PanSim: A Performance-Portable Agent Based Model

István Z. Reguly, Bence Keömley-Horváth, Gábor Szederkényi, Attila Csikász-Nagy

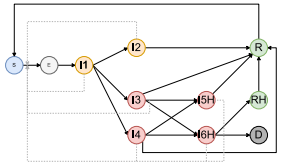
National Laboratory for Health Security, Pázmány Péter Catholic University, Budapest, Hungary

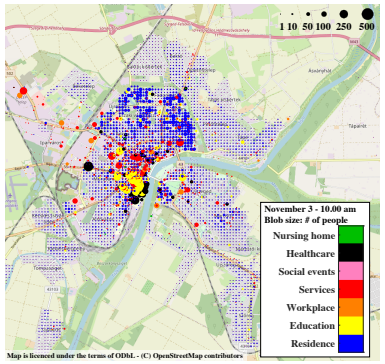
Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, Budapest, Hungary (reguly.istvan@itk.ppke.hu)

PanSim - Agent Based Model

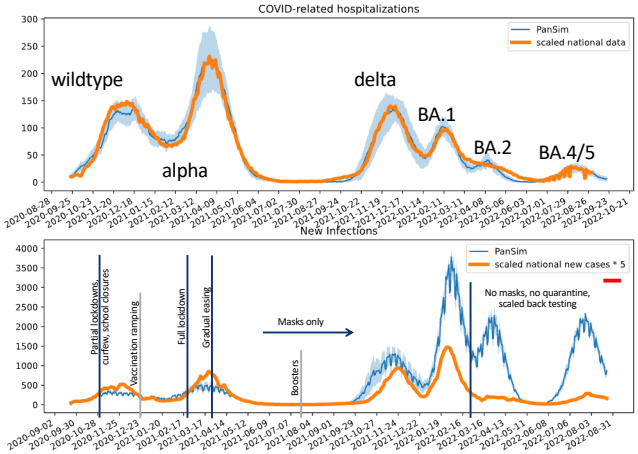
PanSim – massively parallel ABM, used for quantitative and qualitative modelling of pandemic and interventions [1,2]

- Modeling Szeged, HU
  - 180k agents
  - 81k POIs
- Interventions
  - Test, Quarantine
  - Vaccines
  - Curfew, lockdowns
  - School closures
- Used for decision support 2021-2022





Model Calibration



Code Architecture

Key design points: Performance, Productivity, Portability. Statistical evaluation of a range of different interventions and their combinations in limited time - on a local heterogeneous cluster.

C++ codebase with Thrust template library

- Custom OpenMP/CUDA/HIP kernels for critical code paths
- C++11 random/cuRAND/rocRAND

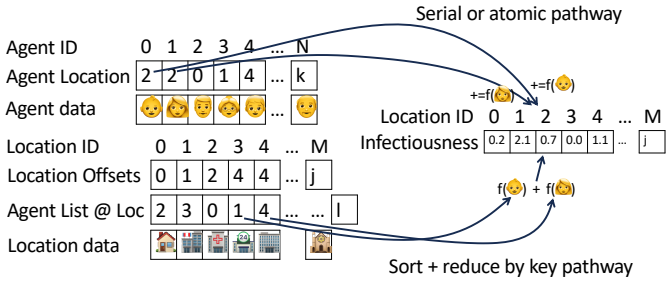
Highly configurable through command line arguments and JSON

Data structures & algorithms

Agent-oriented data (location, behavior, health), Location-oriented data (how many agents, probability of infection).

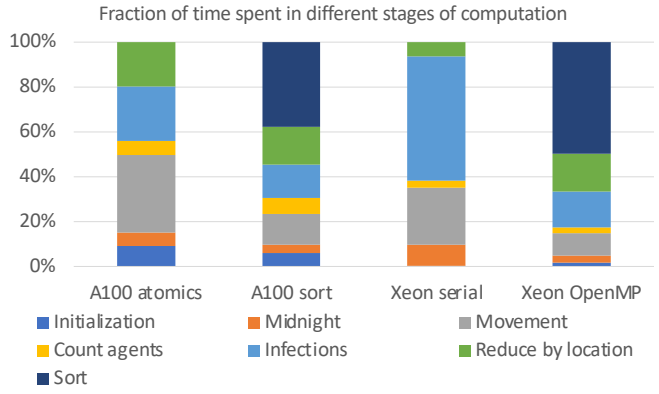
Critical algorithm: Sum of infectiousness of agents at a location

- reduce by key: parallel, but requires agents sorted by location
- atomics (or serial): hardware-dependent



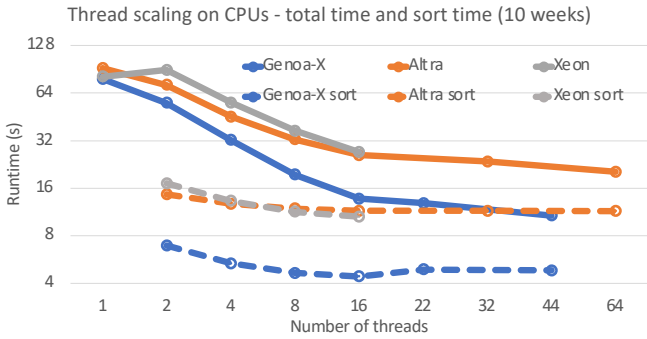
Performance

Szeged case, 2 years, 10 minute timestep. NVIDIA A100 40GB PCI-e, NVIDIA V100 PCI-e, AMD MI100 PCI-e, Intel Xeon Gold 6226R (1S, 16C), AMD EPYC 9V33X (Genoa-X, 2S, 176C), Ampere Altra (1S, 64C). GCC 11.3, CUDA 11.6, RoCM 5.4.2, ARM Clang 23.04.1



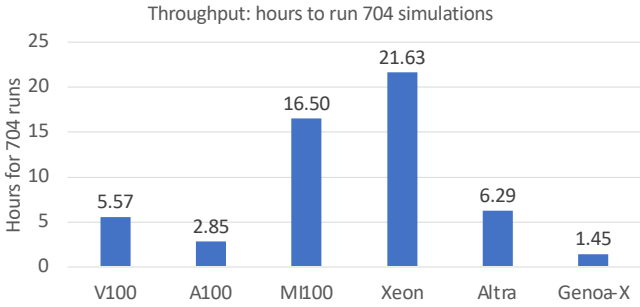
	A100 atomics	A100 sort	Xeon serial	Xeon OMP 16C
Runtime	39.9	66.5	1121.6	266.3
	V100 atomics	V100 sort	Genoa-X serial	Genoa-X 44C
Runtime	54.7	97.7	1129.5	106.4
	MI100 atomics	MI100 sort	Altra serial	Altra 64C
Runtime	98.1	84.4	1324.3	191.65

Thread scaling

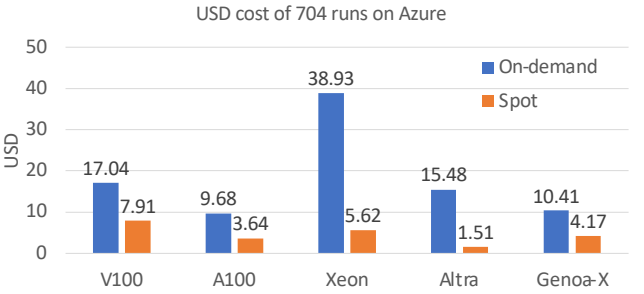


Throughput & Price

Intervention design and parameter search runs require 10s or 100s of experiments – each with 20-30 runs to average over. Can run many at the same time (each with fewer threads or MPS/MIG on GPU).



How much does it cost? Approximate prices from Azure Cloud (US East). NC6s v3 (V100), ND96asr A100 v4 (A100*8), D32d v4 (Xeon), D64ps v5 (Altra), HB176rs v4 (Genoa-X)



Acknowledgements & References

This research was supported by the Hungarian Academy of Sciences, grant POST- COVID2021-64, and in part by the European Union within the framework of the National Laboratory of Health Security under Grant RRF-2.3.1-21-2022-00006.

1. Reguly IZ, et al. (2022) Microsimulation based quantitative analysis of COVID-19 management strategies. PLOS Computational Biology 18(1): e1009693. https://doi.org/10.1371/journal.pcbi.1009693
2. Keömley-Horváth, Bence, et al. (2022) The design and utilisation of PanSim, a portable pandemic simulator. 2022 First Combined International Workshop on Interactive Urgent Supercomputing (CIW-IUS). IEEE, 2022. https://doi.org/10.1109/CIW-IUS56691.2022.00006