

dr Andrzej Mizera
Załącznik nr 3 do wniosku o przeprowadzenie postępowania habilitacyjnego

Autoreferat

**Metody obliczeniowe służące do analizy i sterowania dużymi
modelami boolowskimi sieci regulatorowych genów**

dr Andrzej Mizera

Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski



Warszawa, 2025

Spis treści

1	Imię i nazwisko	3
2	Posiadane dyplomy, stopnie naukowe lub artystyczne	3
3	Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych	4
4	Omówienie osiągnięcia naukowego	6
4.1	Tytuł osiągnięcia naukowego	6
4.2	Publikacje wchodzące w skład osiągnięcia naukowego	6
4.3	Opis osiągnięcia naukowego	7
4.3.1	Wprowadzenie	7
4.3.2	Metody analizy stanów ustalonych w probabilistycznych sieciach boolowskich	12
4.3.3	Konstruowanie modeli PBN na podstawie danych	19
4.3.4	Wyznaczanie atraktorów w synchronicznych sieciach boolowskich	20
4.3.5	Wyznaczanie atraktorów w asynchronicznych sieciach boolowskich	23
4.3.6	Sterowanie sieciami boolowskimi	23
4.3.7	Wykraczając poza formalizm sieci Boolowskich – złożona długofalowa dynamika systemów reakcji	26
5	Aktywność naukowa albo artystyczna realizowana w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury	31
6	Osiągnięcia dydaktyczne, organizacyjne oraz popularyzujące naukę	31
6.1	Osiągnięcia dydaktyczne	31
6.1.1	Prowadzenie zajęć dydaktycznych	31
6.1.2	Promotorstwo pomocnicze prac doktorskich	32
6.1.3	Promotorstwo prac magisterskich	32
6.2	Osiągnięcia organizacyjne	33
6.3	Osiągnięcia w zakresie popularyzacji nauki	33
7	Dodatkowe informacje	34
7.1	Wykaz publikacji niewchodzących w skład osiągnięcia naukowego	34
7.2	Rozszerzone streszczenia	38
7.3	Raporty naukowe	38
7.4	Narzędzia programowe niewchodzące w skład osiągnięcia naukowego	38

1 Imię i nazwisko

Andrzej Mizera

Instytut Informatyki
Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski
Banacha 2, 02-097 Warszawa
e-mail: amizera@mimuw.edu.pl

Instytut Badawczy IDEAS
Królewska 27, 00-060 Warszawa
e-mail: andrzej.mizera@ideas.edu.pl

2 Posiadane dyplomy, stopnie naukowe lub artystyczne

2011-08-26 Praca doktorska obroniona z wyróżnieniem na Åbo Akademi University, Turku, Finlandia.

- Dyscyplina podstawowa: informatyka
- Dyscyplina dodatkowa: matematyka

Tytuł pracy doktorskiej:

“Methods for Construction and Analysis of Computational Models in Systems Biology. Applications to the Modelling of the Heat Shock Response and the Self-Assembly of Intermediate Filaments.”

Promotorzy:

Prof. Ion Petre, Department of Information Technologies, Åbo Akademi University

Dr hab. Barbara Gambin, prof. IPPT PAN, Instytut Podstawowych Problemów Techniki, Polska Akademia Nauk

2005-09-28 Obrona pracy magisterskiej na kierunku Informatyka na wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego.

Temat pracy magisterskiej: „Modelowanie optymalnej sieci regulacji genów”.

Promotor pracy magisterskiej: prof. dr hab. Jerzy Tiuryn.

3 Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

- 2025-07-01 – obecnie **Instytut Badawczy IDEAS [Główny specjalista badawczo-techniczny] i Uniwersytet Warszawski [Adiunkt]**
Główny specjalista badawczo-techniczny w Instytucie Badawczym IDEAS prowadzącego działalność w zakresie sztucznej inteligencji i gospodarki cyfrowej oraz adiunkt w Instytucie Informatyki na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego.
Tematyka badawcza: Sztuczna inteligencja dla skalowalnego sterowania dużymi modelami sieci regulatorowych genów. Kierownik projektu badawczego EdgeCR finansowanego przez Narodowe Centrum Nauki (Grant 2023/51/I/ST6/02864).
- 2023-01-01 – 2025-06-30 **IDEAS-NCBR Sp. z o.o. [Staż podoktorski] i Uniwersytet Warszawski [Adiunkt]**
Staż podoktorski w centrum badawczo-rozwojowym IDEAS-NCBR Sp. z o.o. działającym w obszarze sztucznej inteligencji i gospodarki cyfrowej oraz adiunkt w Instytucie Informatyki na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego.
Tematyka badawcza: Sztuczna inteligencja dla skalowalnego sterowania dużymi modelami sieci regulatorowych genów.
- 2021-05-01 – 2022-12-31 **Uniwersytet Luksemburski [Kierownik projektu, Staż podoktorski]**
Kierownik projektu „Autonomous Trustworthy Monitoring and Diagnosis of CubeSat Health (ATMonSAT)” współfinansowanego przez Europejską Agencję Kosmiczną (ESA).
Tematyka badawcza: System monitorowania i detekcji anomalii w satelitach typu CubeSat oparty na metodach wytłumaczalnej sztucznej inteligencji (ang. *explainable AI*) oraz uczenia głębokiego.
- 2019-04-15 – 2021-04-30 **Uniwersytet Luksemburski [Staż podoktorski]**
Staż podoktorski w grupie badawczej „Security and Trust of Software Systems” prowadzonej przez prof. Sjouke Mauw na wydziale Faculty of Science, Technology and Medicine, Uniwersytet Luksemburski.
Tematyka badawcza: Data Science: zastosowanie i rozwój metod uczenia maszynowego dla problemu fuzji danych w kontekście analizy wielospektralnego obrazowania satelitarnego oraz do problemu detekcji anomalii.

- 2018-08-01 – 2019-04-14 **Instytut Podstaw Informatyki, Polska Akademia Nauk [*Badacz wizytujący*]**
Tematyka badawcza: Obliczeniowa biologia systemów: opracowanie metod formalnych i algorytmów do analizy dynamiki dużych sieci boolowskich oraz probabilistycznych sieci boolowskich modelujących procesy komórkowe.
- 2017-08-01 – 2018-07-31 **Luxembourg Centre for Systems Biomedicine (LCSB) i Luxembourg Institute of Health (LIH) [*Staż podoktorski*]**
Łączony staż podoktorski w grupie badawczej Computational Biology Group w Luxembourg Centre for Systems Biomedicine (LCSB) oraz w Allergology - Immunology - Inflammation Research Unit, Luxembourg Institute of Health (LIH).
Tematyka badawcza: Obliczeniowa biologia systemów: opracowanie metody obliczeniowej określającej synergiczne zbiory czynników transkrypcyjnych determinujących podtypy komórek na podstawie danych RNA-seq dla pojedynczych komórek.
- 2012-08-01 – 2017-07-31 **Uniwersytet Luksemburski [*Staż podoktorski*]**
Staż podoktorski w grupie badawczej „Security and Trust of Software Systems” prowadzonej przez prof. Sjouke Mauw, Computer Science and Communications Research Unit, Faculty of Science, Technology and Communication, Uniwersytet Luksemburski.
Tematyka badawcza: Obliczeniowa biologia systemów: opracowanie metod formalnych i algorytmów do analizy dynamiki dużych sieci boolowskich oraz probabilistycznych sieci boolowskich modelujących procesy komórkowe.
- 2012-01-01 – 2012-07-31 **Instytut Podstawowych Problemów Techniki, Polska Akademia Nauk [*Adiunkt*]**
Adiunkt w Zakładzie Ultradźwięków w Instytucie Podstawowych Problemów Techniki PAN.
Tematyka badawcza: Zastosowanie hipertermii w medycynie: analiza wstecznie rozproszonych sygnałów ultradźwiękowych pod kątem możliwości ich wykorzystania do nieinwazyjnego pomiaru wzrostu temperatury w procesie nagrzewania tkanek miękkich za pomocą skoncentrowanej wiązki ultradźwiękowej.
- 2011-09-15 – 2011-12-31 **Instytut Podstawowych Problemów Techniki, Polska Akademia Nauk [*Programista*]**
Programista w Zakładzie Ultradźwięków w Instytucie Podstawowych Problemów Techniki PAN.

4 Omówienie osiągnięcia naukowego¹

4.1 Tytuł osiągnięcia naukowego

„Metody obliczeniowe służące do analizy i sterowania dużymi modelami boolowskimi sieci regulatorowych genów”

Osiągnięcie naukowe w formie cyklu powiązanych tematycznie artykułów naukowych, o którym mowa w art. 219 ust. 1 pkt. 2 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2021 r. poz. 478 z późn. zm.), zostało udokumentowane poniżej w rozdziale 4.2. Wkład własny w każdą z publikacji wchodzących w skład osiągnięcia naukowego został przedstawiony w rozdziale I załącznika nr 4 „Wykaz osiągnięć naukowych”. Kopie publikacji wraz z oświadczeniami o wkładzie współautorów zostały dołączone jako odrębne załączniki (odpowiednio załącznik nr 5 i załącznik nr 6).

4.2 Publikacje wchodzące w skład osiągnięcia naukowego

Publikacje

Wkład własny w każdą z publikacji wchodzących w skład osiągnięcia naukowego został przedstawiony w rozdziale I załącznika nr 4 „Wykaz osiągnięć naukowych”.

- [P1] **A. Mizera**, J. Pang, and Q. Yuan. Fast simulation of probabilistic Boolean networks. In Proc. 14th International Conference on Computational Methods in Systems Biology (CMSB'16), *Lecture Notes in Computer Science*, vol. 9859, pp. 216–231. Springer-Verlag, Cham, 2016.
- [P2] **A. Mizera**, J. Pang, and Q. Yuan. GPU-accelerated Steady-state Computation of Large Probabilistic Boolean Networks. *Formal Aspects of Computing*, 31(1):27–46, 2019.
- [P3] **A. Mizera**, J. Pang, and Q. Yuan. Reviving the two-state Markov chain approach. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(5):1525–1537, 2018.
- [P4] **A. Mizera**, J. Pang, and Q. Yuan. Parallel approximate steady-state analysis of large probabilistic Boolean networks. In Proc. 31st ACM Symposium on Applied Computing (SAC'16), pp. 1–8, ACM Press, New York, NY, 2016.
- [P5] P. Trairatphisan, **A. Mizera**, J. Pang, A. A. Tantar, and T. Sauter. optPBN: An Optimisation Toolbox for Probabilistic Boolean Networks. *PLoS ONE*, 9(7):980011–15, 2014.
- [P6] Q. Yuan, H. Qu, J. Pang, and **A. Mizera**. Improving BDD-based attractor detection for synchronous Boolean networks. *SCIENCE CHINA Information Sciences*, 59(8):080101:1–16, 2016.
- [P7] **A. Mizera**, J. Pang, H. Qu, and Q. Yuan. A new decomposition method for attractor detection in large synchronous Boolean networks. In Proc. 3rd International

¹W rozumieniu art. 219 ust. 1 pkt. 2 ustawy z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* (Dz. U. z 2021 r. poz. 478 z późn. zm.).

Symposium on Dependable Software Engineering: Theories, Tools, and Applications – SETTA’17, Lecture Notes in Computer Science, vol. 10606, pp. 232–249. Springer, Cham, 2017. (**Best Paper Award Honorable Mention**)

- [P8] Q. Yuan, **A. Mizera**, J. Pang, and H. Qu. A new decomposition-based method for detecting attractors in synchronous Boolean networks. *Science of Computer Programming*, 180:18–35, 2019.
- [P9] **A. Mizera**, J. Pang, H. Qu, and Q. Yuan. Taming asynchrony for attractor detection in large Boolean networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(1):31–42, 2019.
- [P10] S. Paul, C. Su, J. Pang, and **A. Mizera**. A decomposition-based approach towards the control of Boolean networks. In *Proc. 9th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics – BCB’18*, pp. 11–20. ACM Press, New York, 2018.
- [P11] S. Paul, C. Su, J. Pang, and **A. Mizera**. An efficient approach towards the source-target control of Boolean networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6):1932–1945, 2020.
- [P12] **A. Mizera** and J. Zarzycki. pbn-STAC: Deep Reinforcement Learning-based Framework for Cellular Reprogramming. *Theoretical Computer Science*, 1049:Article 115382, 2025.
- [P13] S. Azimi, C. Panchal, **A. Mizera**, and I. Petre. Multi-Stability, Limit Cycles, and Period-Doubling Bifurcation with Reaction Systems. *International Journal of Foundations of Computer Science*, 28(8):1007–1020, 2018.

Udział w rozwoju narzędzi programowych wchodzących w skład osiągnięcia naukowego

- [NP1] **ASSA-PBN** (<https://satoss.uni.lu/software/ASSA-PBN>) – narzędzie programowe do modelowania, symulacji i analizy probabilistycznych sieci boolowskich
- [NP2] **optPBN** (<http://sourceforge.net/projects/optpbn>) – narzędzie programowe oparte na środowisku Matlab do konstruowania probabilistycznych sieci boolowskich
- [NP3] **pbn-STAC** (<https://github.com/jakub-zarzycki2022/pbn-stac>) – narzędzie programowe oparte na głębokim uczeniu przez wzmacnianie, przeznaczone do sterowania (probabilistycznymi) sieciami boolowskimi

4.3 Opis osiągnięcia naukowego

4.3.1 Wprowadzenie

Dążenie do poprawy zdrowia i wydłużenia życia towarzyszyło ludzkości od zarania dziejów. Jednakże postęp w tych odwiecznych zmaganiach wymaga znacznie lepszego zrozumienia mechanizmów rządzących chorobami. Dotąd opracowane metody leczenia wywodzą się albo z medycyny tradycyjnej, co oznacza, że swoje źródła mają w wiedzy, umiejętnościach i praktykach opartych na teoriach, przekonaniach i doświadczeniach swoistych dla danych kultur, albo są wynikiem medycyny opartej na faktach, tzn. powstały na bazie wiarygodnych dowodów naukowych dotyczących skuteczności i bezpieczeństwa terapii [18,

17]. Zarówno medycyna tradycyjna, jak i medycyna oparta na faktach umożliwiają skuteczne przewyciężanie stosunkowo prostych chorób. Nadal jednak skomplikowane choroby stanowią ogromne wyzwanie. Wyzwanie to wiąże się w dużej mierze z faktem, że geny i produkty genów w komórce nie funkcjonują w odosobnieniu, ale poprzez ściśle powiązaną sieć komponentów molekularnych, takich jak geny, białka, informacyjny RNA (ang. *messenger RNA*, mRNA), cząsteczki mikroRNA (miRNA), które współdziałają ze sobą w celu podtrzymywania procesów wewnątrzkomórkowych. Pojęcie sieci umożliwia ujęcie na wyższym poziomie abstrakcji faktu, że geny wpływają na ekspresję innych genów poprzez sieć molekularnych interakcji regulatorowych zakodowanych w genomie [13]. Sieć ta, nazywana **siecią regulatorową genów** (ang. *gene regulatory network*, GRN) jest wysoce wyrafinowanym i złożonym systemem opartym na dodatnich i ujemnych pętlach sprzężenia zwrotnego (ang. *feedback loop*) oraz sprzężenia w przód (ang. *feedforward loop*). Jej dynamika charakteryzuje się emergentnymi własnościami wyższego rzędu. Sieć regulatorowa genów konstituuje mechanizmy regulacji ekspresji genów, które tworzą centralny aparat sterowania procesami wewnątrzkomórkowymi, odpowiedzią na sygnały zewnętrzne oraz homeostazę, tj. utrzymywanie równowagi przez odpowiednie dostosowanie do zmieniających się warunków środowiskowych. Z punktu widzenia medycyny translacyjnej, wyjaśnienie funkcjonowania sieci regulatorowej genów jest kluczowym krokiem do zrozumienia mechanizmów życia, a docelowo do opracowania skutecznych leków oraz terapii leczenia złożonych chorób.

Pojedyncze geny zmieniają swoją ekspresję wskutek interakcji sieci regulatorowych genów. **Profil ekspresji genów (transkryptom)** rozumiany jako zestaw pomiarów ekspresji dla wszystkich genów, odzwierciedla stan sieci regulatorowej genów. Sieć wykazuje immanentne, swoiste zachowania dynamiczne, które na sposób emergentny wyłaniają się ze strukturalnych powiązań pomiędzy elementami sieci, tzn. profil ekspresji genów ewoluuje w czasie w ściśle określony sposób podporządkowany mechanizmom regulatorowym ukonstytuowanym przez wzajemne interakcje pomiędzy genami, aby ostatecznie osiągnąć i ustabilizować się w jednym z **atraktorów** sieci regulatorowej genów [15]. Pojęcie atraktora wywodzi się z teorii układów dynamicznych i odnosi się do podzbioru stanów, do których dynamika systemu w dłuższej perspektywie czasu ogranicza się i do których powraca w przypadku opuszczenia atraktora na skutek perturbacji spowodowanych czynnikami zewnętrznymi. Atraktor może być **atraktorem stałopunktowym** (ang. *fixed-point attractor*), tj. singletonem składającym się z pojedynczego stanu stabilnego, występującym na skutek wzajemnie równoważących się interakcji pomiędzy genami w sieci, lub **atraktorem złożonym** (ang. *complex attractor*), tj. podzbiorem stanów, do których dynamika sieci zostaje ograniczona bez względu na stochastyczne fluktuacje molekularne [20].

Stany należące do dowolnego z atraktorów określane są mianem **stanów atraktorycznych**. Wykazano, że typowe cechy sieci regulatorowej genów, takie jak rzadkość grafu interakcji oraz obecność zarówno nieliniowych interakcji regulatorowych, jak i pętli dodatniego sprzężenia zwrotnego, skutkują występowaniem wielu atraktorów w obrębie dynamiki danej sieci [16]. Atraktory stanowią repertuar możliwych długofalowych zachowań sieci regulatorowej genów. W szczególności odpowiadają one dyskretnym, odrębnym stanom fenotypowym komórki, tj. zróżnicowanym typom komórek (ang. *differentiated cell types*) [19] lub, bardziej ogólnie, losom komórek (ang. *cell fates*) [12, 14, 32], takim jak podział, różnicowanie czy apoptoza, rozumianymi tutaj jako stany a nie procesy. To drugie zostało po raz pierwszy eksperymentalnie potwierdzone w pracy [14].

Atraktory to szczególne punkty w przestrzeni stanów sieci regulatorowej genów, które odpowiadają „stanom o najniższej energii”. Można je postrzegać jako położone na dnie studni quasi-potencjału, odpowiadającej basenowi przyciągania atraktora [15]. Koncep-

cja quasi-potencjału znajduje odzwierciedlenie w *krajobrazie epigenetycznym* Conrada H. Waddingtona [36], czyli słynnej metaforze wizualnej zaproponowanej w 1957 roku w celu wyjaśnienia procesu różnicowania komórek podczas rozwoju oraz dyskretnej natury typów komórek. W tym ujęciu sieć regulatorowa genów formuje krajobraz określonego kształtu z dolinami oddzielonymi wzgórzami, gdzie każdy typ komórki jest przedstawiony jako dolina, a zmiana fenotypu komórki podczas ustalania się rodowodu komórkowego i różnicowania jest reprezentowana jako przejście z jednej doliny do drugiej. Topografia krajobrazu epigenetycznego jest determinowana przez wzajemne interakcje sieci regulatorowej genów.

Biorąc pod uwagę skomplikowane molekularne współzależności funkcjonalne, złożona choroba rzadko jest wynikiem zaburzenia działania pojedynczego genu. Raczej powszechnie przyjmuje się, że taka choroba wynika z zakłóceń w złożonych sieciach wewnątrzkomórkowych i międzykomórkowych [2]. Zakłócenie subtelnej dynamiki sieci regulatorowej genów prowadzi do zniekształcenia fenotypu. Wiele chorób, w tym liczne zaburzenia rozwojowe, a przede wszystkim nowotwory, jest bezpośrednio wynikiem zaburzenia mechanizmów regulacyjnych na poziomie komórkowym (obszerne omówienie tego zagadnienia zob. np. [23]). W terminach krajobrazu epigenetycznego nie wszystkie doliny są zajmowane przez te komórki, które odpowiadają normalnym typom komórek organizmu wielokomórkowego. Występowanie pozostałych dolin reprezentujących niepożądane profile ekspresji genów jest postrzegane za nieunikniony skutek uboczny złożoności struktury sieci regulatorowej genów. Niektóre z tych profili, tzn. te związane z fenotypem proliferacyjnym, można uważać za „atraktory nowotworowe”. W tym sensie występowanie nowotworów jest immanentną konsekwencją istnienia sieci regulatorowej genów; jest to cena, jaką metazoa płaci za mechanizm rozwoju wielokomórkowego [15]. Chociaż nieprawidłowe atraktory nie ujawniają się w trakcie normalnego rozwoju, zaburzenie mechanizmów regulacyjnych może uczynić niektóre z nich dostępnymi z punktu widzenia dynamiki systemu. Z perspektywy postrzegania złożonej choroby jako nieprawidłowego atraktora w wielowymiarowej przestrzeni stanów sieci regulatorowej genów, leczenie choroby odpowiada wyprowadzeniu dynamiki z nieprawidłowego atraktora i skierowaniu jej w kierunku atraktora docelowego, który odpowiada „zdrowemu” fenotypowi. Taki punkt widzenia pomaga w konceptualizacji eksperymentalnych procedur *reprogramowania komórkowego*, rozumianego jako sztucznie wymuszana zmiana jednego typu komórki w inny.

Poznanie zasad projektowych sieci regulatorowej genów oraz zrozumienie funkcjonowania tychże sieci jest kluczowe dla osiągnięcia ostatecznego celu, jakim jest leczenie złożonych chorób. W szczególności są to elementy kluczowe dla możliwości identyfikacji właściwych celów interwencji oraz strategii manipulacji sieciami regulatorowymi genów, które są w stanie skierować dynamikę sieci w stronę pożądanego, „zdrowego” funkcjonowania. Ponieważ sieci regulatorowe genów są systemami złożonymi, jak wspomniano wcześniej, zrozumienie ich dynamiki wymaga podejścia holistycznego, czyli opartego na metodach i technikach pozwalających na analizę systemu jako całości, w odróżnieniu od klasycznych metod redukcjonistycznych badających poszczególne elementy składowe i wywodzących z wyników tych badań wnioski dotyczące wyjściowego systemu. Opracowywanie odpowiednich ram teoretycznych i technik do holistycznej analizy złożonych systemów biologicznych jest obszarem badań obliczeniowej biologii systemowej – interdyscyplinarnej dziedziny nauki zajmującej się badaniem systemów komórkowych celem dogłębnego zrozumienia istoty zasad projektowych i strategii stosowanych przez naturę do konstruowania mechanizmów realizujących złożone funkcje biologiczne.

Modus operandi obliczeniowej biologii systemowej w badaniach nad systemami lub procesami komórkowymi polega na budowaniu modeli sieci interakcji, zazwyczaj w oparciu o różnego rodzaju formalizmy matematyczne, a następnie analizowaniu ich za pomocą

technik obliczeniowych w połączeniu z weryfikacją eksperymentalną. Działania związane z budową i analizą modeli prowadzą do systematyzacji wiedzy biologicznej w kontekście wyjaśniania badanych mechanizmów. Co istotne, pozwalają one na identyfikację brakujących elementów w naszym rozumieniu rzeczywistości biologicznej: model obliczeniowy lub matematyczny dostarcza formalnych narzędzi do zrozumienia i interpretacji współdziałania różnych części systemu, a proces budowy modelu może ujawnić luki w aktualnym stanie wiedzy biologicznej².

Nawet jeśli do modelowania sieci regulatorowych genów stosuje się różne formalizmy matematyczne, niemal wszystkie uogólniają się do **sieci interaktomowych** (ang. *interactome networks*) w postaci węzłów połączonych skierowanymi krawędziami, gdzie węzeł reprezentuje jednostkę molekularną, taką jak gen lub białko, a krawędź odpowiada bezpośredniemu wpływowi przyczynowemu jednego genu na inny [10]. Zazwyczaj sieci interaktomowe są analizowane pod kątem zrozumienia relacji między ich strukturą a dynamiką. Do badania tych zależności stosuje się różne formalizmy matematyczne, spośród których najczęściej spotykane to sieci boolowskie, probabilistyczne sieci boolowskie, sieci Petriego, sieci Bayesowskie oraz układy równań różniczkowych zwyczajnych.

Z jednej strony, uchwycenie emergentnych własności sieci regulatorowych genów wymaga budowy modeli obliczeniowych obejmujących jeśli nie kompletne genomy, to przynajmniej cały ich rdzeń regulatorowy (ang. *regulatory core*). Modele te charakteryzują się dużymi rozmiarami i złożonością. Z drugiej strony, określenie odpowiednich wartości parametrów biochemicznych dla tak rozległych modeli jest niezwykle trudne, często wręcz niemożliwe. Dlatego konieczne jest znalezienie równowagi między realistycznym odwzorowaniem bio-złożoności a bardziej abstrakcyjnym ujęciem, skutkujące mocno uproszczonym opisem, aczkolwiek nadal umożliwiającym wgląd w kluczowe elementy dynamiki sieci regulatorowej genów. Aby zrozumieć podstawowe zasady projektowe sieci regulatorowych genów i ich funkcjonowanie, należy skupić się na tym co istotne bez naruszania integralności struktury sieci, pomijając w miarę możliwości zbędne szczegóły biochemiczne. Dwa formalizmy są szczególnie odpowiednie do tego celu: sieci boolowskie oraz ich rozszerzenie – probabilistyczne sieci boolowskie [30].

Formalizm **sieci boolowskich** (ang. *Boolean networks*, BNs), wprowadzony przez S. Kauffmaną [21, 19], ma tę zaletę, że jest prosty, a jednocześnie w stanie uchwycić istotne dynamiczne właściwości badanego systemu, co ułatwia modelowanie dużych systemów biologicznych jako całości [4]. Stany sieci boolowskiej są binarnymi wektorami, gdzie każdy element wektora, tj. 0 (OFF) lub 1 (ON), reprezentuje stan odpowiedniego genu w sieci regulatorowej genów. Sieć boolowska ewoluje, tzn. przechodzi z jednego stanu w drugi zgodnie z funkcjami boolowskimi określającymi zachowanie poszczególnych genów w zależności od innych. Użycie tego formalizmu pozwala modelować regulację genów jako sieć elementów o dynamice dyskretniej bez konieczności dopasowywania wartości parametrów ilościowych. Ponadto dynamika sieci boolowskich jest nieliniowa, co czyni ten formalizm odpowiednim do modelowania systemów biologicznych. Wiele prac badawczych potwierdziło skuteczność sieci boolowskich w przewidywaniu zachowania sieci regulatorowej genów, zob. np. [35, 33, 24], oraz zgodność atraktorów sieci boolowskich z określonymi fenotypowymi stanami biologicznymi, zob. np. [4, 6]. Sieci boolowskie pozwalają uchwycić wiele biologicznie istotnych zjawisk, takich jak fenomen przełączania (ang. *switch-like behaviour*), stabilność czy histereza [11].

²Działania polegające na budowie obliczeniowego lub matematycznego modelu systemu biologicznego mogą prowadzić do identyfikacji przyszłych kierunków badań w biologii, np. mogą zasugerować prace eksperymentalne, które pogłębią zrozumienie funkcjonowania badanego systemu. W ten sposób obliczeniowa biologia systemowa może stymulować rozwój biologii.

W ogólnym przypadku wyznaczanie atraktorów w dużej sieci boolowskiej jest zadaniem trudnym obliczeniowo. Udowodniono, że pytanie o istnienie globalnego stanu zgodnego ze wszystkimi funkcjami boolowskimi danej sieci, czyli istnienia atraktora stałopunktowego, jest problemem NP-zupełnym [1], a znalezienie atraktora o najkrótszym okresie w przypadku sieci synchronicznych jest problemem NP-trudnym [39].

Formalizm **probabilistycznych sieci boolowskich** (ang. *probabilistic Boolean networks*, PBNs) stanowi probabilistyczne uogólnienie standardowych sieci boolowskich. Przesłanki przemawiające za ich wprowadzeniem są następujące. Po pierwsze, z conceptualnego punktu widzenia wydaje się prawdopodobne, że obserwowana powtarzalność w funkcjonowaniu genów i interakcji pomiędzy nimi nie jest sztywno zaprogramowana przez reguły logiczne, lecz wynika raczej z wewnętrznej, samoorganizującej się stabilności układu dynamicznego, pomimo istnienia stochastycznych komponentów w komórce. Po drugie, z empirycznego punktu widzenia, zakładanie pojedynczej reguły logicznej dla każdego genu może prowadzić do błędnych wniosków podczas wnioskowania tych reguł na podstawie danych ekspresji genów. Jest to szczególnie istotne, ponieważ pomiary są zazwyczaj silnie zaszumione, a ilość dostępnych danych jest niewielka w odniesieniu do liczby decyzji jakie należy podjąć odnośnie do struktury sieci [29].

Formalizm sieci PBN został zaprojektowany tak, aby zapewnić osiągnięcie zestawu celów modelowania przez konstruowany model, mianowicie: (i) uwzględnienie zależności opartych na regułach między genami; (ii) umożliwienie systematycznego badania globalnej dynamiki sieci; (iii) odporność na niepewności zarówno na poziomie danych, jak i wyboru modelu; oraz (iv) umożliwienie kwantyfikacji względnego wpływu i wrażliwości genów w ich interakcjach z innymi genami [29].

Dynamikę sieci PBN można badać za pomocą teorii łańcuchów Markowa. W szczególności dynamika sieci PBN może być przedstawiona jako łańcuch Markowa z dyskretnym czasem, gdzie macierz przejścia jest w pełni zadana przez możliwe funkcje boolowskie opisujące zachowanie każdego z genów oraz ich prawdopodobieństwa wyboru. Możliwe jest zatem badanie długofalowej dynamiki sieci PBN przy użyciu teorii zachowanie asymptotycznego rozpatrywanego łańcucha Markowa.

Moja praca badawcza przedstawiona w niniejszym omówieniu dotyczy rozwoju metod formalnych do analizy złożonych systemów biologicznych. Koncentruje się ona na opracowywaniu nowych skalowalnych algorytmów do analizy i sterowania długofalowym zachowaniem dużych modeli obliczeniowych opisujące systemy komórkowe charakteryzujące się nieliniową dynamiką. W szczególności rozważane są duże modele BN i PBN sieci regulatorowych genów. Moje osiągnięcia badawcze swoim zakresem obejmują opracowanie:

- wydajnych algorytmów do szybkiej symulacji modeli BN i PBN;
- skalowalnych metod do wyznaczania atraktorów w synchronicznych i asynchronicznych modelach BN;
- wydajnych metod obliczeniowych do estymacji prawdopodobieństw stanów ustalonych w modelach PBN;
- skalowalnych algorytmów do sterowania modelami BN w kontekście reprogramowania komórkowego, tj. kierowania dynamiki modelu do zadanych atraktorów.

Moje poszczególne osiągnięcia badawcze związane z zagadnieniami analizy i sterowania w kontekście dużych modeli BN i PBN złożonych systemów biologicznych, a także wychodzące poza zakres tych dwóch formalizmów, zostały podsumowane w Tabeli 3.

Osiągnięcia badawcze	
Wkład w rozwój dyscypliny informatyka	Publikacje i narzędzia programowe powiązane z osiągnięciem naukowym
Wydajne algorytmy symulowania modeli PBN	P1, P2, NP1
Wydajna estymacja prawdopodobieństw rozkładów stacjonarnych w modelach PBN	P3, P4, NP1
Wnioskowanie struktury modeli PBN na podstawie pomiarów eksperymentalnych	P5, NP2
Algorytmy wyznaczania atraktorów w synchronicznych modelach BN	P6, P7, P8, NP1
Algorytmy wyznaczania atraktorów w asynchronicznych modelach BN	P9, NP1
Sterowanie modelami BN i PBN	P10, P11, P12, NP3
Opis biologicznie istotnych stanów ustalonych w modelu formalnym Systemy Reakcji	P13

Tabela 3: Podsumowanie wkładów w dyscyplinę informatyki wraz z powiązanymi publikacjami (P) i narzędziami programowymi (NP) stanowiącymi omawiane osiągnięcie naukowe.

4.3.2 Metody analizy stanów ustalonych w probabilistycznych sieciach biologicznych

Chociaż formalizm PBN jest odpowiednio zaprojektowany do modelowania złożonych systemów biologicznych, analiza modeli PBN o dużych rozmiarach stanowi istotne wyzwanie, w szczególności trudne do określenia jest długofalowe zachowanie się takich modeli. Wiele wysiłku zostało włożone w rozwój efektywnych metod obliczania prawdopodobieństwa w rozkładzie stacjonarnym, w dalszej części skrótowo określanego mianem prawdopodobieństwa stacjonarnego, znajdowania się modelu PBN w określonym podzbiorze stanów. Możliwość wyznaczania takich prawdopodobieństw ma duże znaczenie zarówno w konstrukcji modelu, np. w estymacji parametrów, jak i w analizie zachowania badanego systemu. Metody te pozwalają nie tylko zidentyfikować najbardziej prawdopodobne stany, tj. stany, które są najczęściej odwiedzane po osiągnięciu przez system dynamiki stacjonarnej i odpowiadające stanom fenotypowym komórki, ale także dostarczają narzędzi do ilościowego określenia średniego wpływu (ang. *average influence*) jednego genu na inny oraz średniej wrażliwości (ang. *average sensitivity*) genu, co w pewnym sensie stanowi miarę autonomii danego genu [29] w sieci regulatorowej genów. Znane są efektywne metody numeryczne do obliczania prawdopodobieństw stanu ustalonego w modelach PBN o małych rozmiarach. Wykorzystują one istotną cechę modeli PBN, mianowicie fakt, że model PBN może być traktowany jako łańcuch Markowa, tj. proces Markowa z czasem dyskretnym (ang. *Discrete-Time Markov Chain*, DTMC) i jego dynamikę można badać przy użyciu bogatej teorii łańcuchów Markowa³. Główna idea polega na wyznaczaniu macierzy

³Dodanie losowych perturbacji o małych prawdopodobieństwach do modelu PBN sprawia, że łańcuch staje się ergodyczny, a w konsekwencji zapewnione jest istnienie dokładnie jednego rozkładu stacjonarnego.

przejsć łańcucha Markowa odpowiadającego analizowanemu modelowi PBN. Metody te sprawdzają się w przypadku modeli PBN o małych rozmiarach, jednak nie są skalowalne i stają się niepraktyczne dla systemów o dużych rozmiarach oraz w przypadku wymagania dużej precyzji estymacji. Aby przezwyciężyć to ograniczenie, jedynym podejściem pozostaje zastosowanie metod statystycznych i próbkowania Monte Carlo łańcuchami Markowa (ang. *Markov Chain Monte Carlo*, MCMC). Metody te wymagają jednak wielu długich symulacji modeli PBN, co czyni szybkość symulacji kluczowym czynnikiem ich efektywności.

4.3.2.1 Szybkie algorytmy symulacji synchronicznych modeli PBN

W ramach omawianych osiągnięć naukowych opracowałem metody obliczeniowe do efektywnej symulacji synchronicznych modeli PBN na architekturach opartych o procesory CPU jak i GPU. Metody te zostały przedstawione w artykułach **P1** i **P2** oraz zaimplementowane w narzędziu programowym **NP1**.

Metoda zrównoleglenia symulacji w oparciu o strukturę modelu. Model PBN z perturbacjami może być symulowany za pomocą dwóch prostych kroków wynikających bezpośrednio z definicji modelu. Po pierwsze, dla każdego węzła (genu) sprawdzane jest czy następuje perturbacja i jeśli tak, to wartość węzła zostaje zmieniona. Po drugie, jeśli w żadnym z węzłów nie wystąpiła perturbacja, stan sieci zostaje zaktualizowany poprzez wybór i zastosowanie funkcji boolowskich dla każdego węzła. Jednak w celu wydajnej symulacji synchronicznych modeli PBN na procesorach CPU, w pracy **P1** proponuję podejście, zwane *metodą zrównoleglenia symulacji w oparciu o strukturę modelu* (ang. *Structure-Based Parallelisation Method*), które przyspiesza proces symulacji w oparciu o analizę struktury modelu PBN, tj. grafu bezpośrednich interakcji regulatorowych między genami. Metoda ta opiera się na dwóch kluczowych pomysłach: *redukcji sieci*, zmniejszającej całkowitą liczbę węzłów poprzez usunięcie nieistotnych elementów, oraz *grupowaniu węzłów*, które umożliwia zrównoleżone sprawdzanie perturbacji, wybór funkcji boolowskich i aktualizację wartości węzłów.

Model PBN jest redukowany poprzez analizę jego struktury w celu zidentyfikowania nieistotnych węzłów. Węzły te, określane mianem *liści* (ang. *leaf nodes*), są zdefiniowane w następujący sposób: są to węzły, które nie są brane pod uwagę przez użytkownika w analizie zachowania modelu PBN w stanie ustalonym i które albo 1) nie mają węzłów potomnych, albo 2) po iteracyjnym usunięciu wszystkich węzłów potomnych będących liśćmi, nie mają żadnych węzłów dzieci.

W przypadku modeli PBN z perturbacjami, aktualizacje stanów przy użyciu funkcji boolowskich są realizowane tylko wtedy, gdy nie występują perturbacje ani w liściach, ani w pozostałych węzłach. W pracy **P1** wykazuję, że występowanie perturbacji w liściach może być sprawdzane w stałym czasie, niezależnie od liczby liści w modelu PBN, oraz że liście mogą być usunięte bez wpływu na symulację pozostałych węzłów w modelu PBN. W szczególności wyniki analizy zachowania systemu w oparciu o symulowane trajektorie zredukowanego i oryginalnego modelu PBN będą identyczne.

Zamiast sprawdzać i przeprowadzać perturbacje węzłów pojedynczo, w pracy **P1** pokazuję, jak realizować perturbacje w sposób zrównoleżony adaptując do tego celu *metodę aliasową* (ang. *alias method*) [37], będącą wydajnym algorytmem losowania z dyskretnego rozkładu prawdopodobieństwa, umożliwiającym generowanie pojedynczej próby w stałym czasie niezależnie od rozmiaru przestrzeni probabilistycznej. Dzięki odpowiednio skonstruowanej tabeli aliasów (ang. *alias table*), dowolny podzbiór spośród określonych k węzłów może zostać poddany na raz perturbacji. Ponieważ prawdopodobieństwo perturbacji p

jest takie samo dla każdego węzła, tabelę aliasów wystarczy skonstruować tylko raz, a następnie może być ona ponownie wykorzystywana dla wszystkich $\lfloor n/k \rfloor$ grup po k węzłów, gdzie n to całkowita liczba węzłów w (zredukowanym) modelu PBN. Ponadto dowodzę, że tabela aliasów może być również użyta dla ewentualnej pozostałej grupy węzłów o rozmiarze $k' = n - \lfloor n/k \rfloor \cdot k$ bez konieczności modyfikacji wcześniej wyliczonego rozkładu prawdopodobieństwa.

Wreszcie, aby przyspieszyć symulacje modelu PBN, w pracy **P1** proponuję metodę jednoczesnej aktualizacji wartości dla grupy węzłów zgodnie z ich funkcjami boolowskimi. Kluczowa idea polega na podziale n węzłów zredukowanego modelu PBN na m grup oraz skonstruowaniu dla każdej grupy listy tzw. *złączonych funkcji boolowskich* (ang. *combined predictor functions*) wraz z odpowiadającą jej tabelą aliasów. Ta inicjalizacja stanowi adaptację metody aliasowej i umożliwia wydajny wybór złączonej funkcji boolowskiej dla każdej grupy na podstawie wygenerowanej liczby pseudolosowej: wszystkie węzły w grupie są jednocześnie aktualizowane za pomocą złączonej funkcji boolowskiej wybranej na podstawie odpowiedniej tabeli aliasów. Ze względów wydajnościowych stany PBN są przechowywane jako liczby całkowite, a aktualizacja wartości węzłów w grupie jest realizowana za pomocą logicznej operacji bitowej OR.

W pracy **P1** pokazuję, jak łączyć funkcje boolowskie oraz jak dzielić węzły na grupy, aby zapewnić odpowiedni kompromis pomiędzy następującymi dwoma wzajemnie wykluczającymi się celami. Z jednej strony korzystne jest posiadanie jak najmniejszej liczby grup, ponieważ pozwala to na aktualizację wartości wszystkich węzłów w minimalnej liczbie kroków. Z drugiej jednak strony, mniejsza liczba grup oznacza większy rozmiar każdej z nich, co może prowadzić do ogromnej liczby złączonych funkcji boolowskich przypisanych do każdej grupy. Dlatego liczba grup musi być starannie dobrana tak, aby zminimalizować ich całkowitą liczbę, jednocześnie zapewniając, że złączone funkcje boolowskie mieszczą się w dostępnej pamięci. W pracy **P1** formułuję ten podział jako zadanie optymalizacyjne i argumentuję, że jest ono podobne do NP-trudnego *problemu podziału multizbioru na możliwie równe jego podzbiory* (ang. *multiway number partitioning problem*). W celu rozwiązania sformułowanego zadania optymalizacji proponuję adaptację istniejącego algorytmu zachłannego dla problemu podziału.

Przedstawione w pracy **P1** wyniki obszernych eksperymentów na losowo wygenerowanych modelach PBN pokazują, że moje podejście znacząco przyspiesza symulacje modeli na procesorach CPU. Jest ono szczególnie efektywne w analizie rzadkich sieci z dużą liczbą liści. Zaproponowana metoda zrównoleglenia symulacji w oparciu o strukturę modelu została zaimplementowana w narzędziu programowym **NP1**.

Metoda zrównoleglenia symulacji na poziomie trajektorii. Rozwój w zakresie mocy obliczeniowej procesorów graficznych do obliczeń ogólnego przeznaczenia (GPU) i ich dostępność umożliwia masowe zrównoleglenie procesu symulacji modeli PBN. W celu urzeczywistnienia tej idei, w pracy **P2** proponuję *metodę zrównoleglenia symulacji na poziomie trajektorii* (ang. *Trajectory-Level Parallelisation Framework*), która wykorzystując możliwości GPU przyspiesza estymację prawdopodobieństw stacjonarnych znajdowania się dużego modelu PBN w zadanym podzbiory stanów. Architektura GPU zasadniczo różni się od architektury procesora CPU, a wydajność programu wykonywanego na GPU jest silnie zależna od efektywności procesów synchronizacji pomiędzy rdzeniami GPU oraz zarządzania dostępem do pamięci. Moje podejście redukuje koszt obliczeniowy synchronizacji poprzez umożliwienie każdemu rdzeniowi GPU symulacji oddzielnej trajektorii modelu PBN. W zakresie zarządzania pamięcią moja metoda wnosi dwa istotne usprawnienia. Po pierwsze, wprowadza mechanizm dynamicznego rozmieszczania danych, który pozwala na

obsługę modeli PBN o różnych rozmiarach i maksymalizuje efektywność obliczeń na GPU w przypadku stosunkowo małych modeli PBN. Po drugie, wykorzystuje specyficzny sposób przechowywania stanów i funkcji boolowskich modelu PBN w pamięci GPU, co zmniejsza zużycie pamięci oraz zwiększa szybkość dostępu do tych elementów modelu w pamięci.

Ogólnie rzecz ujmując, istnieją dwie możliwości zrównoleglenia symulacji modelu PBN. Pierwszym sposobem jest synchroniczna aktualizacja wszystkich węzłów, tzn. każdy wątek GPU aktualizuje tylko jeden węzeł PBN. Drugim sposobem jest jednoczesna symulacja wielu trajektorii. Ponieważ pierwsze podejście wymaga synchronizacji między wątkami, co jest czasochłonne przy obecnej architekturze GPU, w zaproponowanej metodzie wybrałem drugie rozwiązanie, czyli równoczesną symulację wielu trajektorii.

Ponieważ każdy krok symulacji wymaga dostępu do losowo wybieranego fragmentu informacji o strukturze modelu PBN, a koszt opóźnień (ang. *latency*) związanych z częstym dostępem do danych w wolnej pamięci globalnej GPU jest ogromny, buforowanie tych informacji w szybkiej pamięci, tj. pamięci współdzielonej (ang. *shared memory*) i rejestrach (ang. *registers*), oraz pamięci specjalnej, tj. pamięci tekstur (ang. *texture memory*) i pamięci wartości stałych (ang. *constant memory*), prowadzi do bardziej efektywnych obliczeń niż zwiększanie liczby aktywnych grup procesorów, tzw. osnów (ang. *warps*). Dlatego metoda najpierw stara się umieścić wszystkie często używane dane w możliwie największym stopniu w szybkiej i specjalnej pamięci, a następnie, w oparciu o dostępne zasoby, oblicza optymalną liczbę wątków i ich grup, tzw. bloków wątków, do uruchomienia.

Ponieważ rozmiar szybkiej pamięci jest ograniczony, jej zalety zależą od trybu dostępu do danych, oraz zapotrzebowanie na pamięć jest inne dla każdego modelu PBN, konieczne było opracowanie odpowiedniej strategii rozmieszczenia danych w pamięci. Moja metoda zapisuje różne dane w różnych typach pamięci. Mianowicie

- dane tylko do odczytu, które najprawdopodobniej będą jednocześnie wykorzystywane przez wszystkie wątki w osnowie, umieszczane są w pamięci wartości stałych;
- jeśli to możliwe, inne dane tylko do odczytu są umieszczane w pamięci współdzielonej;
- pozostałe dane są w miarę możliwości przechowywane w rejestrach.

Proponuję *dynamiczny proces decyzyjny* określający sposób rozmieszczenia często używanych danych, takich jak indeksy dodatkowych funkcji boolowskich, indeksy węzłów, od których zależą poszczególne funkcje, oraz tablice stanów bieżącego i następnego, w różnych obszarach pamięci GPU dla różnych modeli PBN. Proces ten oblicza ilość pamięci potrzebnej do przechowywania wszystkich danych dla danego modelu PBN i decyduje o ich rozmieszczeniu w zależności od wyliczonego rozmiaru wymaganej pamięci. Jeśli pamięć współdzielona i rejestry są wystarczające, wszystkie dane są przechowywane w tych dwóch szybkich rodzajach pamięci. W przeciwnym razie są one umieszczane w pamięci globalnej. W tym drugim przypadku proponuję dwa sposoby przyspieszenia dostępu. Pierwszy to wykorzystanie pamięci tekstur do szybszego odczytu danych tylko do odczytu, np. indeksów węzłów, od których zależą poszczególne funkcje. Drugi sposób to optymalizacja struktury danych w celu umożliwienia skoordynowanego dostępu, np. do bieżącego stanu. GPU zazwyczaj ma bardzo ograniczoną ilość szybkiej pamięci, a opóźnienia mogą się znacznie różnić w zależności od warunków dostępu do tego samego rodzaju pamięci, np. odczytując pamięć współdzieloną z konfliktem dostępu do tych samych banków (ang. *bank conflict*) lub bez niego. Dlatego wprowadziłem odpowiednie struktury danych do przechowywania funkcji boolowskich oraz stanów modelu PBN w celu zmniejszenia zapotrzebowania na pamięć i maksymalizacji szybkości dostępu; szczegóły dotyczące zaproponowanych struktur danych są przedstawione w pracy **P2**.

Metoda zrównoleglenia symulacji na poziomie trajektorii została zaimplementowana w narzędziu programowym **NP1**. Przeprowadzone eksperymenty na losowo wygenerowanych modelach PBN pokazują, że pozwala ona na symulację modeli PBN na GPU oraz estymację prawdopodobieństw stacjonarnych za pomocą rozszerzonej metody dwustanowego łańcucha Markowa (ang. *Extended Two-State Markov Chain Approach*)⁴ z przyspieszeniem przekraczającym dwa rzędy wielkości, tj. od 102-krotnego do 405-krotnego, w porównaniu do estymacji prawdopodobieństw stacjonarnych przy użyciu rozszerzonej metody dwustanowego łańcucha Markowa bazującej na sekwencyjnych symulacjach na CPU. Jak pokazałem w pracy **P2**, w przypadku estymacji prawdopodobieństw stacjonarnych dla zwalidowanego modelu PBN biologicznego procesu apoptozy przyspieszenie jest około 400-krotne.

4.3.2.2 Efektywne estymowanie prawdopodobieństw w rozkładach stacjonarnych dużych modeli PBN

Rozszerzona metoda dwustanowego łańcucha Markowa. Dla dużych modeli PBN, które często pojawiają się w biologii systemowej, wyznaczanie stacjonarnego rozkładu prawdopodobieństwa PBN stanowi istotne wyzwanie obliczeniowe. Aby rozwiązać ten problem, rozważam metodę dwustanowego łańcucha Markowa (ang. *two-state Markov chain approach*), tj. metodę estymacji prawdopodobieństwa w rozkładzie stacjonarnym dla podzbioru stanów łańcucha Markowa, oryginalnie przedstawioną w pracy [27]. Główna idea tego podejścia polega na skonstruowaniu abstrakcji oryginalnego łańcucha Markowa w postaci dwustanowego łańcucha Markowa, w którym oryginalne stany są zgrupowane w dwa *meta-stany*: jeden reprezentujący interesujący nas podzbiór stanów oryginalnych, a drugi obejmujący wszystkie pozostałe stany. Następnie na podstawie wygenerowanej trajektorii estymowane są prawdopodobieństwa przejść między tymi dwoma meta-stanami. Służą one do określenia w sposób iteracyjny, czy wyjściowa trajektoria dwustanowego łańcucha Markowa jest dostatecznie długa, tzn. czy próba jest odpowiedniego rozmiaru, aby na jej podstawie oszacować prawdopodobieństwo stacjonarne dla określonego podzbioru stanów PBN zadaną dokładnością. Jeżeli nie, to trajektoria jest wydłużana o brakującą liczbę kroków i prawdopodobieństwa przejścia są ponownie estymowane na podstawie wydłużonej trajektorii; w przeciwnym przypadku z użyciem prawdopodobieństw przejść obliczane jest poszukiwane prawdopodobieństwo stacjonarne.

Wykorzystanie tego podejścia do analizy modeli PBN zostało zaproponowane w [31], jednak nie było ono szeroko wykorzystywane w tym kontekście. Dlatego też w pracy **P3** ponownie zwracam uwagę na metodę dwustanowego łańcucha Markowa, dostosowując ją do analizy dużych modeli PBN i demonstrując jej użyteczność w aproksymacji prawdopodobieństw stacjonarnych tych modeli. W dalszej części niniejszego podrozdziału podsumowuję moje główne osiągnięcia w tym zakresie.

Po pierwsze, identyfikuję błędy w oryginalnych wzorach w pracy [27] określające *czas wypalania łańcucha* (ang. *'burn-in' period*) (m), tj. liczbę ignorowanych początkowych stanów trajektorii zanim zostanie osiągnięta zbieżność do rozkładu stacjonarnego, oraz rozmiaru próbki (n). Koryguję je, podając właściwe wyrażenia dla tych dwóch wielkości w **P3**⁵.

⁴Oryginalna metoda dwustanowego łańcucha Markowa (ang. *Two-State Markov Chain Approach*) oraz moje jej rozszerzenie, tj. rozszerzona metoda dwustanowego łańcucha Markowa (ang. *Extended Two-State Markov Chain Approach*), są omówione w rozdziale 4.3.2.2.

⁵Szczegółowe wyprowadzenia poprawnych wyrażeń na wielkości m i n są przedstawione w materiałach dodatkowych pracy **P3** (Appendix A i B).

Po drugie, identyfikuję problem dotyczący inicjalizacji oryginalnej metody dwustanowego łańcucha Markowa, który związany jest z rozmiarem początkowej próby, tj. długością trajektorii modelu PBN generowanej na samym początku. Jak uzasadniam w pracy **P3**, problem ten niestety może prowadzić do niedokładnych wyników estymacji prawdopodobieństw. Aby temu zapobiec, identyfikuję sytuacje, w których może wystąpić błąd estymacji, i formułuję tzw. „warunek krytyczny” (ang. *critical condition*), przy którym oryginalny algorytm może nie osiągnąć zadanej dokładności estymacji. Następnie proponuję trzy heurystyki rozszerzające oryginalną metodę w celu uniknięcia niewłaściwych inicjalizacji.

Pierwsza heurystyka, tzw. „unikanie pułapek” (ang. *pitfall avoidance*), określa bezpieczne wartości długości początkowej trajektorii używanej do wstępnej estymacji prawdopodobieństw przejść między dwoma meta-stanami abstrakcyjnego dwustanowego łańcucha Markowa, aby zapewnić zadaną dokładność i określony poziom ufności estymacji prawdopodobieństw.

Druga heurystyka, tzw. „kontrolowana wstępna estymacja parametrów α i β ” (ang. *controlled initial estimation of α and β*), zapewnia kontrolę nad jakością początkowych oszacowań prawdopodobieństw przejść między dwoma meta-stanami.

Wreszcie trzecia heurystyka, tzw. „prosta heurystyka” (ang. *simple heuristic*), to bezpośrednia, lecz skuteczna strategia zapewniająca minimalną liczbę przejść między dwoma meta-stanami w wygenerowanej trajektorii PBN zanim nastąpi pierwsze oszacowanie prawdopodobieństw przejść.

Metoda dwustanowego łańcucha Markowa implementująca te trzy heurystyki została określona mianem „Rozszerzonej metody dwustanowego łańcucha Markowa” (ang. *Extended Two-State Markov Chain Approach*). Została ona zaimplementowana w narzędziu programowym **NP1**.

Po trzecie, przeprowadzam szeroko zakrojoną ewaluację rozszerzonej metody dwustanowego łańcucha Markowa i porównuję uzyskane za jej pomocą wyniki z wynikami metody Skart [34] na dużej liczbie losowo wygenerowanych modeli PBN. Do analizy uzyskanych wyników stosuję techniki wywodzące się z uczenia maszynowego. Uzyskane wyniki pokazują, że w kontekście sieci regulatorowych genów modelowanych za pomocą PBN, rozszerzona metoda dwustanowego łańcucha Markowa w większości przypadków przewyższa metodę Skart pod względem szybkości obliczeń.

Po czwarte, demonstruję potencjał rozszerzonej metody dwustanowego łańcucha Markowa na przykładzie analizy dużego, 91-węzłowego modelu PBN apoptozy w hepatocytach. W szczególności pokazuję, że metoda umożliwia kwantyfikację zarówno *stacjonarnego wpływu genów* (ang. *steady-state influence of genes*), jak i *stacjonarnej wrażliwości genów* (ang. *steady-state sensitivity of genes*) w dużych modelach PBN⁶. Studium przypadku pokazuje, że metoda umożliwia ilościową analizę dużych modeli PBN rzeczywistych systemów biologicznych w praktyce.

Zrównoleglona przybliżona analiza modelu PBN w warunkach stacjonarności.

W pracy **P4** proponuję podejście przyspieszające analizę zachowania modelu po osiągnięciu przez niego rozkładu stacjonarnego, tj. estymacji prawdopodobieństw stacjonarnych modelu PBN dla zadanych podzbiorów stanów, poprzez generowanie odpowiedniej próby na podstawie wielu trajektorii PBN symulowanych równolegle. Pomysł wywodzi się z próbkowania Monte Carlo łańcuchami Markowa (ang. *Markov Chain Monte Carlo*, MCMC). W tego rodzaju algorytmach zbieżność symulacji MCMC jest kluczowym zagadnieniem. Powszechnym podejściem do oceny zbieżności MCMC jest analizowanie występowania różnic pomiędzy wieloma łańcuchami. W szczególności metoda Gelmana i Rubina [9] jest

⁶Definicje wpływu i wrażliwości genów można znaleźć w pracy [29].

przykładem takiego podejścia. Pozwala ona monitorować zbieżność wielu niezależnych łańcuchów Markowa do rozkładu stacjonarnego. Stąd idea mojego podejścia jest następująca. Równolegle generowane są niezależne trajektorie modelu PBN. Po spełnieniu warunku zbieżności według metody Gelmana i Rubina fragmenty poszczególnych trajektorii uznane za wygenerowane zgodnie z rozkładem stacjonarnym są łączone w jeden ciąg. Na podstawie tego ciągu szacowane są interesujące nas prawdopodobieństwa w rozkładzie stacjonarnym za pomocą rozszerzonej metody dwustanowego łańcucha Markowa przedstawionej w pracy **P3** lub metody Skart [34]. Obie te metody służą do określenia rozmiaru próbki wymaganej do obliczenia prawdopodobieństw z określoną dokładnością.

Dokładniej rzecz ujmując, gdy tylko zostanie osiągnięta zbieżność, drugie połowy⁷ trajektorii łańcuchów są scalane w jeden ciąg na CPU, co nie wymaga synchronizacji. Scalony ciąg jest następnie analizowany za pomocą rozszerzonej metody dwustanowego łańcucha Markowa. Jeśli wymagane jest przedłużenie ciągu, wyjściowe trajektorie zostają równolegle przedłużone na GPU.

Aby uzyskać maksymalne przyspieszenie, modyfikuję rozszerzoną metodę dwustanowego łańcucha Markowa pomijając iteracyjne obliczanie czasu wypalania łańcucha. Ponieważ zbieżność symulowanych trajektorii oryginalnego łańcucha Markowa nie musi implikować zbieżności abstrahowanej trajektorii dwustanowego łańcucha Markowa do jego rozkładu stacjonarnego, wprowadzam dodatkową weryfikację: po spełnieniu kryterium zatrzymania obliczany jest czas wypalania łańcucha dla dwustanowego łańcucha Markowa (M). Założenie o zbieżności jest uznawane za prawdziwe, jeśli M nie jest większe niż czas wypalania łańcucha w metodzie Gelmana i Rubina (ψ). W przeciwnym razie dodatkowe $M - \psi$ stanów zostaje odrzuconych z początkowego fragmentu każdej trajektorii, a zmodyfikowana rozszerzona metoda dwustanowego łańcucha Markowa jest ponownie uruchamiana nałączonych skróconych trajektoriach.

W przypadku wykorzystania metody Skart proponuję podobną strategię zrównoleglenia w celu zmniejszenia kosztu czasowego. Skart to zautomatyzowana sekwencyjna procedura oparta na analizie średnich odcinkowych (ang. *batch means*) służąca do konstrukcji asymptotycznie poprawnego przedziału ufności dla estymowanej średniej rozkładu stacjonarnego. Metoda ta została zaprojektowana tak, aby zapewnić spełnienie określonych wymagań dotyczących precyzji estymacji. W szczególności, metoda Skart dzieli symulowaną trajektorię na niepokrywające się odcinki (określane również mianem partii) (ang. *batch*) i sprawdza, czy liczba ignorowanych początkowych stanów trajektorii, liczba odcinków oraz długość odcinków są wystarczająco duże, aby zagwarantować, że średnie odcinkowe są w przybliżeniu niezależnymi zmiennymi losowymi o jednakowym rozkładzie normalnym – warunek, które musi być spełniony, aby można było obliczyć przedziały ufności.

Moje podejście opiera się na obserwacji, że to założenie pozostaje prawdziwe, gdy odcinki są uzyskiwane z różnych łańcuchów, pod warunkiem że te łańcuchy zbiegły do rozkładu stacjonarnego. Dlatego metodę Skart można zrównoleglić poprzez pobieranie odcinków z wielu niezależnych łańcuchów, które osiągnęły zbieżność. Ponieważ metoda Gelmana i Rubina zapewniła zbieżność trajektorii, podczas obliczania przedziałów ufności w zrównoleglonej wersji metody Skart już żadne dodatkowe fragmenty trajektorii nie są odrzucane. Ze względów efektywnościowych dodatkowo wymagamy, aby liczba odcinków uzyskanych z każdego łańcucha była taka sama, co może skutkować nieco większą liczbą odcinków użytych w zrównoleglonej metodzie Skart niż wymagana w wersji oryginalnej.

Wyniki eksperymentów przedstawione w pracy **P4** pokazują, że zaproponowane metody zrównoleglenia mogą znacząco zmniejszyć koszty czasowe obliczania prawdopo-

⁷Dla trajektorii długości N przez jej pierwszą połowę rozumiany jest ciąg elementów indeksowanych od 1 do $\lceil N/2 \rceil$, zaś przez drugą połowę ciąg elementów indeksowanych od $\lceil N/2 \rceil + 1$ do N .

bieństw stacjonarnych dużych modeli PBN na wielordzeniowych CPU. Z kolei wyniki uzyskane w pracy **P2** wykazują jeszcze większe przyspieszenia uzyskiwane na procesorach GPU poprzez połączenie zrównoleglonej przybliżonej analizy modelu PBN w warunkach stacjonarności z metodą zrównoleglenia symulacji na poziomie trajektorii omówionej w rozdziale 4.3.2.1. Szczegółowe omówienie wspomnianych wyników zaprezentowano w pracach **P4** i **P2**.

Ponadto podczas analizy systemu biologicznego może zaistnieć potrzeba wyznaczenia prawdopodobieństw stacjonarnych modelu PBN dla więcej niż jednego podzbioru stanów, np. w przypadku analizy stacjonarnej wrażliwości genów wspomnianej powyżej. Aby oszacować prawdopodobieństwa dla q różnych podzbiorów stanów, podejście dwustanowego łańcucha Markowa musiałaby zostać uruchomione q razy. Ponieważ proces symulacji jest najbardziej kosztowną częścią procedury pod względem czasu obliczeniowego, w pracy **P4** proponuję znacząco zmniejszyć ten koszt poprzez wielokrotne wykorzystanie równolegle wygenerowanych trajektorii. Kluczowym pomysłem jest jednoczesne abstrahowanie poszczególnych stanów modelu PBN w symulowanych trajektoriach do odpowiednich meta-stanów dwustanowych łańcuchów Markowa odpowiednich dla q poszukiwanych prawdopodobieństw za każdym razem, gdy następuje przedłużenie trajektorii. Ponadto obliczenia wymaganych rozmiarów prób, tj. długości trajektorii dwustanowych łańcuchów Markowa, dla estymacji poszczególnych prawdopodobieństw z zadaną dokładnością są realizowane równocześnie, co skutkuje kolejnym poziomem zrównoleglenia. Długość o jaką należy przedłużyć trajektorie jest określana jako minimalna wartość spośród wszystkich obliczonych długości rozszerzeń dla poszczególnych prawdopodobieństw. Oczywiście użycie wartości minimalnej może prowadzić do zwiększonej liczby rozszerzeń. Niemniej jednak jest to i tak korzystne, ponieważ ze względu na iteracyjny charakter metody dwustanowego łańcucha pozwala często uniknąć przeprowadzania zbędnych operacji abstrahowania trajektorii, które są stosunkowo kosztowne obliczeniowo. Proces przedłużania trajektorii jest zatrzymywany, gdy wszystkie wymagane rozmiary prób są mniejsze niż aktualny rozmiar próby i obliczane są poszukiwane prawdopodobieństwa. Jak pokazano w pracy **P4**, porównanie wydajności zrównoleglonej rozszerzonej metody dwustanowego łańcucha Markowa do wyznaczania siedmiu prawdopodobieństw dla 96-węzłowego modelu PBN apoptozy za pomocą 40-rdzeniowego procesora CPU dało niemal 53-krotne przyspieszenie w porównaniu do sekwencyjnego obliczenia zrealizowanego za pomocą sekwencyjnego użycia rozszerzonej metody dwustanowego łańcucha Markowa.

4.3.3 Konstruowanie modeli PBN na podstawie danych

W pracy **P5** przedstawiam *optPBN* – narzędzie programowe napisane w język programowania pakietu MATLAB przeznaczone do konstrukcji modeli PBN poprzez optymalizację ich parametrów. Stanowi ono rozszerzenie środowiska BN/PBN Toolbox stworzonego przez Lähdesmäkiego i Shmulevicha w 2009 roku do symulacji, wizualizacji i analizy modeli BN oraz PBN [22]. *optPBN* rozszerza oryginalny BN/PBN Toolbox w następujący sposób. Po pierwsze, umożliwia uwzględnienie eksperckiej wiedzy biologicznej, np. znanych interakcji pomiędzy genami. Wiedza ta jest dostarczana na wejściu w postaci reguł boolowskich. Po drugie, wspomaga identyfikację odpowiednich funkcji boolowskich w modelu BN/PBN oraz określanie prawdopodobieństw wyboru funkcji w modelu PBN na podstawie danych eksperymentalnych i dostarczonej wiedzy eksperckiej. Po trzecie, implementuje analizę statystyczną zoptymalizowanych parametrów. Wreszcie, pozwala na przyspieszenie obliczeń poprzez ich wykonywanie na systemach gridowych. Implementacja została zrealizowana jako narzędzie programowe **NP2**.

optPBN ułatwia konstrukcję modelu BN lub PBN, uwzględniając wiedzę ekspercką

w postaci wstępnej struktury modelu, która zazwyczaj pochodzi z literatury. Wiedza ta może być dostarczona poprzez określenie zbioru potencjalnych interakcji w postaci reguł boolowskich dla każdego z węzłów sieci. W przypadku węzła z kilkoma funkcjami boolowskimi (zwanymi również predyktorami), odpowiadające im prawdopodobieństwa wyboru mogą być albo przypisane, albo wyznaczone w kolejnym kroku poprzez optymalizację względem znormalizowanych danych eksperymentalnych.

Problem optymalizacyjny jest tworzony na podstawie integracji wstępnej struktury modelu BN/PBN z dostępnymi danymi eksperymentalnymi. Mogą one obejmować pomiary uzyskane w różnych warunkach eksperymentalnych, takich jak różne stymulacje ligandami, mutacje czy traktowanie inhibitorami. Problem optymalizacyjny jest następnie rozwiązywany za pomocą jednego z dwóch dostępnych optymalizatorów: 1) algorytmu optymalizacji typu *particle swarm* lub 2) algorytmu ewolucji różnicowej (ang. *differential evolution algorithm*) w ramach metaheurystycznego środowiska optymalizacyjnego ParadisEO [5]. Drugi z optymalizatorów jest przeznaczony do wykorzystania na architekturze gridowej.

Narzędzie optPBN oferuje dwa tryby optymalizacji: *tryb dyskretny* oraz *tryb ciągły*. W pierwszym przypadku dla modelu PBN wybierany jest podzbiór składowych sieci BN (ang. *constituent Boolean networks*) z puli rozważanych możliwych sieci i składowym tym przypisywane są równe prawdopodobieństwa wyboru. W drugim przypadku prawdopodobieństwa wyboru dla składowych BN mogą przyjmować dowolne wartości z zakresu od 0 do 1 sumujące się do jedynki. Podczas procesu optymalizacji w trybie ciągłym brzegowy rozkład stacjonarny dla węzłów wyjściowych modelu jest estymowany za pomocą rozszerzonej metody dwustanowego łańcucha Markowa, która została omówiona w rozdziale 4.3.2.2, i porównywany z wynikami pomiarów eksperymentalnych.

optPBN umożliwia analizę skonstruowanych modeli za pomocą rozszerzonej metody dwustanowego łańcucha Markowa. W pracy **P5** pokazuję, że modele boolowskie uzyskane przy użyciu optPBN pozwalają dokonywać jakościowych i ilościowych interpretacji biologicznych, które są w zgodzie z obserwowanymi zjawiskami biologicznymi odzwierciedlonymi w danych eksperymentalnych. W szczególności dwa modele PBN apoptozy w hepatocytach wyznaczone za pomocą optPBN, określane mianem *modelu wstępnego apoptozy* (ang. *initial apoptosis model*) oraz *modelu rozszerzonego apoptozy* (ang. *extended apoptosis model*) w pracy **P5**, umożliwiają poprawne ilościowe określenie wpływu promieniowania ultrafioletowego UVB na poziom aktywacji kompleksu białkowego NF κ B, kaspazy 3 oraz apoptozy w pierwotnych hepatocytach. Mianowicie, przeprowadzona ilościowa analiza modeli poprawnie przewiduje, że silniejsze promieniowanie UVB, tj. 600 lm/m² zamiast 300 lm/m², prowadzi do silniejszej aktywacji szlaku NF κ B przy zmniejszonej aktywacji kaspazy 3 oraz do osłabionej aktywności apoptotycznej.

4.3.4 Wyznaczanie atraktorów w synchronicznych sieciach boolowskich

Jak wspomniano w rozdziale 4.3.1, identyfikacja wszystkich atraktorów dużej sieci boolowskiej jest problemem trudnym obliczeniowo – udowodniono, że w ogólności jest to problem NP-trudny. Niemniej jednak, zadanie to było szeroko badane w literaturze i zaproponowano kilka rozwiązań skutecznych w praktyce. Wśród nich można wyodrębnić ważną klasę algorytmów opartych na wykorzystaniu Zredukowanych Uporządkowanych Binarnych Diagramów Decyzyjnych (ang. *Reduced Ordered Binary Decision Diagram*, ROBDD lub po prostu BDD). BDD to struktura danych, która w kontekście sieci boolowskich (zarówno BN jak i PBN) umożliwia kompresję reprezentacji funkcji boolowskich w postaci skierowanego acyklicznego grafu (ang. *directed acyclic graph*, DAG), a na bardziej abstrakcyjnym poziomie na skompresowaną reprezentację boolowskich relacji przejść pomiędzy

stanami sieci jak i zbiorów stanów sieci (P)BN. Co więcej, w przeciwieństwie do innych skompresowanych reprezentacji, BDD pozwalają na wykonywanie operacji bezpośrednio na skompresowanej reprezentacji obiektów, tj. bez uprzedniej dekompresji.

W pracy **P6** proponuję nowe podejście mające na celu poprawę wydajności wyznaczania atraktorów w synchronicznych sieciach BN w oparciu o BDD, poprzez ulepszenie algorytmu pierwotnie przedstawionego w [8]. Moje podejście opiera się na kilku rozwiązaniach. Po pierwsze, zaproponowany zostaje *algorytm monolityczny*, który jest prostą iteracją punktu stałego relacji przejścia między stanami sieci BN zakodowanej za pomocą BDD. Algorytm monolityczny rozpoczyna działanie od całego zbioru stanów sieci BN, którego rozmiar jest wykładniczy względem liczby węzłów sieci BN. W związku z tym jego zastosowanie jest ograniczone do małych sieci.

Po drugie, aby poradzić sobie z dużymi sieciami, proponuję *strategię wyszukiwania wyliczeniowego* (ang. *enumerative search strategy*) do wykrywania wszystkich atraktorów osiągalnych ze stanów z zadanego zbioru Z . Algorytm rozpoczyna działanie od losowo wybranego stanu $s \in Z$ i wyznacza atraktor A osiągalny ze stanu s ⁸. Po usunięciu ze zbioru Z wszystkich stanów, które są wstecznie osiągalne ze stanów atraktora A , tzn. osiągalne po odwróceniu kierunku tranzycji w grafie stanów, co jest obliczane za pomocą operacji *pre-image* na BDD, algorytm kontynuuje wyszukiwanie innych atraktorów osiągalnych z pozostałych stanów w Z . W porównaniu do podejścia z [8], algorytm zaproponowany przeze mnie w pracy **P6** lepiej wykorzystuje strukturalne regularności występujące w BDD i przyspiesza obliczenia poprzez iteracyjne wyznaczanie bezpośrednich poprzedników każdego z rozpatrywanych osobno stanów atraktora. Takie podejście pozwala na unikanie powtarzania zbędnych obliczeń za pomocą operacji *pre-image*, jak to jest uzasadnione w pracy **P6**.

Po trzecie, proponuję algorytm wykrywania atraktorów polegający na dekompozycji celem zredukowania rozmiaru przestrzeni stanów do przeszukania. Przed rozpoczęciem wykrywania atraktorów sieci BN, algorytm analizuje strukturę sieci i próbuje podzielić BN na podsieci zwane *blokami*. Wprowadzam dwie metody tworzenia bloków: 1) dekompozycję opartą na liściach oraz 2) dekompozycję opartą na silnie spójnych składowych. Następnie wykrywanie atraktorów jest przeprowadzane dla każdego bloku przy użyciu algorytmu monolitycznego lub strategii wyszukiwania wyliczeniowego. Wykryte *atraktory blokowe* są wykorzystywane do ograniczenia przestrzeni stanów oryginalnej sieci BN poprzez eliminację stanów, które są nieistotne dla wyszukiwania atraktorów. Ograniczona przestrzeń stanów jest tworzona poprzez wprowadzenie pojęć *stanów skompresowanych* i *stanów lustrzanych*, identyfikację zbiorów *stanów lustrzanych* dla stanów atraktorów blokowych we wszystkich blokach oraz obliczenie części wspólnej tych zbiorów. Na koniec atraktory całej sieci BN są wykrywane poprzez uruchomienie algorytmu monolitycznego lub strategii wyszukiwania wyliczeniowego na ograniczonej przestrzeni stanów.

Proponuję dwie heurystyki dotyczące ustalania efektywnej kolejności zmiennych w BDD. Wiadomo, że kolejność zmiennych wejściowych w BDD może znacząco wpływać na rozmiar skonstruowanego BDD oraz koszt obliczeniowy operacji na BDD. Znalezienie optymalnej kolejności jest problemem NP-zupełnym [3]. Jednakże w literaturze zaproponowano wiele efektywnych heurystyk umożliwiających ustalenie quasi-optymalnej kolejności zmiennych dla BDD.

W pracy **P6** proponuję dwie statyczne heurystyki porządkowania zmiennych w BDD, oparte na strukturze BN oraz na założeniu, że zmienne, które są topologicznie bliskie w oryginalnej sieci, powinny być również uporządkowane względnie blisko siebie w BDD.

⁸Zauważmy, że w przypadku rozważanych tutaj synchronicznych sieci boolowskich, dla każdego stanu takiej sieci istnieje dokładnie jeden atraktor deterministycznie osiągalny z tego stanu.

Heurystyka (I) umieszcza na początku węzeł z najmniejszą liczbą węzłów potomnych, natomiast heurystyka (II) przeciwnie – umieszcza na początku węzeł z największą liczbą węzłów potomnych.

Ewaluacja nowego podejścia do wyznaczania atraktorów uwzględniającego wszystkie wyżej przedstawione udoskonalenia, została przeprowadzona na kilku przykładach oraz na realistycznym modelu apoptozy w hepatocytach. Uzyskane wyniki wskazują, że zaproponowane strategie mogą znacząco poprawić wydajność oryginalnego algorytmu wykrywania atraktorów opartego za pomocą BDD zarówno dla małych, jak i dużych modeli BN.

W pracy **P7** proponuję nową metodę dekompozycji opartą o silnie spójne składowe (SSS) do wykrywania atraktorów w dużych synchronicznych modelach BN. Metoda ta składa się z trzech kroków.

Po pierwsze, model BN jest dzielona na podsieci, zwane *blokami*, na podstawie grafu struktury poprzez dekompozycję na SSS. Po drugie, wykrywane są atraktory każdego bloku. W końcowym kroku atraktory oryginalnego modelu BN są obliczane na podstawie atraktorów poszczególnych bloków. Jednakże atraktory pełnego, wyjściowego modelu BN są uzyskiwane w odmienny sposób niż w pracy **P6**, gdzie atraktory bloków służyły jedynie do ograniczenia przestrzeni stanów wyszukiwania dla algorytmu monolitycznego lub strategii wyszukiwania wyliczeniowego. Tutaj atraktory modelu BN są konstruowane bezpośrednio na podstawie atraktorów bloków.

W tym celu wprowadzane są w pracy **P7** pojęcia *elementarnych* i *nielelementarnych* bloków oraz *realizacji bloku nielelementarnego*, proponowana jest koncepcja *krzyżowalności stanów*, a także definiowane są operacje *scalania bloków*, *krzyżowania zbiorów stanów* oraz *krzyżowania rodzin zbiorów stanów*. Następnie wykazano, że jeśli bloki są rozpatrywane i scalane w określonej kolejności, atraktory wyjściowego modelu BN mogą być iteracyjnie rekonstruowane poprzez zwykłe wykonywanie operacji krzyżowania atraktorów bloków.

Powyższy algorytm oparty na dekompozycji wykorzystujący reprezentację BDD oraz operacje na BDD został zaimplementowany jako rozszerzenie narzędzia do weryfikacji formalnej (ang. *model checker*) MCMAS [25]. Następnie rozszerzona implementacja MCMAS została włączona do narzędzia programowego **NP1**. Skuteczność nowej metody wykazuję poprzez porównanie jej z innymi metodami dekompozycji opartymi na BDD na modelach BN o różnej liczbie węzłów, losowo wygenerowanych za pomocą narzędzia programowego **NP1**.

Matematyczna poprawność opisanych algorytmów została w pełni wykazana w pracy **P8**, gdzie przedstawione są wszystkie dowody twierdzeń i lematów z pracy **P7**. Ponadto, w pracy **P8** przeprowadzam szeroką ewaluację mojej metody na dziewięciu opublikowanych modelach biologicznych rzeczywistych procesów komórkowych, takich jak rozwój przerzutów, kaskady sygnałowej kinazy białkowej aktywowanej mitogenem (MAPK), przeżycie cytotoksycznych limfocytów T w białaczce z dużych ziarnistych limfocytów typu T (T-LGL), migracja keratynocytów indukowana czynnikiem wzrostu hepatocytów (HGF), proces sterowania różnicowaniem limfocytów Th oraz apoptoza w hepatocytach. Liczba węzłów w tych modelach wynosi od 32 do 97.

Wyniki ewaluacji na tych biologicznych sieciach są zgodne z wynikami uzyskanymi dla losowo wygenerowanych modeli w pracy **P7**, tj. wskazują na znaczące przyspieszenie obliczeń. Co więcej, wykorzystując model apoptozy z różnymi ustawieniami węzłów wejściowych, dodatkowo wykazuję, że zaproponowana metoda pozwala na znaczne przyspieszenie analizy modeli z niewielką liczbą atraktorów. Ta cecha sprawia, że moja metoda jest szczególnie przydatna do analizy modeli BN rzeczywistych sieci regulatorowych genów, ponieważ w takich modelach BN liczba atraktorów powinna być stosunkowo niewielka, aby modele te były zgodne z biologiczną rzeczywistością, a tym samym miały biologiczny sens.

4.3.5 Wyznaczanie atraktorów w asynchronicznych sieciach boolowskich

Metoda wykrywania atraktorów oparta na dekompozycji, przedstawiona w pracach **P7** i **P8**, została rozszerzona na duże asynchroniczne modele BN w pracy **P9**. W tym celu przeddefiniuję pojęcie *realizacji bloku nieelementarnego*, aby odpowiadało trybowi asynchronicznej aktualizacji. Realizacja bloku nieelementarnego jest uzyskiwana poprzez jawne skonstruowanie systemu tranzycji (ang. *state transition system*) tego bloku określonego przez rozważany atraktor innego, elementarnego bloku będącego rodzicem (ang. *elementary parent block*) danego bloku nieelementarnego. Ponieważ blok rodzic bloku nieelementarnego może mieć więcej niż jeden atraktor, dany blok może mieć więcej niż jedną realizację. Konstruowanie realizacji dla bloku nieelementarnego jest kluczowym krokiem w procesie wyznaczenia jego atraktorów. Atraktory bloku nieelementarnego stanowią zbiór atraktorów wszystkich jego realizacji.

W pracy **P9** udowadniam, że bloki elementarne *zachowują atraktory* oryginalnego asynchronicznego modelu BN. Ponadto pokazuję, że atraktory bloku otrzymanego poprzez scalanie dwóch bloków elementarnych lub bloku nieelementarnego z jego elementarnym, pojedynczym blokiem rodzicem można uzyskać poprzez wykonanie operacji krzyżowania rodzin atraktorów dwóch składowych bloków.

Opracowane podstawy teoretyczne pozwalają mi zaprojektować nowe podejście oparte na dekompozycji do wykrywania atraktorów w dużych asynchronicznych modelach BN, którego główne kroki są następujące. Po pierwsze, graf struktury modelu BN jest dzielony na bloki zgodnie z silnie spójnymi składowymi. Po drugie, bloki są sortowane topologicznie, a ich atraktory są wykrywane iteracyjnie, jeden po drugim. Następnie wykonywana jest operacja krzyżowania na dowolnych dwóch blokach elementarnych lub na bloku elementarnym i jednym z takich jego bloków dzieci, który nie ma innych bloków rodziców w celu uzyskania atraktorów połączonych bloków. Powstały scalony blok staje się nowym blokiem elementarnym. Powtarzając iteracyjnie operację krzyżowania do momentu uzyskania pojedynczego elementarnego bloku zawierającego wszystkie węzły modelu BN, algorytm konstruuje wszystkie atraktory oryginalnego modelu BN.

Podobnie jak w przypadku synchronicznym, algorytm dekompozycyjny dla dużych asynchronicznych modeli BN, wykorzystujący reprezentację BDD i operacje na niej, jest zaimplementowany jako rozszerzenie narzędzia do weryfikacji formalnej MCMAS [25], która to implementacja została włączona do narzędzia programowego **NP1**.

W pracy **P9** oceniam efektywność opracowanego algorytmu wykrywania atraktorów opartego na dekompozycji dla asynchronicznych modeli BN na dwóch rzeczywistych systemach biologicznych, tj. logicznym modelu sieci MAPK zawierającym 53 węzły oraz modelu BN apoptozy składającym się z 97 węzłów. Eksperymenty wykazały, że nowa metoda działa szczególnie dobrze w przypadku dużych sieci, które zawierają kilka stosunkowo małych SCC, co prowadzi do znacznych przyspieszeń obliczeń w porównaniu z innymi metodami.

4.3.6 Sterowanie sieciami boolowskimi

Opracowane kluczowe idee i rozwiązania związane z wykrywaniem atraktorów w dużych modelach BN opartym na dekompozycji, które zostały przedstawione w rozdziałach 4.3.4 i 4.3.5, były rozwijane przy opracowywaniu algorytmów dla problemu sterowania w dużych asynchronicznych sieciach boolowskich w pracach **P10** i **P11**. W szczególności sformułowalem następujący problem sterowania.

Zdefiniowałem sterowanie C jako podzbiór $\{1, 2, \dots, n\}$, gdzie n to liczba węzłów w sieci boolowskiej. Dla stanu $\mathbf{s} \in \{0, 1\}^n$ zaaplikowanie sterowania C do $\mathbf{s}$, oznaczone jako $C(\mathbf{s})$,

jest definiowane jako stan $\mathbf{s}'$, taki że $\mathbf{s}'[i] = 1 - \mathbf{s}[i]$, jeśli $i \in C$, oraz $\mathbf{s}'[i] = \mathbf{s}[i]$ w przeciwnym razie. Innymi słowy, wartość i -tego elementu $\mathbf{s}$ jest zmieniana dla wszystkich $i \in C$, a dla pozostałych elementów pozostaje niezmienniona.

Następnie wprowadziłem pojęcia *słabego basenu* (ang. *weak basin*) oraz *silnego basenu* (ang. *strong basin*) atraktora. Słaby basen atraktora to zbiór wszystkich stanów, z których istnieje ścieżka prowadząca do danego atraktora w grafie systemu tranzycji. Jednak ze względu na niedeterministyczny charakter asynchronicznego trybu aktualizacji, niektóre z tych stanów mogą również posiadać ścieżki prowadzące do innych atraktorów.

W odróżnieniu silny basen jest zdefiniowany w taki sposób, że wszystkie ścieżki z jego stanów prowadzą wyłącznie do danego atraktora. W związku z tym silne baseny są zawsze podzbiorami słabych basenów. Co istotne, silny basen atraktora gwarantuje jego ostateczną osiągalność.

Na koniec zdefiniowałem *minimalne jednoczesne sterowanie docelowe* (ang. *minimal simultaneous target control*) dla danego modelu BN, stanu wyjściowego $\mathbf{s}$ oraz atraktora docelowego A_t jako minimalne pod względem liczności sterowanie C , takie że $C(\mathbf{s})$ znajduje się w silnym basenie atraktora A_t .

Podobnie jak w przypadku algorytmów opartych na dekompozycji do wykrywania atraktorów, które zostały przedstawione w rozdziałach 4.3.4 i 4.3.5, rozważam systemy tranzycji poszczególnych bloków. Jednakże rozwiązanie stosowane do wykrywania atraktorów oparte na dekompozycji, gdzie zbiór stanów i przejść dla węzłów kontrolnych bloku nieelementarnego jest wyprowadzany poprzez rzutowanie przejść w atraktorze bloku-rodzica na te węzły kontrolne, nie może zostać w prosty sposób rozszerzone w celu rozwiązania problemu minimalnego jednoczesnego sterowania docelowego. Aby rozwiązać ten problem, należy uwzględnić kompletne zachowanie silnego basenu atraktora w złączonych blokach-rodzicach w celu wygenerowania systemu tranzycji dla bloku nieelementarnego. Dlatego wprowadziłem pojęcie *domknięcia przodkowego* (ang. *ancestor-closure*) bloku elementarnego lub nieelementarnego jako złączenie tego bloku oraz wszystkich jego bloków-przodków określonych przez domknięcie przechodnie relacji bycia rodzicem. Ponadto zdefiniowałem *system tranzycji bloku nieelementarnego*, który jest generowany przez silny basen atraktora docelowego rzutowanego na złączenie domknięć przodkowych poszczególnych bloków-rodziców danego bloku nieelementarnego.

Algorytm sterowania oparty na dekompozycji opiera się na dwóch głównych twierdzeniach, które umożliwiają obliczenie silnego basenu atraktora docelowego na podstawie dekompozycji modelu BN na mniejsze bloki.

Pierwsze twierdzenie, tj. *twierdzenie o zachowaniu atraktorów* (ang. *preservation of attractors theorem*) stwierdza, że rzutowanie (projekcja) atraktora docelowego na blok nieelementarny jest atraktorem w realizacji tego bloku generowanej przez basen atraktora docelowego w systemie tranzycji bloku uzyskanego przez złączenie domknięć przodkowych poszczególnych bloków-rodziców.

Drugie twierdzenie, tj. *twierdzenie o zachowaniu basenów* (ang. *preservation of basins theorem*) stwierdza, że celem wyznaczenia (silnego) basenu atraktora docelowego, wystarczy obliczyć lokalne baseny projekcji atraktora docelowego na każdy z bloków, gdzie zgodnie z twierdzeniem o zachowaniu atraktorów projekcje te są atraktorami poszczególnych bloków, a następnie połączyć te lokalne baseny za pomocą operacji krzyżowania.

Opracowanie przedstawionych powyżej podstaw teoretycznych pozwoliło mi zaproponować efektywną metodę rozwiązującą problem minimalnego jednoczesnego sterowania docelowego. Jej istotą jest oparty na dekompozycji algorytm obliczania silnego basenu docelowego atraktora. Główne kroki tej metody są następujące. Najpierw dany model BN jest dekomponowany na bloki, które są sortowane topologicznie ze względu na relację bycia

rodzicem. Następnie atraktor docelowy jest rozkładany na swoje projekcje na poszczególne bloki. Bloki są następnie rozważane zgodnie z porządkiem topologicznym i dla każdego z nich konstruowany jest system tranzycji: bezpośrednio w przypadku bloku elementarnego lub w przypadku bloku nieelementarnego jako system tranzycji realizowany przez basen skrzyżowanych projekcji docelowego atraktora na wszystkie wcześniejsze w sensie uporządkowania topologicznego bloki bloku nieelementarnego. Operacja krzyżowania projekcji atraktora docelowego prowadzi do uzyskania atraktora dla połączonych bloków wcześniejszych, zgodnie z twierdzeniem o zachowaniu atraktorów. Silny basen w aktualnie rozważanym bloku jest obliczany w systemie tranzycji tego bloku dla atraktora docelowego zrzuconego na połączone wszystkie wcześniejsze bloki oraz bieżący blok. Ostatecznie silny basen atraktora docelowego w połączonych blokach jest uzyskiwana poprzez zastosowanie operacji krzyżowania na silnych basenach wcześniejszych połączonych bloków oraz bieżącego bloku. Zgodnie z twierdzeniem o zachowaniu basenów wynikiem jest silny basen atraktora docelowego w połączeniu bloku bieżącego z wcześniejszymi. Po iteracyjnym rozważeniu wszystkich bloków uzyskiwana jest silny basen atraktora docelowego dla całego modelu BN.

Dysponując algorytmem obliczania silnego basenu, zaproponowałem algorytm rozwiązujący problem minimalnego jednoczesnego sterowania docelowego. Najpierw obliczany jest silny basen docelowego atraktora. Następnie wyznaczany jest zbiór indeksów stanu lub stanów należących do tego basenu, które są najbliższe stanowi źródłowemu w sensie odległości Hamminga. Algorytm sterowania oparty na dekompozycji został zaimplementowany w narzędziu programowym **NP1**, które wykorzystuje narzędzie do weryfikacji formalnej MCMAS [25] do efektywnego reprezentowania sieci BN w postaci BDD.

W pracy **P10** nowe podejście zostało sprawdzone na dwóch rzeczywistych sieciach biologicznych. Po pierwsze, na kompleksowym modelu BN, opisującym mechanizmy decyzyjne w komórkach PC12 determinujące procesy proliferacji i różnicowania. Model ten łączy w sekwencję zdarzeń sygnalizację białkową, aktywność czynników transkrypcyjnych oraz następujące po nich sprzężenie zwrotne w postaci sygnalizacji autokrynnej. Model BN składa się z 32 węzłów [26]. Po drugie, na modelu BN apoptozy, który uwzględnia szlaki regulacyjne przeżywalności komórki, szlaki metaboliczne insuliny, szlaki wewnętrzny i zewnętrzny (receptorowy) apoptozy oraz ich wzajemne oddziaływania. Obejmuje on 97 węzłów [28].

Ponadto podejście zostało przetestowane na trzech losowo wygenerowanych modelach BN zawierających odpowiednio 100, 120 i 180 węzłów. Uzyskane wyniki pokazują, że dzięki podejściu typu „dziel i rządź”, moja metoda oparta na dekompozycji ma wyraźną przewagę pod względem wydajności, szczególnie w przypadku dużych sieci, w porównaniu do podejścia globalnego, w którym nie uwzględnia się dekompozycji.

W pracy **P11** algorytm podejścia opartego na dekompozycji rozwiązujący problem minimalnego jednoczesnego sterowania docelowego został dodatkowo zoptymalizowany poprzez uwzględnienie informacji o stanie źródłowym. Główna idea optymalizacji polega na sprawdzeniu dla każdego rozważanego bloku, czy projekcje stanu źródłowego i docelowego atraktora na ten blok są identyczne. Jeśli tak jest, to nie ma potrzeby obliczania całego lokalnego silnego basenu dla tego bloku, ponieważ węzły stanu źródłowego należące do tego bloku nie wymagają perturbacji.

Efektywność pod względem wydajności i skalowalności podejścia w pracy **P10** oraz zoptymalizowanej wersji została przetestowana na sześciu rzeczywistych modelach biologicznych o liczbie węzłów w zakresie od 32 do 68, oraz na trzech losowo wygenerowanych modelach BN zawierających odpowiednio 100, 120 i 180 węzłów.

Jednakże zastosowanie dokładnych metod opartych na obliczeniach symbolicznych wy-

korzystujących BDD jest ograniczone do sieci, które dają się zdekomponować na bloki i gdzie dynamika poszczególnych bloków może być reprezentowana i analizowana za pomocą BDD. Zazwyczaj wymaga to, aby rozmiary bloków nie przekraczały kilkudziesięciu węzłów.

Aby osiągnąć skalowalność, zaproponowałem w pracy **P12** zastosowanie głębokiego uczenia przez wzmacnianie (ang. *deep reinforcement learning*, DRL) do sterowania modelami sieci boolowskich – zarówno klasycznymi, jak i probabilistycznymi – opisującymi złożone systemy biologiczne.

W tym celu sformułowalem nowy problem sterowania dla modeli BN i PBN sieci regulatorowych genów w trybie asynchronicznym, tzw. Source-Target Attractor Control. Problem ten został dostosowany do potrzeb reprogramowania komórkowego poprzez ograniczenie możliwych interwencji wyłącznie do stanów atraktorowych, co odzwierciedla ograniczenia biologii eksperymentalnej. W przeciwieństwie do wcześniejszych podejść, możliwe jest jednoczesna perturbacja wielu genów w ramach jednej akcji podejmowanej przez agenta DRL.

Identyfikacja atraktorów w dużych sieciach BN/PBN stanowi wyzwanie sama w sobie. W celu osiągnięcia skalowalności opracowanego podejścia wprowadziłem pojęcie pseudo-atraktora oraz procedurę identyfikacji stanów pseudo-atraktorowych podczas treningu agenta DRL (ang. *pseudo-attractor states identification procedure*, PASIP). Dodatkowo, jak zostało to uzasadnione w pracy **P12**, wprowadzone pojęcie pseudo-atraktora i metoda PASIP pozwalają radzić sobie z artefaktami modelowania pojawiającymi się w asynchronicznych sieciach BN/PBN, które objawiają się występowaniem dużych, złożonych atraktorów pozbawionych biologicznego znaczenia. Niemniej pewne często odwiedzane podzbiory stanów w takich atraktorach mogą odpowiadać obserwowalnym stanom fenotypowym. PASIP umożliwia identyfikację właśnie takich istotnych stanów w obrębie tych „sztucznych” atraktorów.

Na koniec opracowałem pbn-STAC – obliczeniowe narzędzie do rozwiązywania zaproponowanego problemu sterowania. Mając dane (pseudo-)atraktorowe stany źródłowy i docelowy, pbn-STAC znajduje odpowiednie strategie sterowania, które przeprowadzają dynamikę sieci od stanu źródłowego do docelowego, podejmując akcje sterujące wyłącznie w pośrednich stanach atraktorowych odpowiadającym obserwowalnym stanom fenotypowym komórek. Ocenilem skuteczność PASIP i pbn-STAC na szeregu sieci biologicznych o różnych rozmiarach, a uzyskane wyniki potwierdziły efektywność zaproponowanego podejścia. Następnie porównałem wyniki pbn-STAC z dokładnymi, optymalnymi rozwiązaniami wszędzie tam, gdzie jest to możliwe, pokazując, że pbn-STAC jest w stanie identyfikować optymalne lub suboptymalne strategie sterowania.

pbn-STAC stanowi krok naprzód w kierunku opracowania skalowalnych metod sterowania dużymi sieciami biologicznymi.

4.3.7 Wykraczając poza formalizm sieci Boolowskich – złożona długofalowa dynamika systemów reakcji

Modelowanie komputerowe umożliwiające analizę systemu i dokonywanie predykcji opartych na formalnym rozumowaniu jest nieodzowne dla pełnego zrozumienia mechanizmów biologicznych, a w szczególności dla uchwycenia, w jaki sposób interakcje między poszczególnymi elementami komórki prowadzą do powszechnego występowania różnych typów zachowań emergentnych. W pracy **P13** rozważam następujące przykłady takich zachowań:

- *bistabilność* (ang. *bistability*), gdzie system przełącza się między dwoma alternatyw-

nymi stabilnymi stanami;

- *oscylacje cyklu granicznego* (ang. *limit cycle oscillations*), które charakteryzują okresową dynamikę biologiczną i ekologiczną;
- *bifurkacja podwojenia okresu* (ang. *period-doubling bifurcation*), będąca powszechnym mechanizmem deterministycznego chaosu w organizmach żywych, gdzie niewielka zmiana parametrów systemu powoduje, że początkowo stabilny cykl o długości T staje się niestabilny i dochodzi do powstania nowego stabilnego cyklu o długości $2T$.

Analiza tych zjawisk w przypadku ogólnie pojętych układów reakcji biochemicznych jest zazwyczaj przeprowadzana za pomocą modeli ilościowych, w szczególności modeli opartych na równaniach różniczkowych zwyczajnych. Jednakże w tego rodzaju modelach cechy emergentne systemu są wynikiem ilościowej interakcji między konkretnymi wartościami liczbowymi stałych kinetycznych oraz początkowych wartości zmiennych (stężeń). Jak argumentuję w pracy **P13**, jest to w pewnym stopniu niezadowalające, gdyż stanowi konsekwencję konkretnych wartości numerycznych, które nie dostarczają wglądu w strukturę (naturę) badanego systemu. Aby rozwiązać ten problem, w pracy **P13** przyjmuję perspektywę jakościową badania różnych zachowań układu dynamicznego. Usuwałem parametry liczbowe z modelu poprzez zastąpienie ich elementami jakościowymi, które usuwane parametry liczbowe modelują, dążę do zidentyfikowania mechanizmów czysto strukturalnych, które mogą być odpowiedzialne za występowanie zjawisk, takich jak bistabilność, oscylacja cyklu granicznego czy bifurkacja podwojenia okresu.

W literaturze naukowej wielokrotnie wykorzystywano formalizm sieci Boolowskich do opisywania takich zachowań. Aby jeszcze bardziej przybliżyć się do czysto jakościowego opisu wspomnianych zjawisk, w pracy **P13** zajmuję się tym zagadnieniem na poziomie elementarnym wykorzystując formalizm *systemów reakcji* (ang. *reaction systems*) pierwotnie zaproponowany przez Ehrenfeuchta i Rozenberga ([7]), który to formalizm sprowadza opis systemu wyłącznie do elementarnych operacji teoriomnogościowych. Systemy reakcji oferują wgląd przyczynowy, ułatwiając lepsze zrozumienie relacji przyczynowo-skutkowych między reakcjami, a co za tym idzie, całego modelu. Osiąga się to poprzez skupienie się na samym procesie biologicznym i analizę badanych zjawisk wyłącznie poprzez mechanizmy umożliwiania (ang. *facilitation mechanism*) i inhibicji (ang. *inhibition mechanism*).

W pracy **P13** zdefiniowałem pojęcie *wielostabilnego systemu reakcji* i dowiodłem istnienia systemów reakcji posiadających dowolną liczbę stanów stacjonarnych. Następnie zaprezentowałem odpowiednik najmniejszego chemicznego systemu reakcji, który składa się z minimalnej liczby substratów i reakcji, a odpowiadający mu układ równań różniczkowych zwyczajnych posiada minimalną liczbę składników w równaniach. Przeprowadziłem analizę jakościową zachowania tego systemu i sprawdziłem, że rzeczywiście charakteryzuje się on bistabilnością. Ponadto wykazałem, że system ten jest minimalny pod względem liczby substratów niezbędnych do uzyskania wielostabilności.

Zdefiniowałem *monostabilny system reakcji* oraz *system reakcji z cyklem granicznym*. Konstrukcja systemu reakcji z cyklem granicznym opiera się na minimalnym układzie reakcji chemicznych z cyklem granicznym zaproponowanym w [38].

Następnie wprowadziłem pojęcie *okresowego systemu reakcji* oraz przedstawiłem system, w którym występuje zjawisko podwojenia okresu.

W pracy **P13** pokazałem, że złożone zachowania dynamiczne, takie jak dwustabilność, istnienie dowolnej liczby stanów stacjonarnych, oscylacje cyklu granicznego oraz bifurkacje podwojenia okresu, mogą być modelowane za pomocą formalizmu systemów reakcji. Zazwyczaj takie zjawiska są przedstawiane za pomocą modeli ilościowych z odpowiednio

dobranymi parametrami numerycznymi. Praca **P13** wprowadza naturalne odpowiedniki pojęć teorii układów dynamicznych dla formalizmu systemów reakcji i pokazuje, że te złożone zachowania dynamiczne mogą być odzwierciedlane przez proste modele systemów reakcji. Wyniki te są zgodne z wcześniejszymi badaniami, które wskazują, że takie zachowania mogą być modelowane za pomocą mechanizmów opartych na algebrze Boole’a.

Bibliografia

- [1] T. Akutsu, S. Kuhara, O. Maruyama i S. Miyano. Identification of genetic networks by strategic gene disruptions and gene overexpressions under a Boolean model. *Theoretical Computer Science*, 298(1):235–251, 2003.
- [2] A.-L. Barabási, N. Gulbahce i J. Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, 2011.
- [3] B. Bollig i I. Wegener. Improving the variable ordering of OBDDs is NP-complete. *IEEE Transactions on Computers*, 45(9):993–1002, 1996.
- [4] S. Bornholdt. Systems biology. Less is more in modeling large genetic networks. *Science*, 310:449–451, 2005.
- [5] S. Cahon, N. Melab i E.-G. Talbi. ParadisEO: A Framework for the Reusable Design of Parallel and Distributed Metaheuristics. *Journal of Heuristics*, 10(3):357–380, 2004.
- [6] M. I. Davidich i S. Bornholdt. Boolean Network Model Predicts Cell Cycle Sequence of Fission Yeast. *PloS One*, 3(2):Article e1672, 2008.
- [7] A. Ehrenfeucht i G. Rozenberg. Reaction systems. *Fundamenta Informaticae*, 75(1):263–280, 2007.
- [8] A. Garg, I. Xenarios, L. Mendoza i G. DeMicheli. An Efficient Method for Dynamic Analysis of Gene Regulatory Networks and *in silico* Gene Perturbation Experiments. W T. Speed i H. Huang, redaktorzy, *Proc. 11th Annual International Conference on Research in Computational Molecular Biology, RECOMB’07*, tom 4453 z *Lecture Notes in Bioinformatics*, strony 62–76, Berlin Heidelberg. Springer, 2007.
- [9] A. Gelman i D. B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472, 1992.
- [10] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren i R. Guthke. Gene regulatory network inference: Data integration in dynamic models – A review. *BioSystems*, 96(1):86–103, 2009.
- [11] S. Huang. Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery. *Journal of Molecular Medicine*, 77(6):469–480, 1999.
- [12] S. Huang. Regulation of Cellular States in Mammalian Cells from a Genomewide View. W J. Collado-Vides i R. Hofestädt, redaktorzy, *Gene Regulation and Metabolism: Post-Genomic Computational Approach*, strony 181–220. MIT Press, Cambridge, MA, 2002.
- [13] S. Huang. The molecular and mathematical basis of Waddington’s epigenetic landscape: A framework for post-Darwinian biology? *BioEssays*, 34(2):149–157, 2012.
- [14] S. Huang, G. Eichler, Y. Bar-Yam i D. E. Ingber. Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. *Physical Review Letters*, 94(12):Article 128701, 2005.

- [15] S. Huang, I. Ernberg i S. Kauffman. Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in Cell & Developmental Biology*, 20:869–876, 2009.
- [16] S. Huang i S. A. Kauffman. *Complex Gene Regulatory Networks – from Structure to Biological Observables: Cell Fate Determination*. W *Encyclopedia of Complexity and Systems Science*. R. A. Meyers, redaktor. Springer, New York, NY, 2009, strony 1180–1213.
- [17] J. P. Hughes, S. S. Rees, S. B. Kalindjian i K. L. Philpott. Principles of early drug discovery. *British Journal of Pharmacology*, 162(6):1239–1249, 2011.
- [18] A. W. Jones. Early drug discovery and the rise of pharmaceutical chemistry. *Drug Testing and Analysis*, 3(6):337–344, 2011.
- [19] S. A. Kauffman. *The Origins of Order*. Oxford University Press, New York, 1993.
- [20] S. A. Kauffman. Homeostasis and Differentiation in Random Genetic Control Networks. *Nature*, 224(5215):177–178, 1969.
- [21] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3):437–467, 1969.
- [22] H. Lähdesmäki i I. Shmulevich. BN/PBN Toolbox. <https://code.google.com/archive/p/pbn-matlab-toolbox/>, 2009. Accessed on 6 May 2024.
- [23] T. I. Lee i R. A. Young. Transcriptional Regulation and its Misregulation in Disease. *Cell*, 152(6):1237–1251, 2013.
- [24] F. Li, T. Long, Y. Lu, Q. Ouyang i C. Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences*, 101(4):4781–4786, 2004.
- [25] A. Lomuscio, H. Qu i F. Raimondi. MCMAS: A model checker for the verification of multi-agent systems. *International Journal on Software Tools for Technology Transfer*, 19:9–30, 2017.
- [26] B. Offermann, S. Knauer, A. Singh, M. L. Fernández-Cachón, M. Klose, S. Kowar, H. Busch i M. Boerries. Boolean modeling reveals the necessity of transcriptional regulation for bistability in PC12 cell differentiation. *Frontiers in Genetics*, 7:Article 44, 2016.
- [27] A. E. Raftery i S. M. Lewis. How many Iterations in the Gibbs Sampler? W J. M. Bernardo i inni, redaktorzy, *Bayesian Statistics*. Tom 4, strony 763–773. Oxford University Press, Oxford, 1992.
- [28] R. Schlatter, K. Schmich, I. A. Vizcarra, P. Scheurich, T. Sauter, C. Borner, M. Ederer, I. Merfort i O. Sawodny. ON/OFF and Beyond – A Boolean Model of Apoptosis. *PLOS Computational Biology*, 5:Article e1000595, 2009.
- [29] I. Shmulevich, E. R. Dougherty, S. Kim i W. Zhang. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, 18(2):261–274, 2002.
- [30] I. Shmulevich i E. R. Dougherty. *Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks*. SIAM, Philadelphia, 2010.
- [31] I. Shmulevich, I. Gluhovsky, R. F. Hashimoto, E. R. Dougherty i W. Zhang. Steady-state analysis of genetic regulatory networks modelled by probabilistic Boolean networks. *Comparative and Functional Genomics*, 4(6):601–608, 2003.

- [32] S. Sinha, B. M. Jones, I. M. Traniello, S. A. Bukhari, M. S. Halfon, H. A. Hofmann, S. Huang, P. S. Katz, J. Keagy, V. J. Lynch, M. B. Sokolowski, L. J. Stubbs, S. Tabe-Bordbar, M. F. Wolfner i G. E. Robinson. Behavior-related gene regulatory networks: A new level of organization in the brain. *PNAS*, 117(38):23270–23279, 2020.
- [33] Z. Szallasi i S. Liang. Modeling the Normal and Neoplastic Cell Cycle With “Realistic Boolean Genetic Networks”: Their Application for Understanding Carcinogenesis and Assessing Therapeutic Strategies. W *Proc. Pacific Symposium on Biocomputing (PSB’98)*, tom 3, strony 66–76, 1998.
- [34] A. Tafazzoli, J. R. Wilson, E. K. Lada i N. M. Steiger. Skart: a skewness-and autoregression-adjusted batch-means procedure for simulation analysis. W S. Mason, R. Hill, L. Mönch i O. Rose, redaktorzy, *Proc. 40th Conference on Winter Simulation (WSC’08)*, strony 387–395. Winter Simulation Conference, 2008.
- [35] R. Thomas, D. Thieffry i M. Kaufman. Dynamical behaviour of biological regulatory networks—I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bulletin of Mathematical Biology*, 57(2):247–276, 1995.
- [36] C. H. Waddington. *The Strategy of the Genes*. Allen i Unwin, London, 1957.
- [37] A. J. Walker. An Efficient Method for Generating Discrete Random Variables with General Distributions. *ACM Transactions on Mathematical Software*, 3(3):253–256, 1977.
- [38] T. Wilhelm. The smallest chemical reaction system with bistability. *BMC Systems Biology*, 3(1):Article 90, 2009.
- [39] S.-Q. Zhang, M. Hayashida, T. Akutsu, W.-K. Ching i M. K. Ng. Algorithms for Finding Small Attractors in Boolean Networks. *EURASIP Journal on Bioinformatics and Systems Biology*, 2007:1–13, 2007.

5 Aktywność naukowa albo artystyczna realizowana w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury

- W latach 2012-2017 odbyłem staż podoktorski na Uniwersytecie Luksemburskim. W ramach stażu powstała duża część publikacji wchodzących w skład omawianego osiągnięcia naukowego.
- W latach 2021-2022 odbyłem staż podoktorski na Uniwersytecie Luksemburskim, w ramach którego byłem kierownikiem projektu „*Autonomous Trustworthy Monitoring and Diagnosis of CubeSat Health (ATMonSAT)*” współfinansowanego przez Europejską Agencję Kosmiczną (ESA).
- Obecnie jestem kierownikiem czteroletniego projektu „*EdgeCR: Efficient and Efficacious Cellular Reprogramming with Edgetic Perturbations*”. Projekt jest finansowany przez Narodowe Centrum Nauki (NCN) w ramach współpracy Lead Agency Procedure (LAP) w programie Weave z Uniwersytetem Luksemburskim. Przyznane finansowanie: 1 398 596 PLN (NCN, Polska), 656 870,00 EUR (FNR, Luksemburg). Grant nr 2023/51/I/ST6/02864.

6 Osiągnięcia dydaktyczne, organizacyjne oraz popularyzujące naukę

6.1 Osiągnięcia dydaktyczne

6.1.1 Prowadzenie zajęć dydaktycznych

- Prowadzący zajęcia w ramach przedmiotu „Kurs Programowania w Pythonie” na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego | Semestr zimowy w latach 2023 i 2024.
Język wykładowy: polski
- Prowadzący zajęcia w ramach przedmiotu ”Statistical Machine Learning” na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego | Semestr zimowy w latach 2023 i 2024.
Język wykładowy: angielski
- Wykładowca przedmiotu „Języki i Narzędzia Programowania 3: Wstęp do Ucznia Maszynowego” na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego | Semestr zimowy 2023 r.
Język wykładowy: polski
- Wykładowca przedmiotu „Języki i Narzędzia Programowania 2: Wstęp do Ucznia Maszynowego” na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego | Semestr letni 2023 r.
Język wykładowy: polski
- Prowadzący zajęcia w ramach przedmiotu „Wstęp do Informatyki” na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego | Semestr letni w latach 2023, 2024, 2025.

Język wykładowy: polski

- Wykładowca przedmiotu “Space Informatics” na wydziale Faculty of Science, Technology and Medicine, Uniwersytet Luksemburski prowadzonego w ramach kierunku studiów “Interdisciplinary Space Master” | Semestr zimowy w latach 2019, 2020 i 2021.

Język wykładowy: angielski

- Wykładowca części “Introduction to Artificial Neural Networks and Deep Learning” przedmiotu “Theoretical Computer Science 3” prowadzonego na wydziale Faculty of Science, Technology and Medicine, Uniwersytet Luksemburski | Semestr zimowy 2020 r.

Język wykładowy: angielski

- Wykładowca przedmiotu “Advanced Systems Biology. Stochastic modelling” prowadzonego na wydziale Faculty of Science, Technology and Communication, Uniwersytet Luksemburski prowadzonego w ramach kierunku studiów “Master in Integrated Systems Biology” | Lata 2013–2017.

Język wykładowy: angielski

- Wykładowca przedmiotu “Automata and Formal Languages” prowadzonego na wydziale Faculty of Science, Technology and Communication, Uniwersytet Luksemburski | Lata 2013–2017.

Język wykładowy: angielski

- Prowadzący zajęcia w ramach przedmiotu “Computational Modelling Techniques” prowadzonego na Åbo Akademi University | Semestr letni 2011 r.

Język wykładowy: angielski

6.1.2 Promotorstwo pomocnicze prac doktorskich

1. Promotor pomocniczy pracy doktorskiej **Jakuba Zarzyckiego**. Tytuł dysertacji: “*Scalable deep reinforcement learning-based approaches for control of Boolean models of complex biological networks*”. Planowany termin obrony na Uniwersytecie Warszawskim: 2026 r.
2. Promotor pomocniczy pracy doktorskiej **Qixia Yuan**. Tytuł dysertacji: “*Computational Methods for Analysing Long-run Dynamics of Biological Systems*”. Praca doktorska obroniona w listopadzie 2017 r. na Uniwersytecie Luksemburskim | Ocena: celująca.

6.1.3 Promotorstwo prac magisterskich

1. Promotor pracy magisterskiej **Wojciecha Drochomireckiego**. Tytuł dysertacji: “*Exploration of Limitations in Transformer Models for Time Series Analysis*”. Praca obroniona w lipcu 2025 r. na Uniwersytecie Warszawskim | Kierunek: Machine Learning, Uniwersytet Warszawski.
2. Promotor pracy magisterskiej **Rui Zong**. Tytuł dysertacji: “*Cloud Removal from Satellite Imagery using Generative Adversarial Networks*”. Praca obroniona w lutym 2022 r. na Uniwersytecie Luksemburskim | Kierunek: Interdisciplinary Space Master, Uniwersytet Luksemburski.

3. Promotor pracy magisterskiej **Farzaneh Gholami Boroujeni**. Tytuł dysertacji: *“Machine Learning-based Financial and Market Analysis of Gradel’s Space Department”*. Praca obroniona w sierpniu 2021 r. na Uniwersytecie Luksemburskim | Kierunek: Interdisciplinary Space Master, Uniwersytet Luksemburski.

6.2 Osiągnięcia organizacyjne

Członek komitetów organizacyjnych:

- Computer Algebra Applications in the Life Sciences (CASinLife): sesja w ramach konferencji 30th International Conference on Applications of Computer Algebra (ACA 2025), 14 – 18/07/2025, Heraklion, Kreta, Grecja
- Computer Algebra Applications in the Life Sciences (CASinLife): sesja w ramach konferencji 28th International Conference on Applications of Computer Algebra (ACA 2023), 17 – 21/07/2023, Warszawa, Polska
- Przewodniczący komitetu organizacyjnego konferencji 16th International Conference on Formal Engineering Methods (ICFEM 2014), 03 – 07/11/2014, Luksemburg, Luksemburg
- Workshop on Computational Models for Cell Processes: wydarzenie towarzyszące konferencji 15th International Symposium on Formal Methods (FM 2008), 26 – 30/05/2008, Turku, Finlandia

6.3 Osiągnięcia w zakresie popularyzacji nauki

- Uczestnik panelu dyskusyjnego „Od laboratorium do kliniki – jak wygląda ta droga?” w ramach wydarzenia „Dzień Komórek Macierzystych – UniStem Day” na Wydziale Biologii Uniwersytetu Warszawskiego skierowanego do uczniów szkół średnich, mającego na celu popularyzację wiedzy o komórkach macierzystych, terapiach komórkowych i genowych. | 14/03/2025, Warszawa
[<https://www.biol.uw.edu.pl/dzien-komerek-macierzystych-unistem-day-na-wydziale-biologii-uw>]
- Wygłoszenie prezentacji zatytułowanej „Czy algorytmy mogą pomóc w leczeniu złożonych chorób?” w ramach wydarzenia „Dzień Komórek Macierzystych – UniStem Day” na Wydziale Biologii Uniwersytetu Warszawskiego skierowanego do uczniów szkół średnich, mającego na celu popularyzację wiedzy o komórkach macierzystych, terapiach komórkowych i genowych. | 14/03/2025, Warszawa
[<https://www.biol.uw.edu.pl/dzien-komerek-macierzystych-unistem-day-na-wydziale-biologii-uw>]
- Wygłoszenie prezentacji zatytułowanej “Deep Reinforcement Learning-based Approach Towards Effective Cellular Reprogramming” w ramach 21. epizodu cyklicznego wydarzenia Warsaw.AI. | 18/01/2024, Warszawa
[<https://www.youtube.com/watch?v=dIrCMFiyGj0>]
- Uczestnik panelu dyskusyjnego „Sztuczna inteligencja – transhumanizm i wizja nowego człowieka” w ramach Jubileuszowego Kongresu Polonii Medycznej zorganizowanego przez Federację Polonijnych Organizacji Medycznych. | 12-14/09/ 2024, Kraków
[<https://kongrespoloniimedycznej.org/o-kongresie/program-kongresu/12-wrzesnia-2024-czwartek>]

- Wygłoszenie prezentacji zatytułowanej “Deep Reinforcement Learning-based Approaches Towards Effective Cellular Reprogramming” w ramach inicjatywy „Noc Innowacji” festiwalu “Digital Festival”. | 20/10/2023, Warszawa
[<https://konferencja2023.ideas-ncbr.pl>]
- Udział w podcaście „Szerszy kontekst AI” tworzonym przez IDEAS NCBR Sp. z o.o. w roli gościa odcinka zatytułowanego „Ciało z laboratorium. Jak sztuczna inteligencja pomaga reprogramować komórki”. Rozmowa z dr hab. Iwoną Grabowską-Kowalik z Zakładu Cytologii Wydziału Biologii Uniwersytetu Warszawskiego dotycząca zagadnień związanych z biologią obliczeniową, reprogramowaniem komórek i rolą sztucznej inteligencji w badaniach biologicznych. | 2023, Warszawa
[<https://www.youtube.com/watch?v=zcSBJB7RPhE>]

7 Dodatkowe informacje

7.1 Wykaz publikacji niewchodzących w skład osiągnięcia naukowego

Okres po uzyskaniu stopnia doktora:

1. A. Wang, T. P. Tong, **A. Mizera**, and J. Pang. Benchmarking Structural Inference Methods for Interacting Dynamical Systems with Synthetic Data. In Proc. 38th Annual Conference on Neural Information Processing Systems (Datasets and Benchmarks Track) – NeurIPS’24, 2024.
CORE: A* (CORE2023) | MNiSW: 200 | Wkład własny: 15%
2. R. Horne, S. Mauw, **A. Mizera**, A. Stemper, and J. Thoemel. Anomaly detection using deep learning respecting the resources on board a CubeSat. *Journal of Aerospace Information Systems*, 20(12):794–922, 2023.
CiteScore: 3.7 (Scopus, 2023) | JIF: 1.3 (WoS, 2023) | MNiSW: 100 | Wkład własny: 50%
3. C. Hasan, R. Horne, S. Mauw, and **A. Mizera**. Cloud Removal from Satellite Imagery using Multispectral Edge-filtered Conditional Generative Adversarial Networks. *International Journal of Remote Sensing*, 43(5):1881–1893, 2022.
CiteScore: 7.0 (Scopus, 2022) | JIF: 3.4 (WoS, 2022) | MNiSW: 70 | Wkład własny: 70%
4. **A. Mizera**, J. Pang, C. Su, and Q. Yuan. ASSA-PBN: A toolbox for probabilistic Boolean networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(4):1203–1216, 2018.
CiteScore: 3.9 (Scopus, 2018) | JIF: 2.9 (WoS, 2018) | MNiSW: 70 | Wkład własny: 30%
5. **A. Mizera**, J. Pang, H. Qu, and Q. Yuan. ASSA-PBN 3.0: Analysing context-sensitive probabilistic Boolean networks. In *Proc. 16th International Conference on Computational Methods in Systems Biology (CMSB’18)*, Lecture Notes in Computer Science 11095, pp. 277–284. Springer, Cham, 2018.
CiteScore: 1.6 (Scopus, 2018) | Wkład własny: 35%

6. **A. Mizera**, J. Pang, and Q. Yuan. ASSA-PBN 2.0: A software tool for probabilistic Boolean networks. In E. Bartocci, P. Lio, and N. Paoletti (Eds.), *Proc. 14th International Conference on Computational Methods in Systems Biology (CMSB'16)*, Lecture Notes in Computer Science 9859, pp. 309–315. Springer, Cham, 2016.
CiteScore: 1.6 (Scopus, 2016) | Wkład własny: 40%
7. **A. Mizera**, J. Pang, and Q. Yuan. GPU-accelerated steady-state computation of large probabilistic Boolean networks. In M. Fränzle, D. Kapur, and N. Zhan (Eds.), *Proc. 2nd International Symposium on Dependable Software Engineering: Theories, Tools, and Applications (SETTA'16)*, Lecture Notes in Computer Science 9984, pp. 50–66. Springer, Cham, 2016.
CiteScore: 1.6 (Scopus, 2016) | Wkład własny: 40%
8. P. Żarnowiec, **A. Mizera**, M. Chrapek, M. Urbaniak, and W. Kaca. Chemometric analysis of attenuated total reflectance infrared spectra of *Proteus mirabilis* strains with defined structures of LPS. *Innate Immunity*, 22(5):325–335, 2016.
CiteScore: 5.3 (Scopus, 2016) | JIF: 2.3 (WoS, 2016) | MNiSW: 70 | Wkład własny: 20%
9. H. Qu, Q. Yuan, J. Pang, and **A. Mizera**. Improving BDD-based attractor detection for synchronous Boolean networks. In H. Mei, J. Lü, X. Ma, Q. Wang, G. Yin, and X. Liao (Eds.), *Proc. 7th Asia-Pacific Symposium on Internetware (Internetware'15)*, pp. 212–220, ACM, New York, 2015.
Wkład własny: 25%
10. X. Chen, **A. Mizera**, and J. Pang. Activity tracking: A new attack on location privacy. In S.D.C. di Vimercati, D. Du, G. Noubir, and P. Samarati (Eds.), *Proc. 3rd IEEE Conference on Communications and Network Security (CNS'15)*, pp. 22–30. IEEE, 2015.
Wkład własny: 35%
11. **A. Mizera**, J. Pang, and Q. Yuan. ASSA-PBN: An approximate steady-state analyser for probabilistic Boolean networks. In B. Finkbeiner, G. Pu, and L. Zhang (Eds.), *Proc. 13th International Symposium on Automated Technology for Verification and Analysis (ATVA'15)*, Lecture Notes in Computer Science 9364, pp. 214–220. Springer, Cham, 2015.
CiteScore: 1.5 (Scopus, 2015) | Wkład własny: 40%
12. **A. Mizera**, J. Pang, and Q. Yuan. Model-checking based approaches to parameter estimation of gene regulatory networks. In É. André, L. Zhang, J.S. Dong, and Z. Feng (Eds.), *Proc. 19th International Conference on Engineering of Complex Computer Systems (ICECCS'14)*, pp. 206–209. IEEE Computer Society Conference Publishing Services, 2014.
CiteScore: 0.5 (Scopus, 2014) | MNiSW: 70 | Wkład własny: 35%
13. P. Trairatphisan, **A. Mizera**, J. Pang, A. A. Tantar, J. Schneider, and T. Sauter. Recent development and biomedical applications of probabilistic Boolean networks. *Cell Communication and Signaling*, 11:46, 2013.
CiteScore: 5.7 (Scopus, 2013) | JIF: 4.7 (WoS, 2013) | MNiSW: 140 | Wkład własny: 30%

14. **A. Mizera**, J. Pang, T. Sauter, and P. Trairatphisan. A balancing act: Parameter estimation for biological models with steady-state measurements. In A. Gupta and T. A. Henzinger, (Eds.), *Proc. 11th International Conference on Computational Methods in Systems Biology (CMSB'13)*, Lecture Notes in Bioinformatics 8130, pp. 253–254. Springer, Berlin Heidelberg, 2013.

CiteScore: 1.5 (Scopus, 2013) | Wkład własny: 35%

15. **A. Mizera**, J. Pang, T. Sauter, and P. Trairatphisan. Mathematical modelling of the Platelet-Derived Growth Factor (PDGF) signalling pathway. In I. Petre (Ed.), *Proc. 4th International Workshop on Computational Models for Cell Processes (Comp-Mod'13)*, vol. 116 of Electronic Proceedings in Theoretical Computer Science, p. 35, 2013.

SJR: 0.312 (SCImago, 2013) | MNiSW: 20 | Wkład własny: 35%

16. El. Czeizler, **A. Mizera**, and I. Petre. A Boolean approach for disentangling the numerical contribution of modules to the system-level behavior of a biomodel. *Fundamenta Informaticae*, 116(1-4):51–63, 2012.

CiteScore: 1.7 (Scopus, 2012) | JIF: 0.4 (WoS, 2012) | MNiSW: 70 | Wkład własny: 40%

17. Eu. Czeizler, **A. Mizera**, El. Czeizler, R.-J. Back, J. E. Eriksson, and I. Petre. Quantitative analysis of the self-assembly strategies of intermediate filaments from tetrameric vimentin. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(3):885–898, 2012.

CiteScore: 3.1 (Scopus, 2012) | JIF: 1.6 (WoS, 2012) | MNiSW: 70 | Wkład własny: 45%

18. **A. Mizera**, El. Czeizler, and I. Petre. Computational methods for quantitative submodel comparison. In E. Katz (Ed.), *Biomolecular Information Processing. From Logic Systems to Smart Sensors and Actuators*, pp. 323–346. Wiley-VCH Verlag GmbH, Weinheim, DE, 2012.

Wkład własny: 45%

19. **A. Mizera**, Eu. Czeizler, and I. Petre. Self-assembly models of variable resolution. In C. Priami, I. Petre, E. Vink (Ed.), *Transactions on Computational Systems Biology XIV*, Lecture Notes in Computer Science 7625, pp.181–203. Springer, Berlin, Heidelberg, 2012.

CiteScore: 1.4 (Scopus, 2012) | Wkład własny: 45%

20. E. Kruglenko, **A. Mizera**, B. Gambin, R. Tymkiewicz., B. Zienkiewicz, and J. Litniewski. Nagrzewanie ultradźwiękami tkanek miękkich *in vitro* i własności akustyczne wytworzonych wzorców tkanek miękkich. 59th Open Seminar on Acoustics, 2012-09-10/09-14, Boszkowo (PL), pp.129-132, 2012.

Wkład własny: 20%

Okres przed uzyskaniem stopnia doktora:

1. **A. Mizera**, El. Czeizler, and I. Petre. Methods for Biochemical Model Decomposition and Quantitative Submodel Comparison. *Israel Journal of Chemistry*, 51(1): 151–164, 2011.

CiteScore: 1.5 (Scopus, 2011) | JIF: 1.5 (WoS, 2011) | MNiSW: 70 | Wkład własny: 45%

2. I. Petre, **A. Mizera**, C. L. Hyder, A. Meinander, A. Mikhailov, R. I. Morimoto, L. Sistonen, J. E. Eriksson, and R.-J. Back. A simple mass-action model for the eukaryotic heat shock response and its mathematical validation. *Natural Computing*, 10(1):595–612, 2011.

CiteScore: 2.4 (Scopus, 2011) | JIF: 1.6 (WoS, 2012) | MNiSW: 40 | Wkład własny: 40%

3. **A. Mizera** and B. Gambin. Modelling of ultrasound therapeutic heating and numerical study of the dynamics of the induced heat shock response. *Communications in Nonlinear Science and Numerical Simulation*, 16(5):2342–2349, 2011.

CiteScore: 5.0 (Scopus, 2011) | JIF: 2.8 (WoS, 2011) | MNiSW: 100 | Wkład własny: 60%

4. **A. Mizera** and B. Gambin. Stochastic modelling of the eukaryotic heat shock response. *Journal of Theoretical Biology*, 265(3):455–466, 2010.

CiteScore: 4.1 (Scopus, 2011) | JIF: 2.4 (WoS, 2010) | MNiSW: 70 | Wkład własny: 65%

5. **A. Mizera** and B. Gambin. The dynamics of heat shock response induced by ultrasound therapeutic treatment. In J. Awrejcewicz, M. Kaźmierczak, J. Mrozowski, and P. Olejnik (Eds.), *Proc. 10th Conference on Dynamical Systems - Theory and Applications (DSTA-2009)*, vol. 2, pp. 847–852. Left Grupa, Łódź, Poland, 2009.

Wkład własny: 60%

6. B. Gambin, T. Kujawska, E. Kruglenko, **A. Mizera**, A. Nowicki. Temperature fields induced by low power focused ultrasound in soft tissues during gene therapy. Numerical predictions and experimental results. *Archives of Acoustics*, 34(4): 445–459, 2009.

CiteScore: 1.3 (Scopus, 2011) | JIF: 0.3 (WoS, 2009) | MNiSW: 100 | Wkład własny: 20%

7. I. Petre, **A. Mizera**, C. L. Hyder, A. Mikhailov, J. E. Eriksson, L. Sistonen, and R.-J. Back. A new mathematical model for the heat shock response. In A. Condon, D. Harel, J. N. Kok, A. Salomaa, and E. Winfree (Eds.), *Algorithmic Bioprocesses*, pp. 411–425. Springer, Dordrecht Heidelberg London New York, 2009.

Wkład własny: 50%

8. I. Petre, **A. Mizera**, and R.-J. Back. Computational heuristics for simplifying a biological model. In K. Ambos-Spies, B. Löwe, and W. Merkle (Eds.), *Proc. 5th Conference on Computability in Europe: Mathematical Theory and Computational Practice (CiE 2009)*, Lecture Notes in Computer Science 5635, pp. 399–408. Springer, Berlin Heidelberg, 2009.

CiteScore: 1.3 (Scopus, 2011) | Wkład własny: 60%

9. N. Dojer, A. Gambin, **A. Mizera**, B. Wilczyński, and J. Tiuryn. Applying dynamic Bayesian networks to perturbed gene expression data. *BMC Bioinformatics*, 7:249, 2006.

JIF: 3.6 (WoS, 2006) | MNiSW: 100 | Wkład własny: 20%

7.2 Rozszerzone streszczenia

1. A. Mizera and J. Zarzycki. Deep Reinforcement Learning for Controlled Traversing of the Attractor Landscape of Boolean Models in the Context of Cellular Reprogramming. 32nd Conference on Intelligent Systems for Molecular Biology (ISMB 2024), 2024.

7.3 Raporty naukowe

1. R. Horne, S. Mauw, **A. Mizera**, A. Stemper, and J. Thoemel. Autonomous Trustworthy Monitoring and Diagnosis of CubeSat Health (AtMonSat). Project Report, 2022.

Raport końcowy podsumowującego realizację projektu AtMonSat współfinansowanego przez Europejską Agencję Kosmiczną (ESA) w ramach platformy Open Space Innovation Platform (OSIP) oraz przez Uniwersytet Luksemburski. Zaprezentowany podczas [The Software Systems Division \(TEC-SW\) and Data Systems, Microelectronics & Component Technology Division \(TEC-ED\) Final Presentation Days – Autumn 2022](#)

7.4 Narzędzia programowe niewchodzące w skład osiągnięcia naukowego

1. ATMonSAT: Implementacja algorytmu detekcji anomalii opracowanego w ramach projektu AtMonSat (<https://github.com/andre-stemper/ATMonSAT>).

.....
(podpis wnioskodawcy)