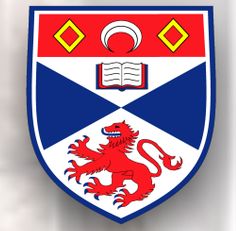
Large-Scale Hierarchical *k-means* for Heterogeneous Many-Core Supercomputers

Liandeng Li, **Teng Yu**, Wenlai Zhao, Haohuan Fu,
Chenyu Wang, Li Tan, Guangwen Yang, John Thomson

School of Computer Science
University of St Andrews

ty33@st-andrews.ac.uk




国家超级计算无锡中心
National Supercomputing Center in Wuxi


Collaborators

Liandeng Li,
Wenlai Zhao,
Haohuan Fu,
Guangwen Yang

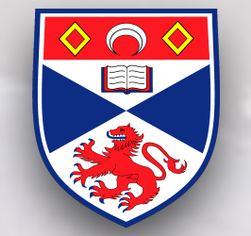
Tsinghua University
National Supercomputer Centre in Wuxi, China

Teng Yu
John Thomson

School of Computer Science
University of St Andrews
UK

Chenyu Wang,
Li Tan

Beijing Techology and Business University, China



University of St Andrews

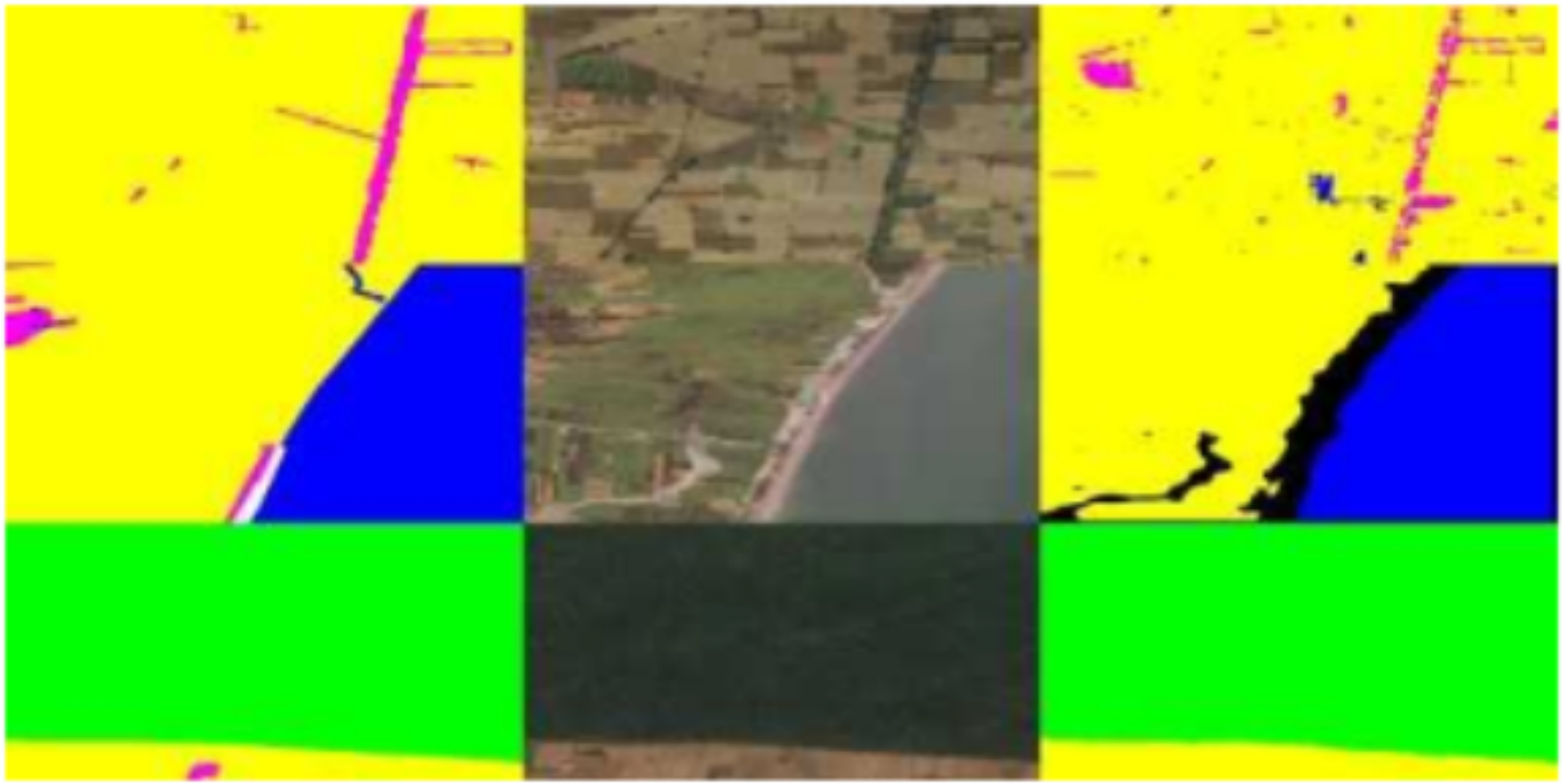
- Cutting edge research, within the top 100 world universities for at least 10 years (QS ranking)

- First university in Scotland, founded 1413

- Beautiful small town on the east coast of Scotland and home of golf

Large-Scale Hierarchical *k-means*

Motivation



- Source of original image:
  Demir, Ilke, et al. "Deepglobe 2018:
  A challenge to parse the earth through
  satellite images." *CVPR 2018*
  Satellite Challange.

State-Of-The-Art

Parallel K-Means:

-> Data Partition by Multiple Procressing Units (***n-partition***)

-> Achieves large-scale centroids: up-to 512 centroids

[Kumar, 2011][Rossbach, 2013][Cai, 2015]

Two-level Parallel K-Means:

-> Data Partition using Multiple Procressing Units and Centroids Partition using Shared Memory (***nk-partition***)

-> Achieves large-scale dimensions: up-to 140,256 dimensions

[Bender, 2015]

Complicated Data

Two examples:

Scientific Benchmark Image:
-> ILSVRC2012 (ImgNet): ***1.3E6*** data samples with ***196,609*** dimensions of each, targeting ***160,000*** centroids

Real-world Remote Sensing Image:
-> Deep Globe 2018: ***5.8E5*** data samples with ***4,096*** dimensions of each,targeting ***7*** centroids

State-Of-The-Art

Critical Problem:

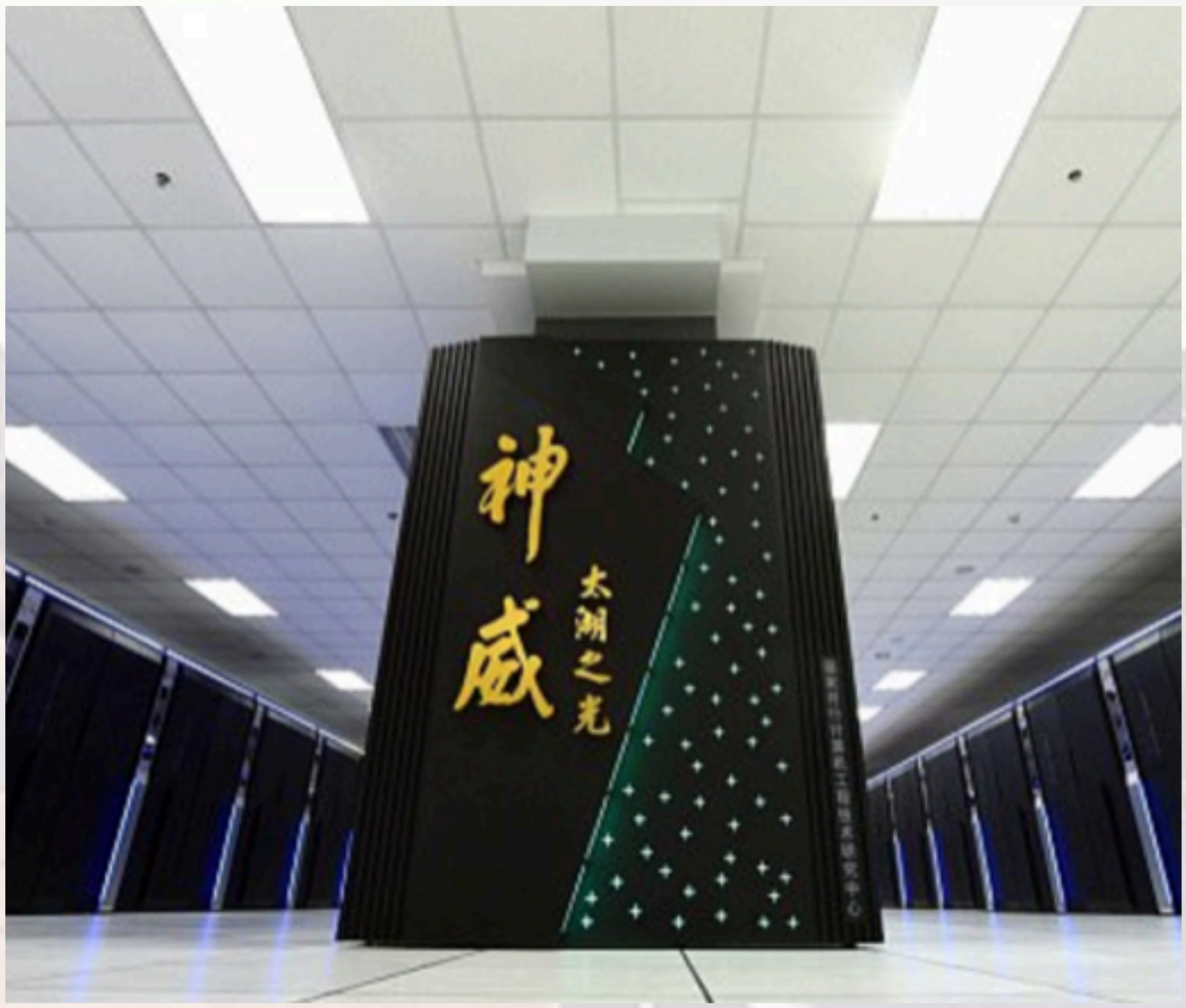
-> Not possible to large-scale data samples (n), centroids (k) and dimensions (d) simultaneously with simple data partitioning

-> ***n*d*** or ***n*k*** is limited by the size of local memory

-> Not possible to large-scale both centroids (k) and dimensions (d) simultaneously with two-level partitioning

-> ***k*d*** is limited by the size of shared memory

Multi-level Data Partitions targeting Heterogeneous Hardware

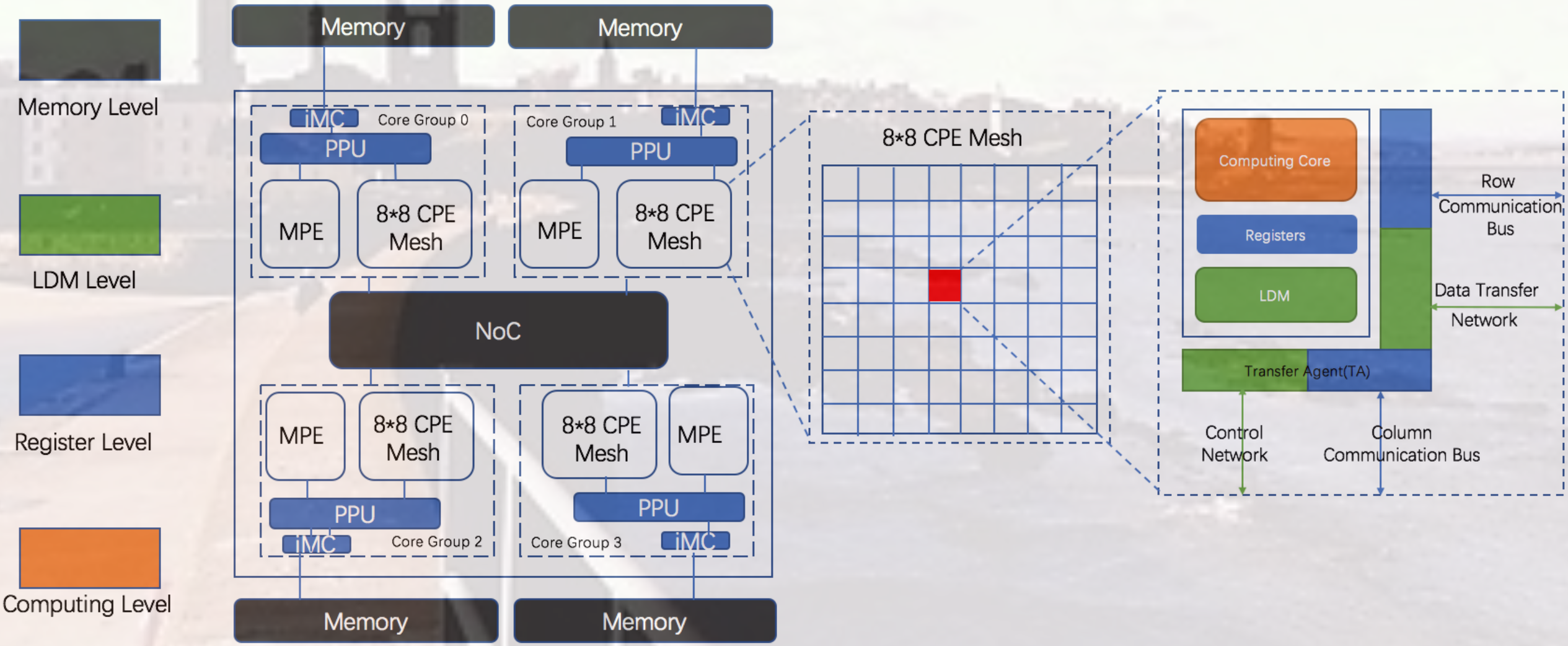


Sunway TaihuLight

- *Top1 Supercomputer from June 2016 - June 2018*
- *Up to 40,960 processors with 10,649,600-bit RISC cores*
- *Up to 20 PB storage*
- *Up to 125 pflops peak perfromance*

Multi-level Data Partitions targeting Heterogeneous Hardware

SW26010 Manycore Processor



Multi-Level Data Partitions

Level 1

DataFlow (***n***)
Partition

n d-dimensional data samples
k d-dimensional target centroids

· · · ***m*** CPEs

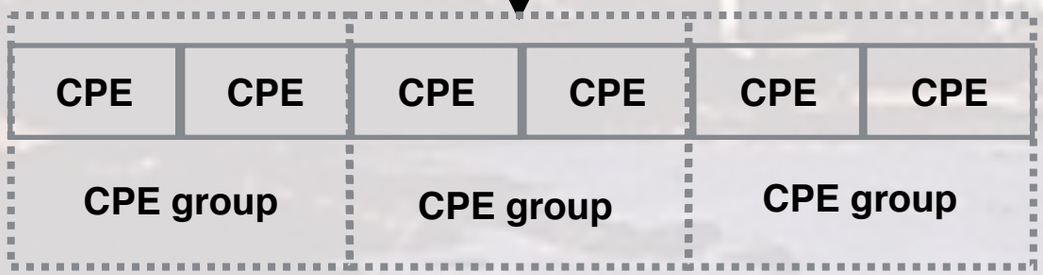
n/m d-dimensional data sample for each CPE
k d-dimensional centroids for each CPE

Multi-Level Data Partitions

Level 2

DataFlow (***n***)
Centroids (***k***)
Partition

n d-dimensional data samples
k d-dimensional target centroids

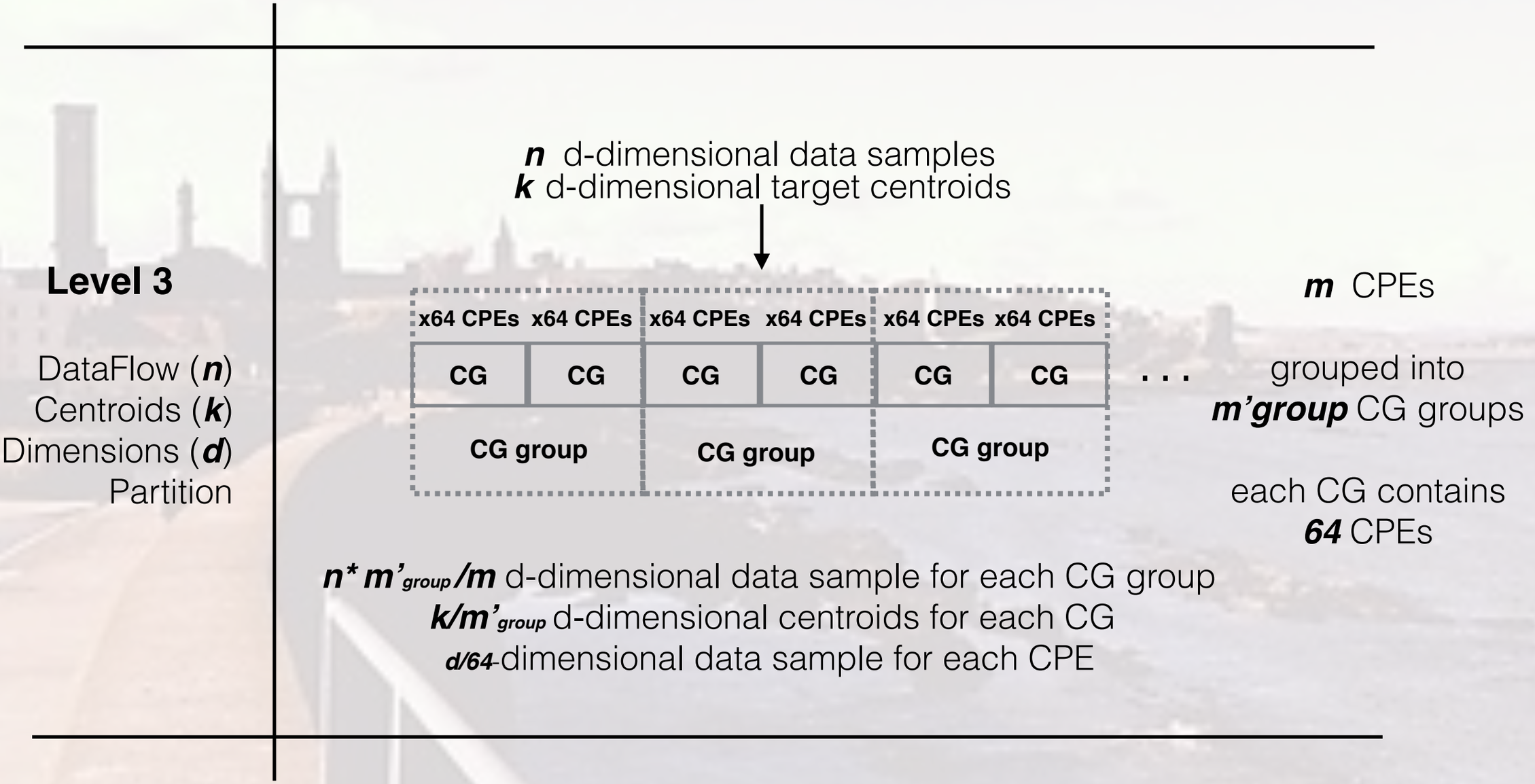


m CPEs

. . . grouped into
mgroup CPE groups

n* mgroup /m d-dimensional data sample for each CPE group
k/mgroup d-dimensional centroids for each CPE

Multi-Level Data Partitions



Multi-Level Data Partitions

The size of centroids and dimensions are
no longer constrainted by shared memory

Level 1:	$\mathbf{d(1 + k + k) + k \leq LDM}$
Level 2:	$\mathbf{d(1 + k + k) + k \leq m_{group} * LDM\ (m_{group} \leq 64)}$
Level 3:	$\mathbf{d(1 + k + k) + k \leq m * LDM}$

Multi-Level Data Partitions

Hierarchical data communication approaches
to guarantee high performance

Memory Access	**DMA**	32 GB/s
Multi-CPE Communication (Intra-CG)	**Register Communcation**	46.4 GB/s
Multi-Node Communcation (Inter-CG)	**MPI**	16 GB/s

Contributions

Fully Partition of Dataflow, Centroids and Data dimensions

Handle big dataset with both large-scale dimensions and large-scale centroids

Multi-level Large-scale

Handle both high dimensional and low dimensional dataset efficiently on supercomputer.

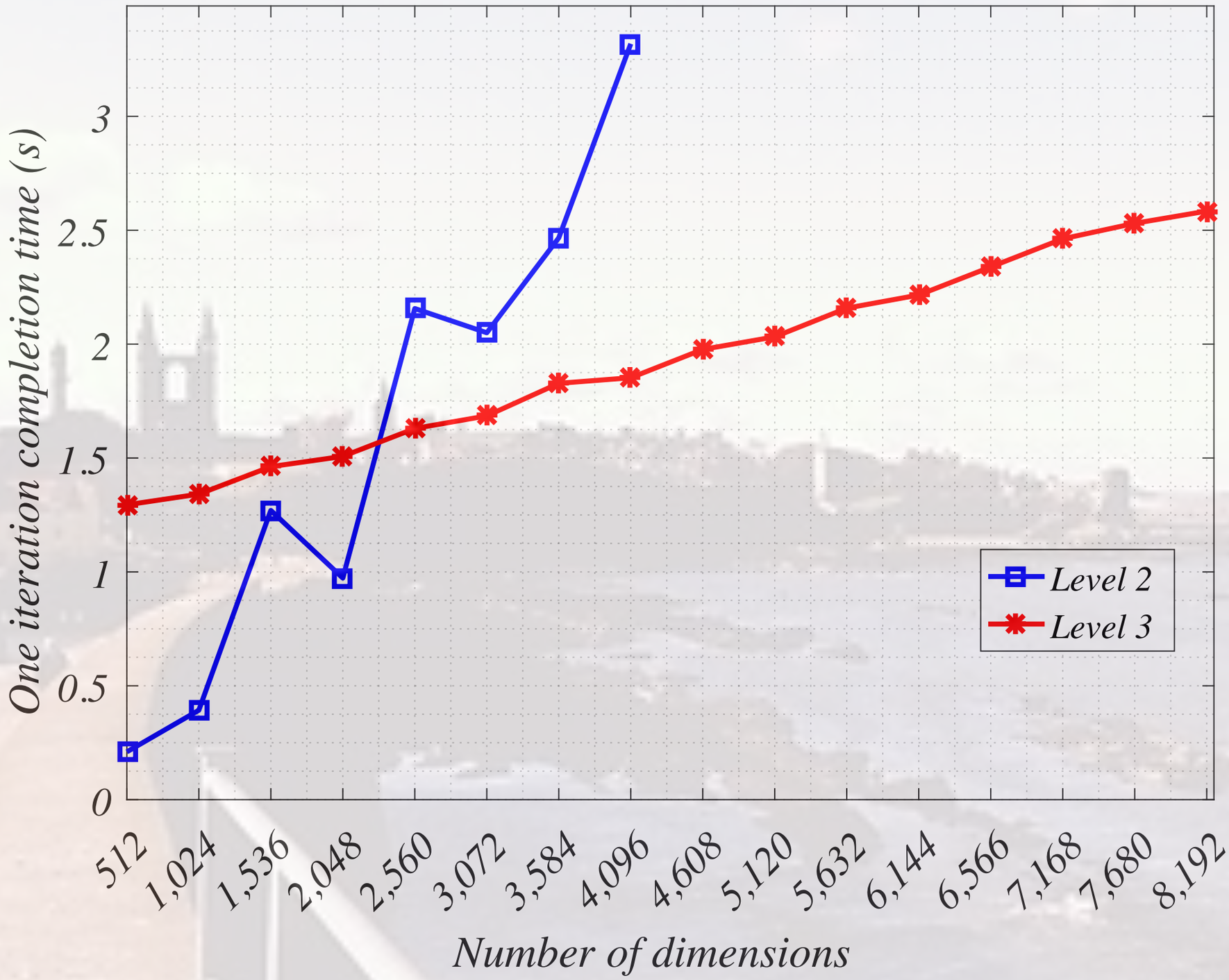
Experimental Evaluation

Scientific Benchmarks

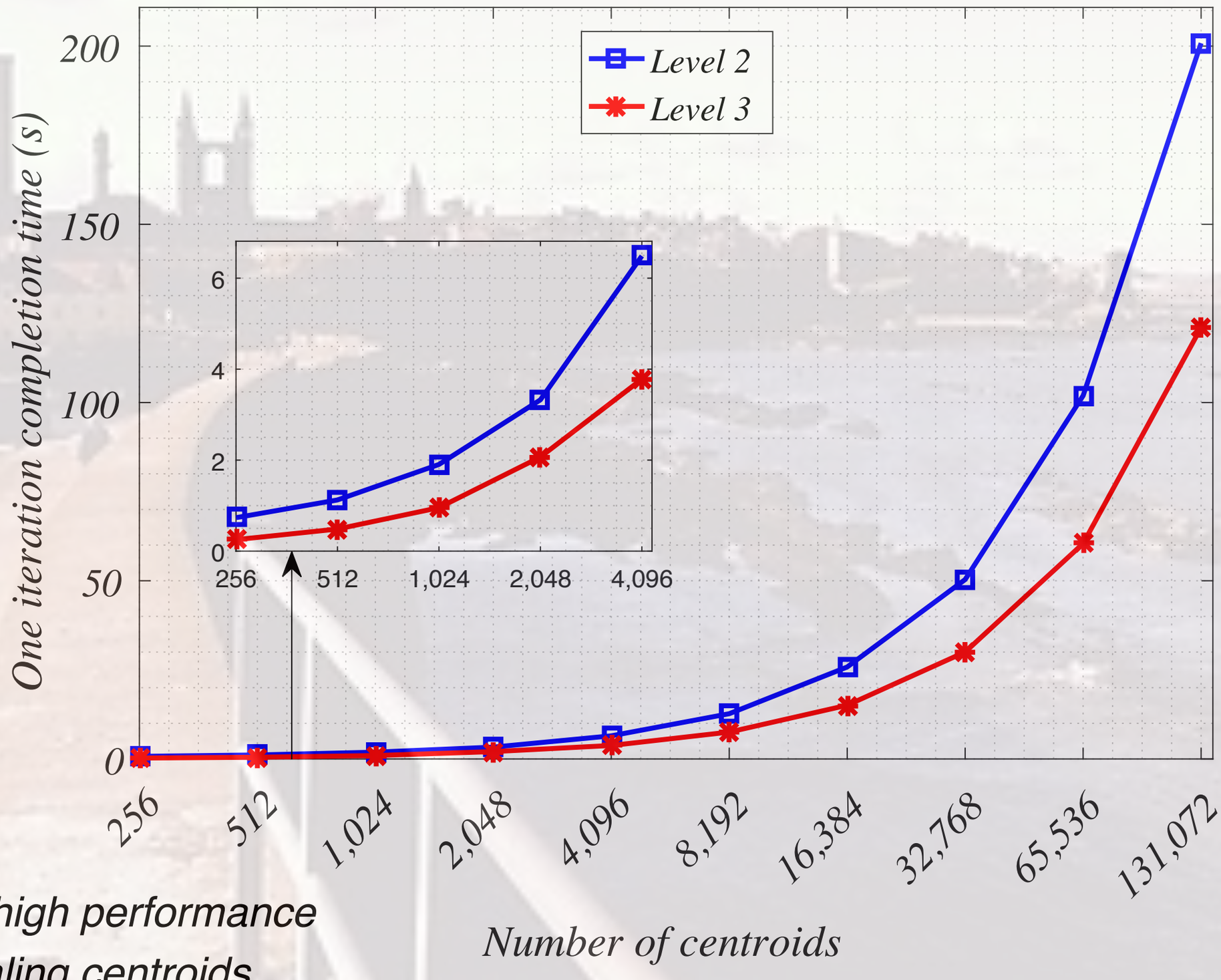
Data Set	n	k	d
Kegg Network	6.5E4	256	28
Road Network	4.3E5	10,000	4
US Census 1990 (*UCI Machine Learning Repository*)	2.5E6	10,000	68
ILSVRC2012 (*ImgNet*)	1.3E6	160,000	196,608

High Scalability on Dimensions

Level-2 **VS** Level-3



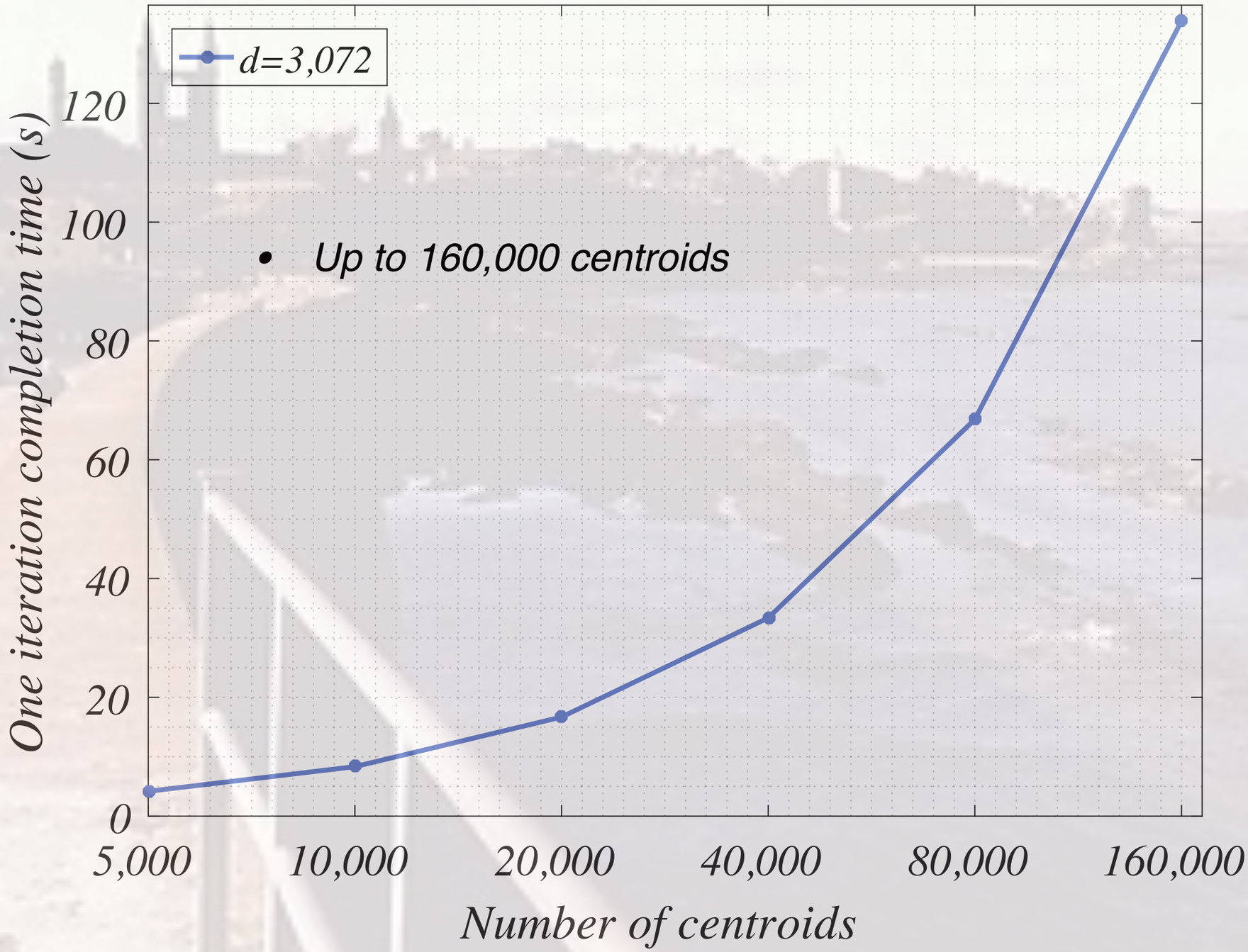
High Scalability on Centroids



- *Keep high performance on scaling centroids*

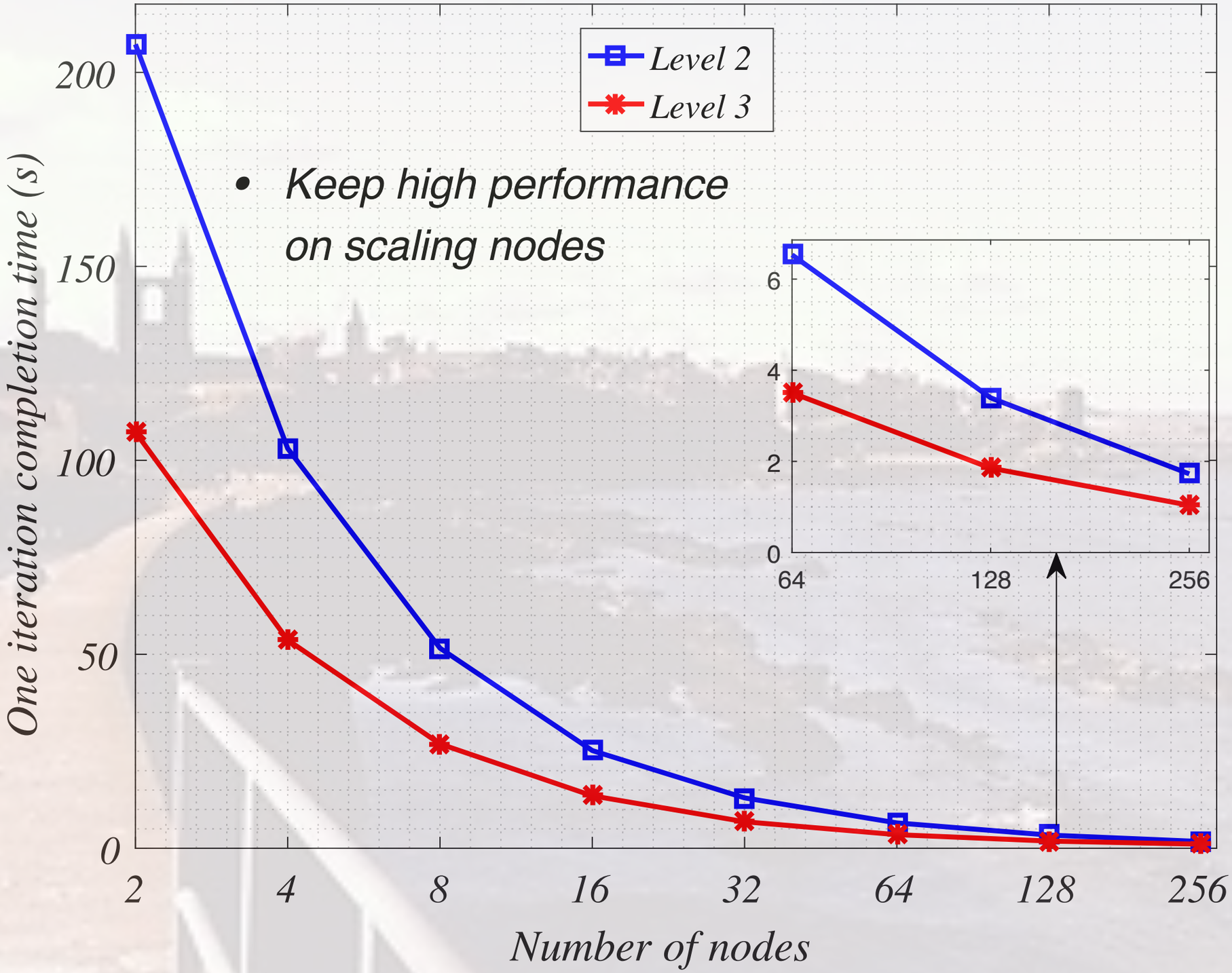
High Scalability on Centroids

Large-scale on Level-3 with multiple nodes



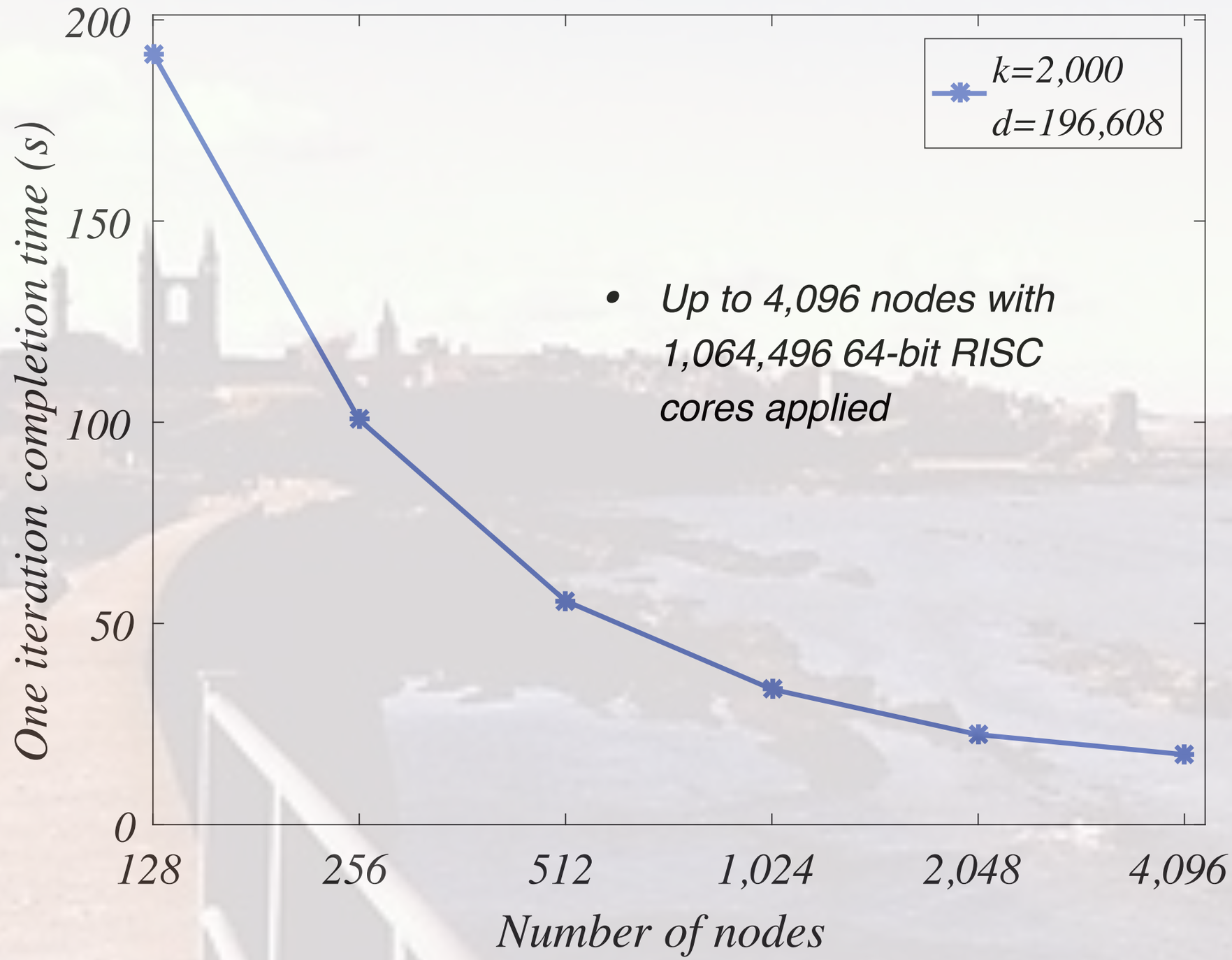
High Scalability on Parallel Processers

Level-2 **VS** Level-3



High Scalability on Parallel Processers

Large-scale on Level-3 with multiple nodes



Conclusion

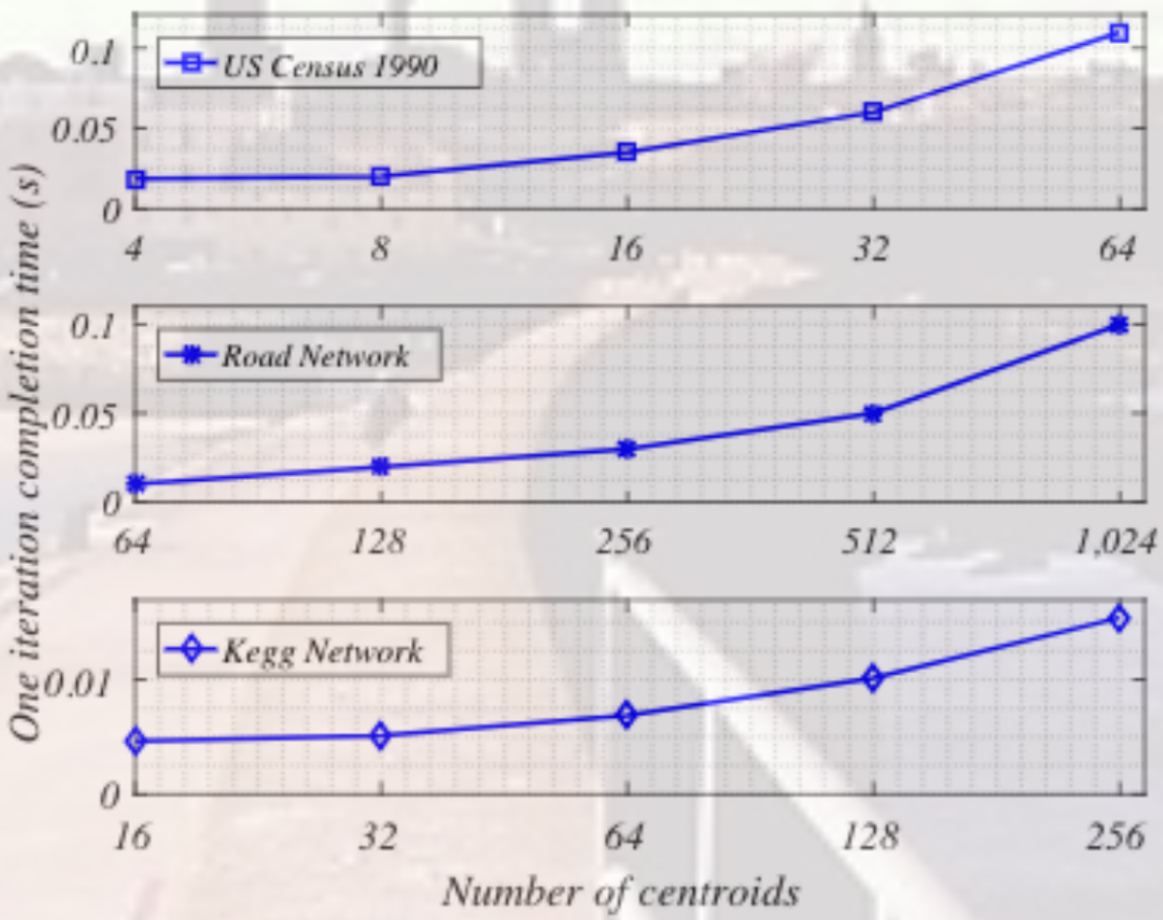
- **We provide the first ever 3-level fully data partition and large-scale the data size to 196,608 dimemsons with 160,000 centriods.**
- **We implement a systematic approach with flexibe multi-level design targeting heterogeneous manycore architectures / supercomputers**
- **We achieve the high performance and high scalability on both Scientific Benchmarks and Real Applications**

Thanks

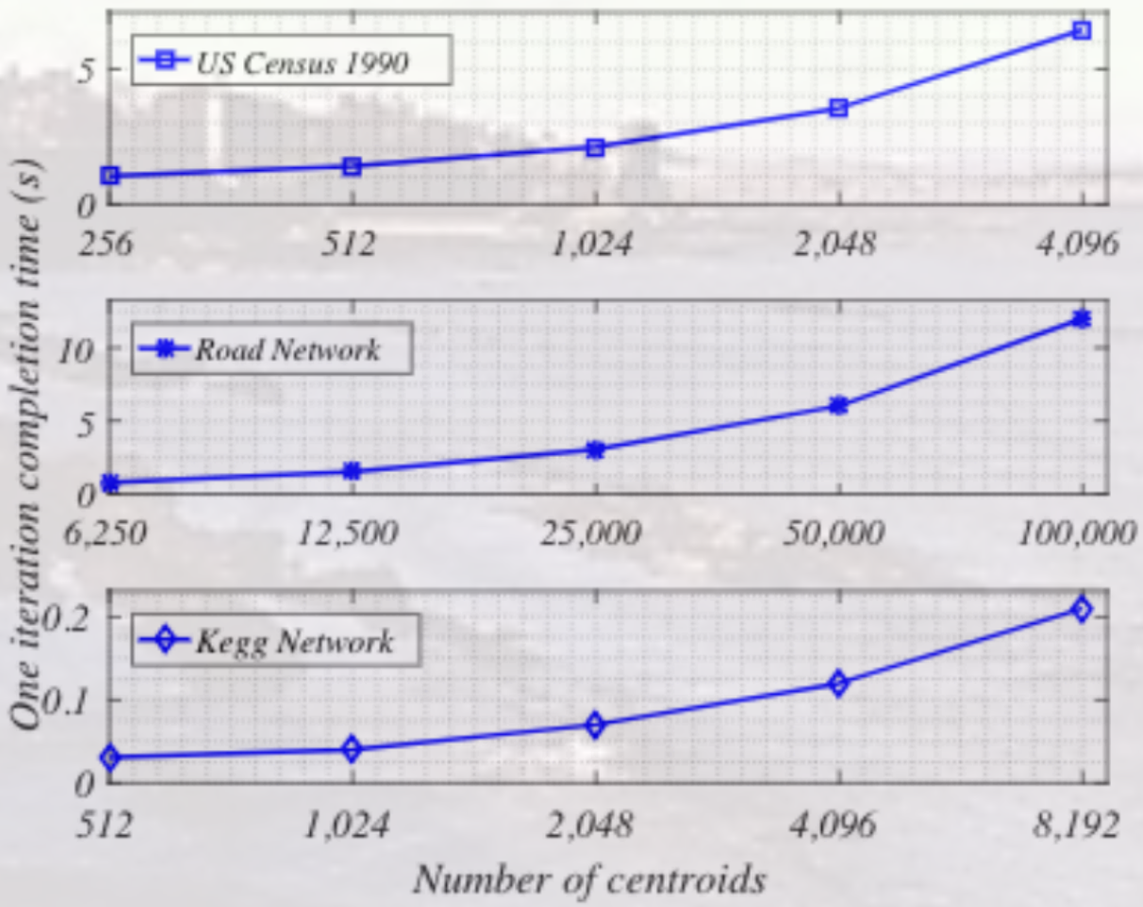
Back Up

Linear-scale on basic levels with single node

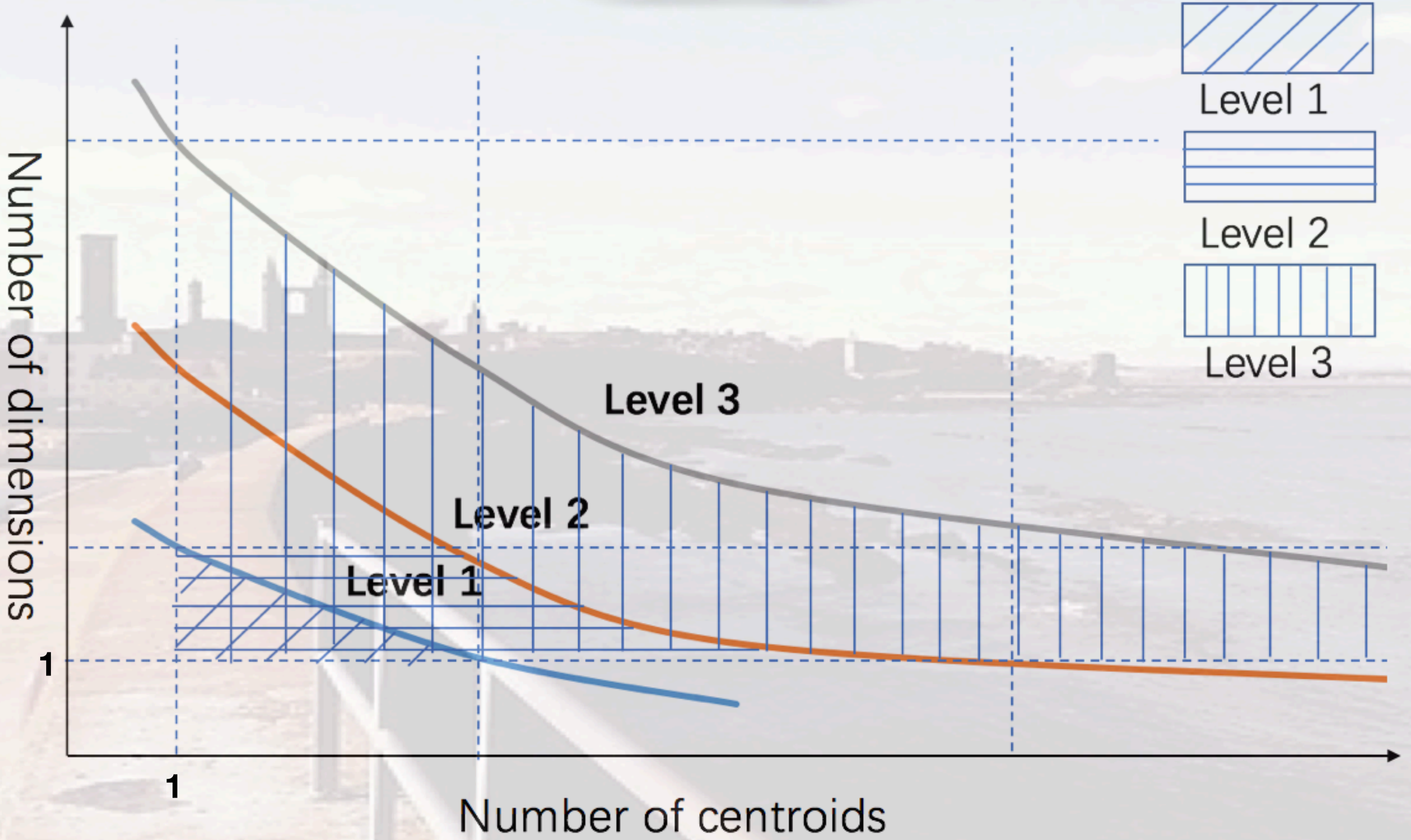
Level-1



Level-2



Back Up


Level 1
Level 2
Level 3
Number of dimensions
Level 3
Level 2
Level 1
1
1
Number of centroids