Cost and Utility Trade-offs on IaaS Clouds, Grids, and On-Premise Clusters

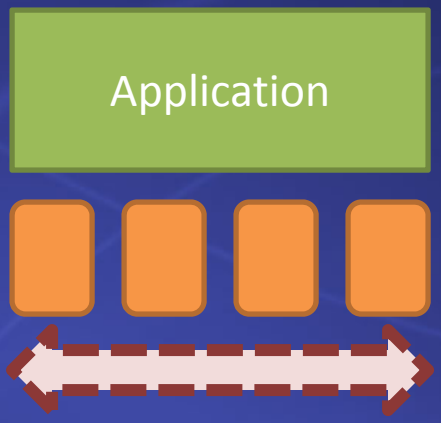
Tiziano Passerini, Jaroslaw Slawinski, Umberto Villa, Sofia Guzzetti
Alessandro Veneziani, Vaidy Sunderam
Mathematics & Computer Science
Emory University, Atlanta, USA

EMORY
UNIVERSITY

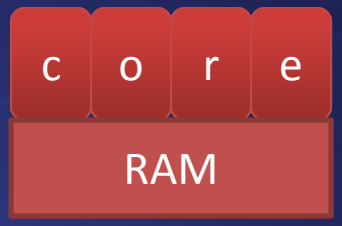
CCDSC 2016
10/4/2016

Resource heterogeneity

- Equivalent platforms for unmodified application

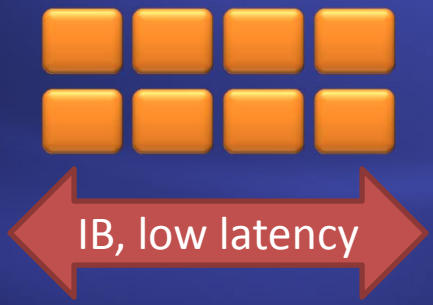


Logical view
- I have my application
- I need some CPU(s)
- Do I care about comm/io? Maybe

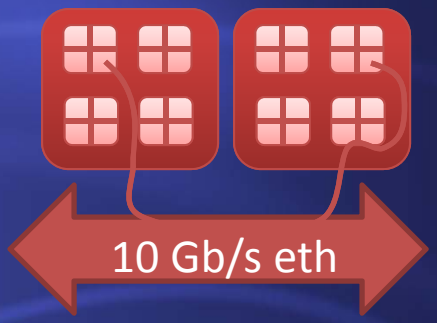


SMP
- Single OS
- Parallel (threads, OpenXYZ, MPI)

VO, P2P, etc.
- Heter. CPUs
- Distributed computing



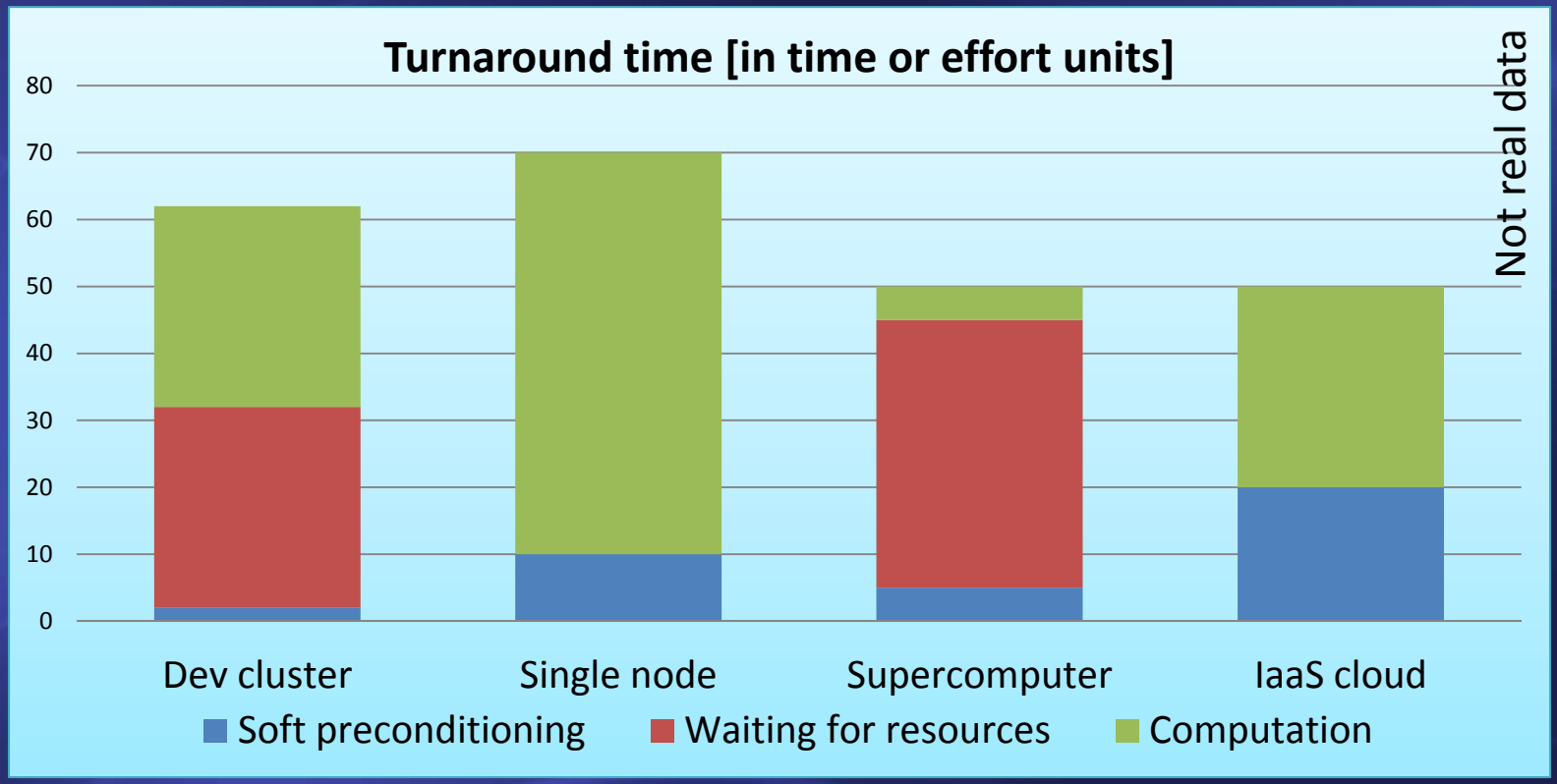
Cluster, supercomp
- Homogen. CPUs
- Soft precon'd
- Good network



Virtualization, IaaS clouds
- Look: soft condition to have a resource like above
- Feel: depends (on coupling)

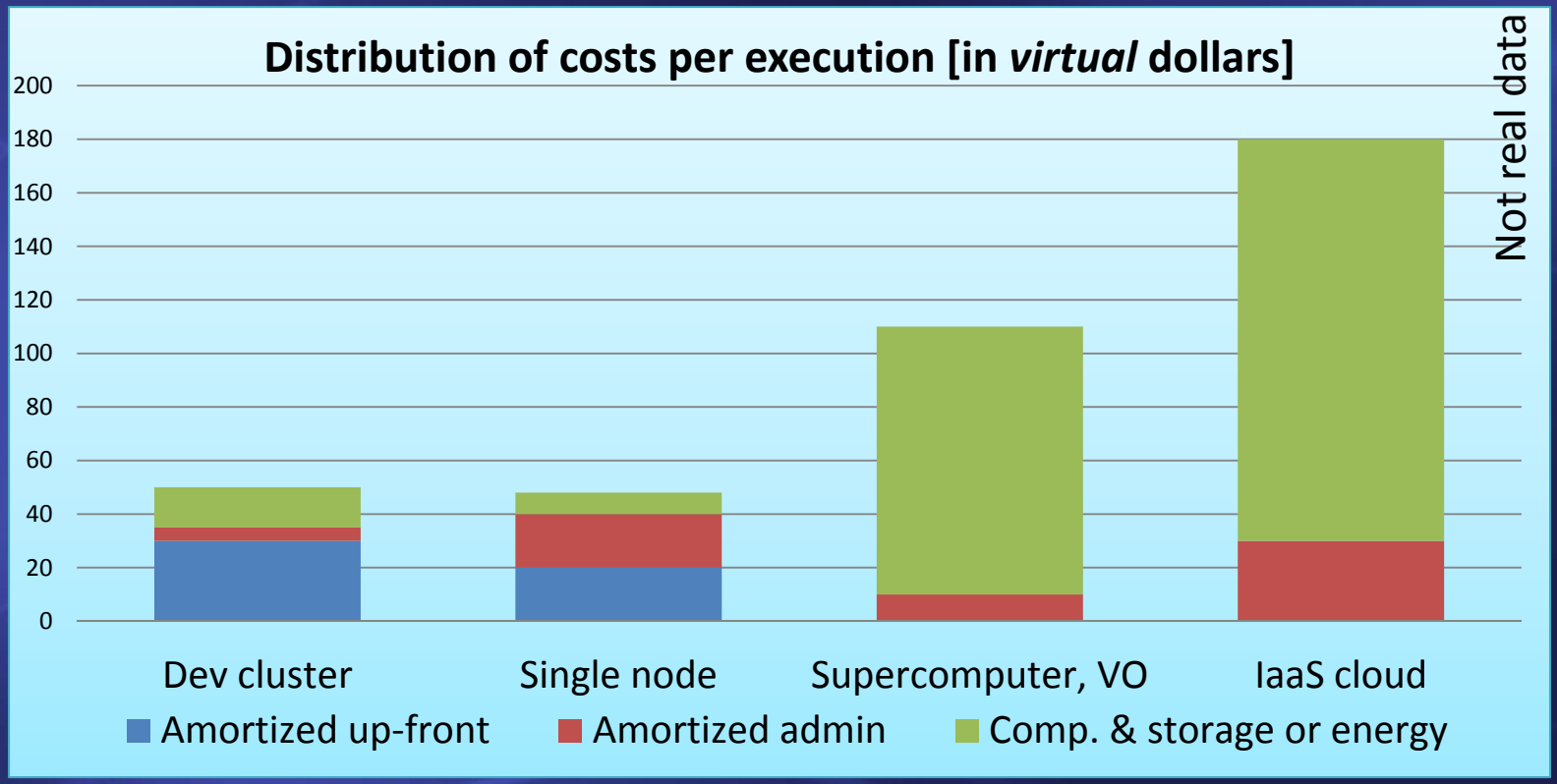
Trade-offs: (1) utilization

- If different computational platforms may be used interchangeably …



Trade-offs: (2) costs

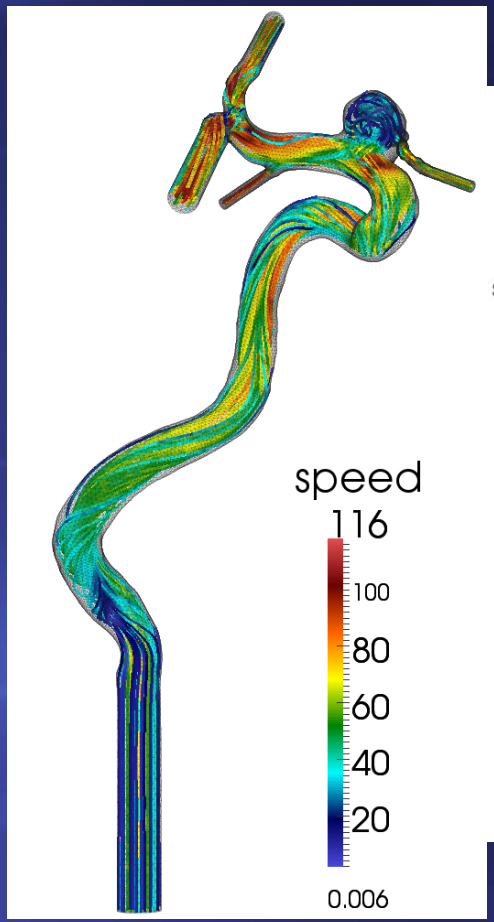
- Dev environment – no soft conditioning
- "Rented" resources – no up-front costs

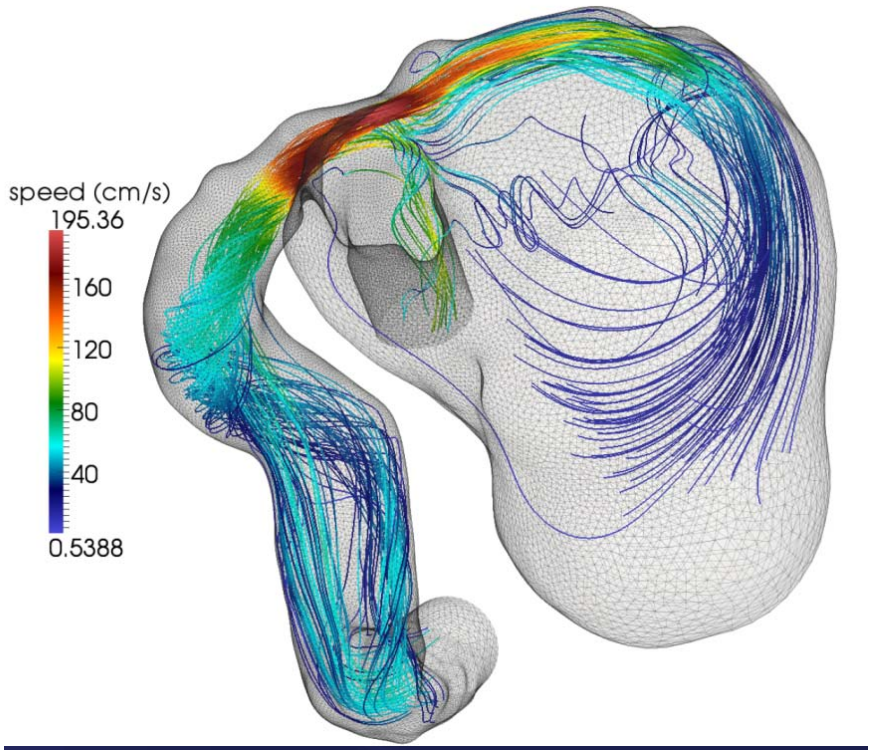


Use-case

Case study: LifeV-based hemodynamic simulation

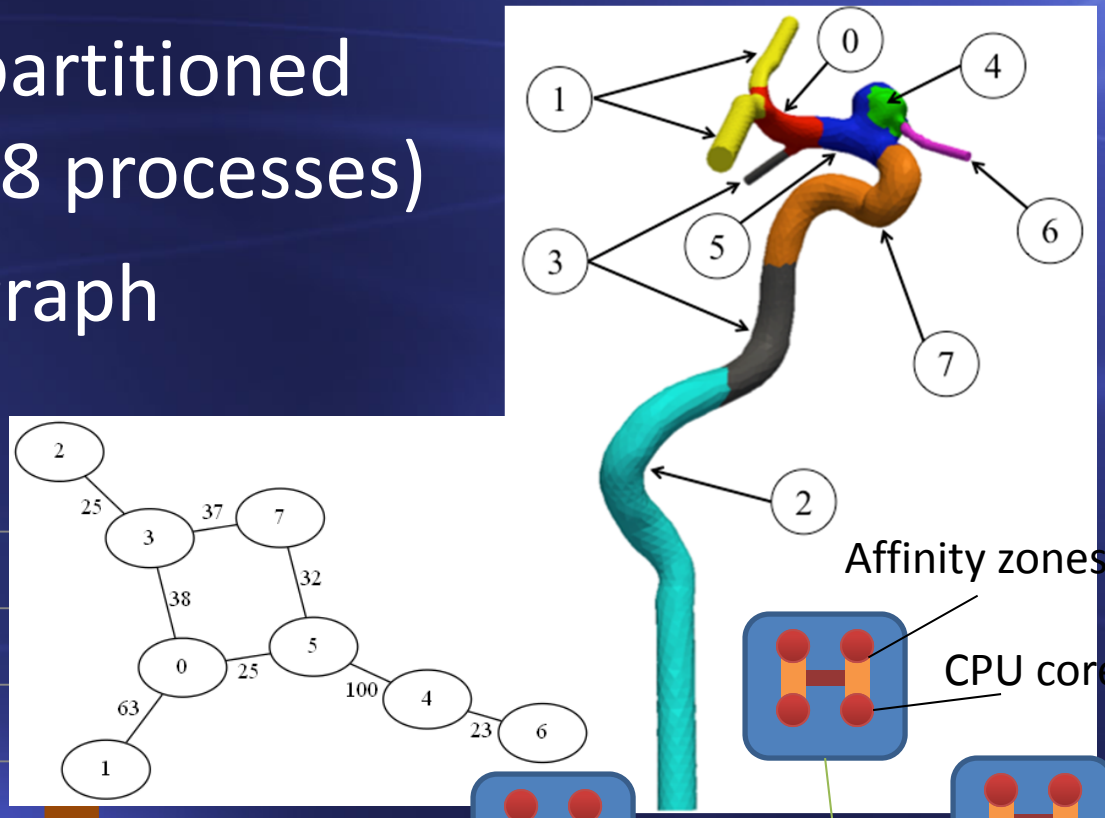
- CFD/FEM MPI parallel code
- LifeV library
- Issues
  - Process placement
  - Turnaround
  - Cost
- Utility





Communication patterns

- FEM input mesh partitioned into 8 partitions (8 processes)
- Logical topology graph
- Physical topology
- How to match?



Scotch

Affinity zones

CPU cores

Internode connection

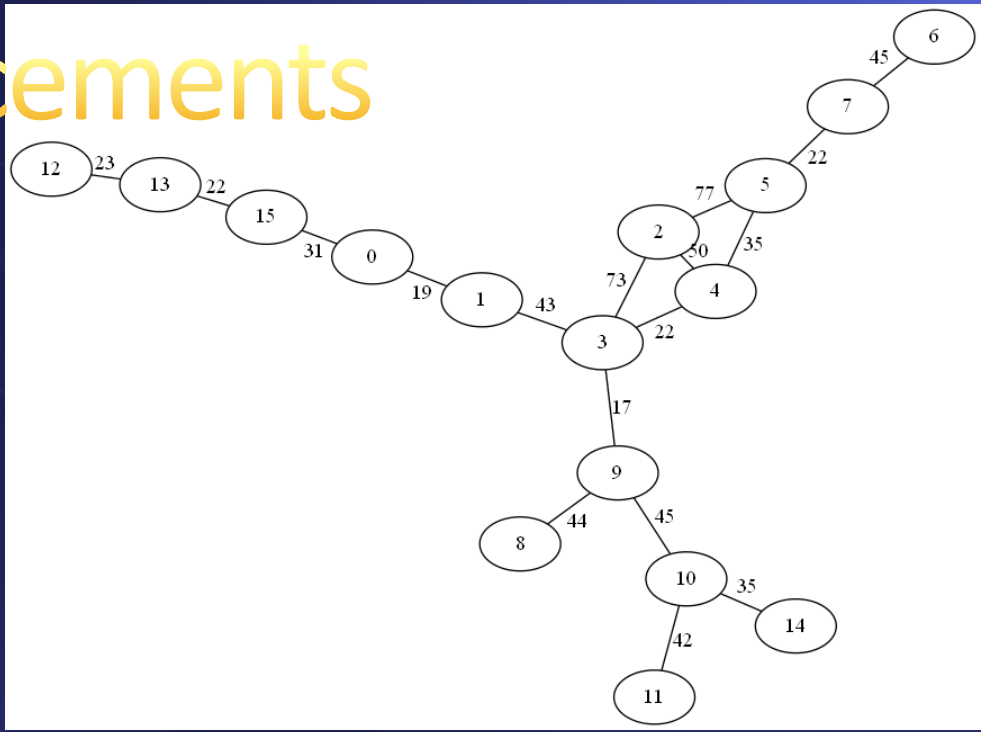
100Gb/s

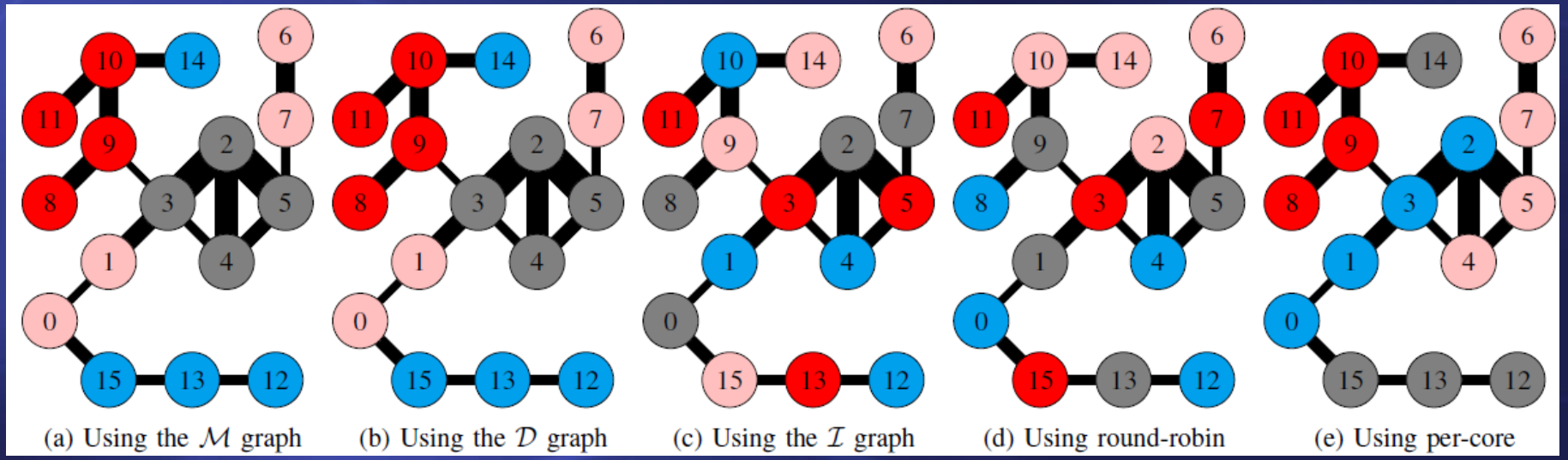
50Gb/s

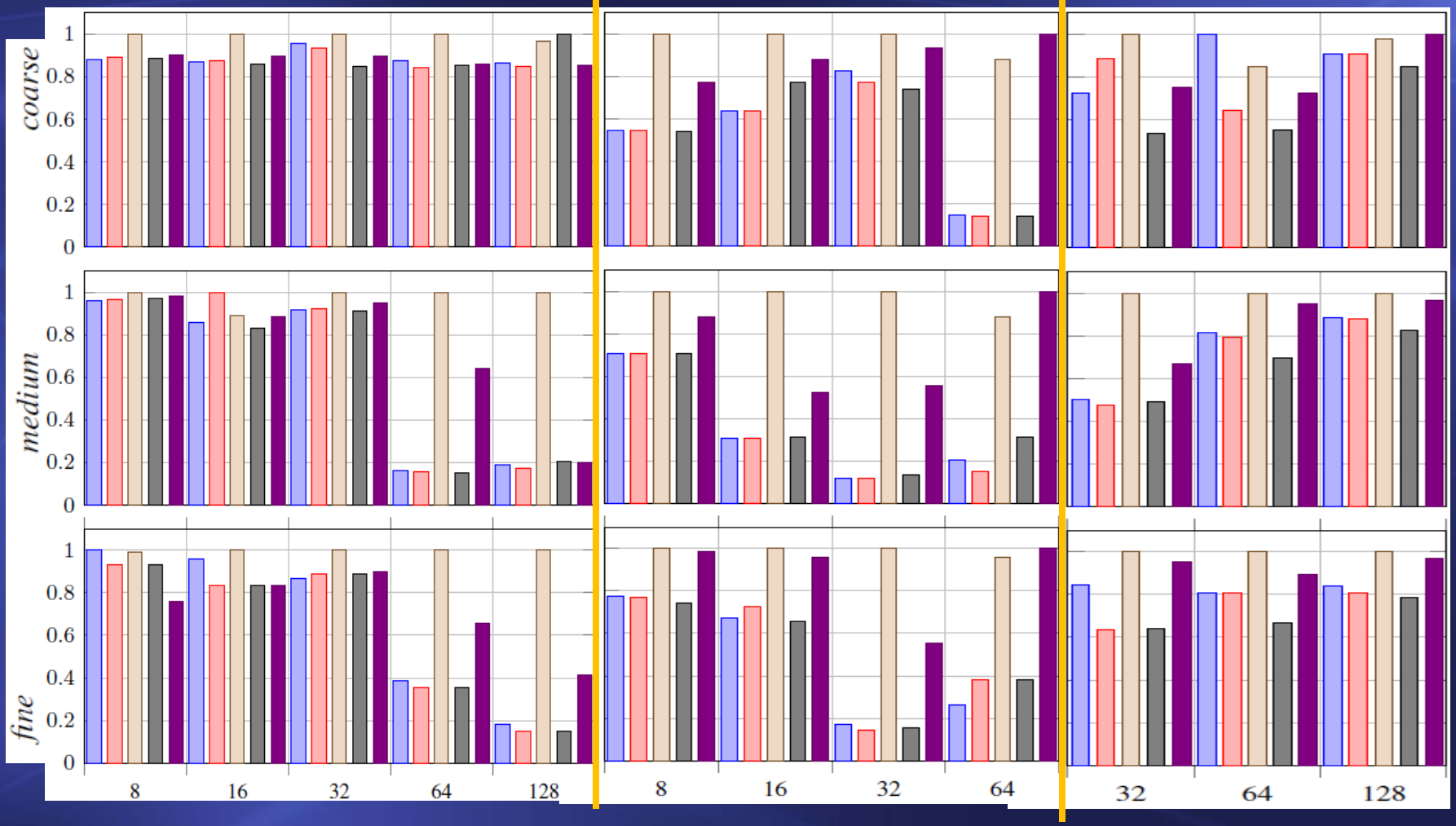
1Gb/s

4core node placements

- M – data from the partitioner
- D – data from benchmarks
- I – inverted D
- Round-robin and per-core – input-agnostic allocation

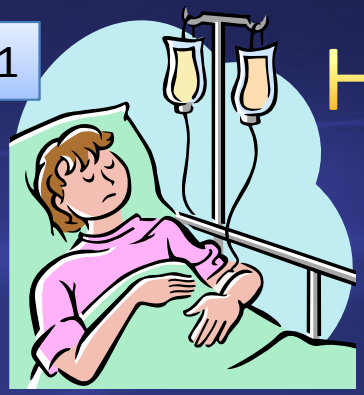




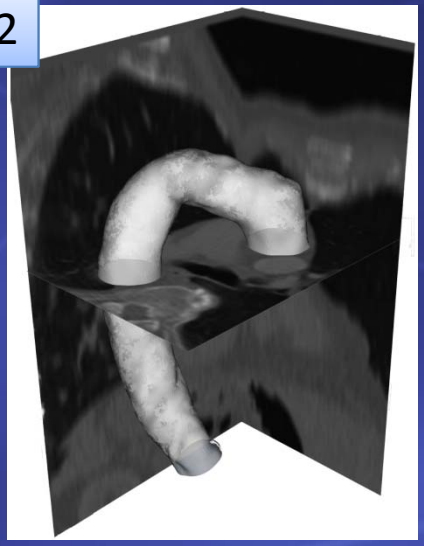

cluster
now
ec2
coarse
medium
fine
0
0.2
0.4
0.6
0.8
1
8
16
32
64
128
NP
data
part
invert
pcore
rr

Hemodynamics in the Clinic

1

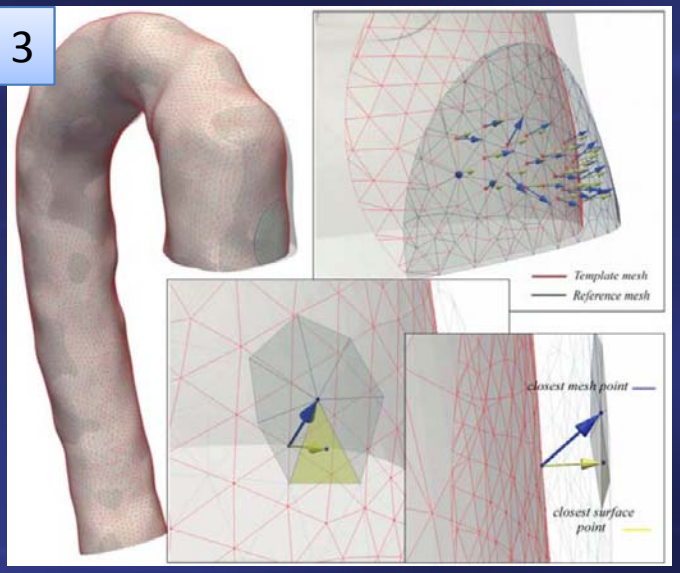


2



Reconstruction from Computed Tomography

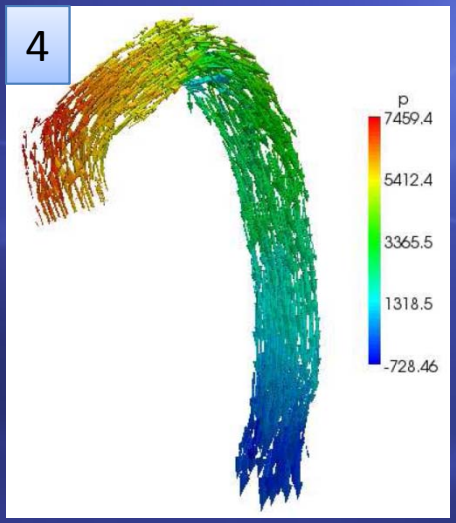
3



Digitalization, mesh generation

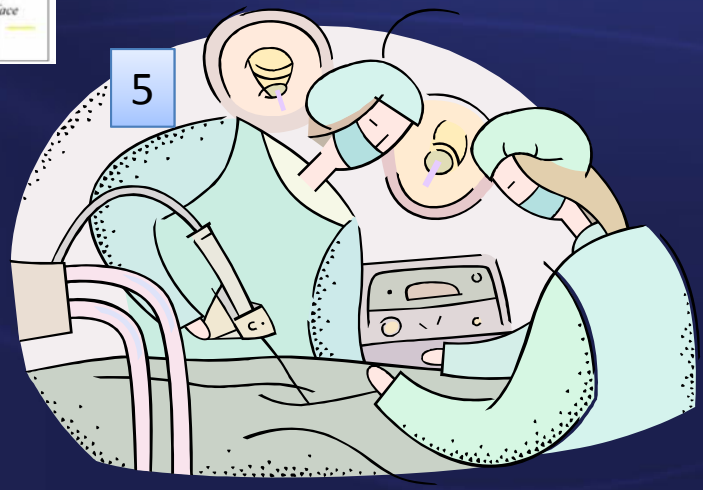
- Diagnosis
- Bypass or stent placement
- ***Cost vs. turnaround***

4



Solution (velocity vectors and pressure)

5



Surgery

EMORY UNIVERSITY

Experiences with Cost and Utility Trade-offs

Trade-off experiment setting

1. Ellipse: university cluster

   256-node 1k-core; 1Gb/s ethernet; queue SGE

2. Puma: dev environment cluster

   32-nodes 128-core; IB SDR; queue PBS

3. Lonestar: XSEDE supercomputer

   IB QDR; queue PBS

4. Rockhopper cluster: On-Demand HPC Cloud Service, Penguin Computing

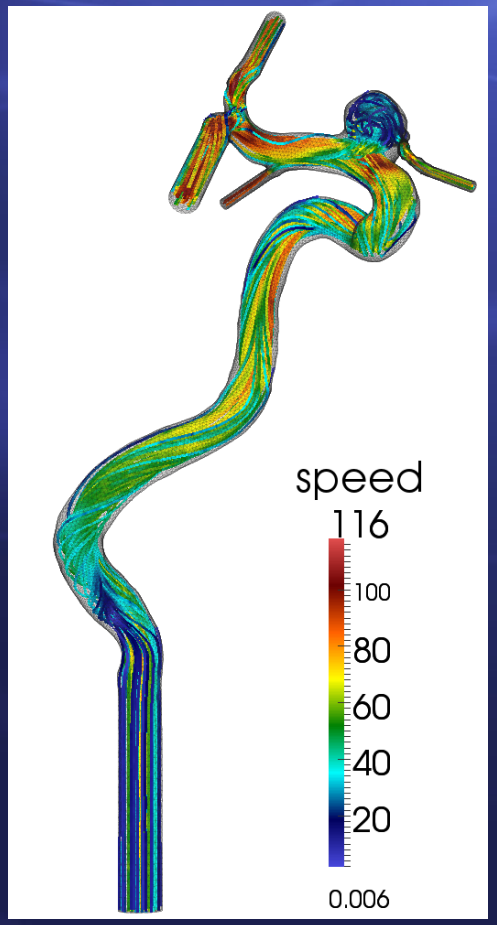
   IB QDR; queue PBS

5. Amazon EC2; 1-16 nodes

   cc2.8xlarge 16-core per node; 10Gb/s ethernet

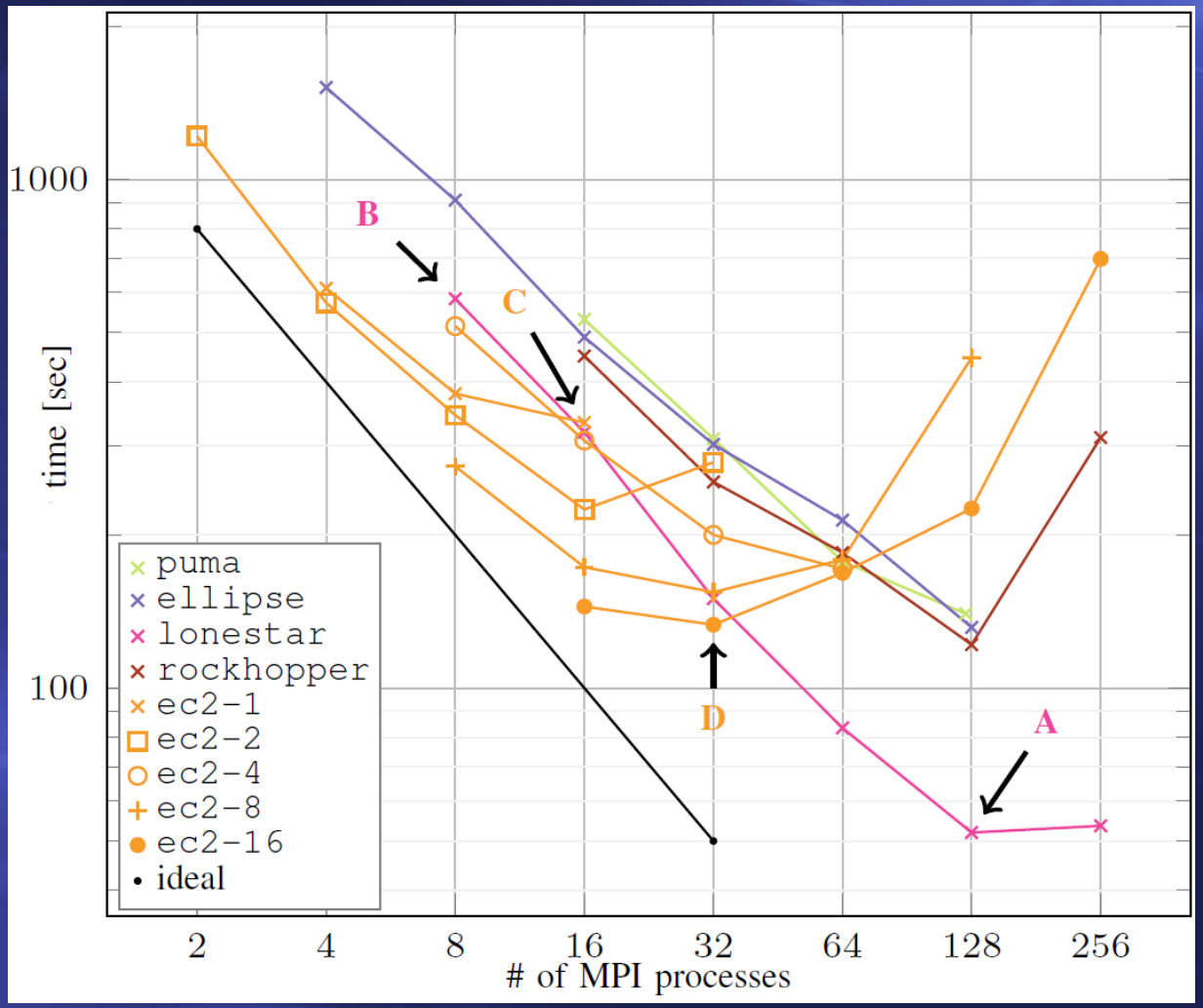
Application

- Aneurism simulation
- About 1 million elements (FEM)
- Computes pressure and velocity for each 0.01 sec
- Same problem, various number of processes (strong scalability test)
- One MPI process per computing core in round-robin placement



Cost of computation

- A – fastest overall
- B – supercom-puter nodes are not the fastest
- C – single EC2 = 16 processes on supercomputer
- D – fastest EC2 configuration
- EC2 scalability…



Fastest time on each platforms
Single simulation frame time (proxy)

rank	time to completion [s]	target	# of MPI proc.
1	1h 31m	lonestar	128
2	3h 33m	rockhopper	128
3	3h 50m	ellipse	128
4	3h 53m	ec2-16	32
5	4h 05m	puma	124
6	4h 30m	ec2-8	32
7	5h 00m	ec2-4	64
8	6h 33m	ec2-2	16
9	9h 43m	ec2-1	16

Our dev machine

Slowest time = 6.4 x fastest. Avg is 4h 44m

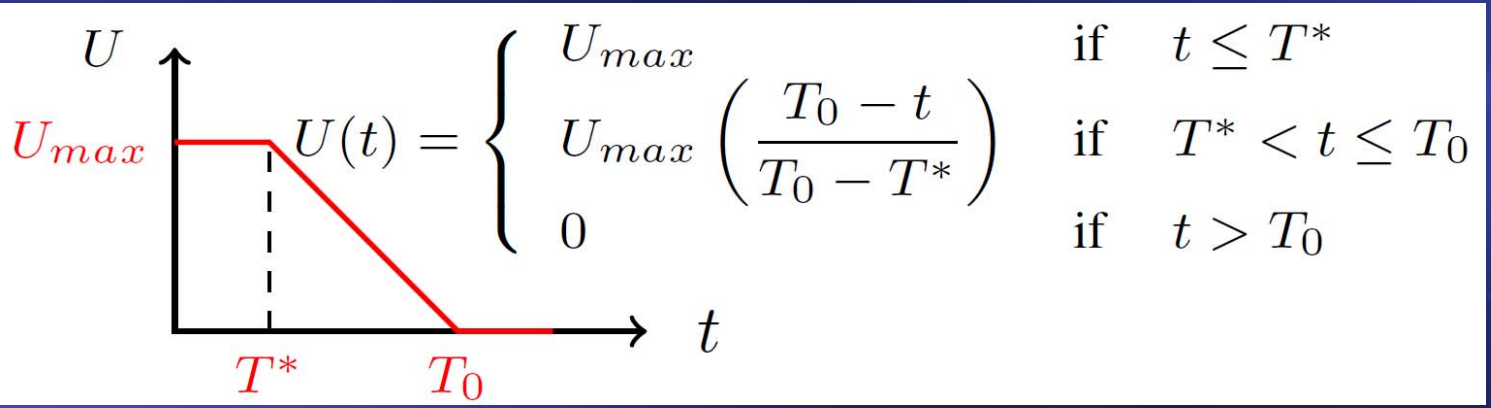
Expenditure calculation

- Puma and Lonestar – estimated cost based on hardware/ operational expenses; typical figures reported in literature
- Ellipse – university pricing
- Rockhopper – actual charges
- EC2 – we used as many cheap spot-request (bid-based) instances as possible (about 6 times cheaper than regular instances)

target	cost per core-hour
puma	2.3¢
ellipse	3¢
lonestar	7¢
rockhopper	10¢
ec2	2.27-15¢

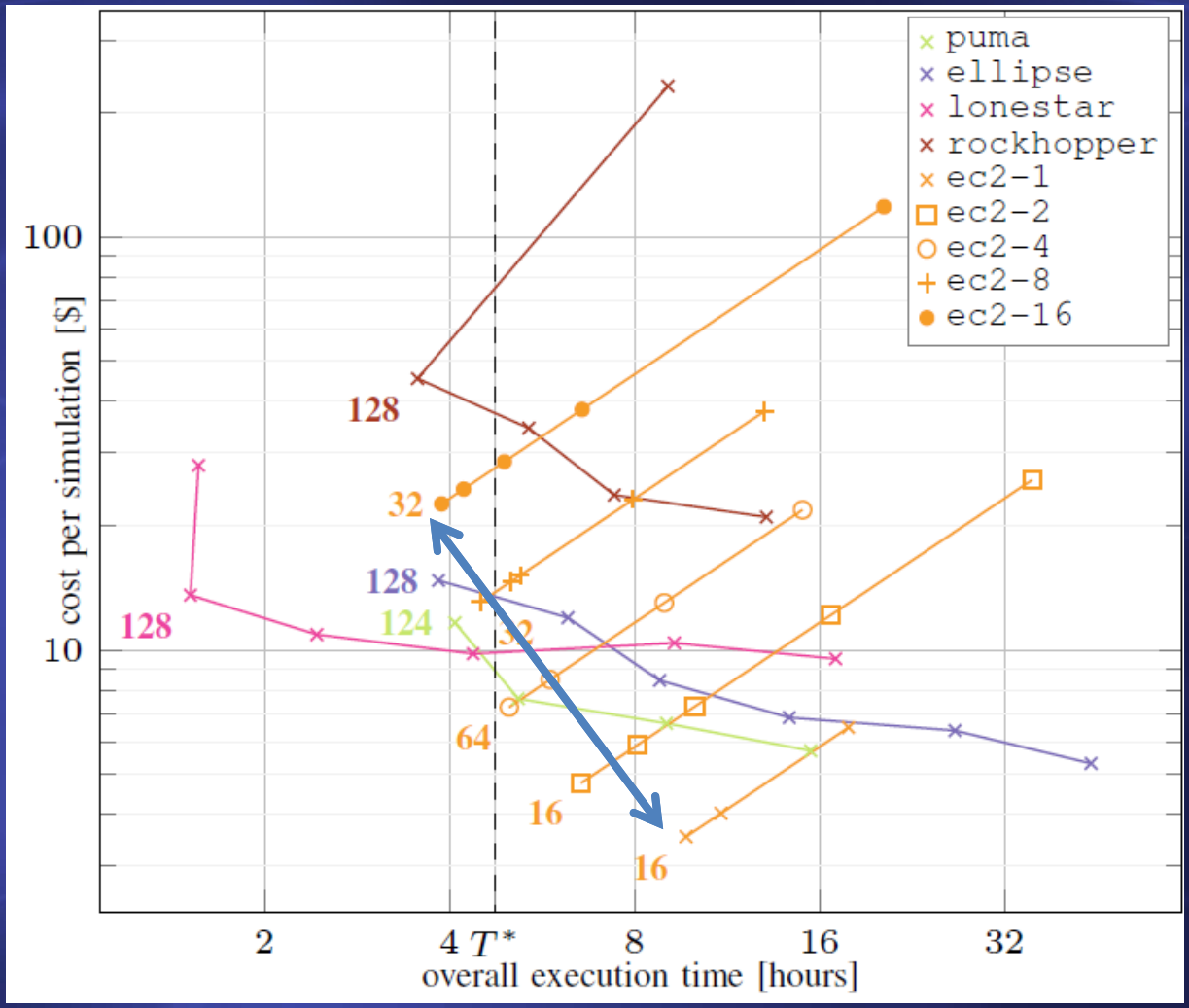
Utility function

- Value of simulation results to user over time



- U – utility value (e.g., in $)
- U_{max} – the max value the user is willing to pay (importance of the task)
- T^* - expected completion time
- $|T^*-T_0|$ - delay tolerance
- T_0 – latest completion time

Relation between time and cost



Range of min. prices per simulation for all architectures:
$3.53 - $22.59
Avg. $10.30

Results

Low (3), high (1), average (2) priority jobs

T* = 4.44 hrs

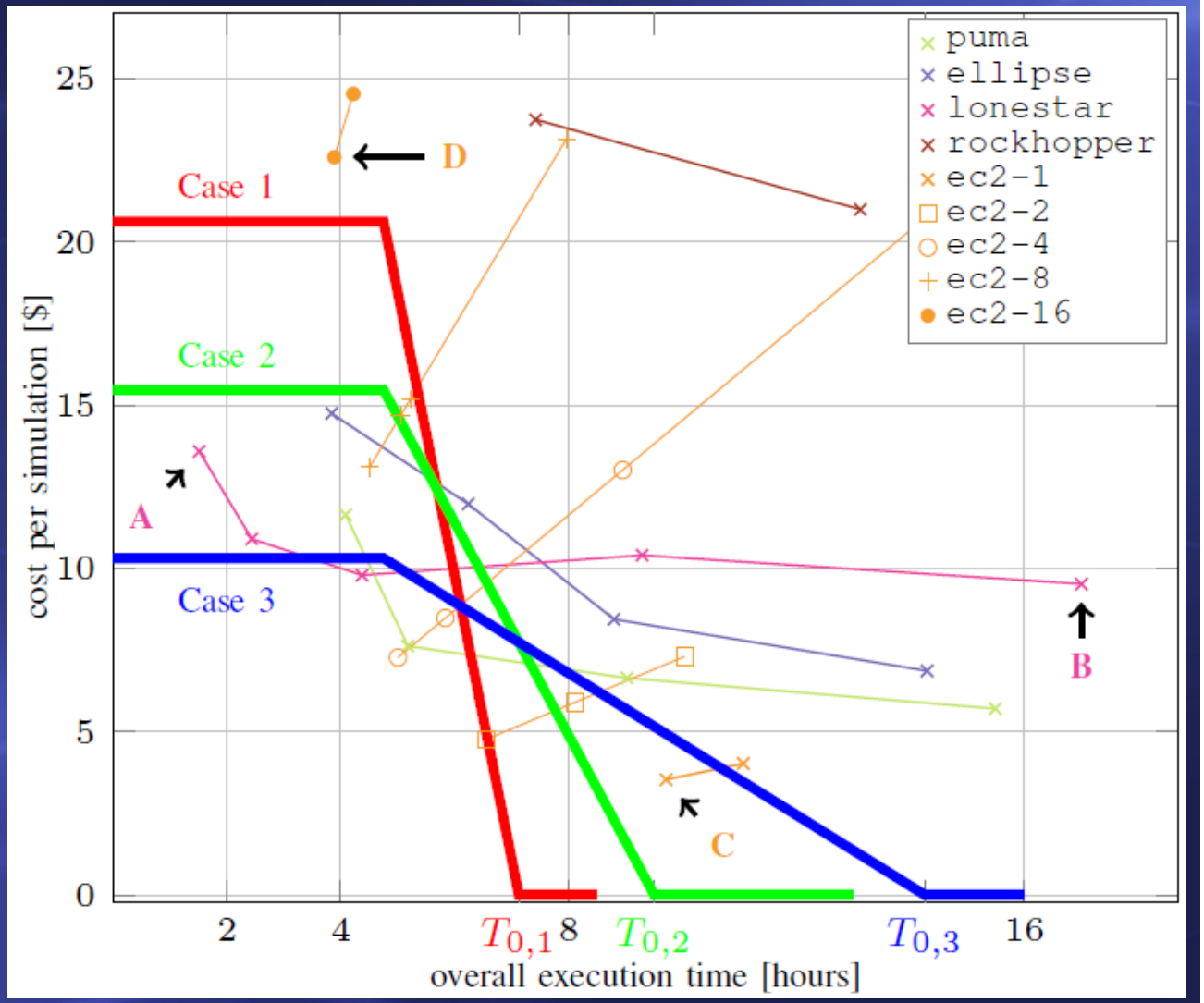
#3 = $10.31

#1 = $20.62

A – overall fastest execution

C – overall cheapest execution

D – fastest time for EC2



Conclusions and future work

- Turnaround vs. cost tradeoffs vary considerably across platforms (multiplied by parameter sweeps)
- Some IaaS cloud resources offer superior capabilities compared to cluster/supercomputer nodes (large single instances vs. local clusters)
- Queue waiting time is not considered in this study, but it may significantly change selection decisions for time-critical computation (e.g., medical diagnosis)