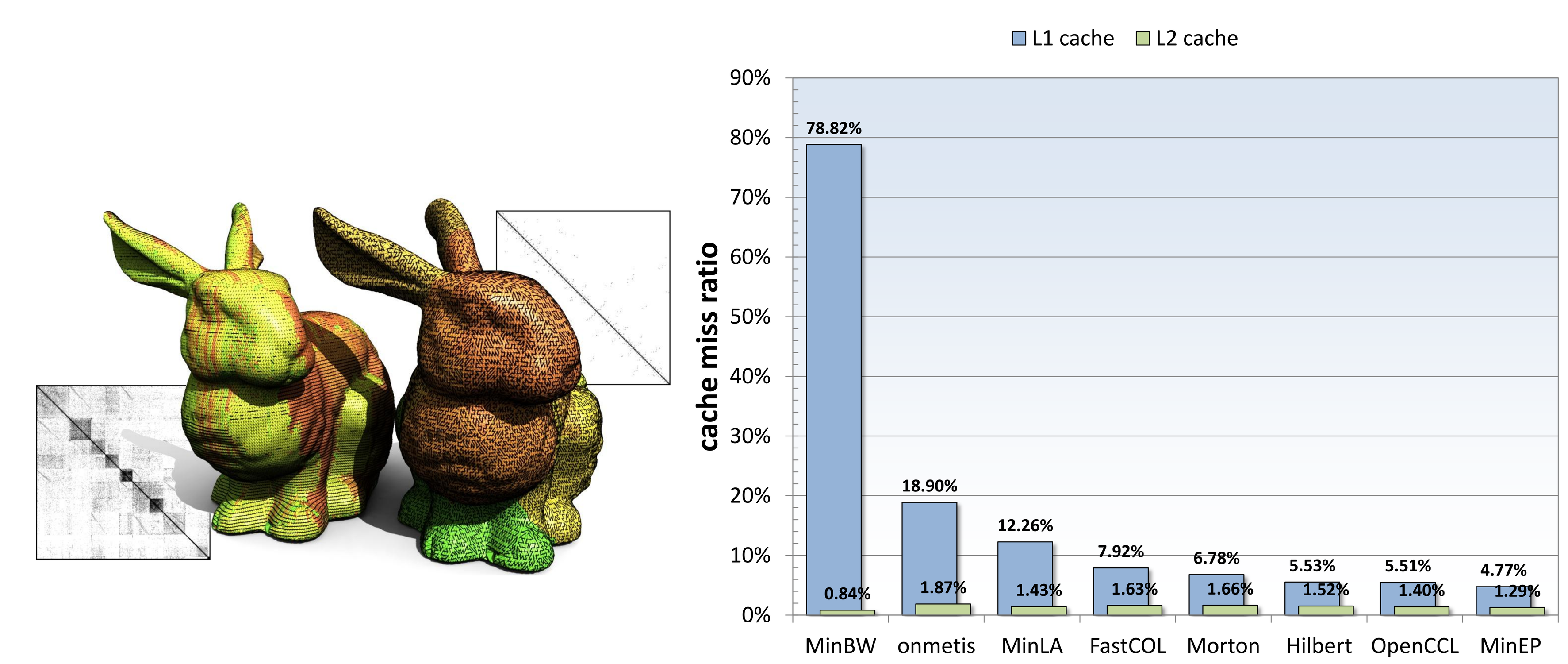
4. MinEP layouts of regular grids appear "fractal" and resemble ***space-filling curves***. For unstructured graphs, MinEP layouts tend to be ***near-Hamiltonian*** and ***hierarchical***.



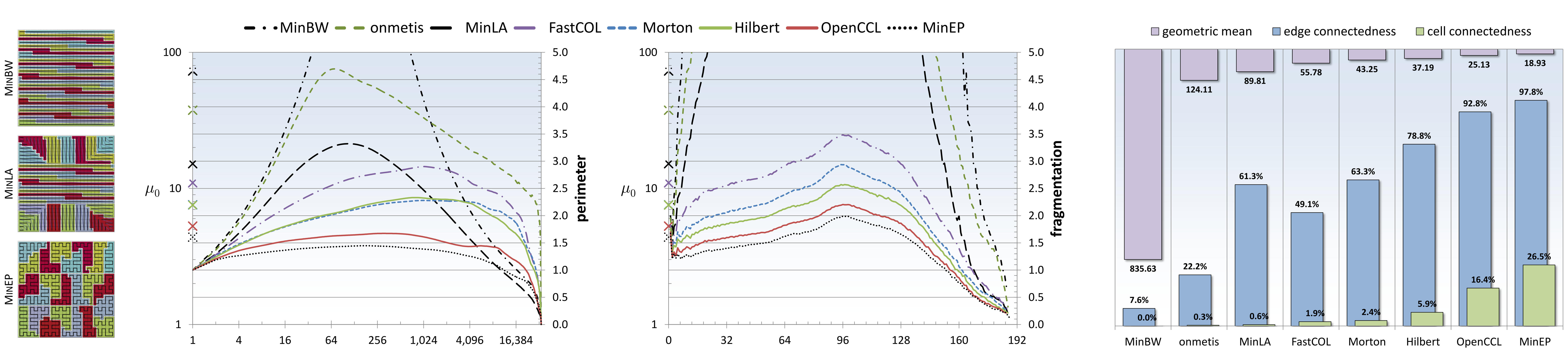
5. MinEP layouts may be computed efficiently using ***algebraic multigrid*** techniques.



7. MinEP is equivalent to minimizing the edge cut at multiple scales. Hence, our layouts support fast ***graph partitioning*** by dividing the nodes into contiguous equal-sized and well-ordered blocks.



8. MinEP layouts accelerate linear algebra routines such as ***sparse matrix-vector multiplication***. Non-compulsory L1 cache misses may be reduced by 15x or more over MinBW layouts.



6. Our $O(|E|)$ measure μ_0 correlates with more expensive $O(|V|^2)$ locality measures, such as ***domain perimeter*** and ***range fragmentation***. Consequently, MinEP layouts score well on such measures.

LLNL-POST-505071

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.