# GPU-powered simulation methodologies for biological systems

**Please check the document version of this publication:**

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

# GPU-powered Simulation Methodologies
# for Biological Systems

Daniela Besozzi[a]    Giulio Caravagna[b]    Paolo Cazzaniga[c]
Marco Nobile[b]    Dario Pescini[d]    Alessandro Re[b]

[a]Università degli Studi di Milano, Dipartimento di Informatica
Via Comelico 39, 20135 Milano, Italy

besozzi@di.unimi.it

[b]Università degli Studi di Milano-Bicocca, Dipartimento di Informatica, Sistemistica e Comunicazione
Viale Sarca 336, 20126 Milano, Italy

giulio.caravagna/nobile@disco.unimib.it, a.re4@campus.unimib.it

[c]Università degli Studi di Bergamo, Dipartimento di Scienze Umane e Sociali
Piazzale S. Agostino 2, 24129 Bergamo, Italy

paolo.cazzaniga@unibg.it

[d]Università degli Studi di Milano-Bicocca, Dipartimento di Statistica e Metodi Quantitativi
Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy

dario.pescini@unimib.it

## Aims and motivation

In the last decades, the study of biological systems witnessed a pervasive cross-fertilization between experimental investigation and computational methods, thanks to the convergence of the so-called high-throughput era with the easy access to high performance computing resources. This combination gave rise to the development of new methodologies, able to tackle the complexity of biological systems in a *quantitative* manner. For instance, given a mathematical formalization of complex biological networks, it is nowadays possible to determine their emergent dynamical and structural properties with whole-system based approaches, in order to make predictions on the way these systems behave in normal conditions or how they react to different perturbations [18].

In this context – according to the system under investigation, to the experimental data at hand and to the biological questions one is expected to unravel – the choice of the most suitable modeling approach is fundamental to properly elucidate the underlying physics of the target system. Standard modeling and simulation approaches span from the stochastic to the deterministic formalization of chemical kinetics [20], also considering the possibility to combine these two approaches into hybrid methods [13, 10, 17]. The conditions of applicability of each method are strictly related to the characteristics of the biological system under investigation, as briefly presented in the next section.

Given either a stochastic, deterministic or hybrid mathematical model, computer algorithms can then be exploited to investigate the temporal evolution of the corresponding biological system: starting from different initial conditions of the model (given in terms of distinct values for the kinetic constants, the initial molecular amounts, etc.), different emergent behaviors of the system can be determined. This fact highlights the relevance of setting a good parameterization for the model; as a consequence, the exploration of high-dimensional parameter spaces will allow to investigate the system functioning across a wide spectrum of natural conditions, as well as to derive statistically meaningful properties. These issues play a fundamental role in the computational analysis of biological systems, which usually exploit

methodologies based on parameter sweep analysis, sensitivity analysis, parameter estimation and reverse engineering of model topologies (see [1, 9] and references therein).

All these methods rely on the repetition of a large number of simulations, therefore demanding the reduction of the computational costs for everyday applicability. An emergent technology suitable to tame this problem is the General-Purpose Graphics Processing Unit (GPGPU) paradigm, in which the parallel computation capabilities of modern video cards are exploited for general purpose computations. Indeed, GPGPU allows to perform multiple analysis in parallel, using cheap, diffused and highly efficient multi-core devices. Despite the remarkable advantages in terms of the achievable computational speedup, computing with GPUs requires specific programming skill, since GPU-based programming substantially differs from CPU-based computing; as a consequence, scientific applications on GPUs risk to remain a niche for few specialists. To avoid such limitations, we are developing GPU-powered simulation algorithms for stochastic, deterministic and hybrid modeling approaches, so that also users with no knowledge of GPUs hardware and programming can easily access the computing power of graphics engines.

## Methods

To investigate the behavior of a given biological system, the choice of the most adequate modeling and simulation approach can be carried out, in general, according to the molecular amounts of the chemical species occurring in the system, or even to the temporal scale of the biological phenomena that one wishes to reproduce and analyze.

When the biological system is characterized by chemicals species occurring in small molecular amounts (in the order of units, tens, or a few hundreds of molecules), accounting for the *intrinsic random fluctuations* is fundamental to correctly mimic the behavior of the target system [11]. In this case, stochastic approaches should be adopted. Under the assumption of spatial homogeneity and thermal equilibrium, the system states can be formally described by a continuous-time discrete-value random variable $\mathbf{X}(t)$, that denotes the number of molecules of each species. Changes in the system state are determined by firing chemical reactions, which describe the interactions between the chemical species occurring in the system. Understanding the system dynamics therefore reduces to estimating the probability of $\mathbf{X}(t)$ to be in each possible (chemical) state in time, given some initial condition. This probability changes according to the *Chemical Master Equation*, which is usually unfeasible, especially for large systems, due to the unpractical numerical calculations required to solve it. However, sample paths of $\mathbf{X}(t)$ and numerical estimates of the probability of any chemical state can be obtained algorithmically, whereby $\mathbf{X}(t)$ is a *Continuous-Time Markov Chain*, whose states are the values of $\mathbf{X}(t)$ [12]. To speedup the simulation of $\mathbf{X}(t)$, approximation techniques either based on the *Chemical Langevin Equation* [4] or on quasi-steady-state assumptions [3] are often used. Coupled intrinsic/extrinsic environmental fluctuations can also be considered [7].

When most of the chemical species are present in the system in large amounts (in the order of thousands of molecules or more), the intrinsic fluctuations of $\mathbf{X}(t)$ get averaged out, and $\mathbf{X}(t)$ is well approximated by a set of *Ordinary Differential Equations*, often termed *Reaction-Rate Equations*. Generally, the unique path of $\mathbf{X}(t)$ is numerically evaluated via numerical integration algorithms. It was shown that, when the system is spatially uniform and at the thermodynamic limit, the stochastic and the mean-field model have the same behavior [13].

When a clear separation between the temporal scales of the reactions exists, hybrid – stochastic and deterministic – approaches can be adopted. This is the case when, for instance, kinetic constants or molecular amounts are separated by several orders of magnitude, which is likely to happen in multiscale

modeling of cellular and intercellular dynamics [17]. Hybrid approaches allow to account for intrinsic random fluctuations due to low-amount species, at the same time integrating an appropriate mean-field approximation concerning the abundant chemicals [6, 5]. Technically, here $\mathbf{X}(t)$ describes a *Piecewise Deterministic Markov Process* [10].

Whatever the mathematical representation of the model, the simulations of the system dynamics in a given set of initial conditions – each one characterized by the same or a different parameterization with respect to the other conditions – are generally executed in a serial fashion on standard CPU personal computers, thus causing high computational costs. However, all these simulations are mutually independent, therefore the computational burden can be strongly reduced by exploiting a parallel architecture. Among the existing solutions, we are considering GPGPU computing for the implementation of the aforementioned simulation algorithms, in order to gain a consistent reduction in the overall computational time required to fully analyze a biological model [14, 15]. In particular, we are adopting nVidia's CUDA, a GPGPU library that combines the single instruction multiple data architecture and multi-threading. Using CUDA's naming conventions, the programmer implements a kernel loaded from the host (the CPU) to the devices (one or more GPUs), replicated in many copies named *threads* (where each thread corresponds to a dynamics simulation). Threads are logically organized in structures named *blocks* which, in turn, are organized in *grids*. Threads are executed in groups named *warps* (corresponding to 16 threads), whose scheduling is handled by the driver which dispatches the work on the multiple streaming multiprocessors, thus allowing a transparent scaling of performances on different GPUs.

## Discussion

In living cells, many processes are regulated by feedback mechanisms, which are usually interlaced in complex regulatory networks and can overall function to either attenuate or amplify molecular noise and stochasticity. In this context, a clear example of the need of *ad hoc* tools to properly analyze the functioning of biological systems when different temporal scales and molecular amounts simultaneously occur, is the Ras/cAMP/PKA pathway in yeast. This signalling pathway regulates cell metabolism and cell cycle progression in response to nutritional sensing and stress conditions through multiple feedback loops [19]; because of such complex interplay, it is not easy to predict its behavior in different growth conditions or in response to various stress signals. To understand the role of the negative feedback controls that are present in this system, in [2, 16, 8] we defined and analyzed a mathematical model of the Ras/cAMP/PKA pathway, focusing our attention on the mechanisms that allow the emergence of oscillatory regimes in cAMP dynamics.

The model of the Ras/cAMP/PKA pathway was initially developed according to the stochastic formulation of chemical kinetics [12], and then translated into a "generalized mass-action based" model. The comparison between stochastic and deterministic simulations evidenced different quantitative and qualitative results under different conditions [2], suggesting a functional role of intrinsic noise. The application of hybrid approach on this model is currently in progress, in order to accurately deal with the small molecular amounts of some pivotal proteins (Cdc25, Ira2) and the large amount of guanine nucleotides (GTP, GDP), which both have shown to represent key regulatory factors for the establishment of oscillatory regimes within the pathway.

We have exploited our GPU implementations of stochastic, deterministic and hybrid simulation algorithms to execute the massive number of simulations that were necessary to carry out an extensive analysis of this signaling pathway. Our preliminary tests confirmed that the GPU implementation is from 10 to 100 times faster than the standard CPU implementation of the same simulation algorithms, therefore allowing deeper investigations of the system. The general GPU suite that we are developing

is therefore a promising tool to easily perform intensive and fast computational analysis of biological systems.

# References

[1] B.B. Aldridge, J.M. Burke, D.A. Lauffenburger & P.K. Sorger (2006): *Physicochemical modelling of cell signalling pathways. Nat. Cell Biol.* 8, pp. 1195–1203. doi:10.1038/ncb1497

[2] D. Besozzi, P. Cazzaniga, D. Pescini, G. Mauri, S. Colombo & E. Martegani (2012): *The role of feedback control mechanisms on the establishment of oscillatory regimes in the Ras/cAMP/PKA pathway in* S. cerevisiae. *EURASIP J. Bioinf. Syst. Biol.* 2012(10).

[3] Y. Cao, D.T. Gillespie & L.R. Petzold (2005): *The slow-scale stochastic simulation algorithm. J. Chem. Phys.* 122(1), p. 014116. doi:10.1063/1.1824902

[4] Y. Cao, D.T. Gillespie & L.R. Petzold (2006): *Efficient step size selection for the tau-leaping simulation method. J. Chem. Phys.* 124, p. 044109. doi:10.1063/1.2159468

[5] G. Caravagna, R. Barbuti & A. d'Onofrio (2012): *Fine-tuning anti-tumor immunotherapies via stochastic simulations. BMC Bioinformatics* 13(Suppl 4), p. S8. doi:10.1063/1.1378322

[6] G. Caravagna, A. d'Onofrio, P. Milazzo & R. Barbuti (2010): *Tumour suppression by immune system through stochastic oscillations. J. Theor. Biol.* 265(3), pp. 336 – 345. doi:10.1016/j.jtbi.2010.05.013

[7] G. Caravagna, G. Mauri & A. d'Onofrio (2013): *The interplay of intrinsic and extrinsic bounded noises in biomolecular networks. PLOS ONE* 8(2), p. e51174. doi:10.1371/journal.pone.0051174.t006

[8] P. Cazzaniga, D. Pescini, D. Besozzi, G. Mauri, S. Colombo & E. Martegani (2008): *Modeling and stochastic simulation of the Ras/cAMP/PKA pathway in the yeast* Saccharomyces cerevisiae *evidences a key regulatory function for intracellular guanine nucleotides pools. J. Biotechnol.* 133(3), pp. 377–385. doi:10.1016/j.jbiotec.2007.09.019

[9] I.C. Chou & E.O. Voit (2009): *Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math. Biosci.* 219, pp. 57–83. doi:10.1016/j.mbs.2009.03.002

[10] M.H.A. Davis (1984): *Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. J. Roy. Stat. Soc. B Met.* 46(3), pp. pp. 353–388.

[11] A. Eldar & M.B. Elowitz (2010): *Functional roles for noise in genetic circuits. Nature* 467, pp. 167–173. doi:10.1038/nature09326

[12] D.T. Gillespie (1977): *Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem.* 81(25), pp. 2340–2361. doi:10.1021/j100540a008

[13] T.G. Kurtz (1972): *The relationship between stochastic and deterministic models for chemical reactions. J. Chem. Phys.* 57(7), pp. 2976–2978. doi:10.1063/1.1678692

[14] M.S. Nobile, D. Besozzi, P. Cazzaniga, G. Mauri & D. Pescini (2013): *cupSODA: a CUDA-powered simulator of mass-action kinetics.* In: *Proceedings of 12th International Conference on Parallel Computing Technologies, LNCS* 7979. In press.

[15] M.S. Nobile, P. Cazzaniga, D. Besozzi, D. Pescini & G. Mauri: *Massive parallel tau-leaping stochastic simulations on GPUs for the analysis of biological systems.* Manuscript in preparation.

[16] D. Pescini, P. Cazzaniga, D. Besozzi, G. Mauri, L. Amigoni, S. Colombo & E. Martegani (2012): *Simulation of the Ras/cAMP/PKA pathway in budding yeast highlights the establishment of stable oscillatory states. Biotechnol. Adv.* 30, pp. 99–107. doi:10.1016/j.biotechadv.2011.06.014

[17] H. Salis & Y. Kaznessis (2005): *Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. J. Chem. Phys.* 122, p. 054103. doi: 10.1063/1.1835951

[18] Z. Szallasi, J. Stelling & V. Periwal (2006): *Systems Modeling in Cellular Biology.* The MIT Press.

[19] J.M. Thevelein (1994): *Signal transduction in yeast. Yeast* 10, pp. 1753–1790. doi:10.1002/yea.320101308

[20]  D. Wilkinson (2009): *Stochastic modelling for quantitative description of heterogeneous biological systems*. *Nat. Rev. Genet.* 10, pp. 122–133. doi:10.1038/nrg2509