

# PASC Co-design Project: *Towards the HPC-inference of causality networks from multiscale economical data*

Illia Horenko, Patrick Gagliardini  
Università Svizzera Italiana  
illia.horenko@usi.ch, patrick.gagliardini@usi.ch

William Sawyer  
Swiss National Supercomputing Centre (CSCS/ETH Zurich)  
william.sawyer@cscs.ch

## 1 Focus of the project

Analysis of large amounts of economical data and data-driven inference of causality relations between different components of economical systems is one of the central problems in modern computational finance and economics. The task of proper mathematical description and adequate causality understanding for the economical data is hampered by the multiscale nature of the underlying processes, resulting from the presence of different temporal and spatial, i.e. regional, sectorial and global, scales.

Important challenges are: (i) an investigation of the mutual causality influences of different economic observables and their spatial (e.g., regional) and temporal (e.g., associated with the business cycle) evolution, (ii) identification of the most important exogenous impact factors that play a role in their dynamics, (iii) proper mathematical and statistical description of the influences coming from the unresolved/latent scales and factors. The solution of these problems can be enhanced by analysis of a causality network inferred from the data. This network is a directed weighted graph with edges representing the causality relations between the different economical variables, exogenous factors, etc. (situated at the vertices of this causality graph). Analysis of this graph would allow to understand the most important features of the underlying complex economical system.

### Milestone questions about the targeted economical data:

1. Is there a causality relation between different sectors of the economy with respect to the credit risk migrations?
2. What is the most effective implementation of the multiscale causality inference framework in the embarrassingly-parallel case?
3. Are there any statistically-significant causality impacts from other sectors on the companies inside of the ‘Banking and Finance’ sector?
4. Among all of the considered alternatives and platforms, what is the most scalable implementation for multiscale causality inference?

## 2 Understanding Causality

In its purest deterministic sense, there is a causal relationship between events if one implies that another has occurred. In reality, such causal relationships can generally only be determined from first principles on a small spatial/temporal scale, with little data involvement. For larger scales and/or larger data involvement, mathematical or statistical models are required from whence causality can be inferred.

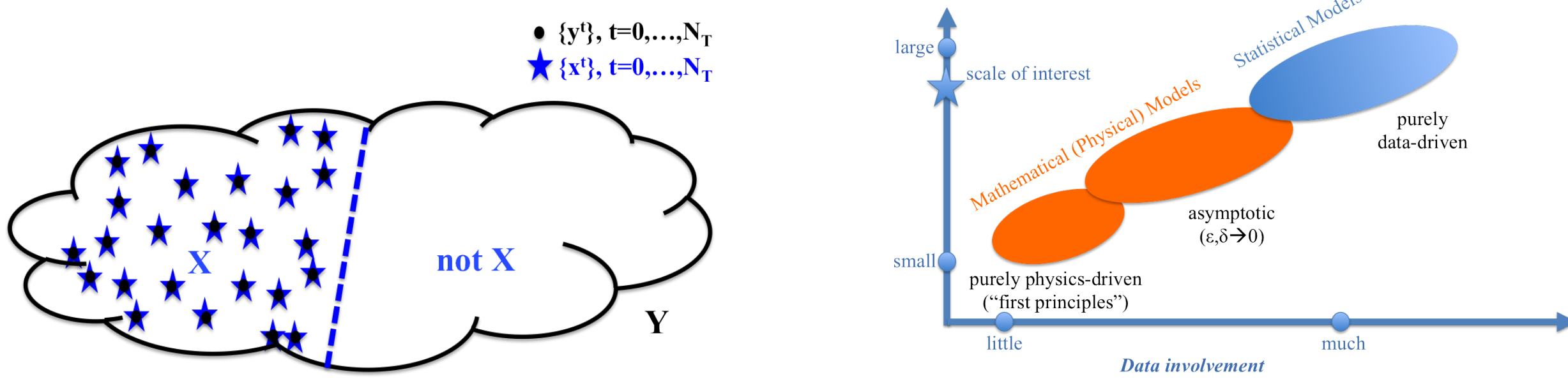


Figure 1: Deterministic Causality (left):  $X$  has a causality impact on  $Y$  if for any  $t$ , event  $y^t$  is happening if and only if event  $x^t$  happened. In real applications a data model is required, but the realm of applicability (right) determines whether the causality inference is driven by the model or by the data.

**Granger Causality:** The standard causality inference measures the ability of predicting the future values of a time series (e.g.,  $y^{t+\tau}$ ) using past values of another time series (e.g.,  $x^t, x^{t-\tau}, \dots, x^{t-q\tau}$ ), where  $\tau$  is a time step and  $q\tau$  is a maximal time lag [2]. The predictive models that have been deployed first to measure such a causality relation between  $y$  and  $x$  were linear **AutoRegressive** models with **eXternal** impact factors (ARX-models) [1]:

$$y^{t+\tau} = \mu + \sum_{i=0}^p A_i y^{t-i\tau} + \sum_{j=0}^q B_j x^{t-j\tau} + \sigma(t), \quad (1)$$

where  $\{\mu, A_0, A_1, \dots, A_p, B_0, B_1, \dots, B_q\}$  are the ARX-model parameters that can be estimated from the available time series for  $x$  and  $y$  (e.g., deploying the maximum likelihood method) and  $\sigma(t)$  is some stochastic noise process [1]. Then, the variable  $x$  has a Granger causality relation to the variable  $y$  if and only if at least one of the  $B_j$  is statistically-significantly different from zero.

**Limitations:** Granger causality may lead to biased results. This bias might be amplified and lead to a completely wrong inference of the causality relations from the data, if the underlying model assumptions of standard causality inference methods (e.g., like the intrinsic linearity of the Granger causality measures) are not fulfilled.

## 3 Probabilistic (Bayesian) Causality

In contrast to typical applications from computational sociology (where the underlying network/graph is mostly available explicitly, e.g., in the form of a social network directly describing the observed connections and relations between the individuals), in a context of economical data these causality network structures are often implicit, difficult to measure, and need to be inferred from the very large sets of available data. The law of total probability can assist us:

**Law of Total Probability:** If  $B_1, B_2, B_3, \dots$  is a partition of the sample space  $S$ , then for any event  $A$  we have  $P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$ .

In this project we investigate the case that both  $X$  and  $Y$  are driven by an (unobserved) process  $U$  with different lags, e.g., unresolved scale impacts. We start by defining the probability vectors  $\Lambda(t) = (\mathbb{P}[y^t|x_1^t \text{ and } u^t], \dots, \mathbb{P}[y^t|x_n^t \text{ and } u^t], \mathbb{P}[y^t \text{ and not one of } x_i^t]), P_x^t = (\mathbb{P}[x_1^t \text{ and } u^t], \dots, \mathbb{P}[x_n^t \text{ and } u^t], 1)$  and a variable  $P_y^t = \mathbb{P}[y^t]$ . Then the probability of observing  $y^t$  together with  $x_1^t, \dots, x_n^t$  and  $u^t$  can be written as a scalar product of these two vectors:

$$P_y^t = \Lambda^\dagger(t) P_x^t, \quad (2)$$

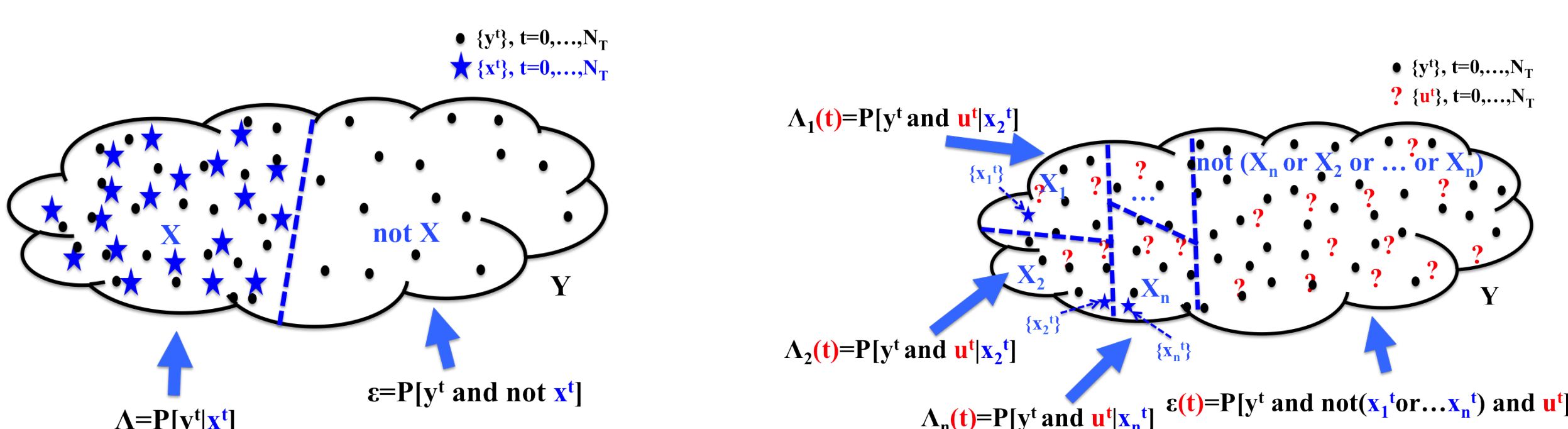


Figure 2: The simplest statistical causality model (left) is  $P[y^t] = \Lambda P[x^t] + \epsilon$ , with  $\Lambda \simeq 0$  implying no causality impact of  $X$  on  $Y$ . The law of total probability allows a partition (right):  $P[y^t u^t] = \Lambda_1(t) P[x_1^t] + \Lambda_2(t) P[x_2^t] + \dots + \Lambda_n(t) P[x_n^t]$ , with  $\Lambda_i \simeq 0$  implying no causality impact of  $X_i$  on  $Y$ .

## 4 Algorithm: approximation through well-posed lower bound

Assuming  $y^t$  being statistically-independent in  $t$  sequence of binary variables or observed probabilities (conditioned on the knowledge of variables  $x$  and  $u$ ), inference of both the unknown causality vector  $\Lambda(t)$  and of the unknown probability process  $P_y^t$  for the discrete state model (2) can be done via a maximisation w.r.t.  $\Lambda(t)$  of the following log-likelihood functional:

$$\mathcal{L} = \max_{\Lambda} \left( \sum_{t=0}^{N_T} \left[ (1 - y^t) \ln (1 - \Lambda^\dagger(t) P_x^t) + y^t \ln (\Lambda^\dagger(t) P_x^t) \right] \right). \quad (3)$$

This problem is unfortunately ill-posed. The central challenge is to formulate the investigations of  $\Lambda_i$  into a well-posed optimisation problem. We start with the approximation:

$$\Lambda(t) = \sum_{i=1}^K \gamma_i^t \Lambda_i \quad \text{with} \quad \sum_{i=1}^K \gamma_i^t = 1, \quad \gamma_i^t \geq 0.$$

In [3] the following lower bound for  $\mathcal{L}$  was proved:

$$\mathcal{L} \geq l^\epsilon = \max_{\gamma_i, \Lambda_i} \left( \sum_{i=1}^K \left[ \sum_{t=0}^{N_T} \gamma_i^t g(t, \Lambda_i) - \epsilon^2 \sum_{j=1}^n \lambda_i^{(j)} \right] \right). \quad (4)$$

with  $g(t, \Lambda_i) = (1 - y^t) \ln (1 - \Lambda_i^\dagger P_x^t) + y^t \ln (\Lambda_i^\dagger P_x^t)$ , and subject to the constraints:

$$\sum_{i=1}^K \gamma_i^t = 1, \quad \gamma_i^t \geq 0 \quad \text{for all } t \text{ and } i, \quad (5)$$

$$0 < \sum_{j=1}^n \lambda_i^{(j)} < 1, \quad 0 \leq \lambda_i^{(j)} \leq 1, \quad \text{for all } i \text{ and } j, \quad (6)$$

$$\sum_{t_1, t_2=0}^{N_T} |\gamma_i^{t_1} - \gamma_i^{t_2}| \leq \bar{C}(N_T), \quad \text{for all } i. \quad (7)$$

The maximisation of  $l^\epsilon$  belongs to the well-posed class of FEM-BV-problems.

## 5 High Performance Computing Implementation

Currently the causality inference algorithm is implemented in Matlab using interfaces to existing optimisation libraries (Fig. below). This sequential approach is not viable for cases of interest ( $> 10,000$  companies/sectors).

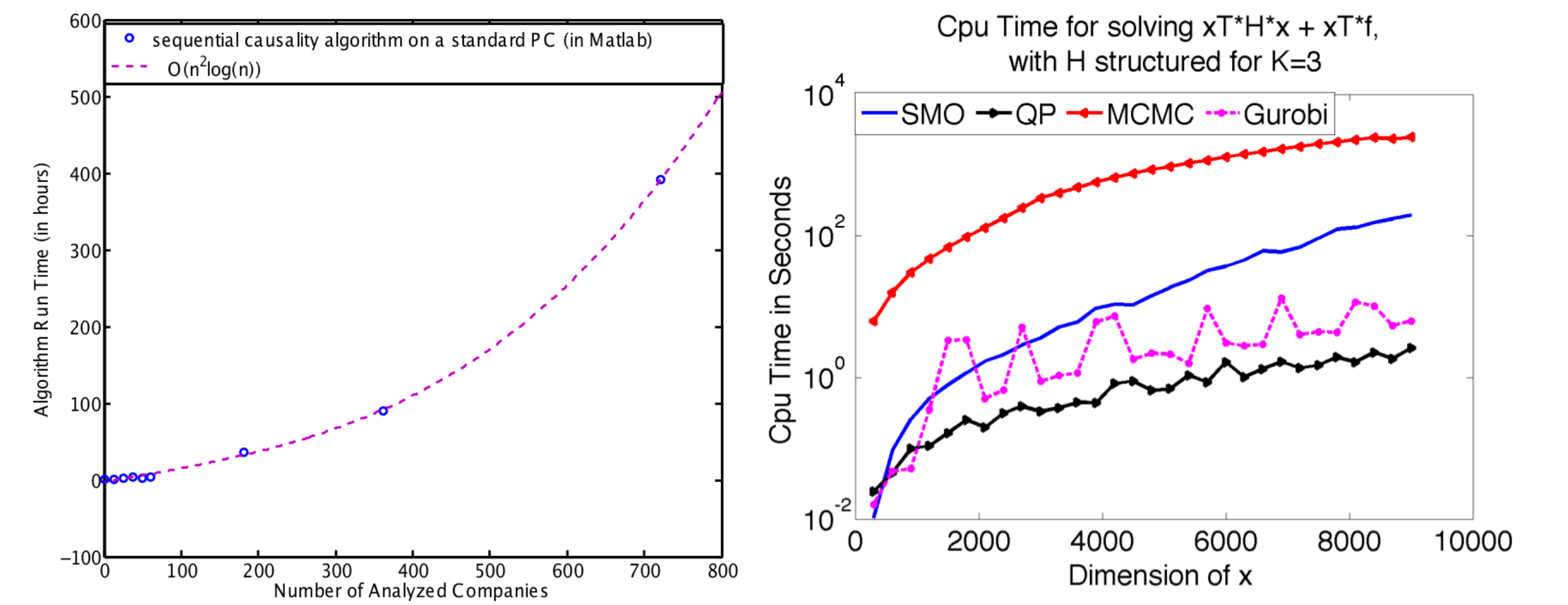


Figure 3: On left, the run-times of causality inference algorithm in its current sequential Matlab implementation on an Apple MacBook are shown. The magenta line is the optimal fit of the  $O(n^2 \log n)$ -function to this sequence of run times. On right (courtesy of O. Kaiser), the run times of several variants of the Quadratic Programming (QP) minimisation problem performed in every annealing step of the causality algorithm illustrate a similar time evolution (note log-linear scale), albeit with different leading coefficients. The dimension of  $x$  is length of the time series and should not be confused with the  $n$ . The magenta Gurobi curve yields a slightly better scaling since this library employs OpenMP multithreading, which may scale better for increasing problem size.

Key challenges in the implementation will be to (i) exploit distributed memory parallelism inherent in the algorithm, and (ii) to identify and utilise existing numerical libraries which hide the complexity of programming emerging architectures such as GPGPUs and Intel Xeon Phi, instead of trying to program these directly. Preliminary investigations (Fig. below) reveal potential strong scaling using a hierarchy of libraries.

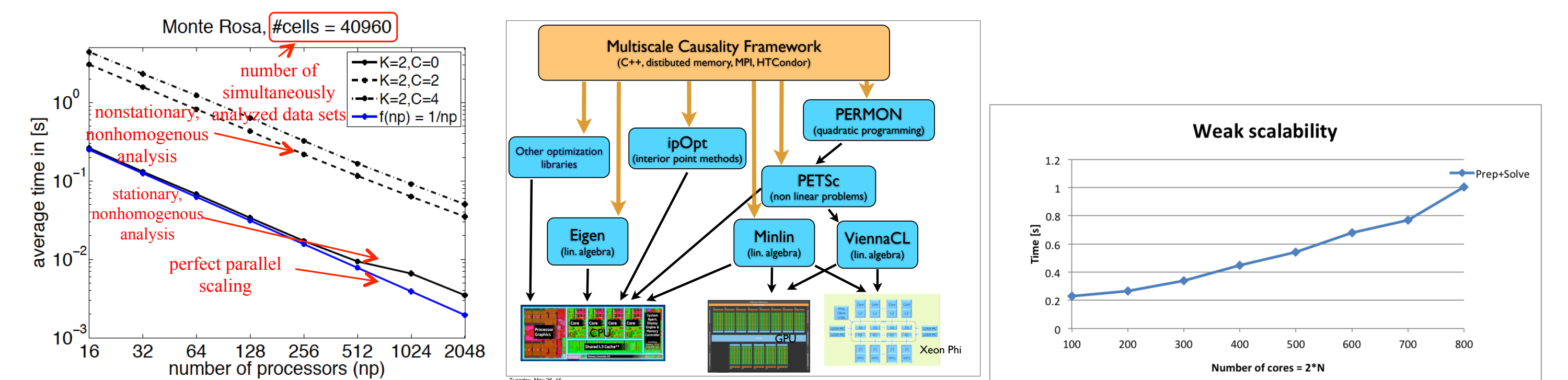


Figure 4: Parallel scaling (left) of the FEM-BV framework in simultaneous (i.e., embarrassingly parallel) analysis of 40960 test data sets (courtesy of P. Metzner). Variable  $K$  denotes the number of locally-stationary graphs models that were identified to be in common for all of the data sets,  $C$  is the persistency (i.e., a maximal allowed number of switches between these local models pro data set). Note that the computational expense increases for the non-stationary case, but the strong scalability also improves. These preliminary results were obtained on the ‘Monte Rosa’ Cray XE6 at CSCS. A hierarchy of state-of-the-art libraries (right) contains, e.g., the templated-C++ minimal linear algebra (Minlin) library for dense linear algebra operations on CPUs (OpenMP) and GPUs (cuBLAS), the PERMON (Parallel, Efficient, Robust, Modular, Object-oriented, Numerical) toolbox for constrained optimisation methods, utilising the well-known Portable Extensible Toolkit for Scientific Computation (PETSc) library for non-linear problems, which in turn utilises the ViennaCL linear algebra backend to offload linear algebra computation to accelerators. On the far right, the above-mentioned initial results for the above-mentioned QP problem solved with PERMON indicate that weak scaling can still be approached (here on dual 8-core Intel Sandybridge nodes), in spite of the dependencies which make the problem difficult to formulate for parallel platforms.

## 6 Acknowledgements

The project PI’s gratefully acknowledge the financial assistance of the Platform for Advanced Scientific Computing (PASC) for one post doctoral researcher for the period of 2 years (position currently open). We would also like to thank Olga Kaiser, Philipp Metzner and the PERMON team for their collaboration in the project.

## References

- [1] P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting*. Springer, Berlin, 2002.
- [2] C. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.



[3] S. Gerber and I. Horenko. On inference of causality for discrete state models in a multiscale context. *Proc. Natl. Acad. Sci. USA (PNAS)*, 2014.