

Panorama

współczesnej nauki



Panorama współczesnej nauki

na podstawie odczytów Konwersatorium PW 2002–2012

pod redakcją Stanisława Janeczko

Centrum Studiów Zaawansowanych
Politechniki Warszawskiej



Centrum Studiów Zaawansowanych
Politechniki Warszawskiej
<http://www.csz.pw.edu.pl>

Skład redakcji:

Stanisław Janeczko
Małgorzata Zielińska
Ewa Stefaniak
Ilona Sadowska

Projekt graficzny i skład:

Danuta Czudek-Puchalska

© Copyright by Centrum Studiów Zaawansowanych Politechniki Warszawskiej,
Warszawa 2012

ISBN 978-83-61993-07-0

Wydrukowano w Polsce

Spis treści

5

Wstęp	
Stanisław Janeczko	7
Wykład 1	
Miniaturowe laboratorium chemiczne	
Zbigniew Brzózka	9
Wykład 2	
Bezemisyjna energetyka węglowa szansą dla Polski	
Jan A. Kozubowski	17
Wykład 3	
Zbiory przybliżone. Nowa matematyczna metoda analizy danych	
Zdzisław Pawlak	27
Wykład 4	
Modelowanie precesji-nutacji jako ważny element badań globalnej dynamiki Ziemi	
Aleksander Brzeziński	35
Wykład 5	
Zapach światła	
Kazimierz Brudzewski	47
Wykład 6	
Nanospintronika	
Tomasz Dietl	53
Wykład 7	
Uczciwość i wiarygodność nauki	
Maciej W. Grabski	59
Wykład 8	
Problem erozji ładu moralnego w świecie	
Andrzej Szostek	73
Wykład 9	
W jakim Wszechświecie żyjemy?	
Kazimierz Stępień	81

Wykład 10	
Wieloskalowe modelowanie molekularne białek	
Andrzej Koliński	87
Wykład 11	
Dynamika regularna i chaotyczna w układach technicznych z tarciami i uderzeniami	
Jan Awrejcewicz i Grzegorz Kudra	95
Wykład 12	
Einstein po stu latach	
Andrzej Kajetan Wróblewski	101
Wykład 13	
Czarne dziury: obiekty odkryte w przyrodzie czy wymyślone przez człowieka?	
Jerzy Kijowski	109
Wykład 14	
Czy krzem może chodzić – czyli o mikrosystemach, które myślą, czują i pracują	
Ryszard Stawomir Jachowicz	121
Wykład 15	
Dokąd zmierza Świat i Polska	
Grzegorz W. Kołodko	131
Wykład 16	
Counting Single Electrons Using a Carbon Nanotube Field-Effect Transistor	
Adrian Bachtold	137
Wykład 17	
Wiara, technika i medycyna	
Henryk Hoser	141
Wykład 18	
Science, Society and Sustainability	
Harold Kroto	147
Wykład 19	
Nauka w świecie islamu	
Janusz Danecki	151
Wykład 20	
Semiconductors in 21 st Century – the First Decade	
Jerzy Rużyłto	159
Wykład 21	
Utrzymanie Życia jako podstawowa wartość przestrzeni Miast	
Marek Budzyński	169
Wykład 22	
Daleka podczerwień (THz) w półprzewodnikach – fizyka i aplikacje	
Marian Grynberg	183
Wykład 23	
Jasne i ciemne strony Wszechświata	
Marek Demiański	185

Wstęp

7

Prezentowany zbiór artykułów powstał na bazie odczytów wygłoszonych przez wybitnych uczonych w pierwszej dekadzie programu Konwersatorium Politechniki Warszawskiej. Dwuczęściowa struktura tego wydarzenia – wykłady i dyskusje w Kawiarni Konwersatorium – stworzyła doskonałe warunki do aktywnego przedstawienia całego spektrum współczesnych osiągnięć nauki i techniki w szczególności metod i kierunków rozwoju ich najważniejszych dyscyplin.

W ramach Konwersatorium odbyło się ponad 65 odczytów, jednak tylko 23 spotkania zaowocowały artykułem w niniejszym tomie. Artykuły popularyzują ważne tematy i zaciekawiają oryginalnością oraz aktualnością prezentowanych zagadnień. Stanowią panoramiczny przekrój, poprzez wiele podstawowych i stosowanych dyscyplin nauki: fizykę, chemię, matematykę, biologię, etykę, ekonomię, socjologię, medycynę, modelowanie matematyczne, kosmologię, architekturę oraz wiele dziedzin techniki i technologii.

Z pierwszej dekady Konwersatorium PW wyłania się obraz złożoności i widocznego rozczłonkowania obszarów i dyscyplin myśli ludzkiej. Skuteczność nieograniczonej specjalizacji i uszczegółowienia wiedzy może być jednak ograniczona. Często sukcesy odnoszą metody oparte na interdyscyplinarności – pojawiające się na styku dziedzin, przełamujące hermetyczność dyscyplin i skuteczne siłą swojej uniwersalności. Wiele z artykułów pokazuje także jak sztuczne jest dzielenie nauki na część teoretyczną i część mającą zastosowania, jak ważne jest rozwijanie dziedzin tzw.

teoretycznych lub badań podstawowych, z których najbardziej niespodziewanie mogą następnego dnia wyrosnąć dziedziny „stosowane” – ważne zastosowania.

Autorzy artykułów to wybitni specjaliści w swoich dziedzinach piszący z możliwie najwyższą precyzją specjalistycznego języka, jednak z myślą o poszukującym i otwartym na nową wiedzę czytelniku. W artykułach Zbigniewa Brzózki, Kazimierza Brudzewskiego, Ryszarda Jachowicza, Adriana Bachtolda i Jerzego Rużyłto, można poznać najnowsze wyzwania zintegrowanych nauk technicznych. Problemy rozumienia wszechświata, jego struktury, jego ładu, również moralnego, znajdujemy w artykułach Kazimierza Stępnia, Andrzeja Kajetana Wróblewskiego, Jerzego Kijowskiego, Harolda Kroto i Andrzeja Szostka. Bardzo obiecujące możliwości zastosowań zaawansowanych metod fizyki można znaleźć w artykułach Tomasza Dietla i Mariana Grynberga. Bezpośrednie zastosowania metod matematycznych, metod obliczeniowych prezentowane są w artykułach Zdzisława Pawlaka, Aleksandra Brzezińskiego, Andrzeja Kolińskiego i Jana Awrajcewicz. Poglądy na naukę, jej historię, wiarygodność i zagrożenia płynące z wykorzystywania nauki przeciwko człowiekowi wyrażone są w artykułach Macieja Grabskiego, Henryka Hosea i Janusza Daneckiego. Wreszcie, koncepcje rozwoju ekonomicznego i troski o architekturę dla utrzymania życia znajdujemy w artykułach Grzegorza Kotodko i Marka Budzyńskiego. Mamy nadzieję, że niejednorodność tematyczna podejmowanych wątków, a jednocześnie siła uniwersalności metody naukowych analiz, pozwoli czytelnikowi dostrzec potrzebę przekraczania barier hermetycznych dyscyplin nauki i wykorzystania jej złożoności w wielodyscyplinarnych projektach badawczych.

Składam gorące podziękowania wszystkim, którzy przyczynili się do powstania tej książki. Autorom artykułów, które sukcesywnie ukazywały się w postaci wkładek do Miesięcznika PW, zespołowi redakcyjnemu Centrum Studiów Zaawansowanych, w szczególności Małgosi Zielińskiej szefowej zespołu za doprowadzenie projektu książki do szczęśliwego zakończenia.

Profesor Stanisław Janeczko

Miniaturowe laboratorium chemiczne

Na podstawie odczytu wygłoszonego w dniu 10 kwietnia 2003 roku

Zbigniew Brzózka

Politechnika Warszawska, Instytut Biotechnologii, Zakład Mikrobiologii

9

Dlaczego miniaturyzacja?

Postęp w dziedzinie systemów analityki chemicznej jest stymulowany przede wszystkim świadomością nowych zagrożeń w otoczeniu człowieka, rozszerzeniem potrzeb kontroli analitycznej, a także coraz ostrzejszymi wymaganiami stawianymi metodom pomiarowym. Wśród nowoczesnych rozwiązań systemów kontroli analitycznej obok zaawansowanych, wyrafinowanych metod instrumentalnych znalazły się miniaturowe sensory chemiczne oraz mikrosystemy do kompleksowej analizy wieloskładnikowych mediów (tzw. μ TAS — ang.: *Micro Total Analysis System*) oraz miniaturowe systemy *Lab-on-Chip*. Układy takie wytwarzane są przy zastosowaniu nie tylko najnowszych osiągnięć mikromechaniki krzemowej bazującej na technologii mikroelektroniki zintegrowanej, ale również z wykorzystaniem wielu innych materiałów i technologii. Do zalet tych mikrosystemów można zaliczyć: niski koszt, możliwość wieloskładnikowej analizy bardzo małych próbek (co jest szczególnie istotne w medycynie), uzyskiwanie wyników pomiarowych w czasie rzeczywistym i możliwość pracy ciągłej (monitorowanie) oraz pomiary *in situ*.

Od ponad dziesięciu lat obserwuje się dynamiczny rozwój miniaturyzacji urządzeń i systemów analitycznych dla olbrzymiej gamy zastosowań, począwszy od analizy krwi, przez monitorowanie elektrolitów w komórkach, a skończywszy na bardzo wydajnej analizie sekwencyjnej DNA. Miniaturyzacja jest kluczem pozwalającym na znaczne zmniejszenie zużycia odczyn-

ników i odpadów, skrócenie czasu analizy oraz produkcję tanich urządzeń, co prowadzi do obniżenia kosztów ekonomicznych wykonywanych analiz. Niezależnie od efektów ekonomicznych, miniaturyzacja przynosi rozwiązania dla problemów badawczych dotyczących miniaturowych obiektów, np. umożliwia kontrolowanie procesów wewnątrz komórki.

Koncepcje miniaturyzacji

Dotychczasowe metody i urządzenia analityczne przez wiele lat narzucały warunki pobierania próbek oraz ich wstępnego przygotowania do analizy, co ograniczało zastosowania w wielu nowych, dynamicznie rozwijających się dziedzinach, takich jak diagnostyka kliniczna czy zintegrowane systemy. W wielu przypadkach miniaturyzację urządzeń analitycznych wymusiły potrzeby kompatybilności urządzeń — zarówno pod kątem pomiarowym (mikroelektronika), jak i obiektowym (biologia, medycyna). Istotnym czynnikiem pozostaje zawsze aspekt ekonomiczny i zapotrzebowanie rynku, które powodują tendencję do obniżania kosztów, nie tylko samych urządzeń, ale i wykonywania analiz. Miniaturyzacja i integracja urządzeń analitycznych wychodzi naprzeciw nowym zapotrzebowaniom, dzięki wykorzystaniu w szerokim zakresie osiągnięć technologii mikrokrzemowych, mikromechaniki, nanotechnologii, nowych materiałów polimerowych oraz czułych metod detekcji.

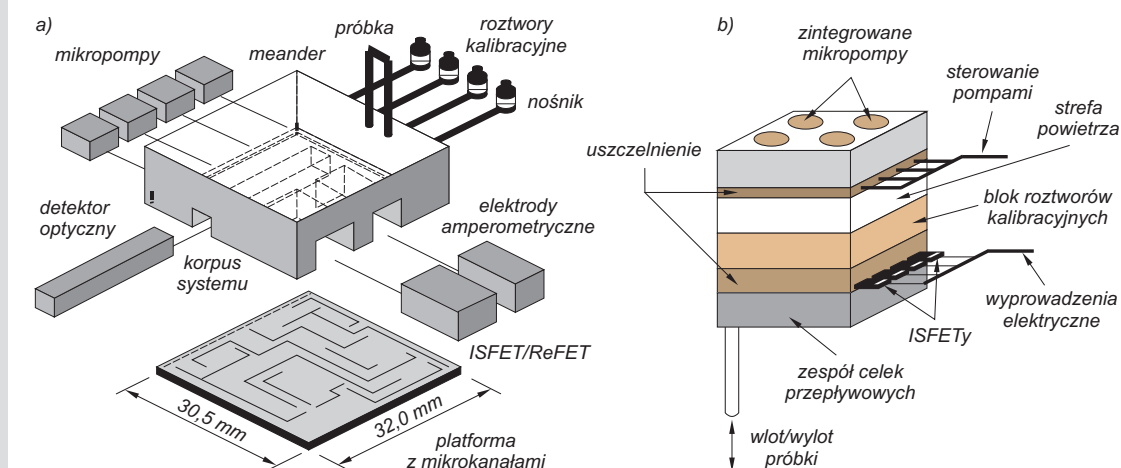
Różnorodne koncepcje miniaturyzacji urządzeń analitycznych i formy ich realizacji są efektem wyko-

rzystania najbardziej zaawansowanych technologii. Początkowe opracowania miniaturowych urządzeń opierały się na **systemach modułowych** lub **zintegrowanych**.

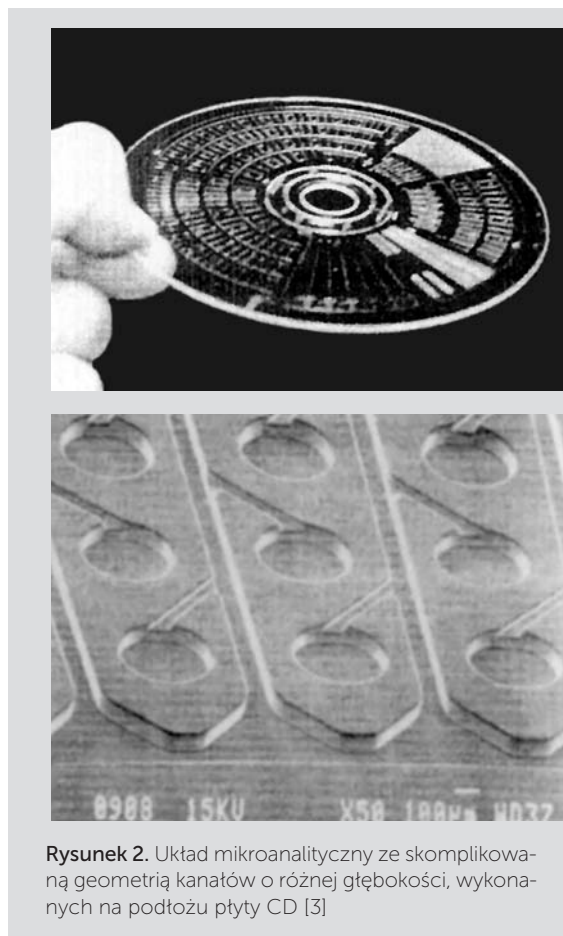
System modułowy zakłada opracowanie zestawu miniaturowych modułów (elementów), czyli pomp, dozowników, reaktorów, detektorów i sensorów o różnej funkcji, które można łączyć jak klocki Lego na podłożu lub płytce z odpowiednio przygotowanymi mikrokanalami. Niewątpliwą zaletą takiej koncepcji jest elastyczność w zestawianiu systemu analitycznego, dopasowanego do określonego zastosowania czy analizy oraz możliwość wymiany uszkodzonego modułu (elementu). Wymaga to jednak dużego doświadczenia potencjalnego użytkownika i nie pozwala na znaczne obniżanie kosztów urządzenia.

Aspekt ekonomiczny oraz wygoda przyszłego użytkownika są głównymi zaletami **systemu zintegrowanego**. W tej koncepcji wszystkie istotne elementy systemu analitycznego są wytwarzane na wspólnym podłożu (krzemowym, szklanym lub polimerowym). Integracja bardzo często obejmuje również umieszczenie układu akwizycji i obróbki sygnału analitycznego na wspólnym podłożu.

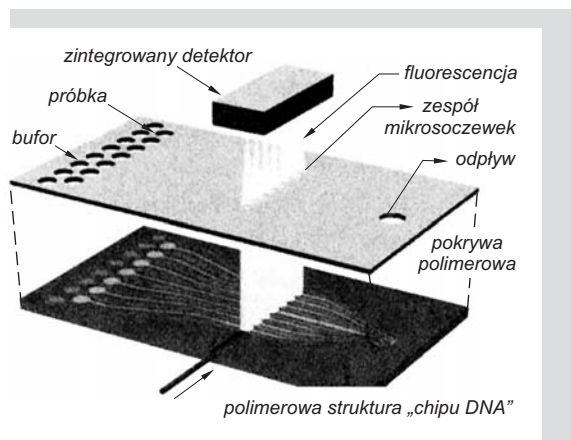
Jeżeli po pobraniu próbki następuje cykl procedur (dodanie reagenta lub/i kalibracja, pomiar, akwizycja i obróbka danych) pozwalający na otrzymanie końcowego wyniku lub oznaczenia dla danej próbki, analizę



Rysunek 1. Schematy miniaturowych systemów analitycznych o konstrukcjach modułowych: a) modułowy μ TAS wg Rossa [1], b) wg Hoffmanna [2]



Rysunek 2. Układ mikroanalityczny ze skomplikowaną geometrią kanałów o różnej głębokości, wykonanych na podłożu płyty CD [3]



Rysunek 3. Projekt „chipu DNA” wg Tabata et al. [4]

próbek można prowadzić sekwencyjnie. W przedstawionym na rysunku 2 układzie, wykonanym na podłożu klasycznej płyty kompaktowej, inicjowanie poszczególnych etapów analizy odbywa się przez różnicowanie częstotliwości obrotów płyty CD z zestawem reagentów. Siły oddziałujące na sam układ wykonany w podłożu CD wywołują określone zjawiska hydrodynamiczne wewnątrz mikrokanatów, pozwalające precyzyjnie sterować całą procedurą analizy.

Alternatywą jest równoległa analiza kilku próbek w jednym cyklu procedur. Rozwiązanie to stosuje się w przypadkach, gdy niezbędne jest wstępne rozdzielanie składników, zwłaszcza dla złożonych próbek środowiskowych i biomedycznych. Powstało wiele prototypowych modeli miniaturowych urządzeń analitycznych, które ze względów marketingowych otrzymały „chwytliwe” nazwy: „ μ GC” (miniaturowy chromatograf gazowy), „DNA chip”, „Lab in Bag”, „Lab on Chip” itp.

Wiele metod rozdzielania, stosowanych w klasycznych procedurach analitycznych, okazało się trudne do miniaturyzacji, a niekiedy całkowicie nieprzydatne do zastosowania w miniaturowych urządzeniach. Metodą, która stała się motorem napędowym rozwoju miniaturowych systemów analitycznych jest elektroforeza kapilarna.

Miniaturowa elektroforeza kapilarna

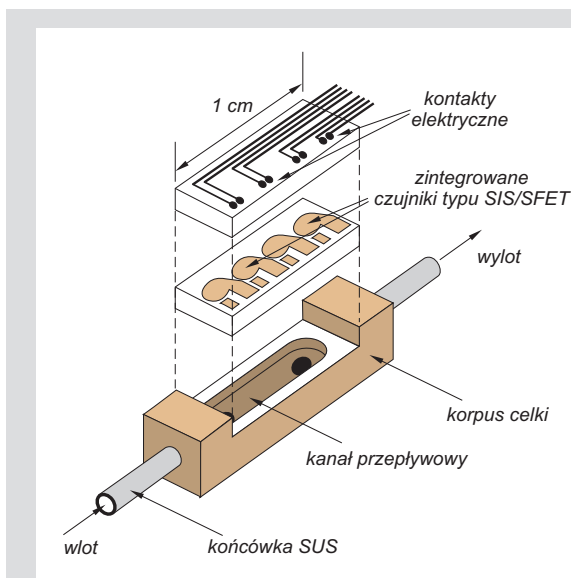
Klasyczna elektroforeza kapilarna, wykorzystująca przepływ elektroosmotyczny, jest stosowana do roz-

działania naładowanych cząsteczek analitów w polu elektrycznym od ponad trzydziestu lat. Jednakże dopiero jej miniaturyzacja i zintegrowanie na podłożu z wykorzystaniem struktury kanałów i kapilar pozwoliło pokazać jej nowe, fascynujące możliwości, zwłaszcza w analizach złożonych próbek klinicznych i biologicznych. Płaski profil analitu w kanale znacznie zwiększa efektywność rozdzielania składników oraz selektywność oznaczeń; kilkucentymetrowej wielkości bloki z mikrokanatami posiadają rozdzielczość odpowiadającą kilkuset tysiącom pótek teoretycznych. Historia miniaturyzacji elektroforezy kapilarnej jest krótka. Pierwszy system został opracowany w koncernie DuPont w 1990 roku [5]. Był on wykonany w płytce krzemowej. Pierwsze wyniki analiz opublikowano rok później dla modelu wykonanego na podłożu szklanym w firmie CIBA-Geigy [6]. Była to płytka szklana o wymiarach 15 cm x 4 cm z kanałami o średnicy kilkunastu μ m, która umożliwiła analizę prostych aminokwasów w czasie poniżej 8 minut. Kolejna wersja tego urządzenia została zaprezentowana w 1993 roku i miała już wymiary 8 cm x 7 cm i pozwalała na analizę próbki 6 aminokwasów o objętości 90–240 μ l w czasie poniżej 20 s. Kolejne lata pokazały niebywałe możliwości tej metody, np. miniaturowy system 15-kanatowy pozwolił na analizę fragmentów DNA w czasie rzędu 2 minut, a inny analizę znaczonych antygenów w czasie poniżej 30 s.

Miniaturowa elektroforeza kapilarna, jako technika rozdzielania do 1 mln pótek teoretycznych, jest znacznie efektywniejsza. Pozwala na szybsze w porównaniu do konwencjonalnej elektroforezy kapilarnej, wymagającej 8–10 minut, wykonywanie analiz (kilka sekund) oraz znaczne zmniejszenie objętości analizowanej próbki (nawet do 50 pL). Istotne jest również to, że umożliwia ona precyzyjne dozowanie i transport analitu, eliminując konieczność stosowania pomp i zaworów w miniaturowych systemach analitycznych.

Metody detekcji

Najczęściej stosowanymi metodami detekcji sprzężonymi ze zintegrowaną elektroforezą kapilarną były: fluorescencja (wzbudzana laserem lub diodami), spektrofotometria w świetle widzialnym oraz metody elektrochemiczne, jakkolwiek należy oczekiwać zastosowania nowych metod detekcji, podatnych na miniaturyzację. Jedną z grup urządzeń spełniających te wymogi są miniaturowe sensory chemiczne. Zasto-



Rysunek 4. Miniaturowy układ analityczny z zespołem zintegrowanych SIS/CHEMFET-ów jako elementów detekcyjnych wg Sakai et al. [7]

sowanie, np. w ochronie środowiska czy w chemii klinicznej, sensorów opartych na zaawansowanych technologiach (chemicznie modyfikowane tranzystory polowe – CHEMFET-y, biosensory, światłowodowe sensory chemiczne) jako detektorów pozwala dodatkowo obniżyć koszt masowej produkcji, przy jednoczesnym spełnieniu wszystkich wymogów związanych ze specyfiką pomiarów przemysłowych.

Z uwagi na technologiczną możliwość produkcji mikrosystemów w strukturze krzemowej, zastosowanie półprzewodnikowych sensorów chemicznych (CHEMFET) oraz elektrod na stałym podłożu krzemowym (*solid-state electrodes*), łączących czułość i se-

lektywność chemicznych warstw z miniaturowością i szerokim wachlarzem rozwiązań konstrukcyjnych przetwornika, pozwoli na opracowanie wielu nowych, miniaturowych systemów analitycznych.

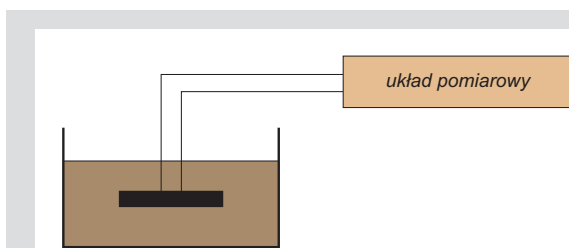
Podstawową zaletą systemów z wysoce selektywnymi sensorami (rys. 5) jest to, że umożliwiają one monitorowanie bądź bezpośrednie oznaczenie wybranego składnika analizowanej próbki z pominięciem etapu przygotowania. Nie zmieniają przy tym pierwotnego składu próbki, co jest szczególnie istotne.

Chemiczne metody modyfikacji powierzchni bramki, a zwłaszcza wprowadzenie polimerowych warstw membranowych i wykorzystanie materiałów supramolekularnych, umożliwiło opracowanie miniaturowych sensorów selektywnych i czułych na kilkanaście jonów oraz substancji organicznych.

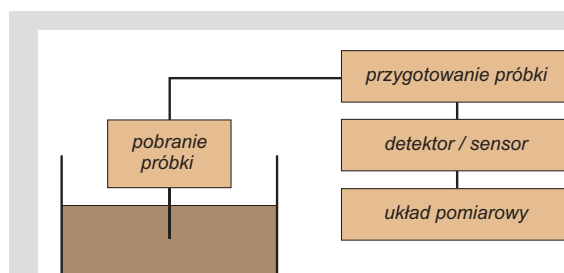
Mikrosystemy do kompleksowej analizy wieloskładnikowej

Skala potrzeb analitycznych, zwłaszcza w chemii klinicznej i ochronie środowiska, jest nieporównywalnie większa i trudno oczekiwać opracowania wysoce selektywnego sensora do oznaczeń wielu ważnych analitów. Ponadto nowe wymagania stawiane analityce wymuszały opracowanie metod pozwalających na równoległe lub sekwencyjne oznaczanie wielu składników przy zastosowaniu coraz mniejszych próbek. Był to jeden z impulsów rozwoju nowej grupy miniaturowych urządzeń analitycznych – mikrosystemów do kompleksowej analizy wieloskładnikowej (μ TAS) [8].

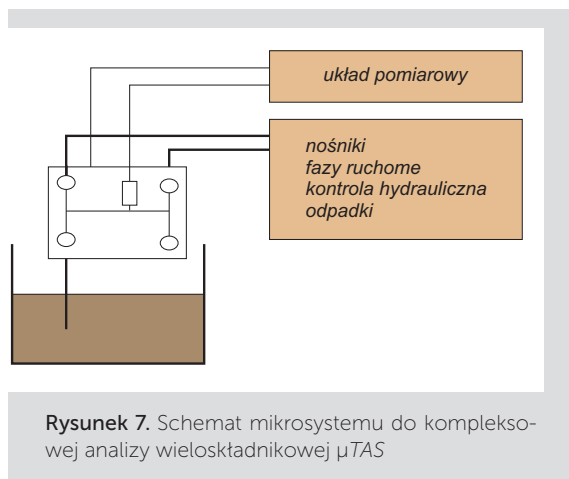
Kompleksowa analiza wieloskładnikowa (TAS – rys. 6) jest wykonywana w laboratoriach dysponujących stacjonarnym i drogim sprzętem analitycznym,



Rysunek 5. Schemat działania wysoce selektywnego (idealnego) sensora chemicznego



Rysunek 6. Schemat systemu do kompleksowej analizy wieloskładnikowej TAS



Rysunek 7. Schemat mikrosystemu do kompleksowej analizy wieloskładnikowej μ TAS

takim jak spektrometry masowe (*MS*), spektrografy z różnymi źródłami wzbudzenia (*MIP*, *ICP*) i inne, w których etap pobrania próbki nie jest zintegrowany z pozostałymi etapami analizy.

Wykorzystując osiągnięcia mikromechaniki i mikroelektroniki krzemowej, technologii wytwarzania mikrostruktur w szkle i w polimerach, technologii chemicznej modyfikacji (uczulania) struktur sensorowych oraz zalety miniaturowej elektroforezy kapilarnej i mikrodializy, opracowano miniaturowe urządzenia analityczne, które staną się wkrótce wyznacznikiem nowej jakości w kompleksowej analizie chemicznej. Są to urządzenia minimalizujące w zasadniczym stopniu konieczne ilości odczynników i powstających odpadków, co prowadzi do znacznego obniżenia kosztów, zwłaszcza że znacznemu skróceniu ulega także czas analizy. Uwzględniając powyższe zalety i spodziewane niskie koszty produkcji urządzeń, należy oczekiwać powstania dużego rynku zbytu na mikrosystemy do kompleksowej analizy wieloskładnikowej.

Opisane zastosowania ograniczają się obecnie do tzw. *screeningu* nowych leków i środków leczniczych, testów osobistych, a przede wszystkim do analizy DNA (*DNA chip*).

Przewiduje się jednak liczne zastosowania takich systemów, m.in. w:

- analitycznej kontroli procesowej *on-line* w biotechnologii,
- diagnostyce i monitorowaniu klinicznym,
- analizie przy łóżku pacjenta (*bed-side analysis*),
- monitorowaniu toksyczności środowiska,
- domowych systemach kontroli (*domestic control systems*).

Powyższe zastosowania nie wyczerpują potencjalnych, nowych możliwości miniaturowych urządzeń analitycznych. Dynamiczny rozwój chemii kombinatorycznej nie byłby możliwy bez tego typu urządzeń, a bieżące osiągnięcia mikrotechnologii w dziedzinie wytwarzania wielofunkcyjnych elementów (modułów) w krzemie, szkle, polimerach i materiałach alternatywnych jeszcze bardziej rozszerzą możliwości i zastosowania miniaturowych urządzeń analitycznych.

W ostatnich kilku latach, równocześnie w Europie, USA i Japonii, nastąpiła prawdziwa eksplozja badań naukowych nad miniaturowymi systemami analitycznymi i diagnostycznymi, połączonych z ich szybkimi wdrożeniami aplikacyjnymi. Szczególnie intensywnie są badane obszary zastosowań w analizie farmakologicznej i chemicznej, analizie próbek środowiskowych i w bezpieczeństwie pracy. Ocenia się, że w perspektywie 20 lat rozwój tych urządzeń może spowodować zmiany cywilizacyjne podobne do tych jakie zaszły w minionym 20-leciu pod wpływem rozwoju układów informatycznych i systemów komputerowych.

Tym samym, badania w rozważanej tematyce oprócz typowego dla nauki wymiaru poznawczego – łatwo publikowanego i przynoszącego znaczący wzrost jakości badań naukowych – posiadają silny aspekt rozwoju cywilizacyjnego.

Stan prac badawczych w Polsce

W Polsce istnieją silne grupy naukowo-badawcze wyposażone w odpowiednie narzędzia metodologiczne umożliwiające od wielu lat utrzymywanie niezwykle wysokiej pozycji w dziedzinie analityki chemicznej. Jednocześnie, szczególnie dzięki opanowaniu wybranych technologii mikromechanicznych, w Polsce istnieje dobrze rozwinięta mikromechanika krzemowa. Badania nad mikrosystemami krzemowymi do analizy składu są najmłodszą dyscypliną naukową stanowiącą połączenie mikroelektroniki i chemii analitycznej. Prace w pełnym wymiarze – również odkrycie podstaw zjawiskowych – rozpoczęto w świecie we wczesnych latach dziewięćdziesiątych, aczkolwiek rozproszone prace związane przede wszystkim z postępującym rozwojem mikromechaniki zintegrowanej na krzemie oraz techniki mikrosystemów były prowadzone od drugiej połowy lat osiemdziesiątych.

Przykładem podejmowanych prac jest między innymi niedawno zakończony grant zamawiany Komitetu

Badań Naukowych *Miniaturowe systemy do kompleksowej analizy mediów wieloskładnikowych*, realizowane przez Wydział Chemiczny Politechniki Warszawskiej, Wydział Elektroniki, Fotoniki i Mikrosystemów Politechniki Wrocławskiej, Instytut Biocybernetyki i Inżynierii Biomedycznej PAN oraz Instytut Technologii Elektronowej w Warszawie. Efektem tego projektu był między innymi pierwszy w Polsce mikrochromatograf gazowy oraz dwa modele miniaturowych systemów do analizy próbek ciekłych. Równie istotnym osiągnięciem było uruchomienie aktywności wielu grup badawczych, które obecnie współpracują nad nowymi projektami.

W Politechnice Warszawskiej powstał także program priorytetowy *Mikrosystemy: konstrukcje, technologie, projektowanie*, który w ciągu dwóch lat przyczynił się do ciekawej i efektywnej współpracy grup badawczych z wydziałów: Elektroniki i Technik Informatycznych, Chemicznego oraz Mechatroniki. Powstało wiele nowych opracowań obejmujących miniaturowe urządzenia oraz metodologię badań mikrosystemów. Niestety, program nie uzyskał odpowiednich środków finansowych (podobnie jak inne programy priorytetowe) na kontynuację prac. Współpraca będzie jednak kontynuowana, aczkolwiek naturalnym są działania poszukujące środków finansowych poza macierzystą Uczelnią.

Kolejnym przykładem inicjatyw jest konsorcjum badawcze skupiające grupy badawcze z Wydziału Chemicznego Politechniki Warszawskiej, Wydziału Chemii Uniwersytetu Warszawskiego, Instytutu Agrofizyki PAN w Lublinie i Instytutu Chemii Fizycznej PAN w Warszawie, które podjęło współpracę w dziedzinie miniaturowych systemów analitycznych obejmujących miniaturowe biosensory, bioogniwa paliwowe, systemy *Lab-on-Chip* dla diagnostyki medycznej i środowiskowej oraz mikroanalitki.

Prowadzone badania nad konstrukcją i technologią bezobstępowych miniaturowych urządzeń (mikrobiosensorów i mikrobioogniw) z przeznaczeniem do domowej diagnostyki medycznej, kontroli stanu środowiska naturalnego (ekosystemu) oraz oceny skażeń chemicznych w warunkach polowych mogą być szansą na ukształtowanie potencjału badawczo-rozwojowego w dziedzinie bioelektroniki i mikroanalitki, rozwój małych przedsiębiorstw bioelektronicznych w Polsce oraz pozyskanie rynku krajów akcesyjnych dla nowej generacji miniaturowych urządzeń.

Sympozja i warsztaty organizowane przez wymienione zespoły badawcze służą promocji i popularyza-

cji tej dziedziny nauki. W listopadzie 2003 roku zorganizowano w Politechnice Warszawskiej trzydniowe warsztaty *Miniaturowe systemy analityczne*

<http://csrg.ch.pw.edu.pl>

a w kwietniu 2004 roku sesję *Mikrosystemy* w ramach konferencji ELTE 2004

<http://elte.imio.pw.edu.pl>

Wydarzeniem powinno być międzynarodowe sympozjum *Miniaturized Analytical Devices*

<http://www.ch.pw.edu.pl/forum>

organizowane w lipcu 2004 roku przez Wydział Chemiczny w Matej Auli Politechniki Warszawskiej, na którym wykłady wygłoszą najwybitniejsi światowi specjaliści w dziedzinie mikrosystemów *Lab-on-Chip*.

Praca *Miniaturowe Laboratorium Chemiczne* jest finansowana przez Fundację na Rzecz Nauki Polskiej w ramach subsydium profesorskiego.

Bibliografia

- [1] B. Ross et al., *Proceedings of Eurosensors XI*, Warsaw, 1997, pp. 607–610.
- [2] W. Hoffmann, R. Rapp, *Sensors and Actuators B*, **34**, 1996, pp. 471–475.
- [3] A.L. Tiensuu, O. Ohman, L. Lundbladh, O. Larsson, *Proceedings of the μ TAS 2000 Symposium*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000, pp. 575–578.
- [4] O. Tabata, H. You, H. Shiraishi, H. Nakanishi, T. Nishimoto, K. Yamamoto, Y. Baba, *Proceedings of the μ TAS 2000 Symposium*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000, pp. 143–146.
- [5] S.J. Pace, *U.S. Patent* **4 908 112**, 1990.
- [6] A. Manz et al., *Trends Anal. Chem.*, **10**, 1991, pp. 144–149.
- [7] T. Sakai, H. Yagi, M. Shiratori, *Sensors and Actuators B*, **20**, 1994, pp. 169–173.
- [8] A. Manz, N. Graber, H.M. Widmer, *Sensors and Actuators B*, **1**, 1990, pp. 244–248.

Abstract

„*Lab-on-chip*” miniature analytical systems enable multi-component analysis of small biological samples by obtaining measurement results in real time and by continuous monitoring. They work out miniature systems containing structures and devices which can execute all chemical analysis components including sample preparation. They are produced with the application of the latest achievements in silicone micro-

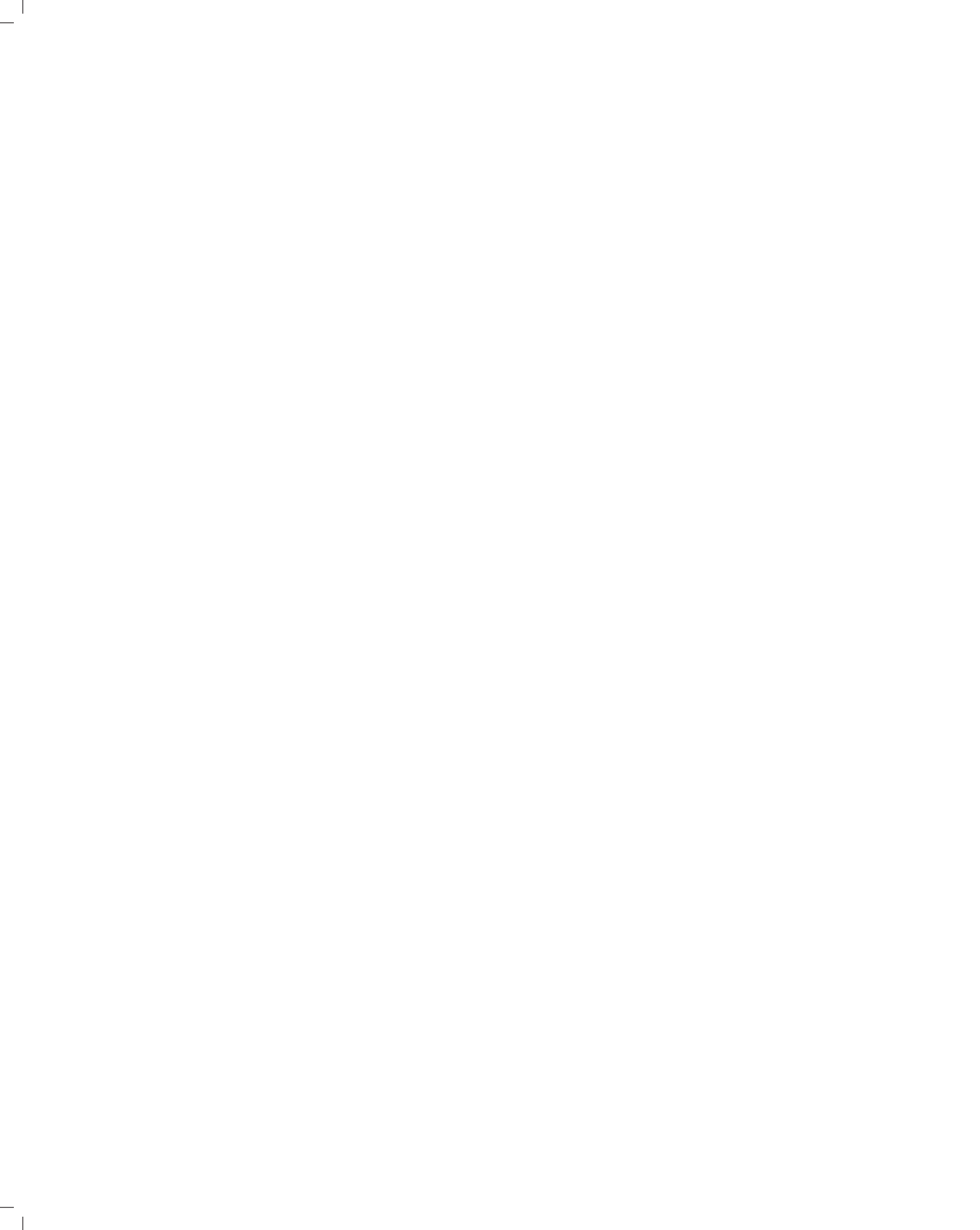
mechanics in relation to all sorts of materials such as: silicone structures, glass and various polymer and ceramic materials.

This publication presents analytical device miniaturization concepts, with particular attention to miniature capillary electrophoresis – the technique enabling dosing, transportation and separation of samples. Methods of detection, indispensable in analytical systems, have been described. The summary has been focused on the presentation of the research stage in the field in Poland, underlining the achievements of research groups from the Warsaw University of Technology.

Słowa kluczowe

miniaturowe systemy analityczne, miniaturowa elektroforeza kapilarna, mikrodetektory, mikroanalitka

Profesor Zbigniew Brzózka chemik, od 2008 roku pełni funkcję dziekana Wydziału Chemicznego Politechniki Warszawskiej. Specjalizuje się w mikrobioanalizie i miniaturyzacji urządzeń analitycznych. Prowadzi badania nad wykorzystaniem chemii supramolekularnej w sensorach chemicznych stosowanych w diagnostyce medycznej i badaniach biologicznych oraz nad miniaturyzacją sensorów chemicznych, biosensorów oraz systemów Lab-on-chip. W 2010 roku został uhonorowany prestiżowym Medalem Wiktora Kemuli. Jest współautorem wielu publikacji, m.in.: *Czujniki chemiczne i bioczujniki* (współautorstwo z Władysławem Torbiczem), *Sensory chemiczne* (współautorstwo z Wojciechem Wróblewskim), *Laboratorium analizy instrumentalnej* (redakcja pracy zbiorowej), *Miniaturyzacja w analizie* (redakcja pracy zbiorowej), *Mikrobioanalitka* (redakcja pracy zbiorowej).



Bezemisyjna energetyka węglowa szansą dla Polski

Na podstawie odczytu wygłoszonego w dniu 26 czerwca 2003 roku

Jan A. Kozubowski

Wydział Inżynierii Materiałowej Politechniki Warszawskiej

I too think the intellectual should constantly disturb, should bear witness to the misery of the world, should be provocative by being independent, should rebel against all hidden and open pressure and manipulations, should be the chief doubter of systems, of power and its incantations, should be witness to their mendacity. For this reason, an intellectual cannot fit into any role that may be assigned to him ... An intellectual essentially doesn't belong anywhere; he stands out as an irritant wherever he is; he does not fit into any pigeonhole completely ... To a certain extent an individual is always condemned to defeat. He is like Sisyphus in

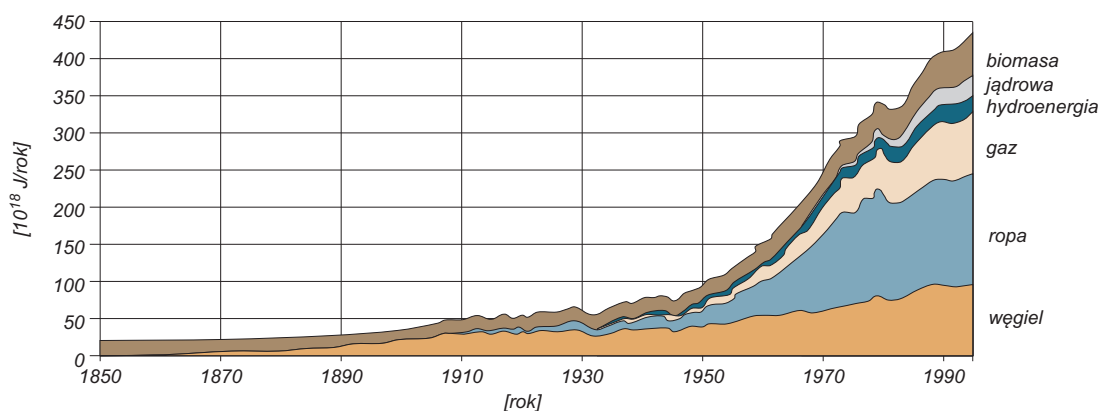
this regard ... And yet in another, more profound sense the intellectual remains, despite all his defeats, undefeated – again like Sisyphus. He is in fact victorious through his defeats.

Vaclav Havel, *Disturbing the Peace*

Perspektywa globalna

Gdy patrzymy na wykresy obrazujące wzrost energii zużywanej przez rozwijającą się ludzką cywilizację, to uderza nas nie tylko gwałtowne przyspieszenie tego wzrostu w ostatnim półwieczu, ale także rosnący udział

17



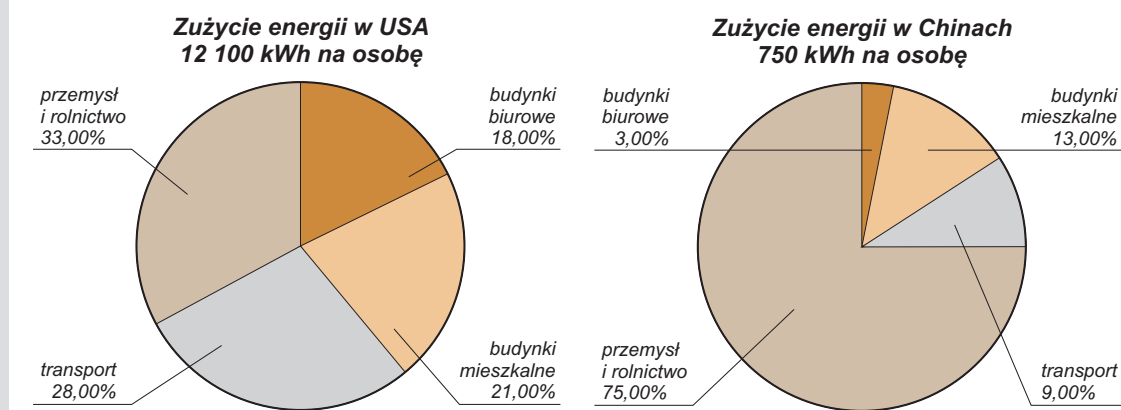
Rysunek 1. Wzrost zużycia energii na świecie w ciągu ostatnich 150 lat, w rozbiciu na różne jej rodzaje (wg materiałów SCEF)

paliiw kopalnych (rys. 1). Obecnie udział paliw kopalnych – węgla, ropy i gazu – jako surowców energetycznych zużywanych na świecie wynosi około 80%, chociaż w poszczególnych krajach spotkać można znaczące różnice.

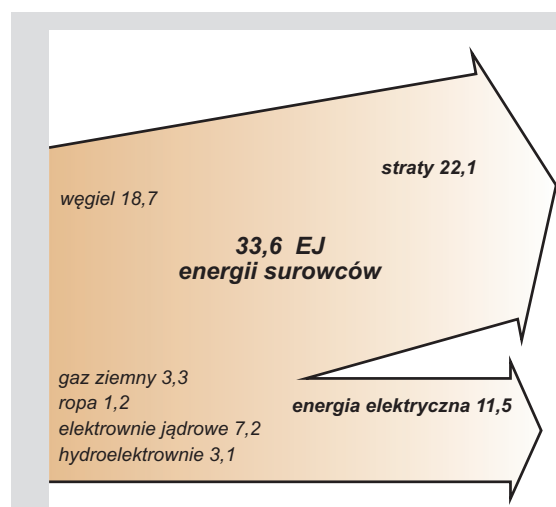
Ten wzrost ilości energii używanej przez ludzkość nie powinien dziwić – zarówno wzrost produkcji przemysłowej, jak wyższy standard życia rosnącej populacji wymagają energii. Widać to szczególnie wyraźnie na przykładzie analizy zużycia energii w krajach rozwiniętych (rys. 2). W ciągu ostatnich 150 lat w Stanach Zjednoczonych wzrosło ono 40-krotnie, a udział samych paliw kopalnych – 350-krotnie! Na drugim krańcu są kraje zacofane w rozwoju, w których zużycie energii niewiele przekracza niski poziom charakterystyczny dla tradycyjnych społeczeństw rolniczych.

Przez ostatnie 300 lat ludzkość opierała swój rozwój cywilizacyjny na zużywaniu coraz większej ilości paliw kopalnych – czerpała z energetycznego kapitału zgromadzonego w przetworzonych przez procesy geologiczne szczątkach organicznych sprzed milionów lat. Dziś problemem jest nie tyle perspektywa rychłego wyczerpania się dostępnych złóż surowców kopalnych, chociaż i z tym trzeba się liczyć, gdyż szacunki ekspertów wskazują na perspektywę wyczerpania się zasobów „łatwej” ropy w ciągu kilkudziesięciu lat, gazu ziemnego – w ciągu około stu lat, a jedynie zapasy węgla są na tyle duże, że przy obecnym zużyciu powinny wystarczyć na kilkaset lat. Problemem są szkodliwe dla środowiska przyrodniczego skutki spalania paliw kopalnych

i niska sprawność przetwarzania zawartej w nich energii na użyteczne dla człowieka jej postaci (rys. 3). Spala się je bowiem w elektrowniach, silnikach i różnego rodzaju piecach, domowych i przemysłowych o niewielkiej wydajności, emitując do atmosfery coraz większą ilość dwutlenku węgla. Obecnie na każdego mieszkańca Ziemi przypada rocznie ponad tona wyemitowanego do atmosfery CO₂, a na mieszkańca



Rysunek 2. Porównanie dystrybucji energii w wysoko rozwiniętym kraju i w kraju rozwijającym się. We Wspólnocie Europejskiej dystrybucja podobna jest do amerykańskiej



Rysunek 3. Na wytwarzanie energii elektrycznej w USA zużywa się obecnie 33,6 EJ energii pierwotnej zawartej w węglu, gazie ziemnym, ropie itp., z tego 2/3 jest tracone w procesach przetwarzania

Stanów Zjednoczonych – ponad 22 tony tego gazu. I chociaż skutki efektu cieplarnianego są trudne do oszacowania – Ziemia jest bowiem zbyt wielkim i zbyt złożonym układem, z trudem poddającym się modelowaniu – to jednak większość zajmujących się nim naukowców nie wątpi w jego istnienie i w to, że może być największym zagrożeniem, z jakim ludzkość będzie musiała się uporać w XXI wieku. Nie zmieniają tego nagłaśniane przez prasę opinie sceptyków (np. petycja oregońska, głośna książka Björna Lomborga *The Skeptical Environmentalist* czy artykuły Williego Soona i Salliego Baliunasa).

Cóż zatem można zrobić? Można zrobić wiele i w krajach rozwiniętych podejmuje się od pewnego czasu takie działania: stosowne regulacje prawne, wspieranie innowacji związanych z oszczędnościami energii, prace nad alternatywnymi źródłami energii, edukacja społeczeństw. W krajach rozwiniętych działania te przynoszą nie tylko doraźne, pozytywne skutki procentujące zahamowaniem lub cofnięciem zniszczeń środowiska, ale przekształcają gospodarkę – wzrost produkcji w tych krajach nie wiąże się już, tak jak kiedyś, ze wzrostem zużycia energii przez przemysł. Bardziej dalekość skutki tych działań odnoszą się do perspektywy „podzielenia się” nowoczesnymi rozwiązaniami z krajami rozwijającymi się w celu zapobieżenia globalnej katastrofie, którą zapewne spowodowałoby kilkudziesięciokrotne zwiększenie spalania surowców kopalnych w XXI wieku przez kraje rozwijające się. Pamiętajmy, że w krajach tych żyje dziś większość ludzi i nie możemy się dziwić ich dążeniu do osiągnięcia poziomu życia współczesnych rozwiniętych społeczeństw. Zwykle zapominamy o tym, że średnio dla prawie każdego z pięciu miliardów mieszkańców krajów rozwijających się musi wystarczyć 1/8 energii komercyjnej, jaka przypada na każdego spośród miliarda ludzi w krajach bardziej rozwiniętych, że około dwa miliardy ludzi nie ma dostępu do elektryczności, czy też gazu używanego do czyszczenia przygotowania posiłków i walczy o przetrwanie postępując się prymitywnymi paleniskami i coraz trudniejszym do zdobycia paliwem organicznym (drewno i suszone odchody zwierząt).

Większość komisji ekspertów powoływanych w różnych krajach w celu rozpatrzenia uwarunkowań i perspektyw zrównoważonego rozwoju ludzkiej cywilizacji jest zgodna co do tego, że w trudnym kilkudziesięcioletnim okresie przejściowym (w który właśnie wchodzimy), oprócz intensyfikacji wysiłków na rzecz eliminacji nieuzasadnionego technicznie, nadmiernego zużycia energii i sprawniejszego jej przetwarzania, a także po-

wszechniejszego wykorzystania odnawialnych źródeł energii, trzeba się oprzeć (z uwagi na skalę dotychczas zaangażowanych środków) na wykorzystaniu paliw kopalnych. Dotyczy to zwłaszcza węgla, bo jego zasoby mają najdłuższą perspektywę czasową. Trzeba je jednak wykorzystywać wydajniej i jednocześnie zredukować emisję zanieczyszczeń (w tym emisję CO₂) – w perspektywie do zera. Jest to możliwe przy stopniowym przechodzeniu do energetyki opartej na elektryczności i wodorze jako głównych nośnikach energii.

Węgiel jako pierwotny surowiec dla energetyki opartej na wodorze

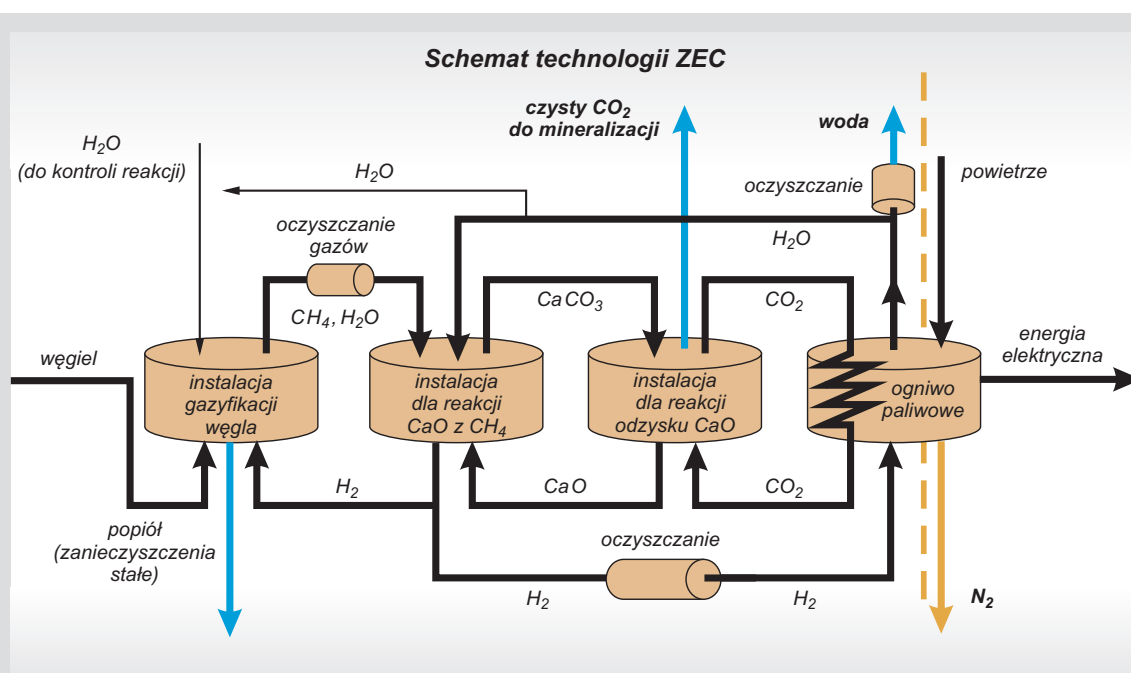
W celu zwiększenia wydajności tradycyjnych elektrowni spalających węgiel kamienny lub brunatny niezbędne jest podjęcie radykalnych kroków technicznych. Stopniowe usprawnianie takich elektrowni, choć kosztowne, nie przynosi znaczącego postępu. Dopiero zastosowanie zintegrowanych systemów energetycznych, w których paliwo kopalne spalane jest w tlenie, a nie w powietrzu i stosowane są turbiny z łożyskami chłodzonymi powietrzem lub parą, pozwala zwiększyć ich wydajność o prawie 10% (z dzisiejszej średniej 35,5% do prawie 44%). Spalanie w tlenie daje na wyjściu skoncentrowany strumień CO₂ – łatwiejszy (tańszy) do oczyszczenia i ewentualnej sekwestracji (wpompowywania w podziemne zbiorniki wodne, osady morskie, pokłady ropo- lub gazonośne, czy też mineralizacji, o której będzie mowa dalej). Pierwszy etap procesu spalania w tlenie stanowi gazyfikacja węgla. Jednym z jej produktów jest wodór, który można ekstrahować, oczyścić i wykorzystać do napędzania pojazdów. Ciepło generowane w takiej instalacji można także wykorzystać w odpowiednio zaprojektowanym, zamkniętym cyklu. W efekcie, wydajność konwersji energii chemicznej węgla w energię elektryczną i cieplną może przekraczać 70%, a przy tym nie ma żadnej (łącznie z emisją CO₂) emisji do atmosfery. Zastosowanie takiego procesu, nawet przy braku sekwestracji CO₂, doprowadza do dwukrotnego zredukowania jego emisji, co przy wprowadzaniu obecnie opodatkowaniu emisji złagodziłoby koszty budowania nowego typu instalacji energetycznych.

Energetyka oparta na wykorzystaniu wodoru, oprócz wszystkich trudności technicznych, które stopniowo są pokonywane, ma jedną zasadniczą wadę –

trzeba znaleźć tanie źródło tego paliwa przyszłości. W oceanicznych wodach są co prawda niewyczerpane zapasy wodoru, ale uzyskanie go z wody metodą elektrolizy wymaga energii elektrycznej. Tą drogą mogą pójść jedynie kraje bogate w naturalne źródła energii, takie jak Norwegia (obfitość hydroenergii) czy Islandia (energia geotermalna). Kiedy w latach 70. ubiegłego stulecia Islandczyk Bragi Arnason mówił o zaletach energetyki wodorowej, traktowano go jak nieszkodliwego maniaka, a dziś Reykjavik i reszta Islandii stają się dla Wspólnoty Europejskiej (przy znacznym udziale kapitału prywatnego) poligonem doświadczalnym, na którym w ciągu najbliższych kilkudziesięciu lat zamierza się całkowicie wyeliminować paliwa kopalne i emisję

CO₂, a już dziś wprowadza się autobusy napędzane silnikami elektrycznymi czerpiącymi prąd z ogniw paliwowych zasilanych wodorem.

Wodór nie jest jednak surowcem energetycznym, chociaż cechuje go największa gęstość energii – 120 MJ/kg (benzyna – 50 MJ/kg). Na powierzchni Ziemi jego zasoby są olbrzymie, występuje jednak w stanie związanym. Aby go pozyskać, trzeba te związania chemiczne zerwać, a do tego potrzebna jest energia czerpana z pierwotnych źródeł. Wodór trzeba zatem traktować jedynie jako czysty nośnik energii. Dotychczasowa produkcja wodoru zużywanego dziś głównie do produkcji nawozów azotowych (2/3 światowej produkcji wodoru), produkcji różnego rodzaju statych



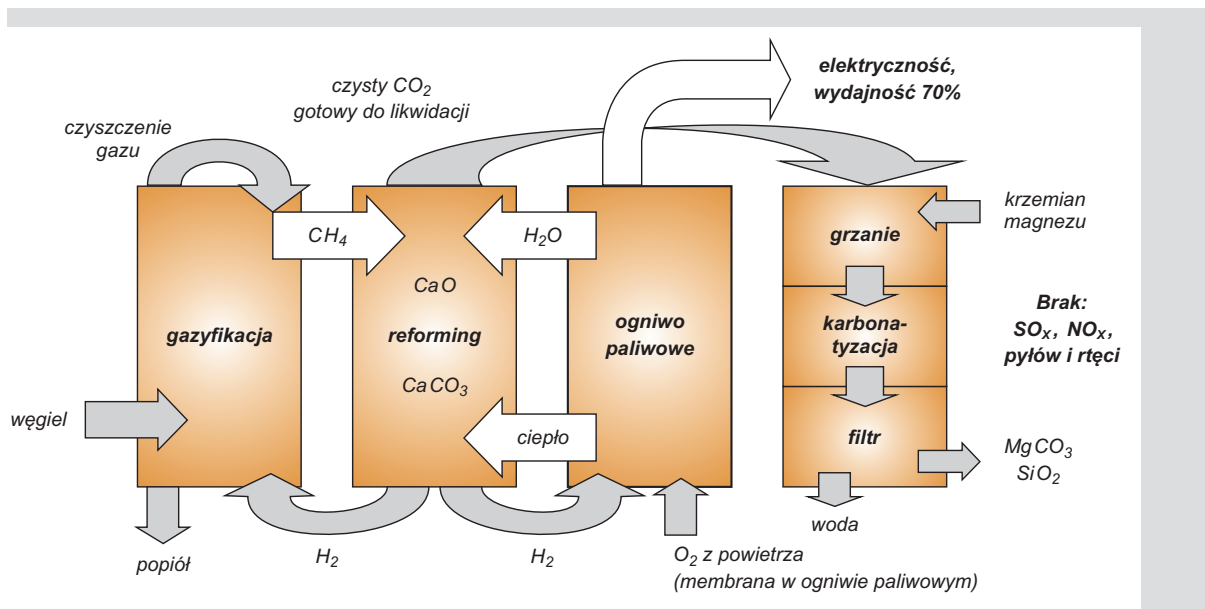
Rysunek 4 a. Schemat nowatorskiego, bezemisyjnego procesu wytwarzania wodoru lub energii elektryczności z węgla kamiennego (technologia ZECA). Proces opiera się na tym, że sumaryczna reakcja $\text{CaO} + \text{C} + 2\text{H}_2\text{O}$ (ciecz) $\rightarrow \text{CaCO}_3 + 2\text{H}_2 + 0,6 \text{ kJ/mol C}$ jest energetycznie prawie neutralna. Chociaż na schemacie wymieniono węgiel kamienny, to zasadniczo proces może być stosowany dla dowolnego paliwa zawierającego węgiel. W procesie wykorzystuje się cykliczną reakcję dwutlenku węgla z tlenkiem wapnia w celu wytworzenia wodoru z węgla i wody. Reakcja CO₂ z tlenkiem wapnia usuwa CO₂ z produktów reakcji i dostarcza dodatkowej energii, niezbędnej do zakończenia produkcji wodoru bez spalania węgla. Reakcja odzysku CaO z węglanu wapnia dokonuje się przy wykorzystaniu ciepła generowanego w wysokotemperaturowym ogniwie paliwowym podczas wytwarzania energii elektryczności z paliwa wodorowego. Przetworzenie ciepła odpadowego w użyteczną energię chemiczną umożliwia osiągnięcie bardzo wysokiej wydajności w przetwarzaniu energii wyjściowego paliwa w energię elektryczną w tym procesie. Ponieważ proces odbywa się zasadniczo w cyklu zamkniętym, to staje się możliwe osiągnięcie zerowej emisji zanieczyszczeń wraz z CO₂, jeżeli skoncentrowany strumień spalin CO₂ zostanie usunięty. Dokonuje się tego poprzez produkcję węglanu magnezu ze skąty ultramorficznej (krzemiany magnezowe). Produkty końcowe procesu mineralizacji CO₂ to trwałe, naturalnie występujące w przyrodzie minerały – węgiel magnezowy, krzemionka i woda

tłuszczów z olei roślinnych, używanego do spawania, do redukcji rud metali i w zastosowaniach kriogenicznych jest oparta w 48% na wykorzystaniu gazu ziemnego (reforming), w 30% na wykorzystaniu ropy (ten wodór jest jednak zużywany głównie w przemyśle petrochemicznym), w 18% na wykorzystaniu węgla i żelaza w 4% na elektrolizie wody.

Przejście od gospodarki opartej na spalaniu paliw kopalnych do gospodarki opartej na elektrochemicznym (głównie „spalaniu” wodoru w różnego rodzaju ogniwach paliwowych będzie trwało kilkadziesiąt lat i będzie się wiązało ze zmianami porównywalnymi z rewolucją przemysłową sprzed dwustukilkudziesięciu lat. Będzie ono wymagało opracowania sposobów taniej, masowej produkcji tego nośnika energii, rozwiązania problemów związanych z jego magazynowaniem i transportem. Obecnie toczą się prace nad optymalizacją (niebagatelna jest bowiem sprawa kosztów) wszystkich możliwych sposobów pozyskiwania wodoru: reformingiem potężnym z sekwestracją CO_2 , fotoelektrolizą, gazyfikacją i pirolizą biomasy, technologiami fotobiologicznymi (naśladującymi fotosyntezę lub wykorzystującymi biotechnologię) itp. W zależności od lokalnych zasobów energii pierwotnej wodór zapewne uzyskiwać się będzie na wiele różnych sposo-

bów, zarówno w małych, rozproszonych instalacjach, jak i w dużych kompleksach energetycznych lub energetyczno-przemysłowych.

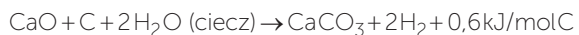
Przykładem koncepcji elektrowni opartej na węglu, ale wykorzystującej zgazowany węgiel do uzyskania wodoru i wysokotemperaturowe ogniwa paliwowe z elektrolitem tlenkowym do generacji elektryczności, jest technologia opracowana przez naukowców z Los Alamos National Laboratory. Utworzone przed kilkoma laty konsorcjum ZECA (*Zero Emission Carbon Alliance*) próbuje zainteresować tą technologią przemysł energetyczny. Stosuje się w niej wieloetapowy proces (rys. 4), w którym węgiel zostaje najpierw zgazowany w obecności wodoru i wody do metanu, a następnie metan poddaje się reakcji z tlenkiem wapnia w celu uzyskania wodoru. Część wodoru jest przekazywana z powrotem do instalacji gazyfikującej węgiel, a część, po oczyszczeniu, kierowana do instalacji wysokotemperaturowych ogniw paliwowych, w których generowana jest energia elektryczna. Ciepło odpadowe przesyłane jest do instalacji odzysku CaO (wyprężanie CaCO_3). Uwolniony dwutlenek węgla poddawany zostaje reakcji z krzemianami magnezu, co prowadzi do otrzymania krzemionki, wody i węglanu magnezu odsyłanego z powrotem do kopalni krzemianów lub użytkowanego jako nawóz



Rysunek 4 b. Schemat technologii ZECA, w której uzyskuje się energię elektryczną z węgla poprzez jego zgazowywanie dla uzyskania wodoru i zastosowanie wysokotemperaturowych, tlenkowych ogniw paliwowych dla uzyskania elektryczności, przy jednoczesnej mineralizacji wytwarzanego po drodze dwutlenku węgla

rolniczy. Cały proces odbywa się w zamkniętej instalacji. Nie wymaga też oddzielnej instalacji służącej do separacji i kompresji tlenu, ponieważ elektrolit tlenkowy w ogniwach paliwowych stanowi barierę między powietrzem atmosferycznym a wodorem.

Proces ZECA opiera się na fakcie, że sumaryczna reakcja



jest energetycznie niemalże neutralna, a wykorzystanie generowanego w niektórych etapach ciepła do napędzania potrzebnych reakcji chemicznych pozwala oczekiwać sprawności przekraczającej 70% w konwersji energii chemicznej węgla na energię elektryczną. W dodatku, z uwagi na znaczną koncentrację produktów reakcji pośrednich, ułatwione jest ich oczyszczenie z niepożądanych składników, a z zanieczyszczeń występujących w węglu i krzemianach można uzyskać pewne ilości tlenków żelaza (rudy żelaza). Tlenki azotu i pyły nie są emitowane do atmosfery, bo nie ma spalania, a związki siarki pozostają w szlamie i mogą być użyte do dalszej przeróbki. Proces ZECA jest więc całkowicie neutralny dla środowiska, jeśli nie liczyć kopalni węgla i kopalni krzemianów (zasypywanej z powrotem obojętnym dla środowiska węglanem magnezu).

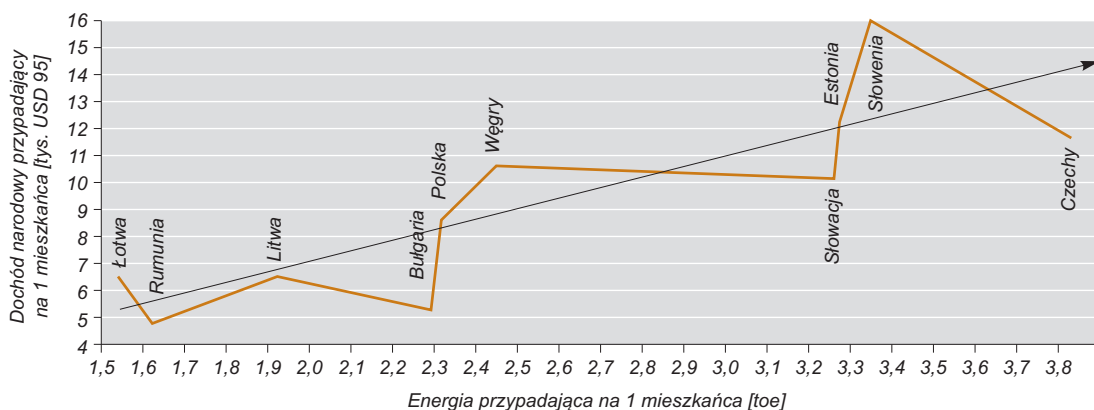
Przykładem innej koncepcji jest opatentowana przez amerykańskie konsorcjum Alchemix technologia uzyskiwania taniego wodoru przy użyciu węgla i wody jako surowców wyjściowych. Skojarzono w niej (w dowcipny sposób) dwa od dawna znane procesy – utlenianie metalu (takiego jak żelazo) dzięki wprowadzeniu pary wodnej w stopiony metal i redukcję tlenku pod wpływem węgla w wysokiej temperaturze (typowy proces metalurgiczny). Metal pełni w tej technologii rolę pośrednika odbierającego tlen cząsteczkom wody i przekazującego go atomom węgla. Nie jest więc w tym procesie zużywany, zaś produktami reakcji są wodór i dwutlenek węgla, które są oddzielane. Wodór po oczyszczeniu może zasilać ogniwa paliwowe lub być wykorzystany w procesach syntezy chemicznej, a dwutlenek węgla może podlegać sekwestracji. Technologia ta została już sprawdzona w skali przemysłowej, bo można było do tego celu wykorzystać wycofywane z eksploatacji instalacje metalurgiczne.

Perspektywa europejsko-polska

Warto może przypomnieć, że gdy przed 50 laty powstała koncepcja wspólnoty europejskiej, u jej podstaw

znajdowała się wspólna polityka dotycząca surowców (Wspólnota Węgla i Stali). Dziś, gdy społeczeństwa Europy uświadomiły sobie zagrożenia mogące wyniknąć z kontynuacji dotychczasowej, niczym nie ograniczanej tendencji do coraz szybszego zużywania (spalania) energetycznych surowców kopalnych, a także rosnącą zależność Europy od importu tych surowców z krajów pozaeuropejskich (głównie z Rosji), Wspólnota Europejska poświęca zagadnieniom energetyki coraz więcej uwagi. Dowodem na to może być szeroka dyskusja wywołana opublikowanym prawie dwa lata temu raportem na temat polityki energetycznej Europy (*Green Paper*). Początkowo dyskusja ta koncentrowała się na zaleceniach odnoszących się z jednej strony do lepszego gospodarowania dostępnymi zasobami energii (budownictwo, transport, rozproszona generacja), z drugiej zaś na położeniu nacisku na zwiększone wykorzystanie odnawialnych źródeł energii (energia wiatru i wody, biopaliwa). Zagadnieniom związanym z energetyką opartą na wodorze poświęcono w tym raporcie niewiele miejsca. Dopiero obawa, że Europa pozostanie w tej dziedzinie w tyle za Stanami Zjednoczonymi, Japonią i Chinami doprowadziła do podjęcia działań zmierzających do skoordynowania wysiłków (*Summary Report "Hydrogen Energy and Fuel Cells – a vision of our future"* przygotowany przez High Level Group for Hydrogen and Fuel Cells dla Komisji Europejskiej w czerwcu 2003 roku) i utworzenia Europejskiej Platformy Wodorowej (*European Hydrogen and Fuel Cell Technology Platform – H/FC TP*), czego dowodem jest raport Komisji Europejskiej z listopada 2003 roku. Celem tej Platformy ma być *ułatwienie i przyspieszenie rozwoju oraz wprowadzanie do użytku europejskich, światowej klasy systemów energetycznych opartych na wykorzystaniu wodoru i ogniwi paliwowych w transporcie i zasilaniu urządzeń stacjonarnych i przenośnych*. Wspólnota Europejska deklaruje również współpracę ze Stanami Zjednoczonymi w ramach ich nowej inicjatywy *The International Partnership for the Hydrogen Economy – IPHE*.

Rozszerzenie Wspólnoty Europejskiej o osiem nowych państw powoduje pewne przesunięcie dotychczasowych akcentów polityki energetycznej Wspólnoty. Nie tylko pojawiają się w jej obrębie państwa ze znaczącymi zasobami węgla (licząc łącznie eksploatawalne zasoby, na Polskę przypada ponad połowa zasobów europejskich), ale też są to państwa o nazbyt energochłonnych gospodarkach (rys. 5), wymagających istotnych przekształceń. Z tabeli zamieszczonej poniżej



Rysunek 5. W krajach Europy Środkowo-Wschodniej obserwuje się (podobnie jak to jest w skali całego świata) wyraźną relację między ilością energii przypadającej na 1 mieszkańca, a przypadającym na niego dochodem narodowym

wynika, że na jednostkę dochodu narodowego przypada w tych państwach o 50% więcej energii niż w dotychczasowych państwach członkowskich (dla Wspólnoty Europejskiej ten wskaźnik wynosi około 0,2).

Dostosowywanie do norm europejskich elektrowni opartych na spalaniu węgla wymaga znacznych nakładów finansowych, a w perspektywie czynne dziś elektrownie czeka wymiana wyeksploatowanych instalacji. Wydaje się, że w tej sytuacji należałoby skoncentrować wysiłki na stopniowym zastępowaniu wyeksploatowanych instalacji energetycznych nowoczesnymi, w których wykorzystuje się spalanie w tlenie i turbiny

nowej generacji. W perspektywie należałoby zmierzać do powstawania instalacji energetycznych wykorzystujących gazyfikację węgla i wytwarzanie wodoru, który mógłby służyć do generacji elektryczności na miejscu przy użyciu wysokotemperaturowych, tlenkowych ogniw paliwowych lub do dalszej dystrybucji.

Nadmierna energochłonność „odziedziczonych po socjalizmie” gospodarek krajów Europy Środkowo-wschodniej, której usuwanie zajmie pewien czas, daje tym krajom szansę na przekształcenie ich infrastruktury energetycznej na bardziej nowoczesną, bo chwilowo ich zapotrzebowanie na energię nie będzie rostało,

Kraj	Energia pierwotna [Mtoe]	Elektryczność [TWh]	Dochód narodowy [Mld USD 95]	Ludność [Mln]	Energia/dochód narodowy [kgoe/USD 95]
Bułgaria	18,8	40,6	44,0	8,2	0,43
Czechy	39,5	73,5	120,6	10,3	0,33
Estonia	4,6	8,5	17,2	1,4	0,27
Litwa	7,1	11,4	24,2	3,7	0,29
Łotwa	3,7	4,1	15,7	2,4	0,24
Polska	90,0	143,2	334,8	38,7	0,27
Rumunia	36,4	51,9	106,9	22,4	0,34
Słowacja	17,6	30,4	55,0	5,4	0,33
Słowenia	6,7	13,6	32,0	2,0	0,21
Węgry	24,8	35,0	107,6	10,1	0,23
Region	249,2	412,2	858,0	104,6	0,29

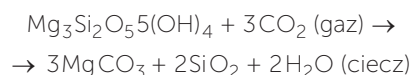
Dane wg M. Jaczewskiego (CENERG) odnoszą się do sytuacji w 2000 roku.

a ich dotychczasowy potencjał energetyczny będzie jedynie wydajniej wykorzystywany.

W najbliższym okresie można by zacząć prace nad wprowadzaniem elementów energetyki wodorowej na obszarze Śląska i sąsiednich terenów Czech, Słowacji i południowo-wschodnich Niemiec w oparciu o wykorzystanie wodoru z gazów koksowniczych. W Polsce wytwarza się rocznie ponad 2 miliardy metrów sześciennych gazów koksowniczych zawierających około 50% wodoru. Dziś gazy te są spalane, bo nie ma na nie zapotrzebowania. Gdyby jednak zainwestować w instalacje separacji wodoru, które i tak będą niedługo potrzebne w instalacjach zgazowywania paliw stałych (różnego rodzaju węgiel, biomasa i odpadki organiczne), to już dziś można by uzyskać znaczące ilości wodoru, pozwalające na rozpoczęcie budowy lokalnej infrastruktury dla gospodarki opartej na wodorze jako nośniku energii.

Obszar Śląska i południowo-zachodniej Polski cechuje nie tylko występowanie złóż węgla kamiennego, ale również złóż minerałów, które można wykorzystać do mineralizacji CO_2 – zbitych, masywnych skał typu serpentynitu (okolice Ślęży, obszar Gogotów–Jordanów i obszar Brzeźnica–Braszowice). Udowodniono (naukowcy z Albany Research Center w Oregonie – W.K. O’Conner, D.C. Dahlin i wsp.), że w warunkach przemysłowych (środowisko wodne z rozdrobnioną poniżej $75 \mu\text{m}$ skałą, ciśnienie CO_2 120–150 atmosfer, tanie katalizatory, temperatura około 180°C) zachodzący w przyrodzie egzotermiczny proces ($-63,6 \text{ kJ/mol}$

CO_2) reagowania krzemianów magnezowych z dwutlenkiem węgla



zostaje przyspieszony i zachodzi w czasie pół godziny, nie wymagając kilkuset tysięcy lat.

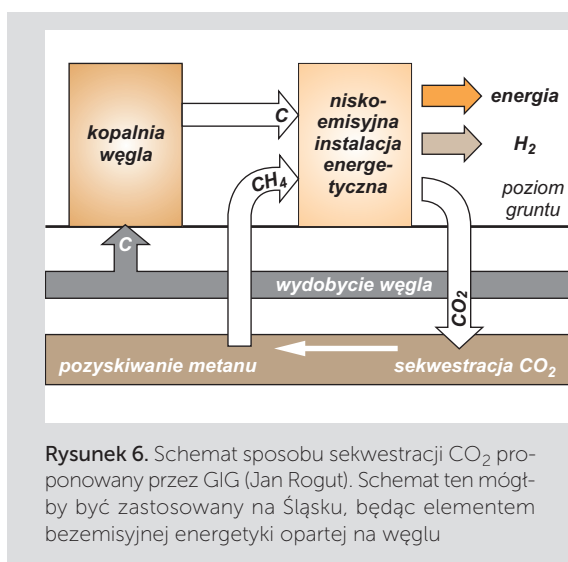
Na terenie Śląska istnieją zatem sprzyjające warunki dla zbudowania pierwszych instalacji energetycznych opartych na zintegrowanej technologii typu ZECA, w której uzyskuje się wodór i energię elektryczną z węgla bez jakiegokolwiek emisji do atmosfery, łącznie z emisją dwutlenku węgla. Przeprowadzone w Stanach Zjednoczonych analizy ekonomicznych aspektów tego typu procesu wskazują na to, że koszt uzyskiwanej energii elektrycznej byłby niewiele wyższy od aktualnych kosztów tradycyjnych elektrowni, a po uwzględnieniu kosztów ekologicznych mógłby być nawet niższy. Warty zainteresowania rozwiązaniem mogłaby być również technologia uzyskiwania wodoru z wykorzystaniem węgla, rozwijana przez amerykańską firmę Alchemix Corporation, gdyby udało się ją potączyć z mineralizacją dwutlenku węgla.

Inną metodą uzyskania „czystej” energii elektrycznej z węgla mogłoby być wykorzystanie proponowanej przez Główny Instytut Górnictwa (Jan Rogut) możliwości sekwestracji dwutlenku węgla poprzez wtłaczanie go w wyeksploatowane lub zbyt głęboko leżące i nie nadające się do eksploatacji złoża węgla (rys. 6). Pojemność tych złóż jest znaczna, a wtłaczany tam dwutlenek węgla wypychałby z tych złóż metan, który można używać do pozyskiwania wodoru.

W ten sposób obszar Śląska wraz z bliskimi mu regionami sąsiednich państw, gdzie występują podobne warunki geologiczne i techniczne, mógłby stać się zagłębiem energetycznym Europy, a Polska, zamiast eksportować, tak jak kiedyś, surowiec energetyczny w postaci węgla, mogłaby się stać eksporterem energii. Pozwoliłoby to jednocześnie na zbudowanie na Śląsku zrębów nowoczesnej infrastruktury ukierunkowanej na produkcję i dystrybucję wodoru dla potrzeb najbliższych regionów, a następnie Europy Środkowej i całej Wspólnoty Europejskiej.

Uwaga: Slajdy prezentowane podczas konwersatorium, w którym wzięt również udział Jan Rogut (GIG) – BEEW_cz1.pdf i BEEW_cz2.pdf, a także dodatkowe informacje można znaleźć na stronach internetowych Wydziału Inżynierii Materiałowej:

<http://www.inmat.pw.edu.pl/zaklady/ZPIM>



Abstract

The rapidly increasing world consumption of fossil fuels, associated with the growing emission of harmful particulates and CO₂, considered to be one of the major greenhouse gases, demands transition to an economy based on electricity and hydrogen as the energy carriers. Various methods of hydrogen production are considered and large effort is underway, focused on increasing their efficiency and lowering their costs. Coal should not be neglected as the primary energy source for hydrogen production, on the condition that CO₂ is isolated and some method of its sequestration is used. An example of emission-free technology of electrical energy production based on coal, developed by scientists from Los Alamos National Laboratory and promoted by ZECA corporation, is described. Natural conditions which make similar technology suitable for Poland – large coal reserves, relatively large beds of magnesium silicates, large industrial and workforce potential in the Silesia region – are stressed. Polish initiative for using its coal reserves as the basis for creating the technological infrastructure for the future hydrogen economy fits well into the European Hydrogen and Fuel Cells Technology Platform recently announced by the European Commission.

Słowa kluczowe

energetyka, wodór, gazyfikacja węgla, ogniwa paliwowe, sekwestracja CO₂

Profesor Jan A. Kozubowski specjalista w zakresie fizyki metali, inżynierii materiałowej i mikroskopii elektronowej. Profesor na Wydziale Inżynierii Materiałowej Politechniki Warszawskiej, ceniony przez młodzież nauczyciel akademicki. W latach 1984–1987 pełnił funkcję zastępcy dyrektora Instytutu Inżynierii Materiałowej, natomiast w latach 1999–2007 kierownika Zakładu Podstaw Inżynierii Materiałowej.

Profesor Jan A. Kozubowski zmarł 1 marca 2008 roku.

Zbiory przybliżone. Nowa matematyczna metoda analizy danych

Na podstawie odczytu wygłoszonego w dniu 30 października 2003 roku

Zdzisław Pawlak

27

Wstęp

Teoria zbiorów przybliżonych [4] jest, z logicznego punktu widzenia, sposobem nowego matematycznego podejścia do pojęć nieostrych, zaś w praktyce jest nową metodą analizy danych.

W teorii mnogości zbiór jest definiowany przez swoje elementy, przy czym nie jest tu potrzebna żadna dodatkowa wiedza o elementach uniwersum, z których tworzymy zbiory. W teorii zbiorów przybliżonych jest przeciwnie. Zakładamy, iż mamy pewne dane o elementach uniwersum i dane te są wykorzystywane do tworzenia zbiorów. Elementy, o których mamy identyczną informację są podobne i tworzą tzw. **zbiory elementarne**. W teorii zbiorów przybliżonych stanowią one podstawę rozumowania. Suma dowolnych zbiorów elementarnych jest nazywana **zbiorem definiowalnym**. Zbiory, które nie są zbiorami definiowalnymi nazywane są **zbiorami przybliżonymi**.

Zbiory definiowalne można jednoznacznie scharakteryzować przez własności ich elementów, natomiast zbiorów przybliżonych nie można w ten sposób określić, dlatego w teorii zbiorów przybliżonych wprowadza się pojęcia **dolnego** i **górnego przybliżenia zbioru**. Pojęcia te pozwalają na scharakteryzowanie każdego zbioru niedefiniowalnego (przybliżonego) za pomocą dwóch zbiorów definiowalnych — jego dolnego i górnego przybliżenia.

Dolnym przybliżeniem zbioru są wszystkie elementy, które w świetle posiadanej wiedzy mogą być jednoznacznie przyporządkowane do rozważanego zbioru,

zaś górnym przybliżeniem zbioru są wszystkie elementy, których przynależności do danego zbioru w świetle posiadanej wiedzy nie można wykluczyć. Różnica między górnym a dolnym przybliżeniem jest nazywana **obszarem brzegowym (brzegiem) zbioru**.

Z matematycznego punktu widzenia zbiór jest przybliżony wtedy i tylko wtedy, gdy jego obszar brzegowy jest niepusty.

Spróbujmy zobrazować różnicę między pojęciami zbioru ostrego i nieostrego. Przykładem zbioru ostrego jest zbiór liczb parzystych, gdyż każdą liczbę naturalną możemy jednoznacznie sklasyfikować jako parzystą lub nieparzystą. Zbiór „zdolnych” studentów jest natomiast pojęciem nieostrym, gdyż nie zawsze możemy jednoznacznie stwierdzić, że dany student jest zdolny lub nie.

Jak już wspomniano, z praktycznego punktu widzenia teoria zbiorów przybliżonych jest nową metodą analizy danych. Umożliwia ona:

- szukanie zależności między danymi,
- redukcję danych,
- określenie wagi danych,
- generowanie reguł decyzyjnych z danych itp.

Za zaletę teorii zbiorów przybliżonych należy uznać fakt, że:

- nie wymaga ona założeń odnośnie danych (np. prawdopodobieństwa czy rozmytości),
- zawiera szybkie algorytmy analizy danych,
- ułatwia interpretację wyników,
- charakteryzuje się znaczną prostotą matematyczną.

Metoda zbiorów przybliżonych znalazła liczne zastosowania, między innymi w:

- medycynie,
- farmakologii,
- bankowości,
- lingwistyce,
- rozpoznawaniu mowy,
- ochronie środowiska,
- bazach danych.

Teoria zbiorów przybliżonych ma wiele związków z innymi dziedzinami, a w szczególności:

- teorię ewidencji Dempstera-Shafera,
- teorię zbiorów rozmytych,
- metodami wnioskowania Boolowskiego.

Mimo to może być ona rozpatrywana jako niezależna (samodzielna) dyscyplina naukowa.

Teoria zbiorów przybliżonych nie jest alternatywą w stosunku do innych istniejących metod – raczej je uzupełnia i może być stosowana łącznie z nimi.

Na temat teorii zbiorów przybliżonych i jej zastosowań opublikowano na świecie do tej pory blisko trzy tysiące prac oraz kilkanaście książek. Wzbudziła ona spore zainteresowanie, głównie w USA, Kanadzie, Japonii, Chinach i Indiach, a prace na jej temat prowadzone są również w wielu innych krajach. Również w Polsce kilka ośrodków badawczych zajmuje się tą teorią i prowadzi prace nad jej zastosowaniami.

Do tej pory odbyło się wiele międzynarodowych konferencji na temat teorii zbiorów przybliżonych oraz jej zastosowań. Ponadto na wielu renomowanych, międzynarodowych konferencjach organizowano specjalne sesje poświęcone tej teorii.

Teoria zbiorów przybliżonych wykazała swą użyteczność w wielu dziedzinach oraz wzbudziła spore zainteresowanie na świecie i to nie tylko wśród informatyków, ale również wśród logików i filozofów. Wymaga jednak dalszych badań, w szczególności w zakresie jej podstaw matematycznych oraz możliwości zastosowania w różnych dziedzinach.

W artykule podane zostaną podstawowe pojęcia teorii zbiorów przybliżonych, przedyskutowane krótko jej zastosowania oraz dalsze perspektywy.

Więcej danych na temat zbiorów przybliżonych i ich zastosowań można znaleźć w Internecie na stronie

<http://www.roughsets.org>

Zbiory

Zanim podamy główne założenia teorii zbiorów przybliżonych, przypomnijmy kilka faktów dotyczących pojęcia zbioru.

Podstawowym pojęciem matematyki jest pojęcie zbioru. Niemal wszystkie konstrukcje matematyczne odwołują się do tego pojęcia.

Sformułowanie pojęcia zbioru oraz stworzenie teorii zbiorów zawdzięczamy niemieckiemu matematykowi Georgowi Cantorowi (1845–1918), który przed około stu laty stworzył podwaliny współczesnej teorii mnogości. Oryginalna, intuicyjna definicja pojęcia zbioru sformułowana przez Cantora [1] brzmi:

Unter einer „Mannigfaltigkeit“ oder „Menge“ verstehe ich nämlich allgemein jedes Viele, welches sich als Eines denken lässt, d.h. jeden Inbegriff bestimmter Elemente, welcher durch ein Gesetz zu einem Ganzen verbunden werden kann.

Jej tłumaczenie według Romana Murawskiego [3] jest następujące:

Pod pojęciem „rozmaitości” czy „zbioru” rozumiem mianowicie ogólnie każdą wielość, która może być pomyślana jako jedność, tj. każdy ogół określonych elementów, które na mocy pewnego prawa mogą być złączone w jedną całość.

Lub w nieco prostszym sformułowaniu:

Pod pojęciem „zbioru” rozumiemy każde zebranie w jedną całość M określonych dobrze odróżnionych przedmiotów m naszego oglądu czy naszych myśli (które nazywane są „elementami M ”) [3].

Jak widać jest to pojęcie intuicyjne i bardzo proste.

W 1902 roku wybitny filozof angielski Bertrand Russell (1872–1970) zauważył, że teoria mnogości jest sprzeczna, czyli prowadzi do antynomii (sprzeczności) logicznych. Antynomia logiczna, w dalszym ciągu rozumowania nazywana po prostu antynomią, powstaje wtedy gdy, prowadząc poprawne rozumowanie logiczne, dochodzimy do sprzeczności, czyli do zdań A i $nie-A$. Podważa to istotę logicznego rozumowania.

Dla przykładu omówimy tzw. antynomię Russella. Rozważmy zbiór X złożony ze wszystkich zbiorów Y , które nie są własnymi elementami. Jeżeli przyjmiemy, że X jest swoim własnym elementem to X , z definicji,

nie może być swoim elementem; jeżeli zaś przyjmiemy, że X nie jest swoim elementem, to zgodnie z definicją zbioru X musi on być swoim elementem. Zatem, przy każdym założeniu otrzymujemy sprzeczność.

Antynomia Russella świadczy o tym, że elementami zbioru nie mogą być dowolne obiekty, tak jak sobie to wyobrażał Cantor.

Mogłoby się wydawać, że antynomie są niewinnymi igraszkami logicznymi, jednakże tak nie jest. Podważają one istotę logicznego rozumowania i dlatego też przez ponad sto lat próbowano „naprawić” teorię Cantora lub zastąpić ją inną teorią zbiorów, jednakże jak dotąd próby te nie doprowadziły do pomyślnych rezultatów.

Jednocześnie, niezależnie od badań matematyków i filozofów, pojęcie zbioru zainteresowało inżynierów. Okazało się bowiem, że używając klasycznego, kantowskiego pojęcia zbioru nie da się sformułować i rozwiązać wielu problemów praktycznych.

W 1965 roku profesor Lotfi Zadeh z Uniwersytetu w Berkely zaproponował inne pojęcie zbioru [6], w którym elementy mogą należeć do zbioru „w pewnym stopniu”, a nie definitywnie, jak to ma miejsce w klasycznej teorii zbiorów. Propozycja ta znalazła bardzo wiele zastosowań i zapoczątkowała całą lawinę badań na temat teorii zbiorów rozmytych (*fuzzy set theory*), jak nazywano teorię Zadeha. Teorię zbiorów rozmytych można uważać za pewną formalizację pojęć nieostrych.

Inne podejście do formalizacji pojęć nieostrych zostało zaproponowane przez autora w artykule *Rough Sets* [4]. Teoria zbiorów przybliżonych (*rough set theory*) znalazła liczne zastosowania i zainteresowała wielu logików na świecie.

Dodajmy, że ani teoria zbiorów rozmytych, ani teoria zbiorów przybliżonych nie są alternatywą dla klasycznej teorii mnogości, gdyż obie zostały sformułowane przy wykorzystaniu pojęć zawartych w tej teorii.

Wnioskowanie z danych

Podstawowe pojęcia teorii zbiorów przybliżonych mogą być sformułowane całkowicie ogólnie, jednakże z punktu widzenia zastosowań tej teorii wygodnie sformułować je w terminologii typowej dla analizy danych, którą można uważać za szczególny przypadek wnioskowania indukcyjnego.

Przypomnijmy, że mamy dwa główne rodzaje wnioskowania — dedukcyjne i indukcyjne. **Wnioskowanie dedukcyjne** daje narzędzia służące do wyprowadzania

zdań prawdziwych z innych zdań prawdziwych. Prowadzi ono zawsze do konkluzji prawdziwych. Teoria dedukcji posiada dobrze ugruntowane i powszechnie przyjęte podstawy teoretyczne. Wnioskowanie dedukcyjne jest głównym narzędziem stosowanym w rozumowaniach matematycznych i poza nią nie znalazło zastosowania.

W naukach przyrodniczych (np. w fizyce) podstawową rolę odgrywa **wnioskowanie indukcyjne**. Cechą charakterystyczną tego typu wnioskowania jest to, że nie wychodzi ono (jak w logice dedukcyjnej) od aksjomatów wyrażających wiedzę ogólną o interesującym nas świecie. Punktem wyjścia tego typu rozumowania są pewne częściowe fakty dotyczące badanej rzeczywistości (przykłady), które następnie są uogólniane, tworząc wiedzę o świecie szerszym niż ten, który stanowił punkt wyjścia wnioskowania.

W przeciwieństwie do wnioskowania dedukcyjnego, wnioskowanie indukcyjne nie prowadzi do wniosków prawdziwych, a jedynie do wniosków prawdopodobnych (możliwych). Nie ma też jednolitych, ogólnie przyjętych podstaw teoretycznych. W logice indukcji rozstrzygnięcie prawdziwości hipotez odbywa się na podstawie eksperymentu, a nie drogą formalnego rozumowania, charakterystyczną dla logiki dedukcji. Fizyka jest tu najlepszą ilustracją.

Powstanie komputerów i ich nowatorskie zastosowania przyczyniły się istotnie do gwałtownego wzrostu zainteresowania wnioskowaniem indukcyjnym. Dzięki informatyce dziedzina ta rozwija się niezwykle dynamicznie. Uczenie maszynowe, odkrywanie wiedzy, wnioskowanie z danych, systemy eksperckie i inne stanowią przykłady nowych kierunków we wnioskowaniu indukcyjnym. Badania nad teorią indukcji zawdzięczają informatyce nowe inspiracje. Jednakże do sytuacji jaką mamy w logice dedukcji jest jeszcze bardzo daleka droga. Nie widać bowiem na horyzoncie zarysu teorii indukcji mającej taki status jak teoria dedukcji.

Wnioskowanie	dedukcyjne	indukcyjne
Zastosowania	matematyka	nauki przyrodnicze i techniczne
Teoria	teoria pełna	częściowe teorie
Wnioski	zawsze prawdziwe	prawdopodobne (możliwe)
Weryfikacja hipotez	dowód	eksperyment

Reasumując, cechy charakterystyczne wymienionych rodzajów wnioskovania podano w tabeli.

Jak widać, analizę danych należy traktować jako szczególny przypadek wnioskovania indukcyjnego.

Zbiory przybliżone i analiza danych

W celu lepszego i bardziej intuicyjnego zrozumienia podstaw teorii zbiorów przybliżonych z punktu widzenia analizy danych zacznijmy od prostego przykładu. Rozpatrzmy zbiór danych podanych w tabeli. Wiersze tabeli opisują pacjentów, kolumny tabeli oznaczone jako **ból głowy**, **ból mięśni**, **temperatura** reprezentują symptomy choroby i nazywane są **atrybutami warunkowymi**. Kolumna oznaczona jako **grypa** definiuje podział pacjentów na dwie klasy – chorujących na grypę i niechorujących – i nazwany jest **atrybutem decyzyjnym**.

Tabele takiego rodzaju nazywamy tablicami decyzyjnymi. Rzeczywiste tablice decyzyjne mogą zawierać dużo więcej atrybutów i przypadków. Ponadto oprócz atrybutów jakościowych mogą także zawierać wartości liczbowe. W pewnych przypadkach niektóre wartości atrybutów mogą być nieznane.

Problem będący przedmiotem naszego zainteresowania może być określony następująco: należy znaleźć zależność pomiędzy występowaniem grypy (lub jej niewystępowaniem), a symptomami opisującymi pacjentów. Inaczej mówiąc, celem jest opisanie zbioru przypadków $\{1, 2, 3, 6\}$ (lub zbioru $\{4, 5\}$) w kategoriach wartości atrybutów warunkowych.

Zauważmy, że analizowane dane są niespójne, gdyż przypadki 2 i 5 dostarczają sprzecznych informacji, tzn. obaj pacjenci opisani są tymi samymi wartościami atrybutów warunkowych, lecz są przydzieleni do różnych klas decyzyjnych. Oznacza to, że zbiór pacjentów cierpiących z powodu grypy nie może być jednoznacz-

nie opisany w kategoriach symptomów **ból głowy**, **ból mięśni**, **temperatura**. Możemy jednak opisać ten zbiór w sposób przybliżony. W oparciu o posiadane dane można stwierdzić, że:

- $\{1, 3, 6\}$ jest maksymalnym zbiorem przypadków, które na podstawie opisu atrybutami warunkowymi są z **pewnością** zaklasyfikowane do klasy pacjentów chorujących na grypę;
- $\{1, 2, 3, 5, 6\}$ jest zbiorem przypadków, których przydzielenie do klasy pacjentów chorujących na grypę jest **możliwe**, na podstawie opisu atrybutami warunkowymi;
- $\{2, 5\}$ jest zbiorem przypadków, które nie mogą być jednoznacznie przydzielone do klasy **grypa** lub do klasy **brak grypy**, ze względu na sprzeczny opis atrybutami warunkowymi.

Dwa pierwsze zbiory są przybliżeniami klasy decyzyjnej **grypa**, nazywanymi odpowiednio jej **dolnym** i **górnym przybliżeniem**.

Przedstawmy obecnie powyższe rozważania bardziej formalnie.

Tablice danych, takie jak podano w powyższym przykładzie, nazywane są często **systemami informacyjnymi**.

System informacyjny jest czwórką (U, A, V, f) , gdzie: U jest niepustym i skończonym zbiorem obiektów nazywanym **uniwersum**; A jest niepustym i skończonym **zbiorem atrybutów**; $V = \bigcup_{a \in A} V_a$, V_a jest **dziedziną atrybutu** $a \in A$; $a f : U \times A \rightarrow V$ jest **funkcją informacyjną**, taką, że $\forall a \in A, x \in U, f(a, x) \in V_a$.

Z każdym podzbiorem atrybutów $B \subseteq A$ związana jest binarna relacja $I(B)$, nazywana **relacją nierozróżnialności**, zdefiniowana jako

$$I(B) = \{(x, y) \in U \times U : \forall a \in B f(a, x) = f(a, y)\}$$

Jeśli $(x, y) \in I(B)$ to obiekty x i y są **nierozróżnialne** ze względu na podzbiór atrybutów B . Relacja nierozróżnialności jest relacją równoważności. Rodzinę wszystkich

Pacjent	Ból głowy	Ból mięśni	Temperatura	Grypa
1	nie	tak	podwyższona	tak
2	tak	nie	podwyższona	tak
3	tak	tak	wysoka	tak
4	nie	tak	normalna	nie
5	tak	nie	podwyższona	nie
6	nie	nie	wysoka	tak

klas abstrakcji relacji $I(B)$, czyli podział zbioru U za pomocą B , oznaczamy jako $U/I(B)$. $B(x)$ oznacza klasę abstrakcji relacji $I(B)$ zawierającą obiekt x i nazywane jest **zbiorem B-elementarnym**.

Jeżeli w systemie informacyjnym wyróżnimy rozłączne zbiory atrybutów warunkowych C i atrybutów decyzyjnych D (gdzie $A = C \cup D$), to system taki nazywany jest **tablicą decyzyjną**.

Niech $S = (U, A, V, f)$ będzie systemem informacyjnym, X niepustym podzbiorem U oraz $B \subseteq A$. Celem jest opisanie zbioru X w kategoriach wartości atrybutów ze zbioru B . Prowadzi to do zdefiniowania dwóch zbiorów $B_*(X)$ i $B^*(X)$, nazywanych odpowiednio **B-dolnym przybliżeniem** i **B-górnym przybliżeniem** X . Definiujemy je jako

$$B_*(X) = \{x \in U : B(x) \subseteq X\}$$

$$B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\}$$

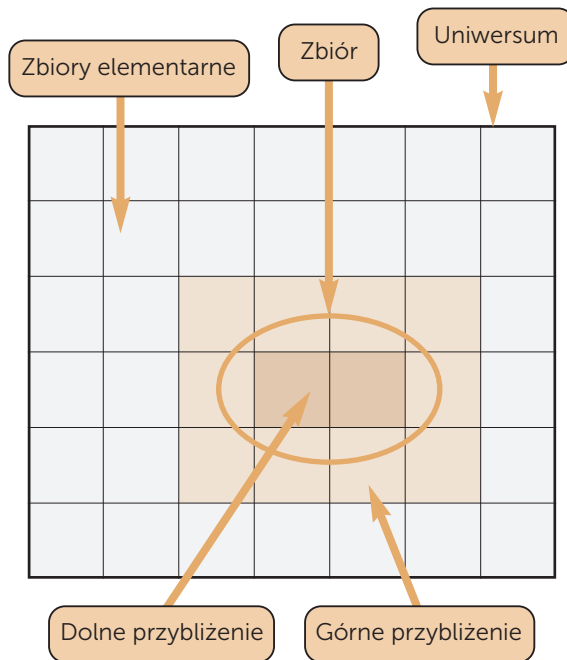
Zbiór $BN_B(X) = B^*(X) - B_*(X)$ jest nazywany **B-brzeżem** zbioru X .

Dolne przybliżenie $B_*(X)$ zbioru X jest zbiorem obiektów, które na podstawie zbioru atrybutów B można z pewnością zaliczyć do zbioru X , podczas gdy obiekty należące do zbioru $B^*(X)$ (w świetle atrybutów B) mogą być jedynie uznane jako **możliwe**, że należą do zbioru X . B -brzeg zbioru $BN_B(X)$ zawiera obiekty, których nie można jednoznacznie przydzielić do zbioru X z uwagi na sprzeczny opis w terminach atrybutów B . Natomiast obiekty ze zbioru $U/B^*(X)$ z pewnością nie należą do zbioru X . Jeśli $BN_B(X) \neq \emptyset$, to o zbiorze X mówimy, że jest on zbiorem **B-przybliżonym**. W przeciwnym razie jest on zbiorem **B-definiowanym** (dokładnym).

Formalne właściwości przybliżeń pokazano poniżej.

- 1) $B_*(X) \subseteq X \subseteq B^*(X)$,
- 2) $B_*(\emptyset) = B^*(\emptyset) = \emptyset$; $B_*(U) = B^*(U) = U$,
- 3) $B^*(X \cup Y) = B^*(X) \cup B^*(Y)$,
- 4) $B_*(X \cap Y) = B_*(X) \cap B_*(Y)$,
- 5) $X \subseteq Y$ implikuje $B_*(X) \subseteq B_*(Y)$ oraz $B^*(X) \subseteq B^*(Y)$,
- 6) $B_*(X \cup Y) \subseteq B_*(X) \cup B_*(Y)$,
- 7) $B^*(X \cup Y) \subseteq B^*(X) \cap B^*(Y)$,
- 8) $B_*(-X) = -B^*(X)$,
- 9) $B^*(-X) = -B_*(X)$,
- 10) $B_*(B_*(X)) = B^*(B_*(X)) = B_*(X)$,
- 11) $B^*(B^*(X)) = B_*(B^*(X)) = B^*(X)$.

Z właściwości tych widać, że przedstawione przybliżenia stanowią wnętrze i domknięcie w topologii generowanej przez dane.



Graficzną ilustrację przybliżenia pokazano na rysunku.

Zbiór przybliżony X może być scharakteryzowany ilościowo za pomocą współczynnika dokładności przybliżenia $\alpha_B(X)$

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|}$$

gdzie $|X|$ oznacza licznosc zbioru X .

Przybliżenia zbiorów są podstawowymi matematycznymi pojęciami teorii zbiorów przybliżonych i służą do opisu nieprecyzyjnej wiedzy o interesujących nas zjawiskach. Są one także używane do znajdowania zależności między danymi, ale tą problematyką nie będziemy się zajmować.

Co dalej?

W wielu ośrodkach badawczych w Polsce, USA, Kanadzie, Japonii, Chinach, Indiach i w Europie prowadzone są intensywne prace dotyczące różnych aspektów teorii zbiorów przybliżonych. Obejmują one między innymi:

- aspekty filozoficzne,
- podstawy matematyczne,
- różne uogólnienia i rozszerzenia,
- związki z innymi podobnymi teoriami,

- różnorodne zastosowania,
- oprogramowanie,
- prace konstrukcyjne.

Teorię zbiorów przybliżonych można uważać za szczególny przypadek realizacji idei Fregego dotyczącej pojęć nieostrych (*vagueness*). Istnieje pogląd, że teoria zbiorów przybliżonych daje tu nowe narzędzie do badania logiki pojęć nieostrych, a w szczególności wyjaśnia niektóre paradoksy nieostrości (*the sorites paradox*). Wielu filozofów i logików prowadzi badania w tym kierunku (patrz [5]).

Pojęcie nierozróżnialności (*indiscernibility relation*), stanowiące podstawę teorii zbiorów przybliżonych, jest ściśle związane z podaną przez Leibniza zasadą **nie-rozróżnialności identycznych obiektów** (*indiscernibility of identicals*). Ten aspekt zbiorów przybliżonych zainteresował również wielu filozofów, którzy prowadzą badania w tym zakresie.

Prace dotyczące podstaw teoretycznych zbiorów przybliżonych obejmują:

- ich algebraiczne i topologiczne własności;
- różne logiki związane ze zbiorami przybliżonymi;
- aspekty probabilistyczne zbiorów przybliżonych, między innymi związki z twierdzeniem Bayesa;
- aspekty teorio-mnogościowe zbiorów przybliżonych, w tym związki z mereologią Leśniewskiego;
- analizę przybliżoną, funkcje przybliżone, przybliżoną ciągłość oraz przybliżone różniczkowanie i całkowanie (w sensie analizy niestandardowej, a nie w sensie metod numerycznych).

W wyżej podanych kierunkach prowadzone są badania. Mają one na celu pełniejsze wyjaśnienie podstaw teorii zbiorów przybliżonych, głównie z punktu widzenia szeroko rozumianych własności matematycznych.

Teoria zbiorów przybliżonych jest uogólniana i modyfikowana na wiele sposobów. Między innymi zamiast relacji równoważności, stanowiącej punkt wyjścia do definicji zbioru przybliżonego, przyjmowane są inne relacje (np. relacja tolerancji). Bardzo dużym zainteresowaniem cieszą się również badania dotyczące sformułowania podstaw teorii zbiorów przybliżonych w oparciu o pewne koncepcje prawdopodobieństwa.

Ważne są również badania nad związkami teorii zbiorów przybliżonych z innymi teoriami dotyczącymi formalnych modeli nieprecyzyjności pojęć, takich jak:

- teoria zbiorów rozmytych,
- teoria ewidencji Dempstera-Shafera,
- statystyczne metody analizy danych.

Teoria zbiorów przybliżonych została również użyta do rozszerzenia niektórych znanych teorii, takich jak np. sieci neuronowe (*rough neural networks*), czy sieci Petriego (*rough Petri networks*). Są to obiecujące kierunki rozwijane przez wielu badaczy.

Bardzo intensywnie prowadzone są prace nad różnymi nowatorskimi zastosowaniami teorii zbiorów przybliżonych. Obejmują one między innymi teorię decyzji, medycynę, ekonomię, teorię sterowania i teorię konfliktów.

Szersze zastosowanie omawianej teorii wymagają odpowiedniego oprogramowania komputerowego. Opracowano już wiele systemów służących do tego celu i w dalszym ciągu prowadzone są w tym zakresie intensywne prace.

Aby w pełni wykorzystać możliwości, jakie daje teoria zbiorów przybliżonych w analizie danych, konieczne jest opracowanie komputerów lepiej przystosowanych do spożytkowania jej zalet. Prace w tym kierunku prowadzone są na uniwersytecie w Osace. Miejmy nadzieję, że doprowadzą one do powstania nowych koncepcji komputerów, lepiej przystosowanych do analizy danych.

Zakończenie

Podane tu sformułowanie podstawowych pojęć teorii zbiorów przybliżonych jest bardzo proste, a tym samym niestety niewystarczające dla celów praktycznych. Jego zastosowanie do rozwiązywania rzeczywistych problemów wymagało wielu rozszerzeń i uzupełnień. Pozwoliły one na stworzenie narzędzia matematycznego, które mogło być z powodzeniem zastosowane do rozwiązywania złożonych problemów. Zainteresowany czytelnik znajdzie odpowiednie dane w Internecie.

Teoria zbiorów przybliżonych, jak już wspomniano, nie jest alternatywą dla klasycznej teorii zbiorów. Jest ona pewną formalizacją pojęć nieprecyzyjnych (nieostrych), tym praktyczniejszą, że wykorzystuje dobrze znaną terminologię pojęć precyzyjnych. Pojęcia nieprecyzyjne są w niej przedstawiane za pomocą dwóch pojęć precyzyjnych (dolnego i górnego przybliżenia), co pozwala na operowanie klasycznym aparatem teorii mnogości przy wyrażaniu i analizie pojęć nieprecyzyjnych.

Teoria zbiorów przybliżonych znalazła liczne zastosowania, jednakże dalszy jej rozwój wymaga nadal badania jej podstaw matematycznych. Konieczne jest

stworzenie ogólnie dostępnego oprogramowania opartego na tej metodzie, a także odpowiednie rozwiązania sprzętowe. Prace w tym zakresie prowadzone są już w wielu krajach.

Literatura

- [1] George Cantor: *Grundlagen einer allgemeinen Mannigfaltigkeitslehre*. Leipzig 1883.
- [2] Gottlob Frege: *Grundlagen der Arithmetik*. Vol. 2. Verlag von Herman Pohle, Jena 1893.
- [3] Roman Murawski: *Filozofia matematyki, antologia tekstów klasycznych*. Seria Filozoficzna I, Logika nr 46, Uniwersytet im. Adama Mickiewicza w Poznaniu, Poznań 1986.
- [4] Zdzisław Pawlak: *Rough Sets*. "Int. J. of Information and Computer Sciences", no. 11, 5, pp. 341–356, 1982.
- [5] Read Stephen: *Thinking about Logic. Introduction to the Philosophy of Logic*. Opus, Oxford University Press, Oxford–New York 1995.
- [6] Lotfi Zadeh: *Fuzzy Sets*. "Information and Control", no. 8, pp. 338–353, 1965.

Abstract

Rough set theory is a new mathematical approach to imperfect knowledge.

The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians and mathematicians. Recently, it has also become a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate imperfect knowledge. The most successful one is, no doubt, the fuzzy set theory proposed by Zadeh.

The rough set theory proposed by the author presents still another approach to this problem. The theory has attracted the attention of many researchers and practitioners all over the world, who contributed essentially to its development and applications.

Rough set theory has an overlap with many other theories. However, we will refrain from discussing these connections here. Despite of the above mentioned connections, rough set theory may be considered as the independent discipline in its own rights.

Rough set theory has found many interesting applications. The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition.

The main advantage of rough set theory in data analysis is that it does not need any preliminary or additional information about data — like probability in statistics or basic probability assignment in Dempster-Shafer theory, grade of membership or the value of possibility in fuzzy set theory.

In this paper basic concepts of rough set theory are outlined, and various applications are briefly discussed.

At the end future problems and prospects of the theory are pointed out.

Słowa kluczowe

zbiory (*sets*), **zbiory przybliżone** (*rough sets*), **zbiory rozmyte** (*fuzzy sets*), **pojęcia nieostre** (*vagueness*), **analiza danych** (*data mining*)

Profesor Zdzisław Pawlak matematyk, informatyk, twórca teorii zbiorów przybliżonych (ang. *rough set theory*), członek Polskiej Akademii Nauk. Na tematy związane z tą teorią odbyło się wiele międzynarodowych konferencji, opublikowano kilka tysięcy artykułów naukowych i kilkadziesiąt książek. Inne znaczące osiągnięcia naukowe to: budowa pierwszego komputera GAM-1 (Grupy Aparatów Matematycznych), opracowanie nowej metody generowania liczb przypadkowych, zaproponowanie nowej metody przedstawiania liczb w systemie pozycyjnym z ujemną podstawą (tzw. system arytmetyki minus dwójkowej ozn. „ -2 ”), zaproponowanie nowej klasy języków beznawiasowych, stanowiących uogólnienie teorii znanej jako beznawiasowa notacja polska Jana Łukasiewicza, przedstawienie nowego modelu formalnego maszyny liczącej nazwanego od jego nazwiska „maszyną Pawlaka”, stworzenie pierwszego formalnego modelu matematycznego kodów genetycznych DNA, a także opracowanie nowego podejścia matematycznego do teorii konfliktów.

Profesor Zdzisław Pawlak zmarł 7 kwietnia 2006 roku.

Modelowanie precesji-nutacji jako ważny element badań globalnej dynamiki Ziemi

Na podstawie odczytu wygłoszonego w dniu 3 czerwca 2004 roku

Aleksander Brzeziński

Centrum Badań Kosmicznych Polskiej Akademii Nauk

35

Wstęp

Ruch obrotowy Ziemi, czyli ruch względem środka mas, nie jest jednostajną rotacją. Występuje szereg zaburzeń będących wynikiem grawitacyjnego oddziaływania Księżyca, Słońca i planet (perturbacje astronomiczne) albo generowanych przez wymianę momentu pędu między Ziemią stałą a jej komponentami ciekłymi i gazowymi, głównie atmosferą i oceanem (perturbacje geofizyczne).

Badanie zmian ruchu obrotowego Ziemi ma ogromne znaczenie praktyczne, polegające na wyznaczaniu i prognozowaniu zmiennych w czasie parametrów transformacji między **układami współrzędnych ziemskich i niebieskich**. Transformacja taka jest dokonywana zawsze, gdy chcemy interpretować obserwacje obiektów opisywanych w przestrzeni inercjalnej (gwiazdy, planety, sztuczne satelity Ziemi) dokonane przez obserwatora znajdującego się na Ziemi. Jest zatem ważnym elementem planowania i realizacji misji satelitarnych, które pełnią coraz bardziej istotną rolę w wielu dziedzinach życia, jak np.: prognozowanie pogody, telekomunikacja, satelitarne systemy lokalizacji i nawigacji.

Badanie perturbacji rotacyjnych ma też istotne znaczenie poznawcze, ponieważ dostarcza cennych informacji dotyczących globalnego rozkładu i transportu

mas w układzie Ziemi obejmującym część stałą (płaszcz i części ciekłe (ocean, atmosferę, hydrosferę lądową oraz jądro). Dostarcza również danych dotyczących kształtu, wewnętrznej budowy, a także reologii i własności dynamicznych naszej planety.

Zainteresowanie naukowe problemami zaburzeń rotacji Ziemi znacznie wzrosło w ostatnich kilku dekadach, głównie za sprawą pojawienia się w latach 70. technik **geodezji kosmicznej** (ang. *space geodesy*), takich jak radiointerferometria długich baz (VLBI — *Very Long Baseline Interferometry*), laserowe obserwacje satelitów (SLR — *Satellite Laser Ranging*) i Księżyca (LLR — *Lunar Laser Ranging*) oraz globalny system lokalizacji (GPS — *Global Positioning System*). Poprawiały one precyzję wyznaczania chwilowego kierunku osi układu ziemskiego w przestrzeni w tempie mniej więcej o jeden rząd wielkości na dekadę, osiągając w ostatnich latach poziom 0,1 milisekundy łuku (mas — *milliarcsecond*)¹ i z rozdzielczością czasową rzędu godzin. Dla porównania, od połowy XIX wieku, kiedy to rozpoczęły się próby wyznaczania parametrów rotacji Ziemi, do lat 70. XX wieku udało się poprawić dokładność tych wyznaczeń mniej więcej o czynnik 10.

W ujęciu matematycznym ruch obrotowy Ziemi może być utożsamiany ze zmienną w czasie macierzą transformacji między dwoma geocentrycznymi układami współrzędnych kartezjańskich — układem niebie-

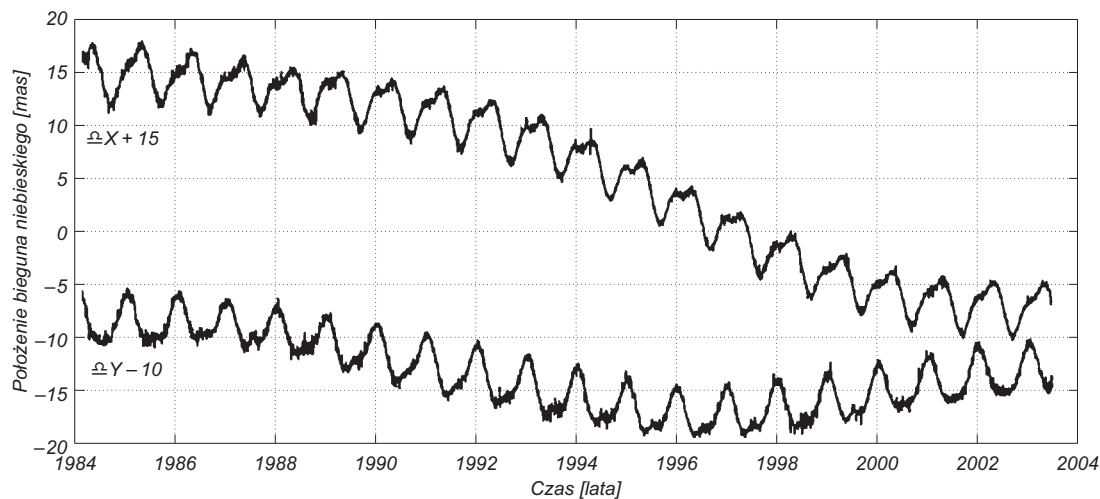
¹ Kąt 1" — sekunda łuku to kąt, pod jakim odcinek 30,9 metra na powierzchni Ziemi widziany jest z jej środka odległego o około 6370 km, zatem kątowi 0,1 mas odpowiada około 3 mm na powierzchni naszej planety.

skim, którego osie mają stały kierunek w przestrzeni, i układem obracającym się wraz z Ziemią. W przypadku tego ostatniego przyjmujemy, że jest sztywno związanym ze skorupą ziemską². Elementy macierzy wyraża się w formie **modeli matematycznych**, które są uzgodnionymi standardami międzynarodowymi (*IERS Conventions*, <http://www.iers.org/iers/products/conv/>), oraz **reziduoów obserwacyjnych**. Te ostatnie są regularnie wyznaczane w ramach międzynarodowych programów i z wykorzystaniem wymienionych wyżej technik, a następnie udostępniane użytkownikom przez Międzynarodową Służbę Ruchu Obrotowego i Systemów Odniesienia (*IERS – International Earth Rotation and Reference Systems Service*, <http://www.iers.org/iers/>).

Najbardziej znaczącym elementem katalogu perturbacji ruchu obrotowego Ziemi jest zjawisko **precesji-nutacji astronomicznej**. Jest to przestrzenny ruch osi figury Ziemi powodowany przez grawitacyjne oddziaływanie Księżyca, Słońca i – w znacznie mniejszym stopniu – planet. Niewielki, choć nie zaniedbywalny przy subcentymetrowym poziomie dokładności, wkład mają również pływy oceaniczne oraz zaburzenia w atmosferze i oceanach, pobudzone przez dobowy cykl

termiczny. Główna składowa tego zjawiska, zauważona już w II wieku p.n.e. przez Hipparcha, to precesja – jednostajny ruch osi po powierzchni prostopadłego do płaszczyzny ekliptyki stożka o kącie rozwarcia $2 \times 23,5^\circ$ i z okresem obiegu 25 400 lat. Precesja jest obserwowana jako przemieszczanie się po sferze niebieskiej punktu równonocy w kierunku zachodnim z szybkością około $1,4^\circ$ na stulecie. Na precesję nakłada się szereg znacznie mniejszych oscylacji eliptycznych osi nazywanych nutacjami (lub zbiorczo nutacją) z głównymi wyrazami o okresie 18,6 roku i amplitudzie $9,2''$, $1/2$ roku – $0,6''$, 13,7 doby – $0,1''$, 9,3 roku – $0,1''$ i 1 rok – $0,06''$.

Modelowanie precesji-nutacji jest ważnym zadaniem badawczym znajdującym się w sferze zainteresowań dwóch światowych organizacji naukowych – Międzynarodowej Unii Astronomicznej (*IAU – International Astronomical Union*) i Międzynarodowej Unii Geodezji i Geofizyki (*IUGG – International Union of Geodesy and Geophysics*). Zadanie polega na stworzeniu modelu matematycznego zjawiska, który byłby w możliwie największym stopniu zgodny z obserwacjami, a jednocześnie wynikał z praw fizyki. Pierwszy nowoczesny model spełniający powyższe kryteria składał się z dwóch czę-



Rysunek 1. Precesja-nutacja Ziemi – rezidua obserwacyjne VLBI w stosunku do modelu precesji IAU1976 i teorii nutacji IAU1980. Wykorzystano rozwiązanie EOP (IERS C04) dostępne pod adresem internetowym <http://hpiers.obspm.fr/eoppc/eop/eopc04>

² W rzeczywistości sprawa jest bardziej złożona, ponieważ punkty skorupy ziemskiej przemieszczają się względem siebie, np. wskutek ruchu płyt kontynentalnych.

ści – modelu precesji IAU1976 i teorii nutacji IAU1980. Ta ostatnia była teorią geofizyczną autorstwa Johna Wahra z USA [Wahr, 1981], obliczoną pod koniec lat 70. XX wieku dla modelu Ziemi sprężystej z ciekłym jądrem, ale bez oceanów i atmosfery. Jednak już w drugiej połowie lat 80. zauważono systematyczne rozbieżności rzędu 10 mas między przewidzianą modelem pozycją bieguna niebieskiego a pozycją zaobserwowaną przy pomocy techniki VLBI (rys. 1).

Podczas Zgromadzenia Generalnego IAU w 1994 roku (Haga) powołano, pod przewodnictwem Veronique Dehant z Belgii, wspólną grupę roboczą IAU/IUGG (*the IAU/IUGG Working Group on "Non-rigid Earth Nutation Theory"*), której zadaniem było opracowanie nowej teorii precesji-nutacji dla Ziemi deformowalnej. Prace grupy roboczej, w których uczestniczyło międzynarodowe grono badaczy w liczbie około 30 osób (w tym autor niniejszego opracowania), trwały 6 lat. Szczegóły dyskusji oraz wstępne wnioski opisane są w obszernym raporcie [Dehant i in., 1999]. Osiągnięto konsensus, w wyniku którego zaproponowano IAU nowy, znacznie lepiej dopasowany do obserwacji, model (por. rys. 1 i 2). Model został przyjęty podczas Zgromadzenia Generalnego IAU w 2000 roku (Manchester) jako **model precesji-nutacji IAU2000**, a następnie zatwierdzony przez Zgromadzenie Generalne IUGG w 2003 roku (Sapporo). Grupa robocza została uhonorowana w 2003 roku prestiżowym wyróżnieniem Unii Europejskiej – **Nagrodą Kartezjusza** – w uznaniu za wybitne osiągnięcie naukowe otrzymane w wyniku współpracy międzynarodowej.

Niniejsze opracowanie jest próbą syntetycznego opisu zjawiska precesji-nutacji Ziemi, ze szczególnym uwzględnieniem współczesnych metod obserwacji i modelowania. Zaczniemy od omówienia kinematyki zjawiska i stosowanych obecnie metod parametryzacji (p. 2). Następnie będą przedstawione wybrane aspekty dynamiczne (p. 3). Punkt 4 zawiera zwięzły opis modelu IAU2000. W końcowej części (p. 5) przedstawiono listę zagadnień i problemów, które w opinii autora opracowania wymagają kontynuacji badań.

Kinematyka zjawiska i współczesne metody parametryzacji

Opis matematyczny ruchu obrotowego Ziemi wymaga zdefiniowania dwóch układów współrzędnych – układu obracającego się wraz z Ziemią (TRS – *Terrestrial*

Reference System) oraz układu nieruchomego w przestrzeni (CRS – *Celestial Reference System*).

W pierwszym przypadku przyjmujemy, że:

- początkiem jest środek mas Ziemi O ,
- oś Oz jest możliwie bliska osi figury Ziemi i skierowana na północ,
- oś Ox leży w płaszczyźnie południka Greenwich, a oś Oy w płaszczyźnie południka $90^\circ E$.

Zgodnie z założeniem, Ziemia powinna być nieruchoma w układzie TRS. Spełnienie tego warunku staje się bardzo złożonym zagadnieniem, jeśli uwzględnimy obecność ciekłych i gazowych części planety oraz deformacje stałego płaszcza prowadzące do zmian wzajemnych odległości jego punktów rzędu decymetrów. Praktyczna realizacja układu TRS [IERS, 2004, chap. 4] polega na przypisaniu pewnemu zbiorowi fizycznych punktów na powierzchni Ziemi, związanych z instrumentami obserwacyjnymi, współrzędnych kartezjańskich. Dla każdego punktu uwzględnia się dwa typy przemieszczeń opisane modelami:

- 1) perturbacje krótkookresowe powodowane przez deformacje płytkowe Ziemi stałej, zmiany obciążeń oceanicznych i atmosferycznych skorupy (tzw. pływ biegunowy) oraz sezonowe przemieszczenia środka mas planety względem jej powierzchni;
- 2) liniowe zmiany współrzędnych, wynikające z obrotu płyt kontynentalnych.

Współrzędne punktów oraz rezidualne prędkości wyznacza się na podstawie obserwacji i przy założeniu, że ich zmiany czasowe nie zawierają wspólnej rotacji względem osi układu (warunek oznaczany skrótem NNR – *No Net Rotation*).

Początkiem fundamentalnego układu niebieskiego CRS jest barycentrum, czyli środek mas Układu Słonecznego, a kierunki osi wyznaczają pozycje na sferze niebieskiej odległych obiektów pozagalaktycznych, głównie tzw. kwazarów obserwowanych w zakresie fal radiowych przy pomocy techniki VLBI [IERS, 2004, chap. 2]. Tak zdefiniowany układ jest niewątpliwie najlepszą z możliwych obecnie realizacji układu inercjalnego. Dla potrzeb modelowania ruchu obrotowego Ziemi dokonuje się translacji układu CRS tak, aby stał się on układem geocentrycznym. Nie zmienia to równań ruchu obrotowego, natomiast w modelu precesji-nutacji pojawiają się niewielkie poprawki relatywistyczne – tzw. precesja geodezyjna $1,9''$ na stulecie oraz nutacja geodezyjna – głównie składowa roczna o amplitudzie $0,06$ mas.

W praktyce, dokonuje się dekompozycji perturbacji ruchu obrotowego Ziemi na składową równikową opisującą zmiany kierunku osi obrotu oraz składową osiową (spin) wyrażającą zmiany szybkości kątowej wokół chwilowej osi obrotu. Z kolei składową równikową wyraża się jako ruch osi w przestrzeni, czyli precesję-nutację, oraz zmiany kierunku osi względem powierzchni Ziemi, czyli ruch bieguna (ang. *polar motion*). Dekompozycję tę można wyrazić, rozpisując transformację między układem niebieskim CRS a układem ziemskim TRS w postaci zaproponowanej przez Brzezińskiego i Capitaine (1993)

$$[TRS] = R_2(-x_p) R_1(-y_p) R_3(\theta) \times R_1(-\delta Y) R_2(\delta X) NP_m(t) [CRS] \quad (1)$$

gdzie:

- $R_\alpha(\alpha)$ – macierz obrotu o kąt α wokół osi Ox_e układu,
- $NP_m(t)$ – macierz wyrażająca konwencjonalny model precesji-nutacji jako funkcję czasu t ,
- $\delta X, \delta Y$ – współrzędne umownego bieguna pośredniego CIP (*Celestial Intermediate Pole*) w układzie $[CRS]^* = NP_m(t) [CRS]$,
- $x_p, -y_p$ – współrzędne³ bieguna CIP w układzie TRS,
- θ – kąt obrotu wokół osi odpowiadającej biegunowi CIP.

Kąt θ wyraża w pierwszym rzędzie jednostajną dobową rotację $\Omega \cdot t$ z częstotliwością kątową $\Omega = 2\pi/\text{sd}$, gdzie sd (skrót – *sidereal day*) oznacza dobę gwiazdową równą 86 164,10 s (sekund SI). Do tego dodaje się model oraz poprawki obserwacyjne w postaci $\kappa \cdot \delta UT_1$, gdzie $\delta UT_1 = UT_1 - UTC$ jest różnicą między czasem uniwersalnym (słonecznym) a jednostajną skalą czasu atomowego, natomiast κ oznacza współczynnik konwersji czasu słonecznego na czas gwiazdowy wyrażony w jednostkach kątowych. Współrzędne bieguna $x_p, -y_p$ w układzie ziemskim są również wyrażane jako suma modelu i obserwacji, jednak w odróżnieniu od precesji-nutacji model stanowi niewielką część całości (poniżej 1 mas, podczas gdy całkowita amplituda zmian sięga 300 mas).

Zbiór $\{x_p, y_p, \delta UT_1, \delta X, \delta Y\}$ to tzw. parametry orientacji przestrzennej Ziemi (EOP – *Earth Orientation Parameters*), które są wyznaczone regularnie z pomiarów geodezji kosmicznej i udostępniane użytkownikom za

pośrednictwem IERS. Ta sama organizacja zajmuje się również opracowaniem prognoz dla parametrów EOP w celu zapewnienia jak najwyższej dokładności transformacji (1) dla przyszłych momentów czasu.

Zauważmy, że oś bieguna CIP, do której odnoszą się parametry EOP, nazwana przez nas wcześniej „osią obrotu Ziemi”, nie pokrywa się z osią wektora $\vec{\omega}$ chwilowej prędkości kątowej układu TRS. Z równania (1) wynika, że obserwacje w połączeniu z konwencjonalnym modelem opisują zmiany czasowe macierzy Q , transformującej układ CRS do układu TRS. Tymczasem z podręczników mechaniki klasycznej [np. Arnold, 1981, rozdz. 6] wiemy, że wektor $\vec{\omega}$ jest zdefiniowany przez 3 niezależne elementy antysymetrycznej macierzy $\dot{Q}Q^{-1}$, gdzie kropka oznacza różniczkowanie względem czasu. Wyznaczenie współrzędnych wektora prędkości kątowej wymaga zatem różniczkowania parametrów obserwacyjnych EOP. Odpowiednie formuły można znaleźć w pracy Brzezińskiego i Capitaine (1993). Pokazano tam również, że biegun CIP jest aproksymacją bieguna chwilowej prędkości kątowej, której dokładność zależy od częstotliwości zaburzenia.

Biegun CIP nie jest jednoznacznie zdefiniowany przez dekompozycję opisaną równaniem (1). Z elementarnych rozważań kinematycznych wynika, że dowolna (mała, aby zachować ważność aproksymacji liniowej) zmiana współrzędnych niebieskich $\delta X, \delta Y$ bieguna może być w pełni skompensowana przez odpowiednią zmianę jego współrzędnych $x_p, -y_p$ w układzie ziemskim. W celu uzyskania jednoznaczności należy określić, kiedy zaburzenie składowej równikowej obrotu Ziemi będziemy traktować jako nutację, a kiedy jako ruch bieguna. Zagadnienie, które jest dość proste z punktu widzenia kinematyki zjawiska, staje się trudnym do rozstrzygnięcia po uwzględnieniu wszystkich aspektów praktycznych, w tym specyfiki technik obserwacyjnych, międzynarodowych programów monitorowania zaburzeń rotacji itd. W wyniku dyskusji przyjęto definicję bieguna CIP zawartą w rezolucji B1.7 Zgromadzenia Generalnego IAU w 2000 roku (Manchester), która stwierdza, że precesja-nutacja obejmuje zaburzenia względem układu niebieskiego o okresach dłuższych niż 2 doby, a wszystkie inne zmiany składowej równikowej rotacji należy traktować jako ruch bieguna [IERS, 2004, sec. A7].

³ Znak „-” przed y_p bierze się stąd, że zgodnie z tradycją współrzędną y bieguna liczy się dodatnio w kierunku zachodnim, a zatem przeciwnie do kierunku osi Oy układu TRS.

Dynamika precesji-nutacji

Precesja-nutacja Ziemi powstaje w wyniku grawitacyjnego oddziaływania Księżyca, Słońca oraz – w niewielkim stopniu – planet na obracającą się Ziemię. Ścisłej mówiąc, chodzi o niewielką składową tej siły zależną od położenia punktu w bryle Ziemi, czyli tak zwaną siłę pływową. W praktyce oblicza się ją, odejmując od całkowitej siły w danym punkcie jej wartość w środku mas Ziemi, która jest odpowiedzialna za ruch orbitalny. Dodatkowym warunkiem powstania momentu sił jest spełnienie wymogu, aby rozkład mas w Ziemi nie był jednorodny, tzn. nie był wyłącznie funkcją odległości od środka mas. Tak jest istotnie, ponieważ na skutek rotacji Ziemia przyjęła kształt bliski elipsoidzie obrotowej lekko spłaszczonej na biegunach⁴. To spłaszczenie ma podstawowe znaczenie w modelowaniu precesji-nutacji, choć w precyzyjnych wyznaczeniach uwzględnia się również bardziej subtelne niejednorodności rozkładu mas w bryle Ziemi.

Izaak Newton był pierwszym, który opisał w swym fundamentalnym dziele *Philosophiae naturalis principia mathematica* z 1687 roku mechanizm fizyczny precesji. Co więcej, znając parametry precesji był w stanie wyciągnąć wnioski dotyczące kształtu Ziemi. Na podstawie wyprowadzonych przez siebie praw mechaniki i przy założeniu, że Ziemia ma kształt elipsoidy obrotowej wywnioskował, że ruch precesyjny powinien odbywać się w kierunku wstecznym (zgodnym z ruchem wskazówek zegara), o ile elipsoida jest spłaszczona na biegunach, i w kierunku przeciwnym, jeśli elipsoida jest wydłużona w kierunku biegunów. Już od czasów Hipparcha (II w. p.n.e.) astronomowie wiedzieli, że punkt równonocy porusza się po sferze niebieskiej w kierunku zachodnim, co Newton potraktował jako obserwacyjny dowód spłaszczenia Ziemi na biegunach [Ekman, 1993]. Newton zrobił krok dalej i dokonał oszacowania tempa precesji. Przy założeniu hydrostatycznej równowagi pod działaniem sił grawitacji własnej mas Ziemi i sił odśrodkowych związanych z jej rotacją policzył najpierw spłaszczenie Ziemi jako 1/231 (dla porównania, współczesne oszacowanie wynosi 1/298,25642), a następnie wykorzystał swą wiedzę dotyczącą sił pływowych od Księżyca i Słońca, otrzymując około 50"/rok, zatem wartość bliską współczes-

nym oszacowaniom precesji księżycowo-słonecznej wynoszącym 50,38"/rok.

Współczesny Newtonowi włoski astronom Cassini, pracujący w Paryżu, otrzymał z pomiarów geodezyjnych długości południka ujemne spłaszczenie Ziemi. Doprowadziło to do trwającej ponad 50 lat dramatycznej kontrowersji. Dopiero ekspedycja naukowa zorganizowana w latach 1735–1743 przez Francuską Akademię Nauk potwierdziła ostatecznie, na podstawie porównania długości stopnia południka w Peru i w Laponii, słuszność tezy Newtona.

Newton przewidział również istnienie nutacji, ale dopiero w 1748 roku Bradley potwierdził obserwacyjnie istnienie jej głównej składowej o okresie 18,6 roku. Dalszy postęp w modelowaniu precesji-nutacji przy założeniu, że Ziemia jest bryłą sztywną, przyniósł prace matematyków d'Alamberta, Eulera i Laplace'a. Pierwszą próbę zbadania zależności precesji i nutacji Ziemi od jej wewnętrznej budowy podjął Hopkins w 1839 roku. Założył on istnienie ciekłego jądra we wnętrzu Ziemi i na podstawie obserwowanego tempa precesji oszacował miąższość płaszczka. Hough (1895) stwierdził, że wzajemne oddziaływanie płaszczka i ciekłego jądra prowadzi do rezonansu, który może zaburzać nutację. Rezonans ten, oznaczany skrótem FCN (*Free Core Nutation*), odgrywa istotną rolę we współczesnych modelach nutacji (patrz p. 4). Elegancki opis nutacji Ziemi sztywnej z ciekłym jądrem zaproponował Poincaré w 1910 roku. Bardziej złożone modele budowy Ziemi, uwzględniające sprężystość płaszczka oraz istnienie stałego jądra wewnętrznego, wykorzystywali później w swoich pracach dotyczących nutacji m.in. Jeffreys, Molodensky i Wahr. Najnowszy model precesji-nutacji, opracowany przez Mathewsa i innych (2002), będzie opisany bardziej szczegółowo w następnym punkcie.

Moment sił działający na obracającą się Ziemię można wyrazić jednoznacznie jako funkcję potencjału pływowego u , który opisuje siły zewnętrzne, oraz pola grawitacyjnego Ziemi U odzwierciedlającego rozkład mas w bryle naszej planety. Obie te wielkości są na ogół przedstawiane w postaci rozwinięcia w szereg harmonik kulistych

$$U = \sum_{e=0}^{\infty} \sum_{j=0}^e U_{ej}, \quad u = \sum_{e=0}^{\infty} \sum_{j=0}^e u_{ej} \quad (2)$$

⁴ Według najnowszych wyznaczeń średni promień równikowy Ziemi wynosi $a = 6\,378\,136,6 \pm 0,1$ m, a promień biegunowy $b = 6\,356\,751,9 \pm 0,1$ m, co daje różnicę około 21 km.

Tabela 1

Perturbacje lunisolarne w składowej równikowej ruchu obrotowego Ziemi. Zestawienie uwzględnia wszystkie składowe, dla których całkowity efekt przekracza poziom 0,1 μs . Ostatnia kolumna pokazuje oszacowanie sumy wartości bezwzględnych amplitud większych niż 0,01 μs

Potencjał pływowy	Geopotencjał	Nutacja	Ruch bieguna	Efekt sumaryczny [μs]
$u_{\ell 1}$ dla $\ell = 2, 3, \dots$	$U_{\ell 0}$ dla $\ell = 2, 3, \dots$	długookresowa*	dobowy wsteczny	nutacja $>10^7$ + + precesja 50,29"/rok
u_{30}	U_{31}	dobowa prosta	długookresowy*	91,3
u_{40}	U_{41}			1,0 + dryft 5,7 $\mu\text{s}/\text{yr}$
u_{21}	U_{22}	półdobowa prosta	dobowy prosty*	51,6
u_{31}	U_{32}			0,2
u_{32}	U_{32}	1/3 doby prosta	półdobowy prosty*	0,1
u_{32}	U_{31}	dobowa wsteczna	półdobowy wsteczny*	0,8
u_{33}	U_{32}	półdobowa wsteczna	1/3 doby wsteczny*	0,1

w którym ℓ oznacza stopień a j rząd składowej harmonicznej. Znajomość współczynników rozwinięcia (2) wystarcza do obliczenia momentu sił zewnętrznych działających na Ziemię. Na podstawie momentu sił można obliczyć za pomocą równu obrotowym Ziemi. Jest to proste zadanie, opisane w podręcznikach mechaniki klasycznej, o ile założymy, że nasza planeta jest bryłą sferoidalną. Uwzględnienie rektoryki Ziemi jest za zadaniem nie zwykłego złożonym i stałymi grotami przy czym ograniczenia do kadencji współczesnych modeli precesji-nutacji.

Tabela 1 zawiera zestawienie wszystkich składowych zaburzeń składowych równikowych ruchów nutacji przekraczających poziom 0,1 mikrosundy tuku (μs – *micro-arc second*). Po prostu ogólne wierzenie przedstawia efekt oddziaływania składowej $u_{\ell j}$ potencjału pływowego na małe, której rozkład w bryle Ziemi wyraża się przez składową $U_{\ell j}$ geopotencjału, po której to gra w bryle Ziemi. Każdy wiersz wyraża się w postaci dwuwymiarowego rozwinięcia trygonometrycznego względem czasu, o częstotliwościach bliskich określonej krotkości częstotliwości dobowej Ziemi względem gwiazd $k\Omega$; dla $k = 0$ mówi o zaburzeniu dobowym, dla $k = -1/+1$ o dobowym wstecznym/prostym, dla $k = -2/+2$ o półdobowym wstecznym/prostym itd. Połowem niekończącym się i 4 tabelki po prostu, że wartość k , a zatem częstotliwość zaburzenia, zależy od tego, czy jest to nutacja, czy ruch bieguna. Gwiazdki zaznaczają opcję zgodną z przyję-

tą niedawno definicją bieguna CIP (zob. opis w ostatnim akapicie poprzedniego punktu).

Dominujący efekt w zestawieniu z tabeli 1, przedstawiony w pierwszym wierszu, obejmuje precesję i długookresową nutację opisywane przez konwencjonalne modele IAU. Efekt precesyjno-nutacyjny jest wynikiem oddziaływania składowych tesseralnych $u_{\ell 1}$ (dobowych) potencjału pływowego na składowe strefowe $U_{\ell 0}$ potencjału grawitacyjnego Ziemi, dla $\ell \geq 2$, przy przeważającym udziale (prawie 99%) harmonik drugiego stopnia ($\ell = 2$). Składową U_{20} geopotencjału opisuje współczynnik J_2 wyrażający spłaszczenie elipsoidy bezwładności Ziemi, czyli współczynnik dynamiczny figury Ziemi, który jest jednym z fundamentalnych parametrów geodezyjnych [IERS, 2004, table 1.1]. Precesja i długookresowa nutacja Ziemi są proporcjonalne do J_2 , a zatem obserwacja tego zjawiska dostarcza informacji o spłaszczeniu Ziemi. Jak wspomniano w punkcie 3, związek między tempem precesji i spłaszczeniem Ziemi opisał już w swoim fundamentalnym dziele Newton.

Współcześnie najczęściej stosowanym sposobem liczenia modelu nutacji jest metoda dwustopniowa, polegająca na wykonaniu najpierw obliczeń dla Ziemi sferoidalnej, a potem pomnożeniu amplitud przez tzw. funkcję przenoszenia (ang. *transfer function*), obliczoną z uwzględnieniem realnej budowy i deformacji Ziemi. W ten sposób rozdziela się aspekty astronomiczne, które ograniczają się do pierwszego kroku, od aspektów geofizycznych. W przypadku precesji występuje brak

zależności rozwiązań od deformacji Ziemi, zatem procedura ogranicza się do pierwszego kroku. Tak też skonstruowany jest model IAU2000, który będzie opisany poniżej.

Model IAU2000

Precesja-nutacja Ziemi sztywnej

W latach 1994–2000 powstały 3 szeregi nutacyjne o wysokiej precyzji, obliczone przy założeniu modelu Ziemi sztywnej:

- SMART97 [Bretagnon i in., 1998],
- REN2000 [Souchay i in., 1999],
- RDAN97 [Roosbeek i Dehant, 1998].

Ich autorzy uwzględnili wszystkie efekty, których wielkość przekracza poziom 1 μ s. Jedną z ważnych konkluzji Grupy Roboczej IAU/IUGG „*Non-rigid Earth Nutation Theory*” było stwierdzenie, że precyzja szeregów nutacyjnych dla Ziemi sztywnej jest dostatecznie wysoka w porównaniu z dokładnością obserwacji. Różnice w dziedzinie czasu między wymienionymi trzema rozwiązaniami sięgają kilkudziesięciu μ s, a zatem są zdecydowanie większe niż oszacowana ich wewnętrzna dokładność. Z drugiej strony różnice te są jednak niższe od poziomu dokładności obserwacyjnych, zatem niemożliwe jest stwierdzenie, który z szeregów jest najlepszy. Widać to również w rezolucji B1.6 IAU dotyczącej przyjęcia modelu IAU2000 [IERS, 2004, sec. A6], która wymienia wszystkie trzy szeregi. Procedura opisana w *IERS Conventions 2003* [IERS, 2004, chap. 5] i dostępna poprzez witrynę internetową *IERS Convention Center*, wykorzystuje szereg REN2000, dodatkowo przeskalowany przez czynnik 1,000 012 249 w celu uwzględnienia zmiany sptaszczenia dynamicznego wynikającej z obserwacyjnej poprawki do precesji równika ziemskiego.

Model budowy Ziemi

Policzenie funkcji przenoszenia dla modelu nutacji wymagało przyjęcia modelu budowy wewnętrznej Ziemi. Punktem wyjścia był model PREM (*Preliminary Reference Earth Model*) Dziewońskiego i Andersona (1981), który jest modelem warstwowym skonstruowanym przy założeniu hydrostatycznej równowagi. Model PREM wymagał modyfikacji w celu otrzymania zgodności obliczonych i wyznaczonych z obserwacji parametrów, takich jak: główne momenty bezwładności, dynamiczna eliptyczność Ziemi oraz sptaszczenie

powierzchni granicznych, a przede wszystkim tej między płaszczem a jądrem. Odstępstwa od założenia równowagi hydrostatycznej można było wytłumaczyć procesami konwekcji w lepkosprężystym płaszczu [Dehant i in., 1999].

Sptaszczenie granicy jądro–płaszcz obliczone na podstawie okresu rezonansowego FCN otrzymanego z wyznaczonych techniką VLBI parametrów nutacyjnych różni się w istotnym stopniu od sptaszczenia obliczonego przy założeniu hydrostatycznej równowagi. Zgodność z obserwacjami otrzymuje się poprzez wzrost sptaszczenia odpowiadający zwiększeniu różnicy między promieniem równikowym a promieniem biegunowym jądra o około 500 m. Tę różnicę otrzymuje się przy założeniu, że sprzężenie między ciekłym jądrem i płaszczem wynika wyłącznie z sił bezwładności. Jeśli dodatkowo uwzględnimy oddziaływania elektromagnetyczne na granicy jądro–płaszcz, różnica redukuje się do około 380 m [Mathews i in., 2002]. Dla porównania, precyzja wyznaczeń kątów nutacyjnych z pomiarów VLBI przekłada się na 2,5 m w sptaszczeniu granicy jądro–płaszcz.

Funkcja przenoszenia

Grupa Robocza „*Non-rigid Earth Nutation Theory*” zarekomendowała IAU model funkcji przenoszenia MHB2000 [Mathews i in., 2002]. Kilka parametrów modelu zostało dopasowanych do obserwacji VLBI, niemniej sposób parametryzacji jest zgodny z prawami fizyki, co z kolei umożliwia interpretację fizyczną rezultatów. W uzasadnieniu stwierdzono: *W chwili obecnej jest to najbardziej zgodny z obserwacjami model skonstruowany na prawach geofizyki*. Model MHB2000 uwzględnia: elektromagnetyczne sprzężenie na granicach płaszcz–ciekłe jądro oraz ciekłe jądro–stałe jądro wewnętrzne, oddziaływanie płytów oceanicznych, lepkość płaszczka, efekty atmosferyczne i wreszcie poprawki do globalnego dynamicznego sptaszczenia Ziemi oraz geometrycznego sptaszczenia jądra.

Punktem wyjścia jest układ dynamiczny złożony z czterech liniowych równań różniczkowych dla zmiennych zespolonych. Rozwiązania układu szukano w dziedzinie częstotliwości. Aby opisać rozwiązanie zgodnie z oryginałem, oznaczymy częstotliwość kątową składowej nutacyjnej, odniesioną do układu TRS, przez $\sigma = \bar{\sigma}\Omega$, gdzie $\bar{\sigma}$ jest bezwymiarowym parametrem, który będziemy w dalszym ciągu nazywać częstotliwością. Jeśli częstotliwość nutacji chcemy odnieść do układu nieruchomego w przestrzeni, wówczas $\bar{\sigma}$ należy zastą-

pić przez $\tau = \bar{\sigma} + 1$. Niech dalej $\tilde{\eta}(\bar{\sigma})$ oznacza amplitudę nutacji o częstotliwości $\bar{\sigma}$ (tzn. składowa harmoniczna nutacji jest zespoloną sinusoidą $n(t) = \tilde{\eta}(\bar{\sigma}) \exp(i \bar{\sigma} \Omega t)$, $i = \sqrt{-1}$), a $\tilde{\eta}_r(\bar{\sigma})$ jest wartością amplitudy dla modelu Ziemi sztywnej (indeks r jest skrótem od ang. *rigid*). Funkcja przenoszenia przyjmuje następującą postać

$$T(\bar{\sigma}; e) = \frac{\tilde{\eta}(\bar{\sigma})}{\tilde{\eta}_r(\bar{\sigma})} = R + R'(1 + \bar{\sigma}) + \sum_{\alpha=1}^4 \frac{R_\alpha}{\bar{\sigma} - \bar{\sigma}_\alpha} \quad (3)$$

gdzie:

$e = (C - A)/A$ – tzw. dynamiczna eliptyczność Ziemi,
 $A = 1/2(I_{xx} + I_{yy})$ – średni równikowy moment bezwładności,

$C = I_{zz}$ – moment biegunowy,

R, R', R_α dla $\alpha = 1, \dots, 4$ – zespolone współczynniki.

Częstotliwości $\bar{\sigma}_\alpha$ są częstotliwościami drgań własnych w składowej równikowej rotacji Ziemi: kotysanie Chandlera (CW – *Chandler Wobble*), swobodna nutacja jądra wsteczna (FCN) oraz prosta (PFCN – *Prograde Free Core Nutation*), swobodne kotysanie stałego jądra wewnętrznego (ICW – *Free Wobble of the Inner Core*). Podobnie jak współczynniki $R, \bar{\sigma}_\alpha$ są wielkościami zespolonymi, w których część urojona wyraża rozpraszanie energii. Jedną z konsekwencji jest to, że w wyrażeniu (3) nigdy nie ma osobliwości ($\bar{\sigma}$ jest wielkością rzeczywistą, a zatem nie może być równa $\bar{\sigma}_\alpha$).

Funkcja przenoszenia $T(\bar{\sigma}; e)$ przyjmuje wartość 1 dla $\bar{\sigma} = -1$, czyli dla $\tau = 0$, co oznacza, że precesja jest zawsze taka sama jak dla Ziemi sztywnej, o ile jest zachowana wartość dynamicznej eliptyczności e . Dodatkowo, $T = 0$ dla $\bar{\sigma} = e$. Istnieją również inne postacie równania (3) uwzględniające wymienione dwa warunki [patrz Mathews i in., 2002].

Interesujące jest porównanie wpływu poszczególnych członów rezonansowych na funkcję przenoszenia T wyrażoną równaniem (3). Wypiszmy w tym celu numeryczną wartość parametrów:

$$\begin{aligned} \text{CW: } & \bar{\sigma}_1 = 0,002601 - i0,0001361489, \\ & R_1 = -(5,46425 + i0,79322) \times 10^{-4}; \\ \text{FCN: } & \bar{\sigma}_2 = -1,0023181 + i0,0000250, \\ & R_2 = -(1,13686 + i0,02555) \times 10^{-4}; \\ \text{PFCN: } & \bar{\sigma}_3 = -0,99903 + i0,000078, \\ & R_3 = (3,62650 + i1,37204) \times 10^{-7}; \\ \text{ICW: } & \bar{\sigma}_4 = 0,000413471 + i0,000000318, \\ & R_4 = -(4,32089 + i0,30306) \times 10^{-8}. \end{aligned}$$

Porównanie współczynników R_α pokazuje, że zdecydowanie ważniejsze są wyrazy związane z rezonansem Chandlera i FCN. Współczynnik R_2 jest co prawda około 5 razy mniejszy niż R_1 , jednakże różnica ta może być w pełni skompensowana przez mianownik. Wynika to stąd, że dla wyrazów nutacyjnych $\bar{\sigma} \approx -1$, czyli w przypadku FCN ($\alpha = 2$) mianownik może przyjmować bardzo małe wartości. Dla rezonansu Chandlera ($\alpha = 1$) mianownik jest bliski -1 i niewiele się zmienia w paśmie nutacyjnym.

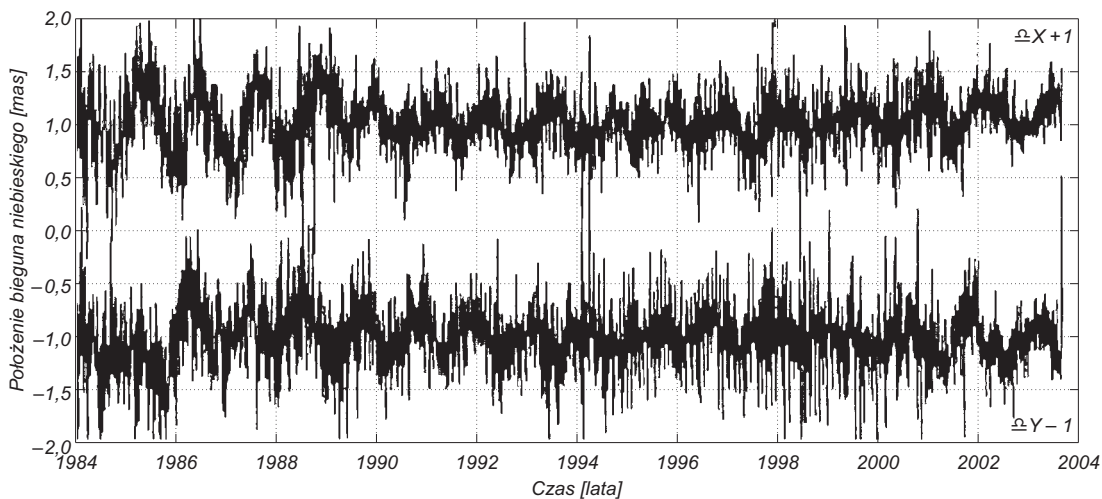
Model nutacji jest przedstawiany w formie szeregów trygonometrycznych dla współrzędnych X, Y biegunu CIP w układzie niebieskim⁵. Każdy obserwowany wyraz nutacji jest złożeniem dwóch ruchów jednostajnych po okręgu o tym samym okresie lecz przeciwnych kierunkach, co daje ruch eliptyczny. Stosując wprowadzone wcześniej oznaczenia mamy nutację wsteczną (ang. *retrograde*) o częstotliwości $-|\tau|$, której kierunek jest zgodny z ruchem wskazówek zegara, i nutację prostą (ang. *prograde*) o częstotliwości $|\tau|$ i kierunku przeciwnym. Ich wspólny okres w układzie niebieskim wynosi $|\tau|^{-1}$ dob gwiazdowych. Ponieważ funkcja przenoszenia (3) przyjmuje różne wartości dla τ i $-\tau$, zatem dla jej zastosowania konieczne jest rozłożenie każdego wyrazu nutacyjnego na sumę przeciwnych ruchów okrężnych [Bizouard i in., 1998, appendix C; IERS, 2004, chap. 5].

Na zakończenie zauważmy, że ponieważ funkcja przenoszenia przyjmuje wartości zespolone, więc zmienia ona nie tylko amplitudę składowej nutacyjnej Ziemi sztywnej, ale również jej fazę.

Inne efekty

Opisana powyżej funkcja przenoszenia wyraża reakcję Ziemi stałej z ciekłym jądrem na moment sił zewnętrznych. Oszacowania pokazują, że ocean i atmosfera mają niewielki, choć dobrze mierzalny wpływ na precesję-nutację Ziemi. Największe jest oddziaływanie pływów oceanicznych – rzędu 1 mas. Mathews i in. (2002) wykorzystali model pływu oceanicznego i odpowiadających mu zmian momentu pędu oceanu obliczony przez Chao i in. (1996) do wprowadzenia poprawki w funkcji przenoszenia MHB2000. Niepływowe zaburzenia nutacji pochodzące z układu atmosfera-ocean, związane z dobowym cyklem termicznym, są na

⁵ Ścisłej mówiąc, są to szeregi Poissona, w których zarówno amplituda, jak i argument funkcji trygonometrycznej są wielomianami. Dla interwałów czasu rzędu lat rozwinięcia te można jednak traktować jako szeregi trygonometryczne.



Rysunek 2. Precesja-nutacja Ziemi – rezidua obserwacyjne VLBI w stosunku do modelu precesji IAU1976 i teorii nutacji IAU1980. Wykorzystano rozwiązanie EOP (IERS C04) dostępne pod adresem internetowym <http://hpiers.obspm.fr/eoppc/eop/eopc04>

poziomie 0,1 mas [Bizouard i in., 1998; Brzeziński i in., 2004]. W modelu MHB2000 uwzględniono je poprzez wprowadzenie do amplitudy i fazy nutacji rocznej prostej poprawki empirycznej nazwanej przez autorów *Sun-synchronous*.

Model IAU2000 opisuje zjawisko precesji astronomicznej i długookresowej nutacji przedstawiony w pierwszym wierszu tabeli 1. Składowe wymienione w drugim i dalszych wierszach tabeli, nazywane często *subdiurnal nutations*, zostały opisane w formie oddzielnego modelu [Brzeziński i Mathews, 2003], który jest elementem realizacji bieguna CIP [IERS, 2004, tab. 5.1].

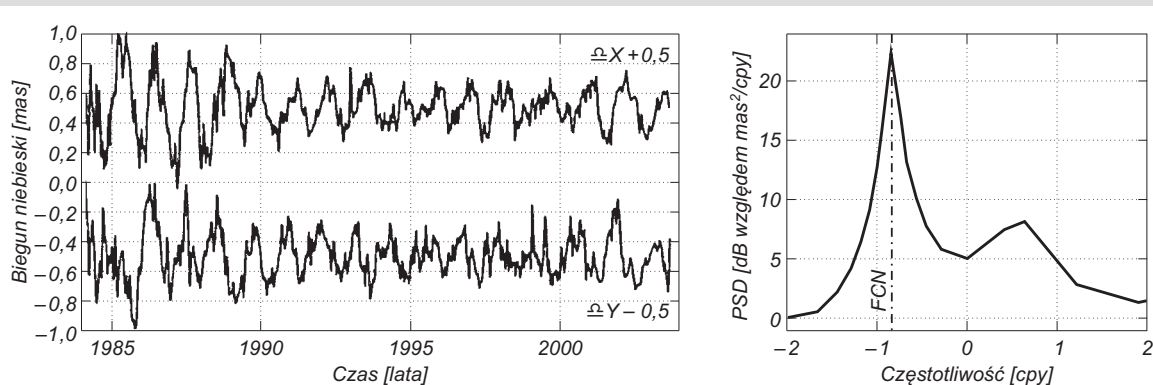
Modelowanie i monitorowanie parametrów nutacyjnych – co dalej?

Przyjęcie przez IAU i zatwierdzenie przez IUGG modelu precesji-nutacji IAU2000 było zwieńczeniem pewnego etapu prac badawczych i wysiłku organizacyjnego, ale z całą pewnością nie oznacza zamknięcia tematu. Poniżej przedstawiony będzie wykaz problemów, które w opinii autora niniejszego opracowania wymagają kontynuacji badań. W celu zapoznania się z innymi zagadnieniami patrz [Dehant i Brzeziński, 2005; Dehant i in., 2003; Brzeziński, 2004].

Na pierwszym miejscu należy wymienić efekty atmosferyczne i oceaniczne. Bizouard i in. (1998) pokazali, że nieregularne zmiany momentu pędu atmosfery (AAM – *Atmospheric Angular Momentum*), tzn. po odjęciu modelu harmonicznego, pobudzają zmiany kątów nutacyjnych na poziomie 0,1 mas, czyli zbliżonym do szacowanej precyzji nowego szeregu precesyjno-nutacyjnego. Zrozumienie i modelowanie tych nieregularnych zaburzeń jest zatem niezbędnym elementem działań, których celem byłaby dalsza poprawa precyzji standardowego modelu. To z kolei można osiągnąć dzięki stworzeniu i regularnemu wyznaczaniu modelu zmian czasowych momentu pędu układu atmosfera–ocean. Model ten powinien:

- 1) mieć dużą rozdzielczość czasową (okres próbkowania 6 godzin lub krótszy),
- 2) być wewnętrznie zgodny, czyli te same globalne pola zmian ciśnienia atmosferycznego i prędkości wiatrów powinny być wykorzystywane do wyznaczania AAM oraz do pobudzania modelu globalnej cyrkulacji oceanu, który z kolei dostarcza danych do liczenia momentu pędu oceanu (OAM – *Oceanic Angular Momentum*),
- 3) wykorzystywać jak największą ilość obserwacji.

Od pewnego czasu dysponujemy regularnymi wyznaczeniami szeregów czasowych AAM z okresem



Rysunek 3. Rezidua nutacyjne z rysunku 2 po odjęciu poprawek empirycznych do modelu IAU2000 i wygładzeniu. Po prawej stronie pokazano oszacowanie gęstości widmowej (PSD – *Power Spectral Density*) funkcji $P = \delta X + i\delta Y$. Częstotliwość jest wyrażona w cyklach na rok (cpy – *cycles per year*) [Brzeziński i Kosek, 2004].

próbki 6 godzin. Przykładem jest jednorodny szereg NCEP/NCAR zaczynający się w 1948 roku i uaktualniany na bieżąco, który został wykorzystany w pracy Bizouarda i innych (1998). W przypadku szeregów OAM, które opisują dynamiczną reakcję oceanu na oddziaływanie atmosfery, sytuacja jest znacznie gorsza. Wyznaczenia o dużej rozdzielczości czasowej mają ciągle jeszcze charakter eksperymentalny i wykorzystują uproszczone modele dynamiki oceanu. Oszacowania dokonane w pracy Brzezińskiego i innych (2004) wskazują, że niepływowe zmiany w oceanie mogą mieć istotny wpływ na nutację Ziemi. Porównanie z obserwacjami VLBI w przypadku nutacji rocznej prostej pokazało jednak, że dodanie OAM do AAM pogarsza zgodność. Istnieje potrzeba dalszych prac mających na celu wiarygodne wyznaczenie OAM z wysoką rozdzielczością czasową.

Znacznie prostszym zadaniem, które powinno być wykonane w najbliższej przyszłości, jest policzenie na nowo modelu pływowych zmian momentu pędu oceanu. Model z pracy Chao i innych (1996) wykorzystany do konstrukcji funkcji przenoszenia MHB2000 wydaje się być zdecydowanie przestarzały. Obecnie dysponujemy wielokrotnie dłuższymi ciągami obserwacji zmian powierzchni oceanu, otrzymanymi z altimetrii satelitarnej. Wykorzystanie tych danych z całą pewnością poprawi wiarygodność modeli pływu oceanicznego i odpowiadających mu zmian OAM.

Istotnym problemem jest modelowanie swobodnej nutacji jądra FCN, pseudoharmonicznego sygnału

obecnego w reziduach kątów nutacyjnych wyznaczonych techniką VLBI (rys. 2 i 3). Całkowanie funkcji gęstości widmowej (rys. 3) pokazuje, że oscylacja FCN wnosi ponad 60% całkowitej wariacji przedstawionego na rysunku 3 szeregu czasowego. Ważne jest wyznaczenie parametrów FCN, okresu T i współczynnika dobroci Q na podstawie przebiegu sygnału oraz zdefiniowanie źródła jego pobudzenia. Te wyznaczenia są niezwykle cennym źródłem informacji na temat kształtu granicy między płaszczem a ciekłym jądrem oraz oddziaływań zachodzących między tymi dwoma ośrodkami. Wyznaczanie mechanizmu pobudzenia swobodnego sygnału może być z kolei istotnym elementem modelowania globalnej cyrkulacji atmosfery i oceanu w zakresie częstotliwości dobowych. Oprócz badań interpretacyjnych należy również tworzyć modele matematyczne, które mogłyby być wykorzystywane zarówno do opisu *a posteriori* przebiegu FCN, jak i do prognozowania przyszłych wartości [patrz Brzeziński i Kosek, 2004].

Obecność składowych nieregularnych w obserwacjach nutacji, takich jak sygnał FCN oraz nutacja roczna o zmieniającej się w czasie amplitudzie, wskazuje na konieczność dalszego monitorowania zmian czasowych kątów nutacyjnych. Model IAU2000 wraz z poprawkami empirycznymi zapewniają najwyższą dokładność transformacji między układem związanym z Ziemią i układem niebieskim. Prowadzone regularne obserwacje nutacji (rys. 2) uzupełnione komplementarnymi danymi geofizycznymi stanowią również niezwykle cenny materiał do badań interpretacyjnych.

Na zakończenie należy wspomnieć o możliwości wykorzystania bogatych doświadczeń uzyskanych przy modelowaniu precesji i nutacji Ziemi do badania innych ciał Układu Słonecznego. Proponowane są misje satelitarne, które pozwalałyby wyznaczać parametry ruchu obrotowego innych planet. Porównanie obserwacji z modelem nutacji planety dałoby możliwość „zajrzenia do jej wnętrza”, np. stwierdzenie, czy posiada ona ciepłe jądro. Zaawansowane prace studialne były przeprowadzone dla Marsa [patrz np. Dehant i in., 2000].

Podziękowania:

Praca naukowa finansowana ze środków Ministerstwa Nauki i Informatyzacji jako projekt badawczy Nr 5 T12E 039 24.

Literatura

- Arnold W.I. (1981). *Metody matematyczne mechaniki klasycznej*. Państwowe Wydawnictwo Naukowe.
- Bizouard Ch., Brzeziński A. and Petrov S. (1998). *Diurnal Atmospheric Forcing and Temporal Variations of the Nutation Amplitudes*. "Journal of Geodesy", **72**, 561–577.
- Bretagnon P., Francou G., Rocher P. and Simon J.-L. (1998). *SMART97: A New Solution for the Rotation of the Rigid Earth*. "Astron. Astrophys.", **329**, 329–338.
- Brzeziński A. (2004). *Nowy model precesyjno-nutacyjny*. Instytut Geodezji i Kartografii, Seria Monograficzna Nr 10 „Nowe obowiązujące niebieskie i ziemskie systemy i układy odniesienia oraz ich wzajemne relacje”, red. J. Kryński, IGiK, Warszawa, 145–161.
- Brzeziński A. and Capitaine N. (1993). *The Use of the Precise Observations of the Celestial Ephemeris Pole in the Analysis of Geophysical Excitation of Earth Rotation*. "J. Geophys. Res.", **98**, 6667–6675.
- Brzeziński A. and Kosek W. (2004). *Free Core Nutation: Stochastic Modeling Versus Predictability*. "Proc. Journées Systèmes de Référence Spatio-Temporels 2003", eds. A. Finkelstein and N. Capitaine, Inst. of Applied Astronomy of the Russian Acad. of Sciences, St. Petersburg, 99–106.
- Brzeziński A. and Mathews P.M. (2003). *Recent Advances in Modelling the Lunisolar Perturbation in Polar Motion Corresponding to High Frequency Nutation: Report on the Discussion of the IAU Comm. 19 WG on Nutation*. "Proc. Journées Systèmes de Référence Spatio-Temporels 2002", eds. N. Capitaine and M. Stavinschi, Ars Docendi, Paris, 101–108.
- Brzeziński A., Ponte R.M. and Ali A.H. (2004). *Non-tidal Oceanic Excitation of Nutation and Diurnal/semidiurnal Polar Motion Revisited*. "J. Geophys. Res.", **109**, No. B11, doi: 10.1029/2004JB003054.
- Chao B.F., Ray R.D., Gipson M.J., Egbert G.D. and Ma Ch. (1996). *Diurnal/semidiurnal Polar Motion Excited by Oceanic Tidal Angular Momentum*. "J. Geophys. Res.", **101**, No. B9, 20,151–20,163.
- Dehant V. et al. (1999). *Considerations Concerning the Non-rigid Earth Nutation Theory*. "Cel. Mech. Dynamical Astron.", **72**, 245–310.
- Dehant V. and Brzeziński A. (2005). *Working Group on Nutation*. "Proc. XXV IAU General Assembly", ed. O. Engvold, "Transactions of the International Astronomical Union", Vol. XXVB, Astronomical Society of the Pacific, Provo, USA, w druku.
- Dehant V., Feissel-Vernier M., de Viron O., Ma C., Yseboodt M. and Bizouard Ch. (2003). *Remaining Error Sources in the Nutation at the Sub-milliarsecond Level*. "J. Geophys. Res.", **108**, doi: 10.1029/2002JB001763.
- Dehant V., Van Hoolst T. and Defraigne P. (2000). *Comparison between the Nutations of the Planet Mars and the Nutations of the Earth*. "Surveys in Geophysics", **21**, 89–110.
- Dziwowski A.D. and Anderson D.L. (1981). *Preliminary Reference Earth Model*. "Phys. Earth Planet. Inter.", **25**, 297–356.
- Ekman M. (1993). *A Concise History of the Theories of Tides, Precession-nutation and Polar Motion (from Antiquity to 1950)*. "Surveys in Geophysics", **14**, 585–617.
- IERS (2004). *IERS Conventions 2003*, eds. D. McCarthy and G. Petit, *IERS Technical Note No. 32*, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main (wersja elektroniczna dostępna bezpłatnie pod adresem <http://www.iers.org/iers/products/conv/>).
- Mathews P.M., Herring T.A. and Buffet B.A. (2002). *Modeling of Nutation-precession: New Nutation Series for Nonrigid Earth, and Insights into the Earth's Interior*. "J. Geophys. Res.", **107**, doi: 10.1029/2001JB000390.
- Roosbeek F. and Dehant V. (1998). *RDAN97: An Analytical Development of Rigid Earth Nutation Series Using the Torque Approach*. "Celest. Mech. Dynamical Astron.", **70**, 215–253.
- Souchay J., Loysel B., Kinoshita H. and Folgueira M. (1999). *Corrections and New Developments in Rigid Earth Nutation Theory: III. Final Tables "REN-2000" Including Crossed-nutation and Spin-orbit coupling effects*. "Astron. Astrophys.", **318**, 639–652.
- Wahr J.M. (1981). *The Forced Nutations of an Elliptical, Rotating, Elastic and Oceanless Earth*. "Geophys. J. R. Astr. Soc.", **64**, 705–727.

Abstract

Precession-nutation is a spatial motion of the Earth's axis caused by the lunisolar torque on the rotating planet. This paper gives a historical overview and describes recent advances in observation and modelling of the precession-nutation. Emphasis is on how the observations of this phenomenon have been used to infer about the shape, the internal constitution and the rheological properties of the Earth. We start from the intro-

duction (Sec. 1), then discuss the kinematical aspects of Earth rotation and the modern methods of its parameterization (Sec. 2). Sec. 3 describes the dynamics of precession-nutation. In Sec. 4 we present basic properties of the model IAU2000 which has been recently adopted as an international standard. The last part, Sec. 5, outlines a list of problems that, in the author's opinion, need further investigation.

Słowa kluczowe

ruch obrotowy Ziemi, precesja-nutacja, globalne układy odniesienia

Profesor Aleksander Brzeziński specjalista w zakresie teorii ruchu obrotowego Ziemi, aktywny uczestnik grup roboczych organizowanych przez międzynarodowe unie naukowe (Astronomiczną IAU oraz Geodezji i Geofizyki IUGG). Profesor na Wydziale Geodezji i Kartografii Politechniki Warszawskiej oraz w Centrum Badań Kosmicznych PAN w Warszawie. Autor około 100 publikacji naukowych z zakresu teorii ruchu obrotowego Ziemi do prasy fachowej oraz opracowań zbiorowych, współautor książki *The Earth and its Rotation: Low Frequency Geodynamics*. Jest m.in. członkiem: Międzynarodowej Unii Astronomicznej, Międzynarodowej Asocjacji Geodezyjnej, Global Geodetic Observing System, Komitetu Badań Kosmicznych i Satelitarnych PAN, Komitetu Narodowego PAN ds. Współpracy w Międzynarodowej Unii Geodezji i Geofizyki, Towarzystwa Naukowego Warszawskiego oraz Societas Humboldtiana Polonorum. Laureat wielu nagród naukowych Polskiej Akademii Nauk, otrzymał Nagrodę Unii Europejskiej The EU Descartes Prize 2003 oraz Nagrodę im. Mikołaja Kopernika Fundacji Miasta Krakowa za osiągnięcia w dziedzinie geodezji w latach 2000–2004.

Zapach światła

Na podstawie odczytu wygłoszonego w dniu 17 marca 2005 roku

Kazimierz Brudzewski

Wydział Chemiczny Politechniki Warszawskiej

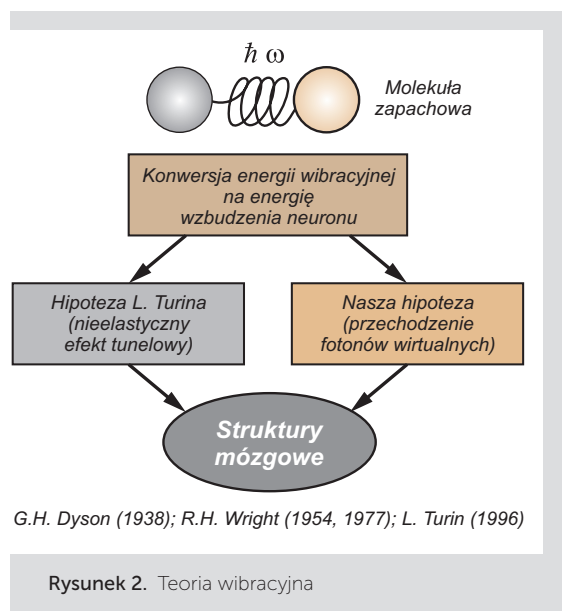
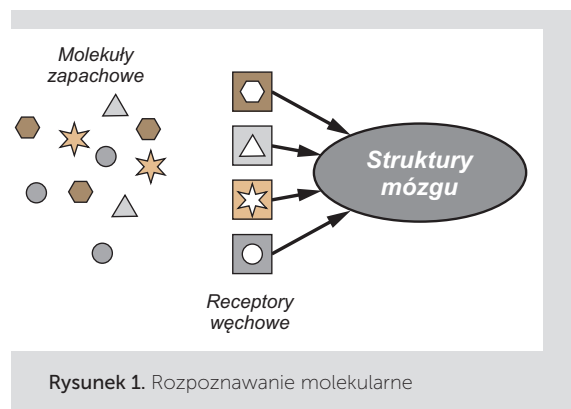
47

Zapach to informacja przekazywana drogą chemiczną nie mniej ważna i wymowna niż słowo i obraz. Truizmem jest stwierdzenie, że człowiek poznaje swoje otoczenie za pomocą swoich zmysłów. Badania psychofizyczne wykazują, że około 66% informacji uzyskujemy przez zmysł wzroku, 11% – przez zmysł dotyku, 8% – przez zmysł słuchu, 8% – przez zmysł smaku, a jedynie 4% – przez zmysł powonienia. W świecie zwierząt, a zwłaszcza owadów, proporcje te są całkowicie odwrócone. Jakkolwiek jedynie 4% informacji czerpiemy przez zmysł powonienia, to jednak zapach odgrywa nadal ważną rolę w wielu dziedzinach gospodarki człowieka. Wymienić tu można takie gałęzie przemysłu jak przemysł spożywczy, kosmetyczny czy farmaceutyczny.

Pomimo znacznego postępu wiedzy w dziedzinie neurobiologii (nagroda Nobla za 2004 rok przyznana

Lindzie Buck i Richardowi Axelowi za badania, które doprowadziły do identyfikacji receptorów węchowych, czyli białek receptorowych, oraz kodujących je genów), **nadal jeszcze nie wiadomo, co tak naprawdę pobudza do działania receptory węchowe**. Nadal są popularne dwie hipotezy sformułowane kilkadziesiąt lat temu, znane pod nazwami **teorii stereochemicznej** oraz **teorii wibracyjnej**.

– **Teoria stereochemiczna**, czyli tzw. **teoria rozpoznawania molekularnego** (popularnie znana jako



teoria klucza i zamka), zapoczątkowana już prawie 100 lat temu przez Emila Fischera, a następnie adaptowana do rozpoznawania zapachów przez Johna Amoore'a.

- **Teoria wibracyjna** zaproponowana przez Dysona, a następnie rozwinięta przez Wrighta, w myśl której receptory węchowe to rodzaj spektrometrów, które wyczuwają częstotliwości drgań molekularnych, zależnych od rodzaju obecnych grup funkcyjnych.

W odróżnieniu od teorii rozpoznawania molekularnego, która jest dobrze ugruntowana w naukach biologicznych i chemii, teoria wibracyjna spotyka się ze znaczną krytyką ze strony „konserwatywnych” naukowców. Krytycy teorii wibracyjnej zazwyczaj formułują zarzut, że nie jest znany biologiczny mechanizm konwersji energii wibracyjnej w energię aktywującą neurony (czyli zmianę potencjału elektrycznego komórki nerwowej).

Pewną próbą rozwiązania tego dylematu jest hipoteza zaproponowana przez L. Turina, zakładająca, że przekaz energii następuje za pośrednictwem nieelastycznego efektu tunelowego.

Enancjomery optyczne

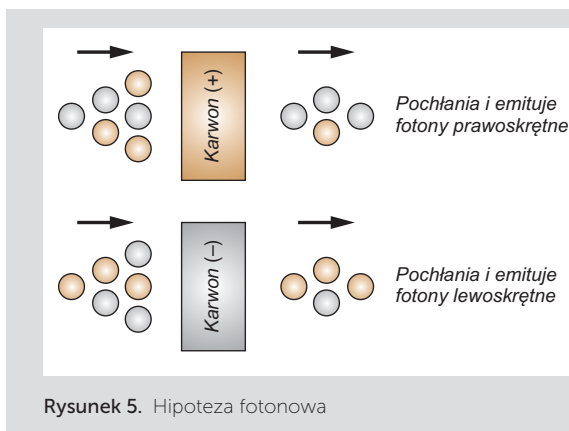
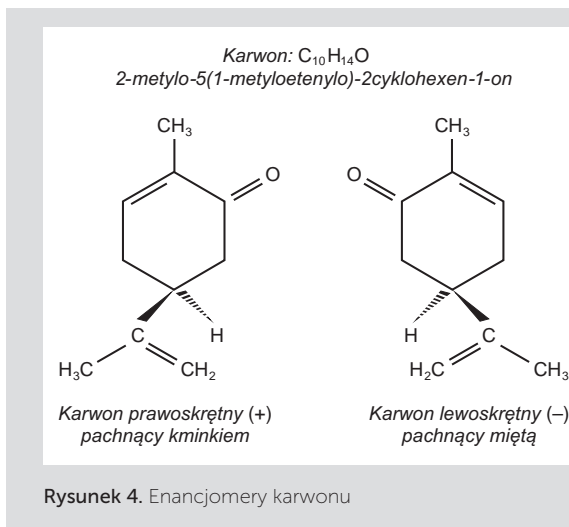
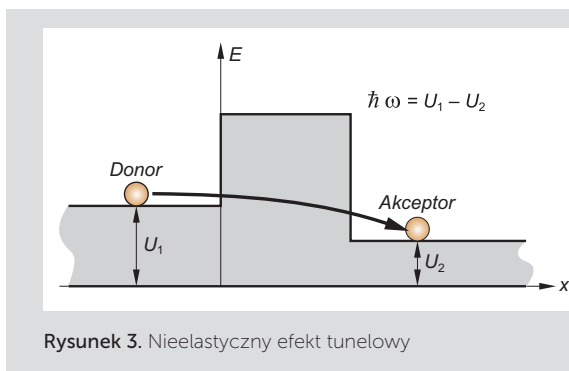
Występowanie cząsteczek związków chemicznych w dwóch nieidentycznych odmianach, będących swymi odbiciami lustrzanymi, nosi nazwę **izomerii optycznej**.

Izomery optyczne noszą też nazwę **enancjomerów**. Enancjomery optyczne wykazują identyczne właściwości fizykochemiczne, z wyjątkiem kierunku skręcania płaszczyzny polaryzacji światła. Ich własności chemiczne w ośrodkach **chiralnych** są natomiast zdecydowanie różne. O przedmiocie, który jest nieidentyczny ze swoim lustrzanym odbiciem mówimy, że jest **chiralny**. Typowym przykładem ośrodka chiralnego są substancje białkowe (białka receptorów węchowych).

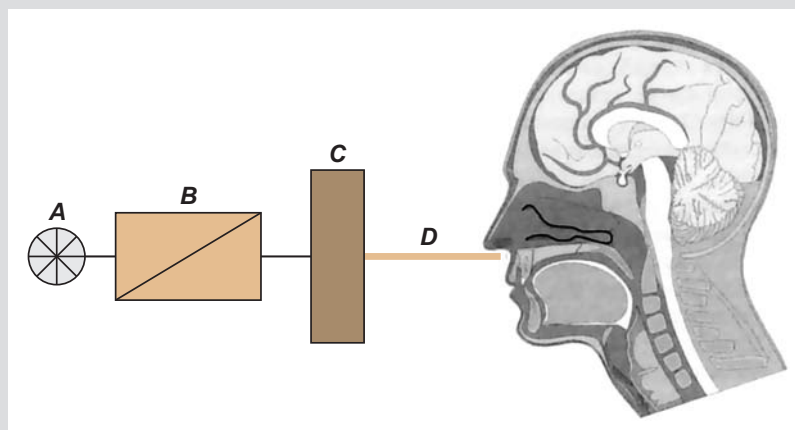
Hipoteza fotonowa

Próba rozwiązania problemu różnego zapachu **enancjomerów optycznych** w oparciu o teorię wibracyjną (te same częstotliwości drgań) musi skończyć się niepowodzeniem.

Rozwiązaniem tego problemu może być przyjęcie hipotezy, że przekaz energii wibracyjnej od molekuly



zapachowej do **chiralnego receptora węchowego** może odbywać się za pośrednictwem fotonów **wirtualnych** (wirtualnych, gdyż ich czas życia jest rzędu 10⁻²² sekundy). Jest to nasza robocza hipoteza oparta



Rysunek 6. Eksperyment psychofizyczny: A — źródło światła, P — polaryzator liniowy, C — achromatyczna ćwierćfalówka, D — światłowód jednomodowy

na doświadczeniach rozpoznawania zapachów enancjomerów karwonu.

Podstawową własnością fizyczną różniącą dwa enancjomery jest skręcanie płaszczyzny polaryzacji światła. Enancjomer skręcający płaszczyznę polaryzacji światła w **prawo** nazywamy **enancjomerem prawoskrętnym (+)**. W przeciwnym przypadku, czyli przy skręcaniu płaszczyzny polaryzacji w **lewo**, mamy do czynienia z **enancjomerem lewoskrętnym (-)**. Skręcenie płaszczyzny polaryzacji jest związane z **dwójtomnością kołową**. Dwójtomności kołowej towarzyszy zazwyczaj również **dichroizm kołowy** (prawo Natansona). Oznacza to, że dichroiczny ośrodek pochłania bardziej fotony o określonym stanie polaryzacji, czyli fotony spolaryzowane kołowo — **prawo-** lub **lewoskrętnie**. Biorąc pod uwagę fakt, że **zdolność absorpcyjna** molekuly jest równa jej **zdolności emisyjnej** widać stąd, że enancjomer prawoskrętny pochłania bardziej oraz emituje fotony spolaryzowane kołowo prawoskrętnie niż fotony spolaryzowane kołowo lewoskrętnie. Odwrotnie zachowuje się enancjomer lewoskrętny. Operując pojęciem fotonu oraz jego stanem polaryzacji, należy pamiętać o statystycznym opisie stanów polaryzacji (np. fakt, że światło spolaryzowane liniowo stanowi spójną superpozycję fotonów spolaryzowanych prawo- i lewoskrętnie) oraz o tym, że foton jako cząstka elementarna (nośnik oddziaływania elektromagnetycznego) może istnieć jedynie w próżni. W ośrodku materialnym foton natychmiast sprzęga się z tzw. quasi-cząstkami (typu wzbudzeń), czyli ekscytonami lub fononami, tworząc **polarytony ekscytonowe** lub **fononowe**.

Jeśli nasza hipoteza fotonowa okaże się słuszna, to oznaczałoby to, że tak naprawdę „**wąchamy foto-**

ny”, a więc można mówić o **zapachu światła**. Eksperymentalną weryfikacją tej hipotezy mogą stać się wyniki eksperymentu psychofizycznego przeprowadzonego według niżej zaproponowanego schematu.

System „elektronicznego nosa”

Na Wydziale Chemicznym Politechniki Warszawskiej od 1995 roku prowadzone są badania dotyczące problematyki **rozpoznawania zapachów**. Badania te dotyczą zarówno problemów natury podstawowej, jak i zagadnień aplikacyjnych. W szczególności pracujemy nad problem **obiektywnej metody zapisu oraz odtwarzania informacji zapachowej** — tzw. profilu zapachowego. W pracach tych chodzi o matematyczne sformułowanie kryteriów **podobieństwa zapachów** na podstawie cyfrowo zarejestrowanych profili zapachowych. Jednocześnie podjęliśmy prace nad konstrukcją urządzenia do zapisu informacji zapachowej — tzw. **elektronicznego nosa**.

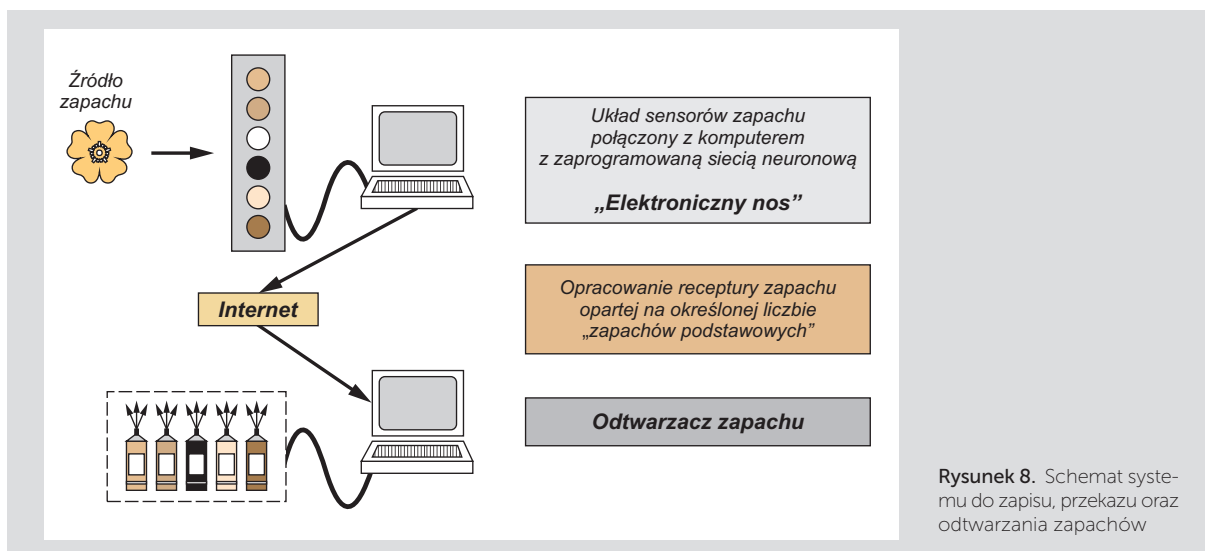
Prototyp elektronicznego nosa ukończono w 1996 roku. Od tego czasu urządzenie to było wielokrotnie ulepszane i modernizowane. Elektroniczny nos jest obecnie wykorzystywany zarówno w pracach badawczych, jak i w dydaktyce (8 magisterskich prac dyplomowych). Podstawowym elementem elektronicznego nosa jest matryca chemicznych czujników gazów oraz system rozpoznawania obrazów wykorzystujący sztuczną sieć neuronową. Zapach, który czujemy, może pochodzić od kilkuset różnych związków, np. zapach piwa czy zapach mleka, a my odbieramy tylko ogólne



Rysunek 7. System „elektronicznego nosa”

wrażenie. Podobnie działa też elektroniczny nos. Czujników w matrycy jest tylko kilkanaście (w naszym biologicznym nosie receptorów węchowych mamy około pół miliona, przy czym wyróżnia się tutaj około 1000 różnych typów receptorów). Podstawową cechą tych czujników jest ich częściowa selektywność, tzn. że dany czujnik reaguje w różnym stopniu na wiele różnych molekuł (podobnie nasze receptory biologiczne są też częściowo selektywne). Właściwą nowość stanowi sposób analizy sygnałów generowanych z matrycy poprzez system oparty na sztucznej sieci neuronowej i tutaj właśnie nos elektroniczny najbardziej przypomina działanie naszego biologicznego organu powonienia. Urządzenia zwane elektronicznym nosem są jeszcze prymitywne w porównaniu z biologicznym

narzędem węchu, ale już obecnie ułatwiają rozwiązywanie różnorodnych problemów z zakresu perfumery, technologii aromatów spożywczych, diagnostyki medycznej, kryminalistyki, kontroli jakości produktów spożywczych, a także badania zanieczyszczeń środowiska naturalnego. Współczesna technika umożliwia swobodny zapis i odtwarzanie wrażeń wzrokowych (rejestracja, odtwarzanie i zapis obrazu), jak również zapis i odtwarzanie wrażeń słuchowych (rejestracja i odtwarzanie dźwięku). Dlatego też zadaliśmy pytanie: **Czy możliwy jest zapis, a następnie odtwarzanie wrażeń węchowych?** Próbę odpowiedzi na tak postawione pytanie stanowią nasze prace dotyczące realizacji tzw. **syntetyzatora zapachów**. W pracach tych chodzi o to, aby zarejestrowane przez elektroniczny nos profile zapachowe przesyłać np. przez Internet do odbiorcy, który wyposażony w syntetyzator (odtworacz) zapachów, może dokonać następnie syntezy zapachu na podstawie przestanej „receptury zapachu”. Pomysł ten ma duże znaczenie użytkowe, zwłaszcza gdy uwzględnimy rozwijający się dynamicznie **rynek internetowy** oraz ogromny **rynek gier komputerowych**. Tak więc staje się możliwe zaangażowanie w grę komputerową trzech zmysłów człowieka – to jest wzroku, słuchu oraz węchu. Oznacza to, że na płycie CD rejestrujemy obok obrazu i dźwięku dodatkowo informacje zapachowe. Informacje zapachowe zostają odtworzone w trakcie gry za pomocą prostego (taniego) syntetyzatora zapachów przypominającego w działaniu drukarkę atramentową.



Rysunek 8. Schemat systemu do zapisu, przekazu oraz odtwarzania zapachów

Abstract

The sense of smell (olfaction) is both a very simple and a very complex sense. The mechanism by which the odor receptor cells interact with odor-causing molecules is still unknown. Several theories of how molecules interact with olfactory cells are currently under investigation. One proposes that odorant molecules vibrate at characteristic frequencies, and that olfactory cells contain molecules that vibrate at similar frequencies (vibration theory of olfaction). Another theory suggests that odorant molecules penetrate the wall of the olfactory cells, disturbing the electrolyte balance between the exterior and interior of the cell, and generating a nerve pulse. The most widely accepted theory emphasizes the importance of the size, shape, and electronic arrangement of the odorant molecules. According to this theory, the olfactory cell responds to the size, shape, and the electronic arrangement of odorant molecules ("lock and key" mechanism of the molecular recognition). One challenge to these theories is the different smells of some enantiomers (optical isomers). We propose a new biological mechanism to convert molecular vibrations into neuronal activation. This mechanism is based on the transfer of virtual photons between molecule of the enantiomer and chiral receptors.

Profesor Kazimierz Brudzewski fizyk, chemik, konstruktor. Profesor na Wydziale Chemicznym Politechniki Warszawskiej. Zatrudniony w Politechnice Warszawskiej od 1966 roku. Prowadzi prace nad metodami rozpoznawania zapachów z wykorzystaniem metod opartych o sztuczne sieci neuronowe oraz metod analizy Falkowej z elementami logiki rozmytej. Jest autorem lub współautorem około 80 publikacji oraz kilkunastu patentów. Jest konstruktorem kilku generacji tzw. Nosów Elektronicznych.



Nanospintronika

Na podstawie odczytu wygłoszonego w dniu 27 października 2005 roku

Tomasz Dietl

Instytut Fizyki Polskiej Akademii Nauk

53

Rewolucja informacyjna, w której uczestniczymy już od kilku dziesięcioleci, zachodzi dzięki systematycznemu zwiększaniu się — zgodnie z wykładniczym prawem Moore'a — ilości informacji, która może być przetwarzana, przechowywana i przesyłana przez jednostkę powierzchni odpowiednio mikroprocesorów, pamięci i światłowodów. Nowoczesny układ scalony zawiera obecnie blisko miliard tranzystorów, a każdy z nich ma rozmiar mniejszy od $100 \text{ nm} = 10^{-7} \text{ m}$, czyli jest kilkaset razy mniejszy od promienia włosa. Przekroczenie symbolicznej granicy 100 nm oznacza, że z początkiem XXI wieku wkroczyliśmy w erę nanotechnologii. Wraz ze zmniejszaniem rozmiarów tranzystorów rośnie ich szybkość działania i spada cena — koszt produkcji jednego tranzystora jest dzisiaj znacznie niższy od kosztu wydrukowania jednej litery.

Czy jednak dalszy postęp przez proste zmniejszanie rozmiarów tranzystorów w mikroprocesorach oraz komórek pamięci ferromagnetycznych twardych dysków i optycznych na płytach DVD jest możliwy? Pomimo czterdziestoletnich sukcesów laboratoriów przemysłowych na drodze pokonywania kolejnych barier technicznych i fizycznych panuje przekonanie, że w **nie - długiej przyszłości nastąpi jakościowa zmiana metod przetwarzania, przechowywania, szyfrowania i przesyłania informacji**. Z tego względu rządy wielu krajów finansują międzydziedzinowe ambitne programy naukowe, których celem jest aktywny udział w tworzeniu nanotechnologii przyszłości.

Wśród wielu propozycji ważne miejsce zajmuje **spintronika**. Jej celem jest poznanie zjawisk fizycznych

związanych ze spinem elektronu oraz zaproponowanie, zaprojektowanie i wykonanie przyrządów, które mogłyby wykorzystywać te zjawiska. U podstaw nadziei związanych ze spintroniką leży dobrze znany fakt, że ze względu na nieistnienie monopoli magnetycznych przypadkowe pola magnetyczne są znacznie słabsze niż pola elektryczne. Z tych względów pamięci magnetyczne są trwałe, natomiast pamięci wykorzystujące nagromadzony ładunek elektryczny (DRAM) wymagają częstego odświeżania.

Można zdefiniować kilka konkretnych problemów cząstkowych, które stoją przed **nanospintroniką** — nanotechnologią wykorzystującą własności spinowe. Intensywne prace badawcze i rozwojowe prowadzone w ostatnich latach doprowadziły do opracowania miniaturowego czujnika pola magnetycznego do odczytu informacji na twardym dysku. Działanie tego stosowanego już we wszystkich komputerach urządzenia spintronicznego wykorzystuje gigantyczny magnetoopór (GMR) naprzemiennych warstw zbudowanych z metali ferromagnetycznych, antyferromagnetycznych i paramagnetycznych. Zjawisko GMR zostało równoległe wykryte przez dwie grupy naukowców — zespół Alberta Ferty z Orsay i zespół Petera Grünberga z Jülich. Zgodnie z teorią rozwiniętą przez Józefa Barnasia z Poznania wynika ono ze wzrostu przewodnictwa elektrycznego w obecności zewnętrznego pola magnetycznego, które porządkuje kierunek namagnesowania sąsiednich warstw. Obecnie prowadzone prace mają na celu zwiększenie magnetooporu przez wykorzystanie

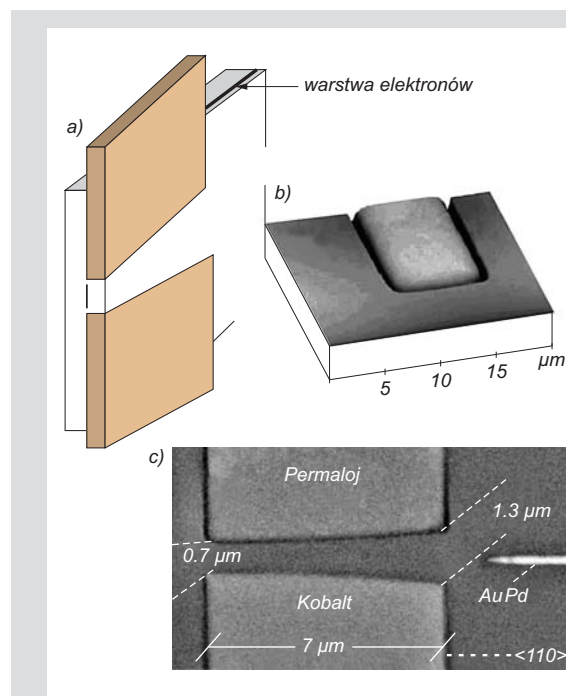
zjawiska tunelowania przez warstwę izolatora między metalami ferromagnetycznymi (TMR).

Znacznie ambitniejszym celem spintroniki jest skonstruowanie magnetycznej pamięci o dowolnym dostępie — MRAM. Urządzenie to łączyłoby zalety pamięci magnetycznej i DRAM. Wymaga to opracowania takich metod namagnesowywania oraz odczytu kierunku namagnesowania, które byłyby całkowicie niezależne od układów mechanicznych. Ważnym krokiem na tej drodze byłoby opanowanie umiejętności kontroli namagnesowania przy zastosowaniu metod izotermicznych — światłem lub prądem bądź polem elektrycznym — podobnie jak w pamięciach półprzewodnikowych DRAM, w których zapis informacji sterowany jest przyłożeniem napięcia do bramki odpowiedniego tranzystora typu MOSFET. W obecnych urządzeniach sterowanie namagnesowaniem (zapis informacji) wymaga stosunkowo dużych energii, gdyż wykorzystuje pole magnetyczne wywołane przez prąd elektryczny.

Opanowanie „inteligentniejszych” metod sterowania namagnesowaniem pozwoliłoby także na skonstruowanie tranzystorów spinowych — urządzeń zbudowanych z dwóch warstw przewodników ferromagnetycznych przedzielonych materiałem niemagnetycznym. Proste rozważania prowadzą do wniosku, że jeśli wstrzyknięte do warstwy niemagnetycznej nośniki zachowują kierunek spinu, to przewodnictwo elektryczne zależy od względnego kierunku wektorów namagnesowania w warstwach ferromagnetycznych. Wykorzystując to zjawisko można stworzyć energooszczędne i szybkie urządzenie przetłaczające, gdyż sterowanie prądem nie wymaga w nim zmiany koncentracji nośników. Oczywistym warunkiem pracy takiego tranzystora jest wydajne wstrzykiwanie spolaryzowanych spinowo nośników z materiału ferromagnetycznego do obszaru niemagnetycznego oraz brak procesów niszczących polaryzację spinową. Równocześnie poszukuje się metod wytwarzania i wykrywania prądów spinowych w przekonaniu, że ruch elektronów o przeciwnych spinach i kierunkach odbywa się bezstratnie, a może przenosić informację. Stanowiłoby to podstawę budowy urządzeń o znacznie zmniejszonym wydzielaniu się ciepła.

Bodaj najważniejszym wyzwaniem intelektualnym elektroniki spinowej jest stworzenie tzw. **informatyki kwantowej**. W światowych wysiłkach budowy podstaw teoretycznych tej nowej dziedziny wiedzy aktywnie uczestniczy m.in. rodzina Horodeckich z Gdańska. Zgodnie z pracami doświadczalnymi grupy Davida Awschaloma, szczególne znaczenie spinowych stopni

swobody wynika z faktu, że znacznie dłużej zachowują one spójność fazową niż orbitalne stopnie swobody. Spin elektronu nadaje się więc znacznie lepiej niż jego ładunek do praktycznej realizacji współczesnych idei dotyczących obliczeń numerycznych przy wykorzystaniu zasady superpozycji stanów kwantowych. Nanostruktury spinowe, w tym wytworzony i badany w Warszawie nanofiltr spinowy Sterna-Gerlacha przedstawiony na rysunku 1, mogą więc zmienić nie tylko podstawy budowy elementów elektronicznych, ale także zasady obowiązującej od półwiecza architektury komputerowej. Warto zauważyć, że już pojawiły się na rynku i są instalowane urządzenia do szyfrowania



Rysunek 1. Nanofiltr spinowy Sterna-Gerlacha z modulacyjnie domieszkowanej heterostruktury GaAs/AlGaAs otrzymany litografią elektronową. Gradient pola magnetycznego wytwarzają namagnesowane w przeciwnych kierunkach mikromagnesy z warstw kobaltu i permaloju (zaznaczone na niebiesko w a). Mikromagnesy umieszczone są w wytrawionych kanałach tak, aby ich środki znajdowały się w płaszczyźnie dwuwymiarowej warstwy elektronów (a i b — obraz z mikroskopu sił atomowych). W obecności gradientu pola magnetycznego prąd elektryczny w górnej i dolnej części nanostruktury (c — obraz ze skaningowego mikroskopu elektronowego) przenoszony jest przez elektrony o przeciwnych kierunkach spinów.

Źródło: J. Wróbel i in., Phys. Rev. Lett. 93, 246601 (2004)

kwantowego. W urządzeniach tych przesyłana informacja jest zakodowana w polaryzacji światła, a jej niezauważone przejście i odczytanie przez intruza jest, jak się wydaje, niemożliwe.

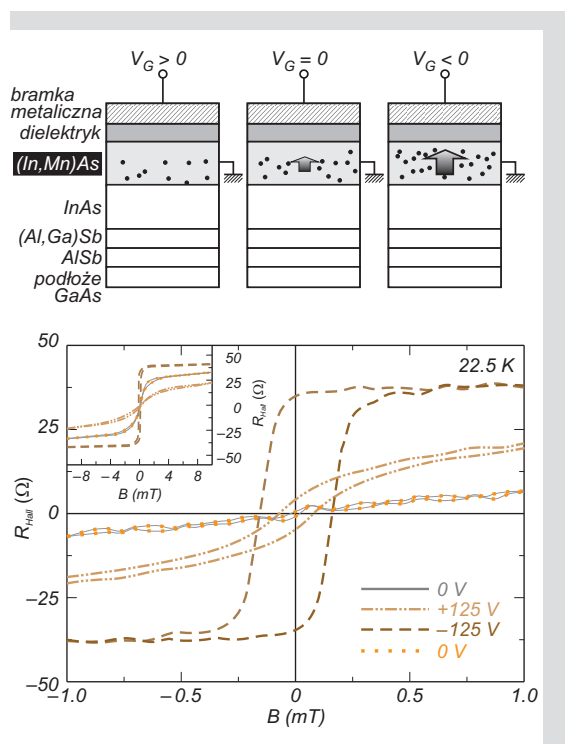
Dzisiejsze badania w dziedzinie spintroniki dotyczą praktycznie wszystkich grup materiałów. Sądzi się jednak, że szczególnie ważne dla rozwoju spintroniki będą półprzewodniki ferromagnetyczne, gdyż łączą one uzupełniające się zalety materiałów półprzewodnikowych i metali ferromagnetycznych. Zasadniczym problemem badawczym jest tutaj stwierdzenie, w jakim stopniu metody z takim powodzeniem stosowane do kontroli gęstości i stopnia polaryzacji spinowej nośników w strukturach półprzewodnikowych mogłyby służyć do sterowania wielkością i kierunkiem namagnesowania. Inne ważne zagadnienie wiąże się z opracowaniem metod wstrzykiwania spinowo spolaryzowanych nośników do półprzewodników. Poza możliwościami budowy omawianych już czujników magnetooporowych oraz tranzystorów spinowych, wstrzykiwanie spolaryzowanych nośników stanowić może metodę szybkiej modulacji laserów półprzewodnikowych. Mogłoby także pozwolić na jednomodową pracę laserów o emisji powierzchniowej.

Ponieważ technologia struktur półprzewodnikowych jest najlepiej opanowana dla związków półprzewodnikowych grupy III–V i II–VI, przetomowym osiągnięciem było wykrycie przez Hideo Ohno i współpracowników ferromagnetyzmu wywołanego nośnikami w $\text{In}_{1-x}\text{Mn}_x\text{As}$ i $\text{Ga}_{1-x}\text{Mn}_x\text{As}$. W przypadku tych materiałów dwuwartościowe jony Mn wprowadzają zlokalizowane spiny i stanowią centra akceptorowe, które dostarczają dziury. Z kolei w innej ważnej technologicznie grupie półprzewodników, w związkach II–VI zawierających jony magnetyczne, gęstości spinów i nośników można zmieniać niezależnie, podobnie jak w materiałach z grupy IV–VI, w których kontrolowany przez dziury ferromagnetyzm został wykryty w Instytucie Fizyki PAN przez Tomasza Storego i Roberta R. Gałązkę już w latach osiemdziesiątych. Badania rozcieńczonych półprzewodników magnetycznych stanowiły od końca lat siedemdziesiątych specjalność warszawskiej szkoły półprzewodników stworzonej przez Leonarda Sosnowskiego, a obecnie rozwijanej przez kilkunastu jego uczniów, dzisiaj profesorów.

W ramach współpracy między laboratoriami w Grenoble i w Warszawie, kierowanymi przez Yvesa Merle d'Aubigné i autora tego artykułu, podjęto kompleksowe badania ferromagnetyzmu wywołanego nośnikami

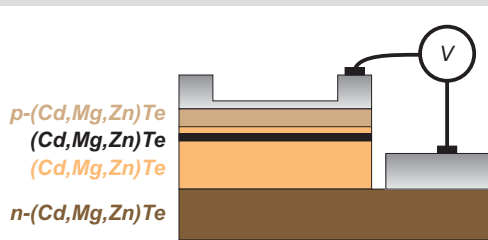
w materiałach grupy II–VI zawierających Mn. Wykonane prace technologiczne i doświadczalne doprowadziły do wykrycia metodami magnetoopiecznymi przewidzianego teoretycznie przez autora ferromagnetyzmu w układzie dwuwymiarowym, czyli w modulacyjnie domieszkowanych studniach kwantowych $\text{Cd}_{1-x}\text{Mn}_x\text{Te}/\text{Cd}_{1-y-z}\text{Mg}_y\text{Zn}_z\text{Te}:\text{N}$, które otrzymano metodą epitaksji z wiązek molekularnych. Spodziewane korelacje ferromagnetyczne wykryto także w trójwymiarowym p- $\text{Zn}_{1-x}\text{Mn}_x\text{Te}$ i p- $\text{Be}_{1-x}\text{Mn}_x\text{Te}$.

Biorąc pod uwagę wyniki doświadczalne, autor tego artykułu opracował model teoretyczny ferromagnetyzmu indukowanego dziurami w półprzewodnikach grupy III–V i II–VI zawierających Mn. Przeprowadzone

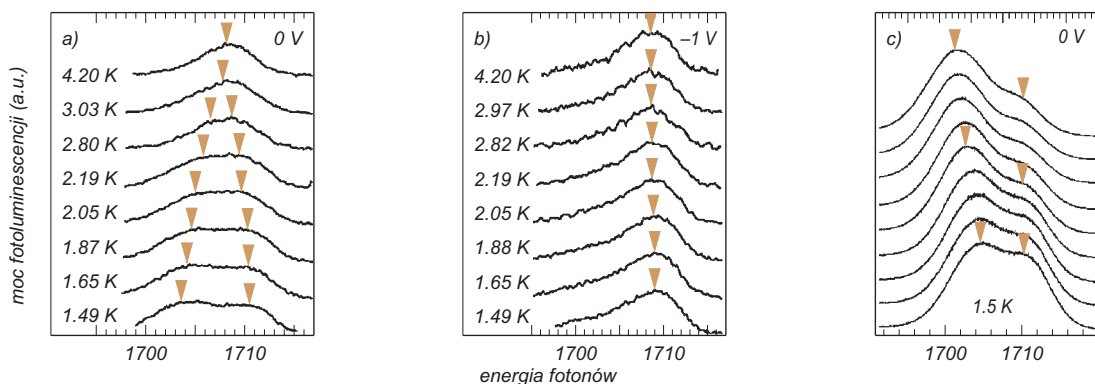


Rysunek 2. Tranzystor polowy z izolowaną bramką z kanałem typu p z $(\text{In,Mn})\text{As}$. W zależności od napięcia bramki, a więc koncentracji dziur, zależność od pola magnetycznego oporności Halla (która ze względu na anomalne zjawisko Halla jest proporcjonalna do namagnesowania) ma charakter liniowy odpowiadający fazie paramagnetycznej lub pętli histerezy odpowiadającej fazie ferromagnetycznej. Wstawka przedstawia oporność Halla w szerszym zakresie pól magnetycznych.

Źródło: H. Ohno i in., Nature 408, 944 (2000)

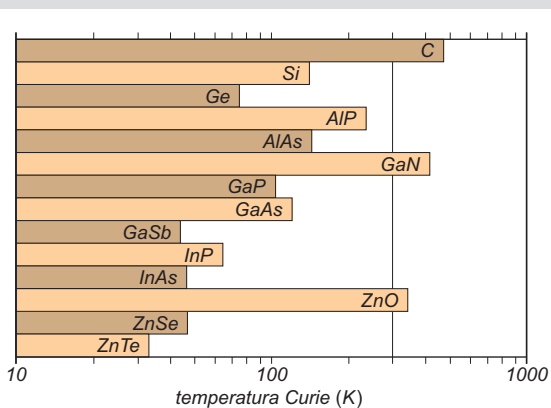


Rysunek 3. Widma fotoluminescencji (energia fotonów w meV) w nieobecności zewnętrznego pola magnetycznego struktury diody świecącej p-i-n, zawierającej studnię kwantową $\text{Cd}_{1-x}\text{Mn}_x\text{Te}$. Rozszczepienie linii (a) dowodzi uporządkowania ferromagnetycznego spinów Mn. Uporządkowanie znika, gdy przyłożone napięcie opróżnia studnię kwantową z dziur (b), a jest wzmacnione (c), gdy koncentracja dziur jest zwiększona przy pomocy dodatkowego oświetlenia.
Źródło: H. Boukari i in., Phys. Rev. Lett. 88, 207204 (2002)



prace dowodzą, że rozcieńczone półprzewodniki ferromagnetyczne typu p łączą problemy fizyki domieszkowanych izolatorów magnetycznych z przeniesieniem ładunku i silnie skorelowanych nieuporządkowanych metali z zagadnieniami defektów i stanów elektronowych wysoko domieszkowanych półprzewodników. Pomimo tej złożonej sytuacji fizycznej, rozwinięty model teoretyczny pozwolił autorowi, a także Allanowi H. MacDonaladowi i współpracownikom z Uniwersyte-

tu Tekszańskiego, na opis bez dobieralnych parametrów własności termodynamicznych, mikromagnetycznych, optycznych i transportowych szerokiej klasy rozcieńczonych półprzewodników ferromagnetycznych. Umożliwił także wykazanie niestuszności konkurencyjnych propozycji teoretycznych. Wyraża się często opinię, że w wyniku tych prac $\text{Ga}_{1-x}\text{Mn}_x\text{As}$ stał się najlepiej rozumianym ferromagnetykiem. Materiał ten służy obecnie do testowania różnych metod numerycznych



Rysunek 4. Przewidywana teoretycznie temperatura uporządkowania ferromagnetycznego różnych półprzewodników typu p, w których podstawiono 5% Mn w miejsca kationowe i w których jest $3,5 \cdot 10^{20}$ dziur w cm^3 .

Źródło: T. Dietl i in., Science 287, 1019 (2000); Phys. Rev. B 63, 195205 (2001)

obliczania własności nieuporządkowanych układów magnetycznych z pierwszych zasad.

Zrozumienie własności rozcieńczonych półprzewodników ferromagnetycznych umożliwiło autorowi i współpracownikom zaproponowanie i doświadczalne wykazanie istnienia wielu nowych zjawisk, które pozwalają na sterowanie namagnesowaniem za pomocą naprężeń, światła, prądu i pola elektrycznego, co przedstawiono na rysunkach 2 i 3. Zjawiska te nie występują w innych ferromagnetykach, co otwiera możliwość konstrukcji nieznanych dotychczas urządzeń technologii informatycznych. W związku z tym poszukiwanie nowych półprzewodników ferromagnetycznych o punkcie Curie powyżej temperatury pokojowej, zapoczątkowane sugestiami teoretycznymi autora, zilustrowanymi na rysunku 4, stanowi jedną z aktywniej rozwijających się gałęzi fizyki materiałowej.

Od lat śledzeniem i przewidywaniem rozwoju technologii informacyjno-komunikacyjnych zajmuje się wptywowe przedsiębiorstwo doradcze *International Technology Roadmap for Semiconductors*, współfinansowane przez wszystkie światowe potęgi przemysłowe w dziedzinie elektroniki półprzewodnikowej. W związku z ogromnym postępowaniem badań w dziedzinie spintroniki półprzewodnikowej w ciągu ostatnich kilku lat, w najnowszej ekspertyzie tej firmy *2004 Update – Emerging Research Devices* po raz pierwszy opisano tranzystory spinowe sugerując, że mogą one zastąpić stosowane z takim sukcesem od lat 60. unipolarne tranzystory krzemowe.

Źródła rysunków

- [1] J. Wróbel, T. Dietl, A. Łusakowski, G. Grabecki, K. Fronc, R. Hey, K.H. Ploog, H. Shtrikman, *Spin Filtering in a Hybrid Ferromagnetic-Semiconductor Microstructure*, Phys. Rev. Lett. **93**, 246601 (2004).
- [2] H. Ohno, D. Chiba, F. Matsukura, T. Omiya, E. Abe, T. Dietl, Y. Ohno, K. Ohtani, *Electric-Field Control of Ferromagnetism*, Nature **408**, 944 (2000).
- [3] H. Boukari, P. Kossacki, M. Bertolini, J. Cibert, S. Tatarenko, D. Ferrand, A. Wasiela, J.A. Gaj, T. Dietl, *Light and Electric Field Control of Ferromagnetism in Magnetic Quantum Structures*, Phys. Rev. Lett. **88**, 207204 (2002).
- [4] T. Dietl, H. Ohno, F. Matsukura, J. Cibert, D. Ferrand, *Zener Model Description of Ferromagnetism in Zinc-Blende Magnetic Semiconductors*, Science **287**, 1019 (2000); T. Dietl, H. Ohno, F. Matsukura, *Hole-Mediated Ferromagnetism in Tetrahedrally Coordinated Semiconductors*, Phys. Rev. B **63**, 195205 (2001).

Literatura uzupełniająca

- [1] I. Žutić, J. Fabian, S. Das Sarma, *Spintronics: Fundamentals and Applications*, Rev. Mod. Phys. **76**, 323 (2004).
- [2] T. Dietl, H. Ohno, *Ferromagnetic III–V and II–VI Semiconductors*, MRS Bulletin, Oct. 2003, p. 714.
- [3] T. Dietl, *Ferromagnetic Semiconductor Heterostructures*, Europhysics News **34**, 216 (2003).
- [4] S.A. Wolf, D.D. Awschalom, R.A. Buhrman, J.M. Daughton, S. von Molnár, M.L. Roukes, A.Y. Chtchelkanova, D.M. Treger, *Spintronics: A Spin-Based Electronics Vision for the Future*, Science **294**, 1488 (2001).

Angielska wersja tego tekstu ukazała się w czasopiśmie "Akademia", no 3, 28 (2005).

Profesor Tomasz Dietl fizyk, kierownik Laboratorium Badań Kriogenicznych i Spintronicznych w Instytucie Fizyki Polskiej Akademii Nauk oraz profesor zwyczajny Wydziału Fizyki Uniwersytetu Warszawskiego. Specjalizuje się w fizyce niskich temperatur, fizyce półprzewodników i spintronice. Od 1998 roku jest członkiem Polskiej Akademii Nauk; w 2002 roku został wybrany do Międzynarodowej Unii Fizyki Czystej i Stosowanej; w 2004 roku został mianowany członkiem Instytutu Fizyki Wielkiej Brytanii. W 2011 roku został powołany przez Komisję Europejską do Rady Naukowej Europejskiej Rady ds. Badań Naukowych (ERC). Jest laureatem wielu nagród, m.in.: Fundacji na Rzecz Nauki Polskiej za prace na temat półprzewodników ferromagnetycznych, Agilent Technologies Europhysics Prize — nagrody Europejskiego Towarzystwa Fizycznego, Narody Naukowej Fundacji im. Aleksandra von Humboldta, Polskiego Towarzystwa Fizycznego — Medal Mariana Smoluchowskiego.

Uczciwość i wiarygodność nauki*

Na podstawie odczytu wygłoszonego w dniu 15 grudnia 2005 roku

Maciej W. Grabski

Wydział Inżynierii Materiałowej Politechniki Warszawskiej

59

Wprowadzenie

Do tej pory byliśmy słusznie przekonani, że nauka stanowi jedną z niewielu przygód ludzkości, która się powiodła i wierzyliśmy, że niesie ona dobro sama z siebie. Nasze samopoczucie poprawia fakt, że naukowcy wciąż cieszą się społecznym uznaniem. Wysoki status uczonego odzwierciedla społeczne zaufanie — wiarę w obowiązywanie w nauce wysokich standardów etycznych i zasad, które gwarantują jej rzetelność. Panuje też ogólne przekonanie, że to właśnie dzięki przestrzeganiu tych standardów oraz swoistej, prawie kastowej, elitarności nauka zachowuje swoją integralność i wykazuje większą odporność na oszustwa i fałszerstwa niż inne obszary działalności ludzkiej. Co więcej wysokie standardy rzetelności i skrupulatne przestrzeganie właściwego dla nauki systemu wartości uznawane są za nieodłączny atrybut pracy uczonego, której główną inspiracją jest poszukiwanie prawdy i dzielenie się nią z innymi.

Nie są to oczekiwania nieuzasadnione. System nauki jest bowiem szczególnie wrażliwy na najmniejszy nawet przejaw nieuczciwości, gdyż prowadząc badania naukowe, czy też wykorzystując ich wyniki, wciąż opieramy się na świadectwie innych, w związku z czym powinniśmy mieć do tego świadectwa zaufanie.

Prawdopodobnie pierwsze sformułowanie odnoszące się do uczciwości i bezinteresowności nauki pochodzi od Francisa Bacona (1561–1626), który napisał, że uczony nigdy nie powinien kierować się osobistymi korzyściami, dążeniem do awansu czy też zyskiem. Podobną przysięgę składamy dzisiaj podczas promocji doktorskiej.

Nierzetelność i nieuczciwość towarzyszą jednak nauce od początku jej istnienia. Największy geograf starożytności Ptolomeusz z Aleksandrii ukradł dane Hipparchowi z Rodos (a więc popełnił plagiat), a ten z kolei wykorzystywał, nie ujawniając tego, źródła babilońskie. Galileusz prawdopodobnie wcale nie wykonał słynnego doświadczenia ze zrzucaniem kulki z wieży w Pizie. Newton aby osiągnąć pożądany wynik bezczelnie manipulował swoimi równaniami, Mendel oszukiwał przy liczeniu groszku, a niezgodne z zapiskami laboratoryjnymi dane prezentowane w publikacji Millikana o ładunku elektronu ratuje tylko to, że ostatecznie okazało się, że miał rację. O nowszych przypadkach nie będę wspominał¹. Zdolność nauki do samoweryfikacji, wynikająca ze skrupulatnego stosowania metody naukowej, powoduje jednak, że przypadki nierzetelności naukowej, szczególnie w aktywnych, „gorących” obszarach badań są zazwyczaj dość szybko wykrywane.

* W wykładzie wykorzystano fragmenty opracowania Zespołu ds Etyki w Nauce przy Ministrze Edukacji i Nauki pt. *Dobra praktyka badań naukowych. Rekomendacje*. Wyd MNil, Warszawa 2004.

¹ Aleksander Kohn: *False Prophets*, B. Blackwell, 1986; William J. Broad: *Betrayers of the Truth*, Oxford Univ. Press, 1985; Horace Freeland Judson: *Great Betrayal: Fraud in Science*, Harcourt 2004; Daniel J. Kevles: *The Baltimore Case*, Norton 1998.

Tak więc sporadycznie ujawniane przykłady fakszterstw, czy nawet oszustw, nie były dotąd w stanie zamącić wizerunku uczciwej nauki, gdyż mieściły się one w granicach „akceptowalnego społecznego błędu”.

Obecnie jednak coś się zmieniło. Obserwowany z niepokojem upadek autorytetu nauki, a także świadome odwracanie się ludzi od głoszonych przez nią prawd, stanowi element szerokiego, światowego procesu, który ogarnął wszystkie społeczeństwa, narastając, ku naszemu zaskoczeniu, w miarę ich demokratyzacji i rozwoju cywilizacyjnego. Pojawiające się w mediach informacje o wciąż wykrywanych nadużyciach na szczytach nauki, albo dokonywanych przy pomocy nauki, trafiają na podatny grunt i przyczyniają się do pogłębienia kryzysu. Uczniowie narzekają więc, że społeczeństwo przestaje im ufać, obwiniając o to jego niedouczenie i ignorancję, naprawdę jednak to nie brak edukacji przyczynia się do dotykającego nas kryzysu, ale rosnący brak zaufania do wiarygodności nauki.

W dzisiejszym niestęchającym się środowisku nauki wypracowane w ubiegłych stuleciach korporacyjne mechanizmy ochronne i korygujące, aczkolwiek wciąż potrzebne, nie stanowią już wystarczającego dobrego zabezpieczenia przed cywilizacyjnymi i społecznymi konsekwencjami narastającej fali nierzetelności naukowej.

Problem ma zasadnicze znaczenie nie tylko dla wewnętrznej spójności i integralności nauki, ale również dla utrzymania jej wiarygodności i społecznego autorytetu, bo niestety ujawniane, niejednokrotnie bulwersujące, przypadki nieuczciwości, mimo że dotyczą pojedynczego uczonego, to odbijają się po stokroć na zaufaniu do całej nauki.

Rozwój nauki jest nieuchronny, gdyż nie tylko generuje ona wiedzę o świecie, ale i tworzy podstawy do działalności człowieka związanej w wykorzystywaniem zasobów natury oraz chronieniem ludzkości przed różnorodnymi zagrożeniami, takimi jak głód, epidemie, ocieplenie globalne, wyczerpywanie się zasobów energetycznych itd., itd. Sama z siebie nauka nie jest jednak w stanie tych innowacji dostarczyć, dlatego główną siłą napędową kierującą do nauki ogromne strumienie pieniędzy jest to, że owe innowacje są potrzebne albo społeczeństwu, albo politykom, albo — w największej mierze — rynkowi.

Wynikające z tych związków coraz bliższe interakcje między polityką, biznesem i społeczeństwem z jed-

nej strony, a nauką ze strony drugiej stały się trwałym elementem współczesnego świata, przyczyniając się niewątpliwie do rozwoju cywilizacyjnego, ale też niosąc ze sobą poważne niebezpieczeństwa i wyzwania.

Dobrze jest bowiem, gdy polityk czy biznesmen działają w interesie publicznym. Tak się zwykle zdarza, ale jednak nie zawsze tak jest. Dobrze jest, gdy naukowiec działa w interesie prawdy, ale to też nie jest regułą.

Dlatego coraz częściej możemy usłyszeć pytanie o to w czym interesie występuje nauka — w interesie prawdy, czy w interesie polityki; w interesie prywatnym, czy też w interesie publicznym?² Samo postawienie tego pytania wskazuje na powstanie wątpliwości co do wiarygodności nauki.

Przekonanie o uczciwości i honorze nauki nie jest już dzisiaj wystarczającą rękojmią na przyszłość, gdyż widzimy jak intensywnie jest ona zaśmiecana i jak gęsto boko sięga manipulacja nauką, czy też manipulacja prowadzona w oparciu o naukę zarówno przez biznes, jak i przez politykę. Sytuacja jest więc zupełnie inna niż jeszcze kilkadziesiąt lat temu, gdy nauka potrafiła sama dać sobie radę z „czarnymi owcami” w swoim własnym gronie.

Problem integralności nauki i jej wiarygodności jako systemu nabrał szczególnego znaczenia i trzeba to zagadnienie podnosić publicznie. To my, jako społeczność naukowa, musimy poszukać odpowiedzi na pytanie o to jak uchronić się przed patologiami i różnego rodzaju nieuczciwościami, które wkradają się do nauki w ślad za jej rozwojem ilościowym oraz nabierającym coraz większego znaczenia wymogiem jej bliskiego związku z otoczeniem.

Wiarygodność nauki

Terminy **nauka** i **badania naukowe** straciły swoją niedawną jednoznaczność, gdyż zostały rozciągnięte ponad wszelką miarę na obszary nie mające wiele wspólnego ani z nauką, ani ze stosowaniem metody naukowej jako systemu poszukiwania prawdy. Łatwo można dostrzec, że nauką nazywa się obecnie praktycznie każdą działalność związaną z badaniem czegośkolwiek. Wiemy jednak, że sam fakt badania nie stanowi jeszcze o tym, że mamy do czynienia z działalnością naukową. Poza tym, zgodnie ze stosowanymi na całym świecie statystycznymi czy budżetowymi nomen-

² Np. Colin Tudge w *New Statesman* (UK) z dnia 26 kwietnia 2004.

kłaturami, pojęcie „nauka” obejmuje się również działalność badawczo-rozwojową (BR), która w rzeczywistości jest działalnością inżynierską, opierającą się co prawda na wykorzystaniu nauki, ale mającą inny od nauki cel. O ile bowiem celem nauki jest poszukiwanie prawdy lub opis rzeczywistości, to działalność badawczo-rozwojowa służy opracowywaniu technologii lub konstrukcji, a ogólniej rzecz biorąc – generowaniu potrzebnych rynków innowacji. Powstałe w ten sposób zamieszanie pojęciowe utrudnia nie tylko profanom zrozumienie istniejącej sytuacji.

Masowy rozwój ilościowy obszaru badawczego, który od czasu II wojny światowej zachodzi w skali całego świata, stanowi bezpośrednią przyczynę załamania się dotychczasowego – korporacyjno-akademickiego – systemu, który dotąd zapewniał nauce jej jakość i integralność. Sytuację utrudnia fakt, że coraz większa część nauki przenosi się ze sfery publicznej do sfery prywatnej, poza bezpośredni obszar owego akademickiego oddziaływania. **Jednak jak wiemy ilość niestety nie przechodzi w jakość, a nadmierny wzrost i związane z nim obniżenie standardów pozwoliły na wnikięcie w naukę trudnych do opanowania zjawisk patologicznych.**

W systemach, w których podstawowym kryterium oceny zespołów naukowych jest ich produktywność, badania naukowe stanowią przedmiot kontraktów finansowych, a dążenie do sukcesu i indywidualnej kariery za wszelką cenę staje się dominującym bodźcem działania. Rodzi to ogromną pokusę nierzetelności badawczej – zarówno tej dużej, jak i małej.

W wyniku takiego podejścia nauka akademicka łatwo przekształca się w nastawioną na produkcję taśmową „McNaukę”, której produktem jest coś co coraz powszechniej nazywa się nauką śmieciową (*junk science*), w której prawda i rzetelność naukowa zajmuje dalekie miejsce na liście priorytetów. Ilość nauki śmieciowej rośnie wykładniczo wraz z liczebnością instytucji naukowych oraz nierecenzowanych, działających poza systemem *peer-review*, czasopism, chociaż i renomowane periodyki nie mogą się przed nią uchronić.

Sprawa nie ma marginesowego charakteru. Trzeba sobie uświadomić, że na świecie ukazuje się corocznie około 10 milionów artykułów naukowych. Blisko 20 procent publikacji w czasopismach pochodzących z tak u nas cenionej *Listy Filadelfijskiej*, nie jest nigdy cytowana. Jeżeli odejmiemy autocytowania i cytowania pojedyncze oraz dodamy to tego publikacje, które ogłoszono w licznie znacznie większej grupie czaso-

pism znajdujących się poza tą listą, to możemy łatwo przyjąć, że większości z owych 10 milionów artykułów nikt nie przeczyta i nikt ich nie zacytuje (być może z wyjątkiem autora lub jego ucznia). Stanowią one po prostu naukowy śmietnik. Niestety *Prawo Sturgeona* mówiące, że *dziewięćdziesiąt procent wszystkiego to bzdety*, daje się zastosować również i do nauki.

Nauka śmieciowa, powstająca poza głównym nurtem nauki, niczego nie wnosi do naszej wiedzy o świecie, tworząc jedynie szum informacyjny, ale, co gorsza, często stanowi użyteczny element manipulacji, gdyż celowo zafałszowując dane oraz naciągając ich interpretację, a także manipulując analizami naukowymi, jest wykorzystywana do wsparcia założonych z góry punktów widzenia. Niewprawny adept nauki, pozbawiony intelektualnej opieki ze strony promotora, co się niestety często zdarza, łatwo wpada w tę pułapkę.

Nie jestem odosobniony w poglądzie, że po długim okresie nieprzerwanych i niezwykle sukcesów nauka znalazła się obecnie w obliczu sytuacji kryzysowej, która, paradoksalnie, jest bezpośrednim skutkiem tych właśnie sukcesów oraz wynikającego z nich uzyskania przez naukę politycznego i ekonomicznego znaczenia.

Dlatego obecnie nauka:

- ma problemy z własną uczciwością,
- ma problemy z utrzymaniem publicznego zaufania,
- a na dodatek (a może w wyniku tego) stała się zarówno podmiotem, jak i przedmiotem, manipulacji.

Obszary krytyczne

Problem nierzetelności ma niejednakową skalę i odmienny charakter w różnych obszarach nauki. Jest w dużym stopniu uzależniony od rygorystyczności metod pozyskiwania danych oraz możliwości ich powtórzenia lub weryfikacji. Badania naukowe możemy z tego punktu widzenia podzielić (z grubsza) na trzy grupy scharakteryzowane swoistą „twardością” danych, na których się one opierają. Są to:

- **dane zwymiarowane** (doświadczalne) – np. większa część fizyki i chemii, większość nauk technicznych, biologia komórkowa itp.;
- **dane niezwymiarowane** (opisowe) i statystyczne – np. ekologia, znaczna część nauk biomedycznych, socjologia;
- **dane historyczne** – np. kosmologia, paleontologia, geologia oraz oczywiście historia.

Szczególną podatność na występowanie nierzetelności wykazują te „miękkie” obszary nauki, w których sprawdzalność wyniku jest trudna do osiągnięcia, a ze względu na wielką złożoność badanych systemów i brak dobrych modeli zjawiska, korzysta się z niedoskonale zazwyczaj zwymiarowanych badań porównawczych i statystycznych.

Omawianie źródeł nierzetelności leży poza tematem dzisiejszej dyskusji. Ma on ogromną i łatwo dostępną literaturę (jeżeli w wyszukiwarce Google wpisze się hasło „scientific misconduct” otrzyma się ćwierć miliona hitów!). Warto jednak przedstawić kilka podstawowych ustaleń. Identyfikując główne źródła pokus skłaniające naukowców do popełniania nierzetelności można wskazać takie, które posiadają charakter wewnętrzny naukowy oraz takie, w których impuls przychodzi z zewnątrz — spoza nauki.

Pokusy wewnętrzne wynikają bezpośrednio ze struktury nauki oraz mechanizmu kariery i pracy naukowej. Szczególną uwagę warto tu zwrócić na nacisk kariery, czyli bezsensowny wymóg publikowania jak najwięcej (*publish or perish*) i na dążenie za wszelką cenę do prestiżu i sławy w wyścigu o pierwszeństwo. Szczególnie podatne na te pokusy są wielkie zespoły badawcze oraz studia doktoranckie, które dzięki malejącej roli lidera czy promotora coraz częściej stają się anonimowe. W takiej sytuacji łatwo o fałszowanie lub naciąganie wyników doświadczeń do spodziewanego rezultatu.

Jeżeli zaś chodzi o **pokusy zewnętrzne** sprawa ma całkiem inny wymiar i wiąże się ze zjawiskiem korupcji intelektualnej oraz konfliktami interesów. W znacznej mierze są one wywołane coraz bardziej poszerzającym się stykiem nauki z gospodarką, gdyż zwykle na tym styku pojawiają się niezwykle groźne dla nauki zjawiska, z których znaczenia jeszcze kilkanaście lat temu nie zdawano sobie sprawy. Jednym z najgroźniejszych odkryć ostatniego półwiecza jest stwierdzenie, że wynik naukowy może być towarem, a badania naukowe — przedmiotem kontraktu, w którym inwestor oczekuje konkretnego wyniku. Co więcej, wynik ten niekoniecznie ma polegać na prawdzie, ale musi zapewnić inwestorowi przewagę nad konkurencją, bo na tym polega racjonalność działania gospodarczego.

W ten sposób pewne szczególne obszary nauki, takie jak na przykład nauki biomedyczne, przekształ-

ciły się *de facto* w niezwykle rentowną dziedzinę gospodarki. Rewolucja w technologii badań genetycznych, będąca następstwem odkrycia struktury DNA, wydukowała nowy rodzaj firm, których jedynym produktem jest wynik naukowy, a właściwie dane naukowe, mogące stanowić przedmiot patentu i obrotu handlowego. Generalnie można stwierdzić, że wprowadzenie powszechnej zasady kontraktowania badań powoduje, że instytucje zajmujące się ich prowadzeniem coraz powszechniej tracą charakter instytucji publicznych, przekształcając się w przedsięwzięcia handlowe, w których kryterium zysku staje się ważniejsze od kryterium prawdy.

W wyniku tych zmian nastąpiło uruchomienie trudnego do opanowania procesu prywatyzowania wiedzy wytwarzanej przez naukę. Nauka akademicka, która zgodnie ze swoją definicją powinna służyć powszechnemu dobru, zostaje obecnie stopniowo zawłaszczona przez pieniądź. Komerccjalizacja osiągnięć naukowych w coraz większym stopniu przekształca się w komercjalizację całych sfer nauki³. Nie oceniam tego zjawiska, ale chodzi mi o zwrócenie uwagi na problem.

Zazębianie się nauki akademickiej z gospodarką powoduje więc powstawanie licznych obszarów, w których pojawiają się ostre wyzwania jakościowe i konflikty etyczne, zagrażające integralności nauki. Wiążą się one już nie tylko z często degenerującą chęcią zdobycia przez uczonych prestiżu i sławy, ale ze zdobywaniem pieniędzy, już nie na badania lecz dla siebie, co zaczęło regulować stosunki między nauką a gospodarką. Nie byłoby w tym nic złego, gdyby można się było ochronić przed pojawiającymi się na tym styku zjawiskami — korupcją intelektualną i sprzeniewierzeniem się pod stawowym zasadom etyki.

Manipulacja nauką

Szczególnym patologicznym zjawiskiem o społecznym wymiarze, które bezpośrednio przyczynia się do obniżenia wiarygodności nauki, a wobec którego środowisko naukowe nie może pozostawać bierne, jest manipulacja nauką. Fakt, że nauka stanowi obiekt manipulacji wynika z jej wciąż wysokiego społecznego autorytetu jako źródła wiedzy o świecie, dzięki czemu odwoływanie się do nauki od bardzo dawna jest

³ Np.: Sheldon Krinsky: *Science in Private Interest*, Rowman nad Littlefield, 2004; Daniel E. Greenberg: *Science, Money and Politics: Political Triumph and Ethical Erosion*, University of Chicago Press, 2001.

wykorzystywane jako rozstrzygający argument w sporach ideologicznych, politycznych, sądowych, czy też w kampaniach handlowych.

Istnienie tak szerokiego zapotrzebowania na autorytet nauki musiało spowodować pojawienie się pokusy manipulowania wynikami badań naukowych i koncepcjami naukowymi w taki sposób, aby stanowiły one uzasadnienie dla partykularnych, nieraz bardzo niejasnych, zamysłów.

Nauką manipuluje więc ten, kto może odnieść z tego korzyść. Grono zainteresowanych jest niezwykle liczne. Wymieńmy tylko niektórych:

- **politycy** – w celu uzyskania wsparcia ze strony określonych grup społecznych, grup interesu lub też ze względu na polityczną poprawność, a także aby przeforsować swoje zamierzenia;
- **instytucje rządowe** – w celu rozszerzenia swoich uprawnień lub podwyższenia budżetów;
- **media** – w celu generowania sensacji lub niejasnych interesów ideologicznych;
- **bizness** – w celu zdobycia przewagi na rynku lub uzyskania zamówień rządowych;
- **aktywiści społeczni i rozmaitego autoramentu ekstremiści** – w celu wymuszenia podporządkowanych ideologii zmian społecznych czy politycznych;
- **prawnicy** – w celu skotowania sądów i uzyskania w ten sposób wielkich odszkodowań dla swoich prywatnych czy instytucjonalnych klientów.

Manipulują również **naukowcy** – dla kariery, rozgłosu, zdobycia fortuny, ze względów ideologicznych czy wreszcie w lobbingu na rzecz swoich kosztownych pomysłów.

Ciekawy jest fakt, że bardzo często grupy te inspirowane wzajemnie i działają wspólnie, w takich czy innych układach, ale zawsze za pomocą mediów.

Aktualny przykład wszechstronnej manipulacji stanowią kontrowersje dotyczące żywności modyfikowanej genetycznie (GM). Ilość generowanej w tym obszarze nauki śmieciowej, wspierającej zresztą obydwie strony kontrowersji, jest gigantyczna. W wyniku działania mediów oraz rozmaitych ekstremistów i polityków społeczeństwo zaczęło się jej obawiać, a przemysł bojąc się strat powoli się wycofuje z produkcji takiej żywności. Rzadki wśród polityków rozsądek wykazał w tej sprawie Tony Blair mówiąc racjonalnie: *ważnym jest, aby cała debata w sprawie GM opierała się na dowodach naukowych (scientific evidence) a nie na uprzedzeniach*. I w rzeczywistości większość uprzedzeń

wynika z niezrozumienia istoty badań naukowych i ich wyników. Ale są i uprzedzenia wynikające nie z braku zrozumienia ale z celowej manipulacji.

Oczywiście w sytuacji takich kontrowersji ustawodawca, czy też sąd, musi sięgnąć do punktu odniesienia dla prawdy, który powinna stanowić nauka, a więc zwrócić się do reprezentujących ją ekspertów o przedstawienie dowodów naukowych. Skąd jednak je wziąć gdy nauka ze swej istoty wiąże się z istnieniem kontrowersji, na których wyjaśnienie trzeba niejednokrotnie długo czekać, a poza tym niejednokrotnie nawet w najpoważniejszych czasopismach naukowych ukazują się artykuły zawierające, jak się okazało, celowo spreparowane dane?

Znawcy przedmiotu wymieniają trzy główne rodzaje ekspertów:

- **Najemnicy** – czyli tacy, którzy są do wynajęcia. To oni właśnie stanowią najliczniejszą grupę odbiorców, a czasem i kreatorów nauki śmieciowej (*junk science*).
- **Wyznawcy** – czyli tacy, którzy są głęboko przywiązani do własnego punktu widzenia i w związku z tym nie mają żadnych wątpliwości. Odpowiednio dobierając spośród nich zawsze można uzyskać zamierzony efekt, tym bardziej, że zwykle są wygadani i medialnie przekonywujący.
- **Pytie** – czyli tacy, którzy posiadają porządną i wszechstronną wiedzę, ale będąc świadomi jej niejednoznaczności nie są skłonni do przedstawiania jednoznacznych opinii, gdyż nie chcą zostać zaplątani w żadną kontrowersję. Tych jest najwięcej, ale paradoksalnie ich użyteczność dla gry politycznej nie jest wielka.

Każdy z nas może przytoczyć przykłady „uznanych ekspertów”, którzy bezkarnie wygłaszają publicznie definitywne opinie w oparciu o niepełne, niejednoznaczne lub trudne do zweryfikowania dane, bez żadnego odniesienia do faktów. Każdy też zna przypadki ekspertów głęboko uwikłanych w **konflikt interesów**. Szczególne miejsce w hierarchii mają ci, którzy autorytarnie wypowiadają się na tematy leżące daleko poza obszarem ich naukowej kompetencji stosując zasadę: *wierz mi – ja jestem profesorem*. Ustugi eksperckie stały się obecnie dochodowym biznesem, co dodatkowo komplikuje sytuację.

Wydaje mi się, że można sformułować coś w rodzaju ogólnego prawa nauki śmieciowej: **każdej eks-**

pertyzie można przeciwstawić równą co do siły i przeciwnie skierowaną kontrexpertyzę.

Dramat o skutkach społecznych pojawia się gdy sprzeczności lub też niejednoznaczności poglądów naukowych musi rozstrzygać nie sama nauka, ale ustawodawca lub sąd, postawiony przed koniecznością rozwiązania praktycznego problemu. A zdarza się to często.

Wróćmy jednak do sprawy manipulacji nauką. Aktywnych obszarów manipulacji nauką jest bardzo wiele, właściwie staje się nią każda dziedzina nauki w chwili, gdy pojawia się jej styk z polityką i rynkiem, a jak widzimy również z prawem, szczególnie zaś gdy znajdujemy się w obszarze „miękkich” dyscyplin, gdzie łatwo może wkroczyć nauka śmieciowa. Dotyczy to również obszarów, w których pojawiają się niemożliwe do rozstrzygnięcia przez naukę problemy etyczne.

Aby uzmysłowić skalę problemu wymienię kilka bardzo aktywnych obszarów. Należą do nich:

- **relacje człowieka z przyrodą** – globalne ocieplenie, ginące gatunki itd.;
- **genetyka** – problemy równości płci, równości rasowej, a także żywności genetycznie modyfikowanej (GM);
- **zdrowie** – przyczyny zachorowań na raka, problemy otyłości oraz np. wpływ telefonów komórkowych na zdrowie i cały kompleks farmakologii;
- **energetyka** – w tym problem energetyki jądrowej i alternatywnych źródeł energii;
- **polityka społeczna** – badania opinii publicznej.

Wymiar biznesowy manipulacji wyraża się niejednokrotnie w miliardach dolarów (np. farmakologia). Wymiar polityczny manipulacji może być również ogromny, czego przykłady można zaczerpnąć m.in. z ostatnich wyborów prezydenckich w USA. Problem pojawia się również w Polsce. Przykład stanowi niedawna dyskusja wokół biopaliw, w której szczególną rolę odgrywali przedstawiciele nauk technicznych.

Oszustwa w nauce

Mówiąc o oszustwach w nauce trzeba zwrócić uwagę na pewną ich specyfikę, która odróżnia je od innego typu oszustw, z jakimi mamy do czynienia w praktyce sądowej.

Czytając pracę naukową możemy zgadzać się lub nie zgadzać z jej wnioskami, lecz musimy mieć zawsze zaufanie do przedstawionego opisu stosowanych procedur oraz do uzyskanych na ich podstawie wyników. Śledząc elegancko uporządkowany wywód rozprawy naukowej nie dowiemy się zwykle nic o poprzedzających sukces chybionych hipotezach, o poczynionych, a następnie zarzuconych, błędnych założeniach i wnioskach prowadzących do poszukiwania właściwej drogi w złym kierunku, o beztładnych próbach i pomyłkach, nieudanych doświadczeniach, odrzuconych wynikach, które uznano za błędne i wreszcie o poniesionych kosztach. Praktycznie we wszystkich publikacjach naukowych znajdują się pominięcia i przeinaczenia dokonywane po to, aby uzyskane wyniki ukazać w korzystnej perspektywie. Podobnie lista współautorów publikacji nie zawsze zawiera nazwiska osób, które przyczyniły się do jej powstania. Takie zachowanie może zostać uznane za naganne lub pożałowania godne, ale nie stanowi ono jeszcze oszustwa. **Z prawdziwym oszustwem mamy do czynienia dopiero wtedy, gdy przedstawione w pracy procedury niezbędne dla powtórzenia wyników, albo też same wyniki, zostały w jakiś sposób świadomie zmanipulowane lub przeinaczone.**

Tak zdefiniowane oszustwo naukowe różni się istotnie od oszustwa objętego prawem cywilnym, ponieważ prawo cywilne przewiduje istnienie poszkodowanego, a więc powoda, który wnosi sprawę do sądu i który musi przedstawić dowody, że dokonano przeinaczenia oraz, że w jego następstwie poniósł szkodę. Natomiast w przypadku oszustwa naukowego może nie być osoby fizycznie poszkodowanej. Nie ma w ogóle potrzeby udowodnienia, że ktokolwiek poniósł szkodę w następstwie tego, że uwierzył w przeinaczone wyniki. To, co ma znaczenie, sprowadza się wyłącznie do stwierdzenia, czy zaprezentowane procedury i wyniki zostały przedstawione rzetelnie, czy też nie⁴. Dlatego też linia odgraniczająca świadome oszustwo od niechlujności, czy zaniedbania, a nawet od nieistotnego merytorycznie upiększenia, jest niezwykle cienka i łatwa do przekroczenia. Prawna kwalifikacja takiego oszustwa ulegnie zmianie, gdy w oparciu o przeinaczone wyniki zostanie przez kogoś podjęta np. procedura medyczna, w wyniku której szkodę poniesie pacjent, albo gdy agencja finansująca badania uzna, że w wyniku oszustwa nastąpiło wyłudzenie lub sprzeniewierzenie powierzonych uczonemu środków.

⁴ Np. David Goodstein: *Conduct and Misconduct in Science*, www.its.caltech.edu/%7Edg/conduct_art.html.

Warto zwrócić uwagę na to, że w dokumentach odnoszących się do problemu nierzetelności naukowej i definiujących związane z nią pojęcia nie używa się zazwyczaj słowa „oszustwo”. Pojęcie „oszustwo” zawiera w sobie intencję celowego wprowadzenia w błąd, natomiast ciała naukowe prowadzące dochodzenia nie chcą koncentrować uwagi na motywach popełnienia nierzetelności, który w tych sprawach nie posiada żadnego znaczenia. Z tego powodu mówi się raczej o **poważnych odstępstwach od praktyki powszechnie przyjętej w ramach społeczności naukowej**. Obecnie jednak, w związku z niejednoznacznością tego sformułowania zaszła konieczność bardzo szczegółowego zdefiniowania i opisanie tej „dobrej praktyki”. Taka jest geneza sformułowania „dobra praktyka badań naukowych”.

Jak dotąd, najbardziej związane i precyzyjne definicje odnoszące się do praktyki badań naukowych przedstawia opracowany w USA przez National Science and Technology Council dokument pod tytułem *Proposed Federal Policy on Research Misconduct*. Został on wydany jesienią 1999 roku przez powołane przez prezydenta USA Office of Science and Technology i odnosi się do federalnych agencji naukowych⁵. Dokument ten określa, że:

1. **Nierzetelność w nauce** (*scientific misconduct*) stanowią występki przeciwko etyce w nauce polegające na fabrykowaniu, fałszowaniu lub plagiatowaniu (FFP) przy aplikowaniu o fundusze, przy prowadzeniu i recenzowaniu badań naukowych lub też przy prezentowaniu ich wyników.
2. **Fabrykacja** (zmyślanie) polega na preparowaniu, rejestrowaniu i publikowaniu wyników nie uzyskanych.
3. **Fałszowanie** polega na manipulacji materiałem badawczym, wyposażeniem lub metodą oraz na zmienianiu lub pomijaniu danych doświadczalnych w ten sposób, że wyniki badań nie zostają prawdziwie przedstawione w raportach.
4. **Plagiaryzm** (plagiatowanie) polega na przywłaszczeniu cudzych idei, metod, wyników lub określeń bez właściwego odniesienia. Plagiatem jest także nieautoryzowane wykorzystanie informacji

uzyskanych w trakcie poufnego recenzowania wniosków i rękopisów.

Nierzetelności naukowa nie obejmuje popełnienia niezamierzonego błędu i nie odnosi się do prawa uczonego do wyrażania rzetelnych różnic w opiniach. Definicje te egzemplifikują, ale nie wyczerpują w pełni, listy zagrożeń dla rzetelności naukowej.

Mimo że stosowane w innych krajach definicje nie odbiegają istotnie od tych, które przyjęto w USA, to jednak możliwość wprowadzenia jednolitej światowej definicji nierzetelności naukowej jest problematyczna. Wiąże się to z trudnościami w ustaleniu ścisłej definicji nierzetelności i odróżnienia jej od zwykłej niechlujności, z trudnościami w określeniu granicy między nierzetelnością naukową a orzekanymi przez sądy przestępstwami czy wykroczeniami (zdefiniowanymi we właściwych dla danego kraju kodeksach karnych, kodeksach postępowania cywilnego, prawach własności intelektualnej itp.) oraz orzekanymi przez komisje dyscyplinarne naruszeniami dyscypliny, czy wreszcie ze stanowiącymi właściwość sądów koleżeńskich sprawami objętymi kodeksami honorowymi. Co więcej, przy orzekaniu spraw związanych z nierzetelnością naukową zachodzi potrzeba uwzględnienia specyfiki poszczególnych dyscyplin. Można np. wskazać na szczególne problemy związane z badaniami na człowieku, gdzie szczególnego znaczenia nabiera poczucie etycznej odpowiedzialności badacza.

Skala problemu

Wbrew wyrażanym częstokroć opiniom i mimo że liczba ujawnianych przypadków w odniesieniu do liczby finansowanych projektów jest rzeczywiście znikoma zjawiska nierzetelności w nauce nie wolno marginalizować. Pewne strukturalne aspekty działania uczelni wyższych skłaniają nawet czołowych uczonych do minimalizowania problemu i do ignorowania możliwości naruszeń etyki⁶. Poza tym uważa się, że ujawnianiu podlegają jedynie przypadki drastyczne, natomiast sprawy drobne (co nie znaczy, że mniej szkodliwe) pozostają niezauważone, a czasem, w niektórych mniej prominentnych laboratoriach, spotykają się z obojętnością

⁵ Proposed Federal Policy on Research Misconduct to Protect the Integrity of the Research Record, http://www.ostp.gov/html/9910_20_3.html.

⁶ C.K. Gunsalus: „Rethinking Unscientific Attitudes about Scientific Misconduct”, *The Chronicle of Higher Education*, Mar. 28, 1997, p. B4.

lub są nawet milcząco tolerowane. Tak więc to, co obserwujemy, stanowi jedynie wierzchołek góry lodowej.

Podkreśla się jednak, że nie ma znaczenia czy prawdopodobieństwo wystąpienia przypadku nierzetelności naukowej wynosi jeden na tysiąc, czy jeden na sto tysięcy. Tak samo nie ma znaczenia prawdopodobieństwo uderzenia piorunu w dom — niezależnie od niego, każdy budynek powinien być wyposażony w odgromniki — bo w przypadku zdarzenia szkody będą ogromne.

W latach 1993–1997 Office of Research Integrity (ORI), powołane w Stanach Zjednoczonych do śledzenia i rozpatrywania nierzetelności w badaniach z obszaru nauk medycznych i biomedycznych finansowanych przez National Institutes of Health (NIH), zarejestrowało 1000 zgłoszonych zarzutów. Zamknięto 150 spraw orzekając, że naruszenie nastąpiło w połowie z nich — czyli w 76 sprawach (7,6%) — w wyniku czego 71 osób odsunięto od możliwości ubiegania się o federalne finansowanie na okres od 18 miesięcy do 8 lat. Należy dodać, że 80 procent przypadków dotyczyło fałszowania wyników badań. W tym okresie NIH przyznał nieco ponad 150 tys. grantów, a więc jedna wykryta nierzetelność przypadała na blisko 2000 grantów. Po ośmiu latach ilość zarejestrowanych przypadków wzrosła do ponad 1500, z tego w 20% spraw było konieczne wszczęcie formalnych procedur, a w około 100 przypadkach (7%) udowodniono winę. Najwyższą dotąd karę stanowiło orzeczone w 2005 roku dożywotnie pozbawienie możliwości ubiegania się o finansowanie ze źródeł federalnych, co w przypadku uczonego stanowi śmierć cywilną. W opublikowanym kilka miesięcy temu raporcie za 2004 rok stwierdzono winę w 12 z pośród 29 przekazanych do tej instancji spraw, przy czym wszystkie dotyczyły fałszowania lub manipulacji wynikami, a żadna plagiaryzmu. Procedura w ORI trwa średnio 9 miesięcy, co trzeba dodać do czasu w instancji pierwszej, który jest z zasady dłuższy.

Ogólnie jednak brakuje dobrych danych odnośnie stopnia wykrywalności przypadków nierzetelności. Badania opublikowane w *American Scientist* w 1993 roku, wskazywały, że około 8% respondentów znało przypadki plagiaryzmu lub fabrykowania danych w swoim

otoczeniu. Badania przeprowadzone w Norwegii w 1995 roku wśród 300 przypadkowo wybranych respondentów wykazały, że 60% z nich wiedziało o naruszeniach zasad rzetelności naukowej w swoim otoczeniu, przy czym 22% znało poważne naruszenia, a 9% przyznało, że osobiście było w takie przypadki zamieszanych. Tym nie mniej Norweski Komitet ds. Nierzetelności Naukowej w ciągu pierwszych pięciu lat swojego istnienia rozpatrywał jedynie 9 przypadków, z czego tylko w dwóch potwierdzono winę.

Badania skali zjawiska są z oczywistych powodów bardzo trudne. W czerwcowym numerze *Nature*⁷ omówiono wyniki pierwszych szeroko zakrojonych badań, które zostały przeprowadzone na kilkudziesięciu populacji pracowników naukowych różnych szczebli w obszarze badań finansowanych w USA przez NIH. Wyniki były szokujące. Okazało się, że aż 33% anonimowych respondentów przyznało się do popełnienia w ciągu ostatnich trzech lat co najmniej jednego z listy 10 zachowań uznanych za naruszenie zasad uczciwości naukowej. Jeżeli zdamy sobie sprawę, że z taką powszechnością zjawiska mamy do czynienia w kraju, w którym już od wielu lat prowadzi się działania zapobiegawcze, to możemy tylko przypuszczać jak sytuacja wygląda tam, gdzie takich działań dotąd wcale nie podjęto, tak jak na przykład w Polsce.

W Chinach, które w badaniach naukowych zaczynają odgrywać coraz większą rolę i poszukują możliwości współpracy naukowej z innymi krajami, sprawę potraktowano poważnie. Ostatnio *Science*⁸ podało, że The National Natural Science Foundation of China (NSFC) zbadała w ciągu ostatnich dwóch lat 542 doniesienia anonimowych *whistleblowerów* potwierdzając winę 59 naukowców, w tym 40% stanowiło fałszerstwo, 34% — plagiaryzm, a 7% — fabrykacja wyników. Inne nierzetelności stanowiły 19% przypadków.

Europejscy przedstawiciele nauk ścisłych i przyrodniczych, uznawanych ze względu na swoje oparcie o sprawdzalne doświadczenia za twarde obszary nauki, wciąż jeszcze lansują pogląd, że znikoma ilość nierzetelności wykrywanych w tych dziedzinach świadczy o ich szczególnej uczciwości, a negatywne przykłady z USA są wynikiem specyficznej sytuacji panującej w tym kraju, która sprzyja popełnianiu nierzetelności⁹.

⁷ „Scientists Behaving Badly”, *Nature*, v. 435, 9 June 2005.

⁸ *Science*, Sept. 16, 2005.

⁹ Np. Georg Kreuzberg: „The Rules of Good Science”, *EMBO Reports*, v. 5, nr 4, 2004, pp. 330–332 (www.nature.com/embor/journal/v5/n4/pdf/7400136.pdf).

Kilka ostatnich skandali podważa jednak tą idealistyczną wiarę w integralność nauk ścisłych. Rzeczywiście ujawnionych przypadków jest mniej niż w innych dziedzinach, ale jeżeli już się pojawiają, to zwykle są bardzo drastyczne, jak np. wykryta dwa lata temu sprawa niemieckiego fizyka Jana Hendrika Schöna z Bell Lab pracującego w grupie zajmującej się nadprzewodnictwem i elektroniką, której wyniki stanowiły podstawę zbiorowych publikacji rozważanych przez Komitet Noblowski, a które okazały się fałszerstwem. Drugi głośny przypadek z ostatnich lat związany jest z osobą Victora Ninova z Berkeley National Laboratory, którego „naciągnięta” interpretacja obserwacji spowodowała ogłoszenie wykrycia pierwiastka o liczbie atomowej 118. Co prawda obydwie przypadki zostały dość szybko zde-maskowane, ale stało się tak głównie dlatego, że dotyczyły bardzo „gorących” tematów badań. W każdym razie nierzetelności nie omijają samych szczytów nauki. Niedawno *New Scientist*¹⁰ doniósł, że profesor biologii Luk van Parijs z MIT został zwolniony z pracy w związku z wykryciem fałszowania i fabrykowania przez niego danych doświadczalnych, przedstawionych w artykułach opublikowanych w drugiej połowie lat dziewięćdziesiątych. Granty, które van Parijs uzyskał na swoje badania ze źródeł federalnych przekroczył kwotę 2 milionów USD. Najgłośniejszym, szeroko opisywanym w prasie codziennej, skandalem z ostatnich miesięcy jest sprawa koreańskiego biologa Woo Suk Hanga.

Musimy jednak pamiętać, że wraz z oddaniem się od głównych i najbardziej aktualnych centrów zainteresowania nauki możliwość wykrycia nierzetelności gwałtownie spada. Jak jest w Polsce? Na ten temat brak jakichkolwiek informacji, poza ujawnianymi przypadkami wulgarnych plagiatów.

Systemy ochronne

W wolnych społeczeństwach — a zwłaszcza w otoczeniu akademickim, gdzie kreatywność i indywidualna myśl stanowią szczególnie pielęgnowaną wartość, której nie można tłumić — nie ma pełnej możliwości

zapobieżenia naruszeniom zasad przez jednostki. Nie-realistyczne jest więc oczekiwanie, że można naukę uchronić przed patologicznymi przypadkami. Konieczne jest zatem tworzenie klimatu, który promuje skrupulatność i przywiązanie do przestrzegania wysokich standardów, bez hamowania produktywności i kreatywności uczonego¹¹.

Pojawia się tutaj dodatkowo nierozwiązany dotąd problem granicy istniejącej między wolnością akademicką a nierzetelnością naukową. Jest on szczególnie istotny w kontekście przyzwolenia na uprawianie w niektórych jednostkach naukowych różnego typu paranauk oraz przyzwolenia na tandetność i pozorność badań naukowych, a więc z uprawianiem nauki śmieciowej, nie mającej nic wspólnego z rzetelnym procesem poznawczym. Ujawnia się w tym miejscu trudny do rozwiązania konflikt między społecznym interesem zapewnienia badaniom naukowym najwyższych standardów a samorządnością jednostek naukowych, a także konflikt między globalnym charakterem nauki a narodowym charakterem instytucji naukowych.

Od wielu lat w różnych środowiskach naukowych, w tym również w Polsce, podnoszono potrzebę opracowania swoistego **Kodeksu Etyki Naukowej**. Takich kodeksów powstało wiele. W Polsce znany jest kodeks etyczny przyjęty przez PAN¹² oraz bardzo piękny dokument Senatu Uniwersytetu Jagiellońskiego¹³. Jednak nie w opracowywaniu kodeksów etycznych określających ogólne zasady działania naukowca leży istota problemu, gdyż to co jest naprawdę potrzebne środowisku naukowemu to są **kodeksy dobrej praktyki naukowej** (*good research practice*)¹⁴, a więc zbiory reguł rzetelnego postępowania, które są powszechnie zrozumiałe i możliwe do wprowadzenia w poszczególnych jednostkach, oraz procedur postępowania w przypadku ujawnienia naruszeń tych reguł. Reguły takie powinny zawierać uzgodnione i zaakceptowane przez środowisko precyzyjne definicje i jasne podstawowe **zasady pracy naukowej** (w tym m.in. zasady kierowania badaniami, uwzględnianie potrzeb młodych badaczy, zasady zapisywania i przechowywania pierwotnych danych

¹⁰ NewScientist.com news service; 28 Oct. 2005.

¹¹ Np. The Maintenance of High Ethical Standards in the Conduct of Research, Association of American Medical Colleges.

¹² *Dobre Obyczaje w Nauce. Zbiór zasad i wytycznych*, wydanie III, PAN, Warszawa 2001.

¹³ www.uj.edu.pl/universytet/wladze/kodeks.pdf.

¹⁴ American Association for the Advancement of Science Professional Ethic Project, Publication 8-R4 1980.

doświadczalnych, zasady autorstwa oraz ujawniania konfliktu interesów), a także, co jest niezwykle ważne, **procedury postępowania** w przypadku pojawienia się zarzutów dotyczących naruszenia tych zasad.

Potrzebę istnienia takich kodeksów dobrej praktyki można uzasadnić postępując się przykładem motoryzacji. Gdy samochodów było mało, wystarczyło ogłosić, że kierowca musi zachowywać się ostrożnie i przyzwoicie, zgodnie z ogólnymi zasadami etyki czy dekalogu. Jednak gdy natężenie ruchu wzrosło, te ogólne zasady przestały wystarczać, w związku z czym pojawiła się konieczność wprowadzenia kodeksu ruchu drogowego, zawierającego precyzyjnie zdefiniowane nakazy i zakazy, procedury niezbędne dla zapewnienia jego skuteczności oraz przewidującego sankcje, bo inaczej prawdopodobieństwo wypadków, w tym również z udziałem tych ostrożnych i przyzwoitych, zagrożiłoby możliwość ruchu.

Taką właśnie sytuację mamy w nauce od kilkunastu lat. Dlatego we wszystkich cywilizowanych krajach wprowadzono systemy, które z jednej strony starają się tworzyć klimat sprzyjający utrzymywaniu zasad dobrej praktyki naukowej, a z drugiej określają procedury postępowania w przypadku ujawnienia naruszenia tych zasad¹⁵. Istnieje powszechne przekonanie, że w interesie społecznym, a również w interesie samej nauki i jej społecznego autorytetu, wszystkie sprawy dotyczące podejrzeń o naruszenie rzetelności naukowej muszą być starannie zbadane i rozstrzygnięte.

Dodatkowy czynnik zmuszający do regulaminowego uregulowania tych problemów stanowi fakt, że większość badań naukowych jest finansowana z grantów pochodzących bądź ze źródeł publicznych, bądź od organizacji prywatnych, a więc nieuczciwość badawcza może być przez fundujące instytucje potraktowana jako sprzeniewierzenie przyznanych środków. Przeważa pogląd, że trzy najważniejsze nierzetelne zachowania, czyli **fabrykacja, fałszowanie i plagiatyzm**, które naruszają fundamentalne zasady badań naukowych, powinny podlegać regulacjom prawnym na poziomie instytucji odpowiedzialnych za dystrybucję środków

finansowych, gdyż za najwłaściwszą sankcją w przypadku udowodnienia winy uznaje się wstrzymanie finansowania lub żądanie zwrotu środków. Natomiast pozostałe podejrzane praktyki, ze względu na ich niższą szkodliwość, mogą znajdować się w bezpośredniej jurysdykcji społeczności naukowej.

Aby zaproponowany system był skuteczny musi się opierać na podstawowej zasadzie — odpowiedzialność za zapobieganie nierzetelności naukowej spoczywa na społeczności naukowej jako całości, a więc zarówno na uczestnikach procesu badawczego (są to studenci, doktoranci, pracownicy oraz kierownicy zespołów i instytucji badawczych), na instytucjach naukowych (uczelnie, instytuty, stowarzyszenia i organizacje naukowe) oraz na agendach rządowych i pozarządowych działających w obszarze nauki. Szczególną rolę mają do spełnienia jednostki posiadające uprawnienia do nadawania stopni i tytułów naukowych, gdyż nierzetelności najlepiej zapobiega wysokie prawdopodobieństwo jej ujawnienia na najwcześniejszym etapie kariery naukowej i publicznego przedstawiania wyników badań.

Pierwszym krajem, w którym systemowo podjęto problem były Stany Zjednoczone, gdzie na początku lat dziewięćdziesiątych przy poszczególnych instytucjach rządowych finansujących badania naukowe utworzone zostały stałe biura odpowiedzialne za monitorowanie postępowań związanych z nierzetelnościami naukowymi wykrytymi w obszarze ich finansowania¹⁶. Do zadań tych biur należy również wprowadzanie dobrych praktyk w badaniach naukowych przez edukację, profilaktykę oraz opracowywanie wytycznych, jak też udzielanie wsparcia. W celu uzyskania prawa dostępu do publicznych źródeł finansowania wszystkie instytucje prowadzące badania naukowe musiały wykazać się wprowadzeniem u siebie odpowiednich regulaminów i procedur wewnętrznych oraz zapewnić odpowiednią edukację wszystkich osób związanych z badaniami naukowymi¹⁷.

Tak więc główna odpowiedzialność za rozpatrywanie spraw związanych z zarzutami nierzetelności naukowej spoczęła na uniwersytetach i instytutach

¹⁵ Klasyczną pozycję stanowi opracowana przez Akademię Nauk USA książeczka: *On Being a Scientist: Responsible Conduct in Research*, National Academic Press, wydana po raz pierwszy w 1988 roku.

¹⁶ Np. w strukturze National Institutes of Health jest nim Office of Research Integrity (www.ori.hhs.gov) a National Science Foundation — Office of Inspector General (www.nsf.gov/oig).

¹⁷ Bardzo dobra strona edukacyjna z Case Reserve University: <http://www.onlineethics.org/>; *Kurs dobrej praktyki naukowej* z Uniwersytetu San Diego: <http://www.sci.sdsu.edu/~smaloy/ethics/#other>.

naukowych, które w przypadku potwierdzenia winy zostały zobowiązane do przedstawienia wyników dochodzenia agencji finansującej badania. Na tej podstawie agencja określa sankcje polegające zwykle na wstrzymaniu finansowania lub zamknięciu dostępu do środków na określony czas (zwykle od kilkunastu miesięcy do dziesięciu lat).

Największą trudność stanowi jednak wykrywanie naruszeń rzetelności. Praktyka pokazała, że na wczesnych etapach badań najważniejszą rolę odgrywają *whistleblowerzy*, czyli osoby, które działając w dobrej wierze i w interesie publicznym ujawniają zaobserwowane w swoim otoczeniu przypadki nierzetelności. **Trzeba tu podkreślić, że whistleblower nie jest donosicielem, ale sygnalizatorem zbliżającego się nieszczęścia, które, jeżeli się zdarzy, uderza w całe środowisko.** Wiąże się z tym bardzo poważny problem ochrony *whistleblowera* przed ewentualnymi represjami otoczenia czy kierownictwa, bez którego żaden system dobrej praktyki nie będzie skutecznie działał.

W ciągu kilku lat podobne zasady wprowadzono w innych krajach¹⁸. W Niemczech i Anglii przyjęto zbliżone do amerykańskiego zdecentralizowane systemy, pozostawiając rozpatrywanie spraw związanych z przypadkami nierzetelności w kompetencji instytucji naukowych z instancją odwoławczą i monitorującą na poziomie rządowych agencji finansujących badania. W krajach skandynawskich natomiast całość procedury została zcentralizowana przez przeniesienie jej na poziom kompetencji specjalnie w tym celu powołanej instytucji rządowej o uprawnieniach śledczych. W wielu krajach sprawę jednak zbagatelizowano, wychodząc z założenia, że nauka z definicji jest uczciwa, a jej niezwykłe zdolności samokorygujące chronią społeczeństwo przed skutkami błędów i nierzetelności, a więc nie ma w ogóle o czym mówić, tym bardziej, że mówienie o tych sprawach publicznie podważa autorytet nauki. Warto wspomnieć, że podobne poglądy dominowały też w Niemczech, aż do serii skandali ujawnionych po 1997 roku, i nie są obce w Polsce.

Jak wspominałem, system niemiecki jest bardzo zbliżony do amerykańskiego, przy czym wprowadzono w nim interesującą modyfikację, polegającą na ustanowieniu instytucji *ombudsmana* (powołały ją zarówno DFG, jak i Towarzystwo Maxa Plancka) jako trzeciej strony w postępowaniu. Osoby, które dostrzegły w swoim otoczeniu przejawy naruszania przyjętych zasad dobrej praktyki mogą się do niego zwracać z zaufaniem o pomoc lub radę. W przypadku uznania zgłoszonego przypadku za dostatecznie poważny *ombudsman* występuje do jednostki naukowej o wszczęcie sprawy. Sam jednak dochodzenia nie prowadzi. W tym systemie osoba *whistleblowera* jest nie tylko dobrze chroniona, ale również zdjęto z niej psychiczne obciążenie działania w izolacji, które niejednokrotnie zniechęca do podjęcia działania.

Zwolennicy modelu zdecentralizowanego podkreślają, że mieści się on w duchu autonomii akademickiej i pozwala dobrze odnieść się do specyfiki powstających problemów. Krytycy uważają, że prowadzenie procedur wewnątrz jednostek badawczych może wywoływać ogromne podziały i zakłócenia w ich pracy, co zwiększa skłonność kierownictw jednostek do „zamiatania spraw pod dywan” i zniechęca ewentualnych *whistleblowerów*, jednak pozostawienie dochodzenia w zakresie kompetencji jednostek badawczych uczyła je na niestety ważny problem prewencji, a więc rygorystyczne przestrzeganie zasad dobrej praktyki. Zwolennicy modelu zcentralizowanego uważają natomiast, że wprowadza on do systemu niezależny od jednostek naukowych bezstronny organ, co ich zdaniem powinno pomóc w przewyżnianiu oporów w wyjaśnianiu spraw związanych z naruszeniami rzetelności. Wydaje się, że przyjęcie jednego z tych dwóch rozwiązań warunkowane jest między innymi również wielkością środowiska naukowego

Podjmując pracę nad określeniem — do zastosowania w Polsce — zasad dobrej praktyki naukowej oraz właściwych procedur postępowania należy uwzględnić prace prowadzone od wielu lat w innych krajach¹⁹.

¹⁸ Np. dokument DFG: *Recommendation of the Commission on Professional Self regulation in Science*, Jan. 1998 (http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/self_regulation_98.pdf); podobne: The Max Planck Society (<http://www.mpg.de/pdf/procedScientMisconduct.pdf>), The Danish Committee on Scientific Dishonesty (1992) (www.forsk.dk) lub Decree on the Research Ethics Council wydany przez Rząd Finlandii (1991) (<http://www.pro.tsv.fi/tenk/guidelines.htm>).

¹⁹ Kompetentne omówienie poglądów odnoszących się do zasad i procedur dobrej praktyki naukowej przedstawiono w Arthur Rörsch (red.): *Good Scientific Practice* (<http://www.stichting-han.nl/Commentaren/algemeen/7.%20Good%20scietific%20Practice1.doc>).

W kontekście naszego członkostwa w Unii Europejskiej szczególną uwagę trzeba zwrócić na ustalenia i dokumenty organizacji europejskich²⁰. Względy praktyczne wynikające ze wzrostu znaczenia finansowania ze źródeł unijnych skłaniają do unifikacji podejścia. Choć nie ulega wątpliwości, że stosowanie w całej Europie zbliżonych zasad stanie się w nieodległej przyszłości koniecznością, to mimo podobnego podejścia do problemu uzgodnienie wspólnego tekstu zasad natrafia na trudności, chociażby ze względu na różne systemy prawne i organizacyjne przyjęte w poszczególnych krajach.

Polskie instytucje naukowe takich dokumentów i procedur jeszcze nie mają. Rekomendacje w tym zakresie zostały przedstawione w opracowanym przez Zespół ds. Etyki w Nauce przy Ministrze Nauki dokumencie *Dobra praktyka badań naukowych. Rekomendacje*²¹.

Wnioski

W Polsce w ciągu ostatnich lat odbyło się wiele dyskusji publicznych oraz konferencji, na których dyskutowano liczne problemy etyki naukowej, jednak raczej w kategoriach ogólnych zasad niż praktyki dnia codziennego. Niektóre uczelnie uchwały swoje *kodeksy etyczne*. Podobnie postąpiła Polska Akademia Nauk. Kodeksy te z punktu widzenia potrzeby wprowadzenia mechanizmów zapewniających integralność badań naukowych są jedynie szlachetnymi ale nieskutecznymi deklaracjami, gdyż nie mają charakteru obowiązującego, nie definiują zasad niezbędnych dla przestrzegania dobrej praktyki oraz nie określają procedur ani sankcji

w przypadku stwierdzenia naruszenia tych zasad. Niestety znaczna część środowiska naukowego traktuje sprawę z obojętnością i lekceważeniem, a mówiąc ogólnie, nie jest ono przygotowane do podjęcia dyskusji. Co więcej, relatywizacja wartości (*co nie jest zabronione przez prawo, jest dozwolone*), a także ogólne nadwyższenie elementarnych zasad uczciwości oraz odziedziczona po PRL społeczna tolerancja dla małych występów, znajdują swoje odbicie również w praktyce życia naukowego. Poza tym, w przeciwieństwie do innych krajów, nie mamy dotąd struktur właściwych do zajmowania się problemem, a z reguły nieprofesjonalne komisje dyscyplinarne istniejące w uczelniach czy innych instytucjach naukowych nie są merytorycznie przygotowane do prowadzenia niezwykle złożonych spraw związanych z niezetelnością naukową. Nie został też podjęty problem ochrony *whistleblowerów*.

Przed środowiskiem nauki w Polsce stoi więc obecnie wyzwanie związane z koniecznością wprowadzenia u nas zasad dobrej praktyki naukowej.

Wynika to z dwóch przestanków:

- po pierwsze — z troski o zapewnienie polskiej nauce integralności,
- po drugie — z potrzeby wprowadzenia w polskich jednostkach naukowych regulacji i procedur, bez których współpraca międzynarodowa stanie się wkrótce niemożliwa.

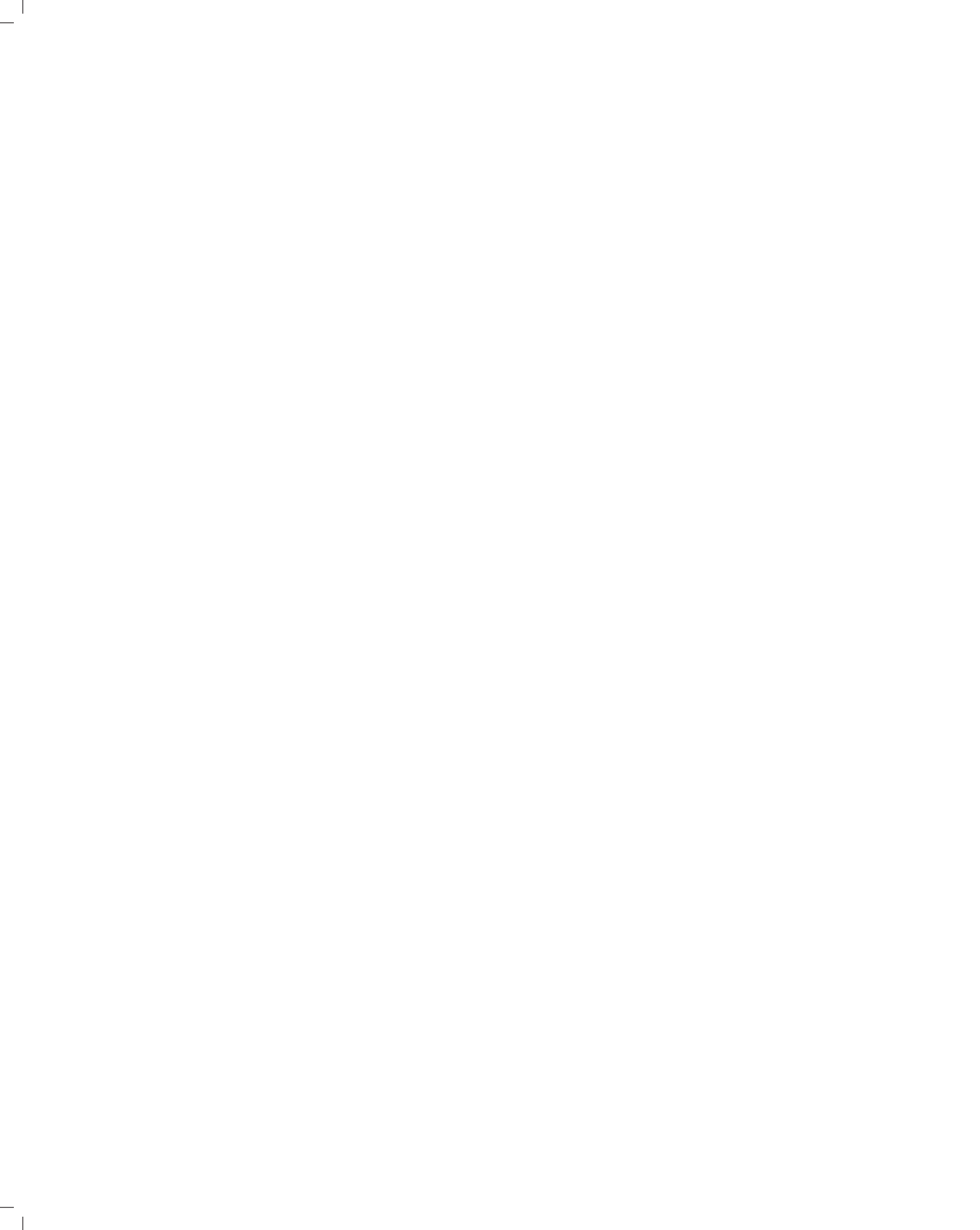
Dlatego kończę swoje wystąpienie nie wnioskami ale apelem, aby Uczelnia, z własnej inicjatywy, nie czekając na zalecenia Ministerstwa, podjęła trud opracowania i wprowadzenia w życie swoich własnych zasad dobrej praktyki naukowej.

²⁰ Np.: *Safeguarding Good Scientific Practice*, A Joint Statement by the Director General of the Research Councils and the Chief Executives of the UK Research Councils, 18 Dec. 1998; dokument DFG: *Recommendation of the Commission on Professional Self Regulation in Science*, Jan. 1998, *Guidelines for the Prevention, Handling and Investigation of Misconduct in Science* wydany w 1994 roku przez National Research Ethics Council of Finland lub *Guidelines* opracowane przez European Science Foundation (<http://www.esf.org/sciencepolicy/170/ESPB10.pdf>) i inne.

²¹ *Dobra praktyka badań naukowych. Rekomendacje*, Wyd. MNiI, Warszawa 2004 (www.mnii.gov.pl/mnii/_gAllery/29/90/2990.pdf).

Profesor Maciej Władysław Grabski

polski metaloznawca, emerytowany profesor Politechniki Warszawskiej. Autor prac z zakresu nauki o materiałach i inżynierii materiałowej, zajmował się głównie defektami sieci krystalicznej i ich wpływem na właściwości metali oraz stabilnością mikrostruktury. W latach 1991–1994 przewodniczył Zespołowi Nauk Technicznych w Komisji Badań Podstawowych Komitetu Badań Naukowych, a w latach 1994–1997 był wiceprzewodniczącym Komisji Badań Stosowanych. Członek Zespołu ds. Etyki w Nauce przy Ministrze Nauki (2000–2008) i jego przewodniczący (2004–2008). W latach 1992–2005 był prezesem Fundacji na Rzecz Nauki Polskiej, a obecnie jest członkiem Rady Fundacji. Został uhonorowany wieloma nagrodami i wyróżnieniami, m.in. Krzyżem Kawalerskim, Krzyżem Oficerskim i Krzyżem Komandorskim Orderu Odrodzenia Polski oraz dwukrotnie tytułem *Doktor honoris causa*.



Problem erozji ładu moralnego w świecie¹

Na podstawie odczytu wygłoszonego w dniu 16 marca 2006 roku

Ks. Andrzej Szostek

Katolicki Uniwersytet Lubelski

73

Ci, którzy mieli łacinę w szkole pamiętają może, że jednym z tekstów zadawanych zwykle do pamięciowego opanowania był początek *Metamorfoz* Owidiusza. Recytowaliśmy i poddawaliśmy gramatycznej obróbce łatwy rytmicznie, ale skomplikowany stylistycznie, heksametr rzymskiego poety: *Aurea prima sata est aetas que vindice nullo, sponte sua sine lege fidem rectumque colebant*. Nie zatrzymywaliśmy się zbyt uważnie na treści analizowanych wersów, lektura tego dzieła kończyła się zresztą zwykle na krótkim wstępnym jego fragmencie. Tymczasem Owidiusz cytowanymi słowami rozpoczyna opis czterech wieków ludzkości, z których pierwszym był wiek złoty, kiedy ludzie nie potrzebowali prawa i jego sankcji, by zachowywać wiarę i obyczaje (*sine lege fidem rectumque colebant*). Po nim nastat wiek srebrny, już gorszy, potem kolejne, znaczone coraz marniejszymi odmianami kruszcu. Owidiusz nie jest historiozofem ani prorokiem interpretującym dzieje ludzkości w jakiś oryginalny, niezwykle sposób. Raczej przeciwnie, daje wyraz dość potocznej już wówczas tendencji do idealizowania przeszłości, w porównaniu z którą teraźniejszość wydaje się błada i kiepska (a warto wspomnieć, że życie i twórczość tego poety przypadły na czas panowania Oktawiana Augusta, kiedy kwitła literatura rzymska, a Owidiusz był już za swego życia uznany za jednego z jej najświetniejszych przedstawicieli).

Narzekać na czasy współczesne — od pogody po moralność — należy do żelaznego repertuaru tematów luźnej konwersacji z sąsiadami lub z przygodnie poznanymi pasażerami w pociągu. Czyż nie jesteśmy świadkami smutnego procesu moralnej degradacji społeczeństwa? Czy za naszych młodych lat ludzie nie zachowywali się bardziej przyzwoicie? Stawiający takie — zwykle retoryczne — pytania nie zawsze umieliby odpowiedzieć, co przez moralność rozumieją, ale nie ulegamy pokusie ich pochopnego dyskredytowania jedynie z powodu kłopotów definicyjnych. Odkładając na później próbę dookreślenia istoty moralności i ładu moralnego zgódźmy się wstępnie, że ci, którzy uważają dzisiejsze czasy za moralnie gorsze od minionych, przez moralny ład rozumieją zazwyczaj akceptowany powszechnie i w praktyce przez większość przestrzegany system norm, które wiążą się jakoś z człowiekiem i jego wyróżnionym miejscem w świecie — stoją na straży wartości kojarzonych z godnością ludzkiej osoby i godnej człowieka spoteczności. Oczywiście, normy te w dawnych, dobrych czasach uznawała większość — nie wszyscy. Przestępców, zbrodniców i innych zbreźników nigdy nie brakowało — ale ich zachowanie było powszechnie potępiane w imię pewnych moralnych oczywistości, których nie trzeba (a bodaj i nie bardzo można) było głębiej uzasadniać. Tak właśnie, jak to krótko ujął Owidiusz: *sine lege fidem rectumque*

¹ Jest to zmodyfikowana wersja odczytu nt. *Problem erozji ładu moralnego*, wygłoszonego 21 stycznia 2005 r. w Studium Myśli Chrześcijańskiej przy Ośrodku Kultury Chrześcijańskiej w Poznaniu.

colebant. Moralne zepsucie polega nie tylko na tym, że zła czyni się więcej, ale także na tym, że nie spotyka się ono z tak zdecydowanym potępieniem jak dawniej. Tu przytacza się przykłady mające ten moralny upadek ilustrować — od coraz krótszych spódniczek, przez chamstwo w życiu publicznym, bezkarność kradzieży i wszelkich oszustw, aż po okrucieństwa wojenne oraz zbyt łagodne traktowanie przestępców, których prawo zdaje się lepiej chronić niż ich ofiary.

Ten typ myślenia łatwo jest krytykować, wskazując na wspomnianą niejasność samej istoty moralności (utożsamianej zbyt pochopnie z nieuchronnie zmieniającymi się obyczajami i zwyczajami), wybiórczo dobrane przykłady (czyż przeszłość także nie obfituje w okrucieństwa?), aż po elementarną niemożliwość uzasadnienia tej pesymistycznej wizji dziejów. Nikt z nas nie jest Panem Bogiem, by móc ocenić intencje, którymi kierują się ci, którzy nas swym zachowaniem gorszą, a których — być może — tłumaczą warunki, w jakich kształtowała się ich osobowość (środowisko wychowawcze, bieda itp.). Przyznając, że opinie takie są dość tendencyjne i stanowią często upust dla osobistych frustracji tych, którzy je wypowiadają, zgódźmy się jednak, że przeświadczenie, iż stacamy się po moralnej „równi pochyłej” nie jest pozbawione podstaw. Często mamy poczucie, że wiele kiedyś uznawanych, a nawet kulturowanych wartości (patriotyzm, poszanowanie autorytetu, zwykła kultura w życiu osobistym itp.) dziś się lekceważy i znieważa, a czasy obecne są trudniejsze od wieków minionych.

Godzi się jednak podkreślić, że takiej pesymistycznej wizji dziejów nie podzielają wszyscy. Nie mniej popularna — choć w nieco innych kręgach — jest wizja przeciwna. Nazwijmy ją „optymistyczną”. Za jej patrona uznać można G.F. Hegla. W swej *Fenomenologii ducha* przedstawia on szeroką panoramę dialektycznego rozwoju ludzkości, która (przez kolejne wielorakie doświadczenia) zmierza od etapu Ducha Subiektywnego, przez kształtowanie Ducha Obiektywnego, aż do

osiągnięcia finału — urzeczywistnienia Ducha Absolutnego². Nie miejsce tu na szczegółową prezentację poglądów Hegla, trzeba jednak podkreślić, że heglowska wizja historii wspierana jest niejako ogólną atmosferą naukową — rozwojem wiedzy o świecie i jego technicznym opanowaniu, a także dość powszechnie przyjętym we współczesnym przyrodoznawstwie paradygmatem kosmicznej ewolucji (co w pewnym sensie oznacza także rozwój) wszechświata, w którą „wpisany” jest człowiek i cała tworzona przezeń cywilizacja. Echa tej wizji znaleźć można także w myśli niektórych teologów, z P. Teilhardem de Chardin na czele³. Myśmy mieli wątpliwe szczęście zapoznać się bliżej z marksofską, materialistyczną odmianą heglizmu, w ramach której przekonywano nas, że ludzkość zmierza nieuchronnie drogą postępu. Jego finalem miała być bezpaństwowa i bezklasowa era komunizmu, uznanego za historyczną konieczność, z którą — jako nieuniknioną — lepiej raczej współpracować, niż się jej bezskutecznie przeciwstawiać⁴. Ten typ myślenia bliższy jest raczej środowiskom akademickim oraz niektórym działaczom społecznym, skłonny do upraszczającego interpretowania trudnych filozoficznych idei na użytek swych politycznych celów. „Optymiści” uważają, że o żadnej erozji moralnego porządku nie należy mówić. Owszem, w krótszej i partykularnej perspektywie możliwe są różnorakie kryzysy. Zapewne — jak chciał Lenin i jego poplecznicy — droga do szczęśliwego finału musi wieść przez nieprzyjemną fazę dyktatury proletariatu, towarzyszący mu terroryzm i inne drobne niedogodności⁵, w skali globalnej jednak nie tylko nie mamy do czynienia z erozją moralnego ładu, ale wręcz z procesem moralnego (jak i ekonomicznego) dojrzewania ludzkości.

Nasze własne doświadczenia z drugiej połowy XX wieku nie nastrajają nas zbyt przychylnie do tej rzekomo optymistycznej wizji rozwoju świata, tym niemniej i ona ma za sobą wiele wartych uwagi atutów. Najważniejszym z nich jest — jak sądzę — proces dojrzewania świadomości moralnej, o jakim zdaje się świadczyć

² Por. G.W.F. Hegel, *Fenomenologia ducha*, przeł. A. Landman, Warszawa 1963, 1065.

³ Por. Cz. Bartnik, *Teilhardowska wizja dziejów*, Lublin 1975.

⁴ Krytyczną analizę tego poglądu szczególnie wnikliwie przedstawia L. Kołakowski. Por. tegoż, *Główne nurty marksizmu. Powstanie — rozwój — rozkład*, Londyn 1988, s. 281–315.

⁵ *Zawsze wiedzieliśmy, mówiliśmy, powtarzaliśmy, że socjalizmu nie można „wprowadzić”, że wyrasta on w toku najbardziej napiętej, najbardziej ostrej, zaciekle, rozpaczliwie ostrej walki klasowej i wojny domowej — że między kapitalizmem a socjalizmem leży długi okres „bólów porodowych” — że przemoc jest zawsze akuszerką starego społeczeństwa — że okresowi przejściowemu od społeczeństwa burżuazyjnego do socjalistycznego odpowiada szczególnego rodzaju państwo (czyli szczególnie system zorganizowanej przemocy w stosunku do pewnej klasy), mianowicie dyktatura proletariatu, W. Lenin, *Dzieła*, t. 26, Warszawa 1956, s. 404.*

historia naszej cywilizacji. Autor *Drugiej Księgi Samuela* rozpoczyna opowiadanie o grzechu Dawida z Bat-szebą od słów: *Na początku roku, gdy królowie zwykli wychodzić na wojnę...*⁶. W owych czasach był bowiem przyjęty miły zwyczaj, by zamiast samemu truduć się hodowlą i uprawą roli, najechać od czasu do czasu na bardziej pracowitych sąsiadów i zebrać ich plony. Nie tak znów dawno temu w strukturze rządów poszczególnych państw jednym z ważnych resortów było ministerstwo wojny, teraz nazywane jest ono z reguły ministerstwem obrony narodowej lub podobnie. Czy nie kryje się za tym świadomość — dziś powszechna, kiedyś bynajmniej nie — że wojna jest wielkim złem, przed którym trzeba się bronić, ale którego nie wolno samemu wszczynać? Jeszcze w czasie II wojny światowej podejmowano działania wojenne, których jedynym celem było złamanie ducha narodu (np. dywanowe naloty bombowe na Drezno). Dziś liczne państwa podpisują umowy zabraniające takich działań, a prezydent J. Bush musi się gęsto tłumaczyć z każdej militarnej akcji w Iraku, która pociągnęta za sobą śmierć cywilów. Także stosowane ongiś powszechnie tortury zostały oficjalnie wykluczone przez państwa, które chcą uchodzić za cywilizowane. Owszem, przypadki stosowania tortur nadal się zdarzają (o czym świadczą m.in. szokujące doniesienia z Iraku), nadal podejmuje się wojny dla osiągnięcia własnych korzyści kosztem słabszych, ale nie robi się tego jawnie i oficjalnie, a ci, którzy dopuszczają się takich aktów przemocy, muszą się liczyć z groźącymi za to surowymi sankcjami. To nie jest mało, bowiem aby złó móc skutecznie zwalczyć trzeba je najpierw rozpoznać i uznać za złó. Jakże nie wspomnieć w tym kontekście o znaczeniu uświadomienia sobie rangi wolności osobistej oraz prawa narodów do samostanowienia! Rozpowszechnione w starożytności niewolnictwo nie budziło zastrzeżeń nawet u świątłych i demokratycznie nastawionych obywateli. Jeszcze w czasach nowożytnych rozwijał się kolonializm, obecnie zaś nikogo nie trzeba przekonywać, że pozbawienie niewinnego człowieka lub narodu wolności uwłacza jego godności. My chlubimy się słusznie królem Zygmuntem Augustem, który na wieść o przyjętej w 1555 roku w Augsburgu zasadzie *cuius regio eius religio* miał zawołać: *Czyż jestem królem ludzkich*

sumień?! Oczywista dla nas zasada wolności sumienia i wyznania bynajmniej nie była taką w XVI wieku. Przykłady takiego postępu świadomości moralnej można długo mnożyć.

Pomimo tych niewątpliwych atutów „optymistycznej wizji dziejów”, także jej zwolennikom można postawić zarzut tendencyjnej interpretacji procesów historycznych. Owszem, trudno niedoceniać procesu dojrzewania moralnego rozumianego jako dojrzewanie wartości moralnych wcześniej nie dostrzeganych. Czy jednak nie towarzyszy mu także proces przeciwny? Oczywiście przez wieki przekonanie, że ludzka istota zasługuje na szacunek i ochronę w każdej fazie swego istnienia, dziś jest dość powszechnie podważane przez ustawodawstwa uznające za legalnie dopuszczalne szereg przypadków aborcji, a ostatnio także i eutanazji. Nie można tłumaczyć tego tylko odróżnieniem porządku prawnego od moralnego. Prawo rzeczywiście nie może przewidywać sankcji karnych za wszelkie czyny moralnie naganne, jednak życie ludzkie jest tym dobrem fundamentalnym, którego ochrona jest niejako pierwszym obowiązkiem prawa⁷. Podobnie wygląda ewolucja wrażliwości w dziedzinie moralności matżeńskiej i — szerzej — seksualnej. Czy tendencja do zrównania związków homoseksualnych z matżeńskimi nie jest świadectwem zatracenia świadomości tego, co jest istotą matżeńskiej miłości i jaka jest rola płci i seksu? Czy nie jesteśmy świadkami takiej apoteozy wolności, że zapomina się, że człowiek jest wolny po to, by móc ze swoją wolnością coś dobrego czynić, a nie tylko ją manifestować, utożsamiając ją praktycznie z dowolnością?

Do niektórych ze wspomnianych przykładów i pytań trzeba będzie jeszcze powrócić, już teraz jednak można powiedzieć, że obie omówione propozycje wizji dziejów — pesymistyczna i optymistyczna — rażą swą jednostronnością. Doświadczenie ludzkości jest tak bogate, że bez trudu można z niego wyławiać przykłady potwierdzające oba te (w istocie ideologiczne) założenia. Czy jesteśmy skazani na taką ideologię, dla której alternatywą jest tylko sceptycyzm powstrzymujący się od jakiegokolwiek odpowiedzi na nasze tytułowe pytanie? Może jednak nie. Oprócz przywołanych dotąd poglądów, wzajemnie przeciwstawnych, choć podobnych w tym, że stanowią pochodną z góry przyjętych

⁶ 2 Sm 11,1.

⁷ Por. T. Styczeń, *Homo homini res sacra. W sprawie aksjologicznej zasady Konstytucji Rzeczypospolitej Polskiej*, w: tenże, *Solidarność wyzwała*, Lublin 1993, s. 137–141.

założeń, znana jest jeszcze co najmniej jedna warta uwagi wizja dziejów ludzkości. Znaleźć ją można w myśli chrześcijańskiej — zwłaszcza w zamykającej kanon *Pisma Świętego Księdze Apokalipsy*. Jest to księga pocieszenia (choć potocznie określenie „apokaliptyczny” wywołuje raczej przerażenie niż otuchę), którą jej autor dedykuje przeżywającym wielki ucisk chrześcijanom. Zapowiada ona ostateczny triumf Chrystusa i Jego wyznawców. Prezentowana w niej interpretacja dziejów ma więc w tym sensie charakter teologiczny. Zawarte są w niej jednak prawdy o człowieku, które są godne uwagi niejako niezależnie od zasadniczego teologicznego przestania *Księgi*. W ostatnim jej rozdziale Anioł mówi do Jana: *Kto krzywdzi, niech jeszcze krzywdę wyrządzi i plugawy niech się jeszcze splugawi, a sprawiedliwy niech jeszcze wypełni sprawiedliwość, a święty niechaj się jeszcze uświęci*⁸. W dzieje świata wpisane są dzieje poszczególnych ludzi, te zaś zmierzają do ostatecznej autodeterminacji człowieka — za lub przeciw temu, co dobre. Proces tej autodeterminacji biegnie często zygzakiem. Bywa, że sprawiedliwy odstępkuje od swej pierwotnej sprawiedliwości, grzesznik zaś — od swego grzechu⁹, ale i ten „zygzak” nie oznacza prostych kroków wprzód i w tył. Przez takie odstępstwa i akty nawrócenia stopniowo kształtuje się charakter człowieka, ku ostatecznemu zbawieniu lub potępieniu.

Biblijna interpretacja dziejów nie sprowadza się więc ani do pesymistycznej, ani do optymistycznej ich wersji. Obie perspektywy są otwarte — człowiek jest bowiem naprawdę wolny, a zarazem sobie „zadany”. Owszem, zmieniają się warunki, w których żyje. Niektóre ich elementy sprzyjają jego moralnemu rozwojowi, inne zaś stwarzają dlań pokusy lub trudne do pokonania przeszkody. Przez próbę przejść musi jednak każdy. Żadne warunki nie skazują człowieka ani na nieuchronną zgubę, ani na rozwój.

Stwierdzenie, że człowiek jest istotą społeczną, to banał, bowiem nie tylko żyje obok innych, ale wespół z innymi organizuje swe życie na Ziemi. Przynależy też do naturalnych społeczności (rodziny i narodu), sam tworzy struktury życia społecznego, spośród których za najbardziej dojrzałą uważa się zwykle państwo, czerpie soki z zastanej kultury i ubogaca ją własną twórczą aktywnością. Tak więc, jak losy poszczególnych ludzi

przebiegać mogą różnymi szlakami, zmierzając ostatecznie do zdeterminowania własnej osobowości, tak też poszczególne społeczności rozwijają swe dzieje, które przebiegać mogą różnymi meandrami — ale w końcu, w ramach poszczególnych tradycji kulturowych, prowadzą do swoistej ich autodeterminacji.

Spojrzenie na dzieje poszczególnych narodów i państw zdaje się potwierdzać tę wizję. Żeby poprzeć na jednym, najbliższym nam przykładzie — historycy dość zgodnie wskazują w dziejach Polski na okresy jej wzlotów i upadków. Okres panowania Sasów uważany jest za czas ciemny, naznaczony stagnacją państwa, fatalną rolą zasady *liberum veto*, złudnym samozadowoleniem symbolizowanym groźnym w politycznych i moralnych skutkach powiedzeniem: *Za króla Sasa jedz, pij i popuszczaj pasa*. Zryw Sejmu Czteroletniego był okresem przebudzenia narodowej świadomości, nawet jeśli okazał się politycznie nieskuteczny, podobnie jak szereg zakończonych klęską powstań. Kolejną szansę odrodzenia, nie tylko politycznego, ale i moralnego, otrzymała Polska po I wojnie światowej — i choć nie uchroniła się od szeregu błędów, to przecież udało się utworzyć niepodległe państwo i wykształcić wysokiej próby patriotyzm, którego heroicznym świadectwem jest Powstanie Warszawskie i całe dzieje Armii Krajowej. Mamy w żywej pamięci inny etap przebudzenia narodowego i moralnego, który wiąże się z NSZZ „Solidarność”. Nie był to przecież tylko ruch narodowo-wyzwoleńczy (choć i on ma istotne znamiona moralne), ale także odrodzenie moralne *sensu stricto*, jak o tym dobitnie świadczy niezrozumiałe dla obco-krajowców hasło rozplakatowane przed wyborami do Sejmu kontraktowego: *Żeby Polska była Polską, 2 + 2 musi być zawsze 4!*. Myśmy jednak wiedzieli, co się za nim kryje, jak gorzka w gruncie rzeczy jest anegdota o Malinowskim, który na pytanie: *Ile jest 2 + 2?* odpowiada: *A ile pan kierownik sobie życzy?* Pytać można, czy dziś przeżywamy okres moralnego odrodzenia, czy raczej głębokiego regresu. Pytanie to stawiać sobie jednak musimy, pamiętając o tym, że Polacy uczestniczą dziś w kulturalnym krwioobiegu obejmującym całą cywilizację euroatlantycką, wywierającą silny wpływ także na inne kultury i państwa ziemskiego globu. Zapewne odpowiedź może nie być prosta. Ewolucja moralna tej cywilizacji — tak przebogatej i zróżnicowanej — ma nie-

⁸ Ap 22,11.

⁹ Por. Ez 18.

jednoznaczny charakter. Tak właśnie, jak czytamy w *Apokalipsie*: *plugawi się jeszcze plugawią, a święci jeszcze uświęcają*.

Pytać można, czy pytać trzeba? Czy nie wystarczy poprzestać na uznaniu, że świat ludzki jest złożony, trudny do objęcia umysłem przez nas — prostych jego obywateli — i skromnie zająć się własnym podwórkiem, pozostawiając ocenę jego moralnego stanu prorokom i politykom mającym szersze horyzonty poznawcze i większe możliwości działania? Otóż — odnosząc się z całym szacunkiem do takiego pokornego myślenia i akceptując szczególne znaczenie „własnego podwórka” — całkowicie uwolnić się od tych pytań nie powinniśmy właśnie dlatego, że jesteśmy obywatelami jednego, zresztą coraz bardziej jednoczącego się gospodarstwo i kulturowo świata, toteż na naszym „podwórku” znaleźć możemy skarby i śmieci podobne, jak u innych. Nie o kreowanie się na sędziego świata więc tu chodzi, ale o lepsze rozpoznanie własnego „małego świata”, a także o odpowiedzialność za rodzinę ludzką, do której należymy i której przyszły los także od nas zależy. Owszem, warto rozglądać się za prorokami, którzy świat ten widzą bardziej przenikliwie. Czyż nie taka jest rola filozofów, poetów i kaptanów?

Jednym ze współczesnych kaptanów, poetów i filozofów wartych postuchania był niewątpliwie Jan Paweł II. Podkreślić warto, że wiele swych encyklik, adhoratacji i innych oficjalnych wypowiedzi rozpoczął on od przedstawienia panoramy współczesnego świata. Jej nachylenie bywa różne, zależnie od tematyki dokumentu. Zawsze jednak przywołuje się w niej szeroki kontekst współczesności i zawsze naznaczona jest ona silnym akcentem moralnym. Zresztą Jan Paweł II nie pierwszy tak czynił. Podobną strukturę mają także niektóre wcześniejsze dokumenty papieskie, choć trzeba przyznać, że szczególną inspiracją była dla Papieża *Konstytucja duszpasterska o Kościele w świecie współczesnym* — *Gaudium et spes*, do której zresztą Jan Paweł II często wprost nawiązywał. Jej pierwsze słowa: *Radość i nadzieja, smutek i trwoga ludzi współczesnych* wskazują na tę właśnie złożoną sytuację i dynamikę świata, której prototyp odnajdujemy w *Księdze Apokalipsy* — świata, w którym dostrzec można zarówno rodzące się i dojrzewające dobro, jak i coraz bardziej potężniejsze zło. Odpowiedni fragment tej *Konstytucji* (pp. 4–10) warto zresztą i dziś uważnie

przeczytać. Nasz świat nie rozwija się aż tak szybko, by uwagi tam zawarte, a wypracowane 40 lat temu, zupełnie straciły na aktualności. Jednak zamiast zawsze niepełnego katalogu zalet i wad współczesności proponuję (zainspirowany nauczaniem Jana Pawła II, ale na własną tylko, nie Papieża, odpowiedzialność) skupić uwagę na **dwóch rysach** znamionujących obecne czasy, istotnych dla rozważenia kwestii postępu lub regresu moralnego.

Pierwszym z nich, niewątpliwie głęboko pozytywnym, jest tendencja do **głębszego poznania i poszanowania ludzkiej osoby**. Tendencja ta ma różne postacie. Jedną z nich jest formułowanie i rozbudowywanie listy praw człowieka, a także usytuowanie ich — wbrew modnemu kiedyś pozytywizmowi prawnemu — u podłoża wszelkich praw stanowionych, których nie tylko nie mogą one naruszać, ale dla których stanowią najważniejszą inspirację. Już preambuła *Konstytucji RP* zawiera wzmiankę o *przyrodzonej godności człowieka*. Wzmianka ta zostaje następnie rozwinięta — na początku rozdziału zatytułowanego *Wolności, prawa i obowiązki człowieka i obywatela* — jako artykuł 30: *Przyrodzona i niezbywalna godność człowieka stanowi źródło wolności i praw człowieka i obywatela. Jest ona nienaruszalna, a jej poszanowanie i ochrona jest obowiązkiem władz publicznych*. Następujące po nim artykuły rozdziału II *Konstytucji RP* wymieniają poszczególne uprawnienia i określają obowiązki władzy do ich poszanowania. Podobne zapisy odnaleźć można oczywiście w ustawach zasadniczych innych państw. Sama moda na demokrację, w historii ludzkości dość przeciwieństwo, czerpie swe inspiracje nie tylko z Rewolucji Francuskiej i oświeceniowej koncepcji władzy pochodzącej nie z boskiego, ale społecznego mandatu. Czerpie je również z przekonania o tym, że każdemu człowiekowi przysługują prawa, które ludzka władza winna z całą mocą respektować. Warto wspomnieć, że amerykańska *Deklaracja Niepodległości* z 4 lipca 1776 roku (starsza więc o 15 lat od *Konstytucji Francuskiej*) wyprowadza prawa człowieka z Boskiego ustanowienia: *Uważamy za oczywiste następujące prawdy: że wszyscy ludzie stworzeni zostali jako równi sobie, że Stwórca wyposażył ich w pewne niezbywalne prawa, że wśród nich są [prawo] do życia, do wolności i do poszukiwania szczęścia*¹⁰. Ta wrażliwość na godność każdego człowieka, zobowiązująca szczególnie

¹⁰ Cyt. za W. Witkowski, *Pierwsze konstytucje europejskie*, Lublin 1996, s. 36.

wobec słabszych i bezbronnych, stanowi inspirację dla bardziej dziś rozpowszechnionej i bardziej zorganizowanej pomocy okazywanej ofiarom tragicznych wydarzeń, chorym, niepełnosprawnym i samotnym. Podkreślić warto, że rozszerzająca się lista praw człowieka, nie poprzestająca już tylko na prawie do życia, wolności i poszukiwania szczęścia (jak we wspomnianej *Deklaracji Niepodległości Stanów Zjednoczonych Ameryki Północnej*) jest świadectwem coraz głębszego poznania człowieka i jego natury. Pod pewnym względem doktryna praw człowieka zajęła miejsce tradycyjnej doktryny o prawie naturalnym. Prawo do wolności sumienia i wyznania, prawo do pracy i odpoczynku, prawo do zrzeszania się, uprawnienia mniejszości narodowych, wspomniany zakaz stosowania tortur i innych form nękania ludności nawet w ramach działań wojennych, odejście od stosowania kary śmierci — wszystko to wkomponowane jest we współczesną strukturę państwa demokratycznego, zarazem jednak uwydatnia te rysy ludzkiej osoby, które coraz wyraźniej sobie uświadamiając mamy obowiązek respektować. Również normy odnoszące się do poszanowania środowiska naturalnego, dziś szczególnie ważne z powodu zagrożeń ekologicznych, które człowiek wywołał i które tak poważnie zagrażają przyszłości rodzaju ludzkiego, zwracają uwagę na głębszy niż kiedyś sądzono związek łączący człowieka z jego otoczeniem. Przykłady wskazujące na to, jak daleko sięga współczesna „wrażliwość personalistyczna” są nader liczne i nie ma potrzeby ich tu wyliczać. Wszystkie one przybliżają człowieka, pomagają dostrzec jego naturę i wewnętrzne bogactwo, pomagają wzbudzić poszanowanie dla innych, umacniają poczucie międzyludzkiej wspólnoty i budzą miłość.

Mimo takich deklaracji trudno nie dostrzec dysonanisu między szlachetnymi hasłami wzajemnej pomocy czy zaożeniami organizacji krajowych i międzynarodowych, których celem jest pomoc świadczona najbardziej potrzebującym (zwłaszcza najuboższym), a faktycznymi kierunkami ekonomicznych przemian, w wyniku których bogaci stają się coraz bogatsi, a biedni coraz biedniejsi¹¹. Zdumieniem i smutkiem napawa głębokie pęknięcie, jakim w rozumieniu tych praw jest coraz bardziej rozpowszechniona akceptacja aktów

aborcji, a ostatnio także eutanazji — najpierw przez zaniechanie ścigania tych aktów, a następnie dokonywanie ich w majestacie prawa. Argumentacja, że przekonanie o przystępującym także nienarodzonemu statusie ludzkiej osoby nie jest powszechne, nie usprawiedliwia legalizacji aborcji. Nie chcę przywoływać całej bolesnej dyskusji na ten temat¹², przypomnę tylko starą formułę prawną, respektowaną bez zastrzeżeń w innych przypadkach, która powiada: *in dubio pro reo* (wątpliwości należy rozstrzygać na korzyść tego, kogo one dotyczą). Tak jak nie wolno w trakcie polowania strzelać do czegoś, co się porusza w krzakach, jeśli nie ma stuprocentowej pewności, że nie jest to współmyśliwy, tak nie wolno zabijać istoty, co do której istnieje bodaj cień podejrzenia, czy nie jest to jednak równy nam w prawach człowiek, w dodatku niewinny, bezbronny i zastępujący przez to na szczególną ochronę. O czym to pęknięcie świadczy? Obawiam się, że o zasadniczej zmianie kryterium wartościowania człowieka i jego życia. Próba legalizacji eutanazji wyraźniej jeszcze to pokazuje. Jej zwolennicy podkreślają, że człowiek sam ma prawo do decydowania czy nadal chce żyć, czy też woli umrzeć, jeśli dojmujące i niemożliwe do przezwyciężenia cierpienie lub pozbawienie możliwości sprawnego poruszania się i czerpania z radości tego świata odbierają mu chęć dalszego życia. Daleki jestem od lekceważenia dramatycznych okoliczności związanych z przypadkami eutanazji. Chodzi mi o samą argumentację — o to, że życie człowieka mierzone jest swoistym „rachunkiem zysków i strat”, jaki przeprowadza się w odniesieniu do perspektywy dalszej egzystencji.

Ta właśnie argumentacja pokazuje siłę **sekularyzmu**. Pojmuje się go różnie (znów nie chcę wdawać się w dywagacje definicyjne). Tu rozumiem przezeń taką „filozofię życia”, która zamyka się w obrębie „tego wieku” (*hoc saeculum*), tego świata. To nie jest ateizm, tym bardziej nie tzw. wojujący ateizm, który ma w sobie coś z religii *a rebour* — walczy z Bogiem i Jego wyznawcami, ale w miejsce dotychczasowej religii dąży do ustanowienia nowej, także angażującej gorliwych wyznawców. Sekularyzm natomiast cechuje się obojętnością na problem Boga. Człowiek zsekularyzowany żyje tym światem — i tylko nim. Jeśli przynależność

¹¹ Godne uwagi są w tym kontekście — dalekie od optymizmu — rozważania o globalizacji Z. Baumana. Por. tenże, *Globalizacja — i co z tego dla ludzi wynika*, Warszawa 2000.

¹² W tej sprawie por. np.: T. Styczeń (red.), *Nienarodzony miarą demokracji*, Lublin 1991.

do jakiejś wspólnoty religijnej mu odpowiada, to nic nie stoi na przeszkodzie, by do niej wstąpił. Religia także może zaspokajać jakieś nasze potrzeby, tak jak staramy się zaspokoić potrzeby materialne i estetyczne. Niezależnie od ewentualnej przynależności do jakiejś wspólnoty religijnej, zsekularyzowany człowiek żyje jednak tak, jak gdyby Boga nie było (*etsi Deus non daretur*), zamknięty w obrębie wartości tego świata, szukający maksimum korzyści, jakie z niego może wyciągnąć. Wyrazem sekularyzmu jest więc **konsumizm** (dążenie do posiadania i zażywania jak największej ilości dóbr, jakie oferuje świat), wygodnictwo (m.in. zaniechanie kłopotów, jakie niesie z sobą rodzina, zwłaszcza wiodzietna), a także karierowiczostwo. Nie wszystkie przecież miejsca w świecie są równie atrakcyjne, trzeba dążyć do zajęcia możliwie najwyższej pozycji w społeczeństwie, a dla osiągnięcia tego celu warto podjąć niematry trud. Świat bowiem wprawdzie obfituje w dobro, ale kandydatów do ich nabycia jest wielu, a aspiracje ludzkie są nieograniczone. Zrozumiałe, że sekularyzm wspiera silnie promocję wolności. Zrozumiałe też, że „nakręcaniu” takiego konsumizmu służy cała produkcja dóbr i towarzysząca jej reklama. Trudno natomiast znaleźć w tym świecie miejsce dla moralnego dojrzewania człowieka, dla kształtowania jego charakteru i rozbudzania gotowości życia dla innych.

Tak pojętym sekularyzmem (tu tylko szkieletowo przedstawianym) zarażeni jesteśmy wszyscy, zwłaszcza w krajach gospodarczo rozwiniętych i zasobnych. Czy jest to przypadłość tylko współczesnego świata? Raczej nie. Już grzech pierworodny zaczął się od tego, iż niewiasta spostrzegła, że drzewo to ma owoce dobre do jedzenia, że jest ono rozkoszą dla oczu i że owoce tego drzewa nadają się do zdobycia wiedzy¹³, a wraz z niewiastą spostrzegł to oczywiście także mężczyzna. I ta atrakcyjność owoców zakazanego drzewa przestroniła im Boga — Jego mitość i Jego prawa. W różnej formie i nasileniu ta sama próba wierności Bogu i ta sama możliwość stania się tymi, którzy żyją według ciała¹⁴ powtarza się w całym biegu historii. To jednak nie zwalnia nas z rozpoznania tej postaci owej pokusy, z którą mamy do czynienia dziś. Dziś — jak sądzę — przybrała ona postać sekularyzmu, który nie

poprzestaje na dostrzeżeniu cudowności tego świata, ale jego dobra — i **tylko** jego dobra — stawia na boskim piedestale, podcinając korzenie religijności wraz z leżącą u jej korzeni świadomością doczesności tego świata, którego ostateczny sens (i sens życia człowieka) leży poza nim.

Ten proces dokonuje się na naszych oczach i to dokonuje się szybko. W swej godnej uwagi książce *Bezład* Zbigniew Brzeziński wprowadza pojęcie „permissywnej kornukopii” (od łac. *Cornu copiae* — róg obfitości), przez które rozumie on obserwowany dziś w bogatych i demokratycznych krajach *stopniowy rozkład regulujących funkcji kryteriów moralnych, współwystępujący z koncentracją uwagi na zaspokojeniu potrzeb materialnych jednostki, tak, że dominantą kulturową staje się jednostkowy i zbiorowy hedonizm. Erozja wartości moralnych w połączeniu z dowartościowaniem dóbr materialnych wywołuje permissywizm na poziomie działania i chciwość na poziomie motywacji*¹⁵. N. Davies natomiast tak charakteryzuje przemiany w świadomości współczesnych Brytyjczyków: *Najbardziej uderzający wydaje się jednak powojenny upadek samej wiary religijnej. W roku 1945 zdecydowana większość Brytyjczyków była praktykującymi chrześcijanami. [...] Pięćdziesiąt lat później wielowiekowe tradycje niemal całkowicie zniknęły. Absolutna większość obywateli Wielkiej Brytanii nigdy nawet nie weszła do żadnej świątyni. [...] Według opublikowanych w roku 1994 danych za niewierzących uważało się 24% Brytyjczyków, rzymskich katolików było 9%, praktykujących anglikanów — 4%, presbiterian — 3%, muzułmanów i metodystów — po 2%, hinduistów — 1%, a pozostałe wyznania chrześcijańskie — 4% (reszta, to tzw. niepraktykujący anglikanie)*¹⁶.

Problem erozji ładu moralnego sięga więc głęboko. Nie można go sprowadzić do nieco jałowych sporów o to czy i jakich dawnych obyczajów trzeba dziś bronić, a od jakich bez żalu odstąpić. Ostatecznie sprowadza się on do zmagania między mitością Boga i człowieka a „zmaterializowanym” i zredukowanym do sfery dóbr konsumpcyjnych zamknięciem w sobie. Ład moralny bowiem to w istocie ład mitości. Przez dostrzeżenie osoby (boskiej i ludzkiej), przez zafascynowanie się nią i życie dla niej, spełnia się człowiek — i tylko tą

¹³ Rdz. 3, 6.

¹⁴ Rz 8, 5.

¹⁵ Z. Brzeziński, *Bezład. Polityka światowa na progu XXI wieku*, Warszawa 1995, s. 64.

¹⁶ N. Davies, *Wyspy*, przet. E. Tabakowska, Kraków 2003, s. 796 n.

drogą spełnia się naprawdę, tak jak to niezwykle zwięźle i trafnie ujęli Ojcowie Soboru Watykańskiego: *Człowiek, będąc jedynym na ziemi stworzeniem, którego Bóg chciał dla niego samego, nie może odnaleźć się w pełni inaczej, jak tylko poprzez bezinteresowny dar z siebie samego*¹⁷. Tylko poprzez bezinteresowny dar z siebie człowiek staje się naprawdę sobą jako człowiek – i to właśnie stanowi istotę dobra moralnego. Czy więc i w jakim sensie mamy do czynienia ze zjawiskiem erozji moralnego ładu w świecie zależy od tego, jak dalece panująca mentalność i związane z nią struktury życia społecznego pomagają człowiekowi wzrastać w miłości, gdzie zaś mu to utrudniają, stawiając przed oczy pozorną wielkość zamkniętą w obrębie tego świata. Nie potrafię ocenić proporcji między oboma tendencjami, choć przyznaję, że „opcja sekularystyczna” wydaje się bardziej natarczywa. Może zresztą „opcja miłości” nie znosi natarczywości. Nie narzekamy jednak zbyt. Jesteśmy nie tylko świadkami tych zmagania, ale także ich aktorami budującymi lub niszczącymi ład moralny w świecie – a przez to kształtującymi własny moralny postęp lub regres.

Ksiądz profesor Andrzej Ryszard Szostek polski duchowny rzymskokatolicki, profesor filozofii. Przez całą swą karierę naukową związany z Katolickim Uniwersytetem Lubelskim, którego w latach 1998–2004 był rektorem. Obecnie pracuje w Katedrze Etyki KUL. Wyświęcony na kapłana 23 czerwca 1974 w Górze Kalwarii. Od 2006 roku jest radnym generalnym zgromadzenia księży Marianów. Jest członkiem Polskiej Akademii Umiejętności, Fundacji Lus et Lex oraz Rady Naukowej Narodowego Centrum Nauki. Został odznaczony Krzyżem Kawalerskim Orderu Odrodzenia Polski.

¹⁷ *Konstytucja Duszpasterska o Kościele w świecie współczesnym*, p. 24 (cyt. za: *Sobór Watykański II. Konstytucje, Dekrety, Deklaracje*, Poznań 1967).

W jakim Wszechświecie żyjemy?

Na podstawie odczytu wygłoszonego w dniu 27 kwietnia 2006 roku

Kazimierz Stępień

Obserwatorium Astronomiczne Uniwersytetu Warszawskiego

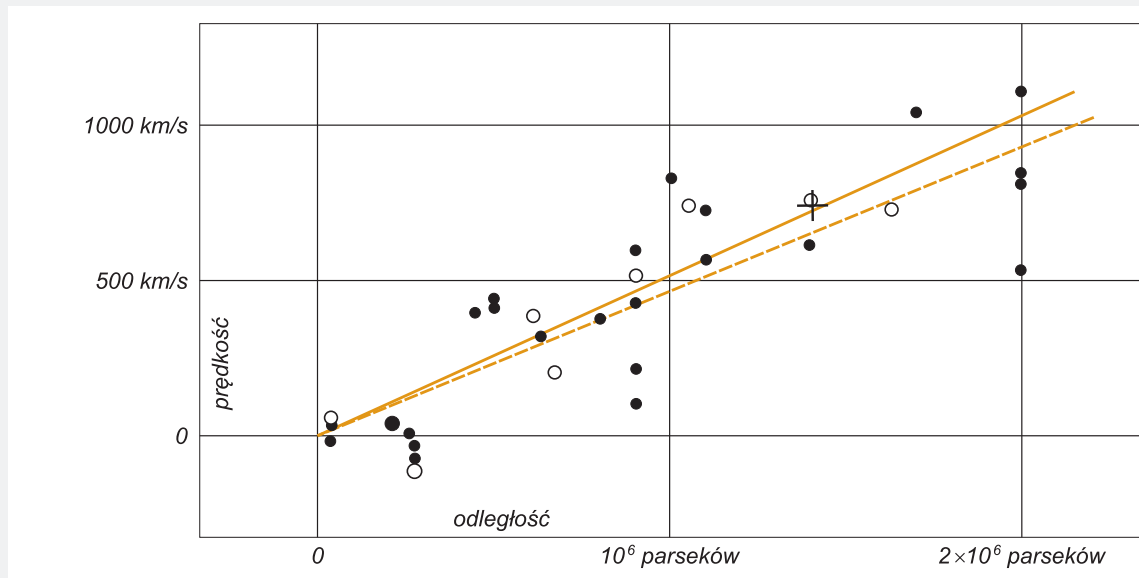
81

Kosmologia, traktowana jako nauka ścisła, nie istniała do początku XX wieku. Do tego czasu nieznanymi były (nawet w grubym przybliżeniu) wiek Wszechświata, ani jego rozmiar. Wiadomo było, że Słońce znajduje się w rozległym układzie gwiazdowym – Galaktyce, ale nie wiedziano nic o jego granicach, a tym bardziej o tym, co znajduje się poza nimi. Stan wiedzy na temat Wszechświata na przełomie XIX i XX wieku tak został podsumowany w monumentalnym dziele encyklopedycznym *Wszechświat i Człowiek*, wydanym (w tłumaczeniu polskim) w 1906 roku pod redakcją Jana Kremera: *O początek i koniec wszechrzeczy nie troszczy się wiedza przyrodnicza. Wie dobrze, że nieumiarkowane takie dążności, jakby mamidła, uwodziły myśl naszą coraz głębiej w pustynię dogmatów jałowych*. Jak widać, autorzy słusznie uznali, że przy zupełnym braku faktów obserwacyjnych, nauka powinna powstrzymać się od zajmowania się problemem powstania i budowy Wszechświata.

W początkach XX wieku dokonano kilku ważnych odkryć, które zasadniczo zmieniły tę sytuację. Odkrycie radioaktywności zaowocowało metodami datowania, które pokazały, że Ziemia, a zatem i cały Wszechświat, istnieje co najmniej kilka miliardów lat. Odkryto, że poza naszą Galaktyką istnieją inne, podobne układy gwiazd wypełniające Wszechświat aż do najdalszych obserwowanych jego granic. W 1916 roku Albert Einstein opublikował ogólną teorię względności – idealne narzędzie do modelowania całego Wszechświata. Autor ogólnej teorii względności od razu sam wykorzystał ją do tego celu. Zgodnie z ówczesnym stanem

wiedzy przyjęto, że Wszechświat jest niezmienny i statyczny. By zbudować taki model, wprowadził *ad hoc* oddziaływanie między galaktykami równoważące grawitację. Odpowiedni człon w równaniach ogólnej teorii względności zawierał tzw. stałą kosmologiczną i miał sens siły odpychającej – rosnącej wraz z odległością. Taka zależność powodowała, że oddziaływanie między bliskimi obiektami (gwiazdami czy nawet sąsiednimi galaktykami) było zaniedbywalne, ale cały Wszechświat nie zapadał się pod wpływem grawitacji. Kolejnego odkrycia, uważanego za najważniejsze w historii kosmologii, dokonał Edwin Hubble, który zmierzył odległości do kilkudziesięciu bliskich galaktyk i zauważył, że (obserwowane wcześniej) przesunięcie ku czerwieni linii widmowych galaktyk jest proporcjonalne do ich odległości (rys. 1). Stała proporcjonalności nazwana została stałą Hubble'a. Interpretując przesunięcie ku czerwieni, jako efekt oddalania się od siebie galaktyk, i uogólniając prawo Hubble'a na cały Wszechświat wnioskujemy, że Wszechświat rozszerza się. Taki model nie wymaga członu kosmologicznego, ale sugeruje, że Wszechświat musiał mieć początek. Wprawdzie nie możemy zbliżyć się z fizycznym opisem dowolnie blisko czasu $t = 0$, gdyż w formalnym opisie natrafiamy na osobliwość – zerowy rozmiar i nieskończoną gęstość – ale możemy pokusić się o opis Wszechświata od momentu, gdy osiągnął tzw. rozmiar Plancka (około 10^{-35} cm), co nastąpiło w wieku 10^{-43} s.

Tworząc ewolucyjny model Wszechświata, czynimy pewne założenia. Zakładamy mianowicie, że znane nam prawa fizyki (w tym ogólna teoria względności)



Rysunek 1. Oryginalny wykres Hubble'a. Czarne kropki odpowiadają indywidualnym galaktykom, a kółka średnim dla kilku galaktyk

obowiązują zawsze i wszędzie we Wszechświecie, oraz że Wszechświat jest jednorodny i izotropowy, a obserwowana przez nas jego część jest reprezentatywna dla całości (tzw. zasada kopernikańska). Z opisu wczesnych stadiów życia Wszechświata wynikało, że w pierwszych minutach życia musiała zachodzić w nim nukleosynteza, podczas której około 1/4 masy zawartej w barionach zamieniło się w hel. Współczesne dane obserwacyjne potwierdzają, że pierwotna materia kosmiczna składała się z wodoru i helu w oczekiwanych proporcjach (pominając znikome ilości litu i berylu). Z modelu wynikało też, że pamiątką po gorącej fazie powinno być wypełniające cały Wszechświat promieniowanie elektromagnetyczne o rozkładzie widmowym odpowiadającym paru kelwinom. Odkrycie w 1965 roku promieniowania o temperaturze 2,73 K, nazwanego mikrofalowym promieniowaniem tła uznano za koronny dowód poprawności ewolucyjnego modelu Wszechświata powstałego w wyniku Wielkiego Wybuchu. Dokładne badanie mikrofalowego promieniowania tła i diagramu Hubble'a dla coraz odleglejszych obiektów, to obecnie dwa najważniejsze źródła naszej wiedzy empirycznej o ewolucji Wszechświata.

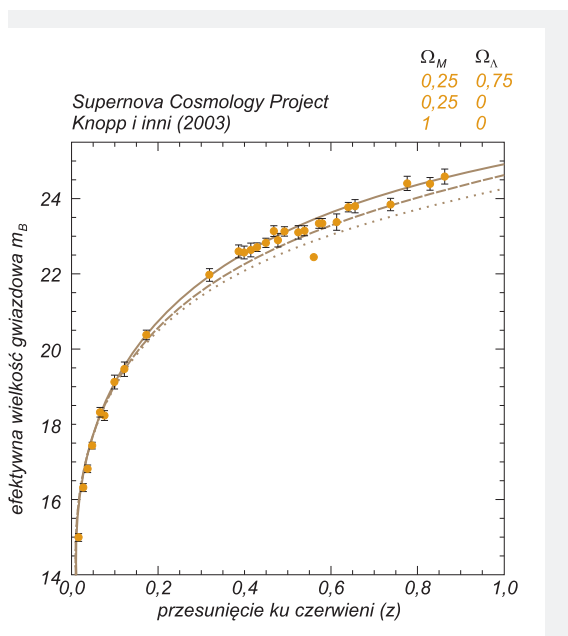
Jednorodność mikrofalowego promieniowania tła w obserwowanej części Wszechświata, czyli w skali rzę-

du kilkunastu miliardów lat świetlnych, wymusiła wprowadzenie szczególnej fazy we wczesnym jego życiu, nazwanej fazą inflacyjną, podczas której w małym ułamku sekundy Wszechświat zwiększył swoje rozmiary około 10^{30} razy. Dzięki temu obszar powiązany przyczynowo (w którym promieniowanie mogło zuniformizować się) może dzisiaj przewyższać wielokrotnie obserwowaną jego część. Jednak cena za wprowadzenie fazy inflacyjnej była wysoka. Po pierwsze, inflacja wymaga, by przestrzeń Wszechświata miała zerową krzywiznę, czyli była płaska (euklidesowa), co automatycznie pociąga za sobą warunek na średnią gęstość masy–energii. Gęstość ta musi być równa pewnej wartości krytycznej. Ponieważ przestrzeń, zgodnie z ogólną teorią względności, rozpięta jest na masie–energii zawartej we Wszechświecie, to gdy gęstość jest dostatecznie duża, przestrzeń ma dodatnią krzywiznę (w dwóch wymiarach byłaby to sfera), a gdy jest zbyt mała, krzywizna jest ujemna. Wartość krytyczna rozdzielająca te dwie rodziny przestrzeni odpowiada przestrzeni płaskiej. Po drugie, by otrzymać inflację, trzeba było wprowadzić do równań ogólnej teorii względności człon podobny do kosmologicznego, który pojawia się w określonym momencie i przyjmuje ogromną wartość pozwalającą gwałtownie rozdmąć Wszechświat,

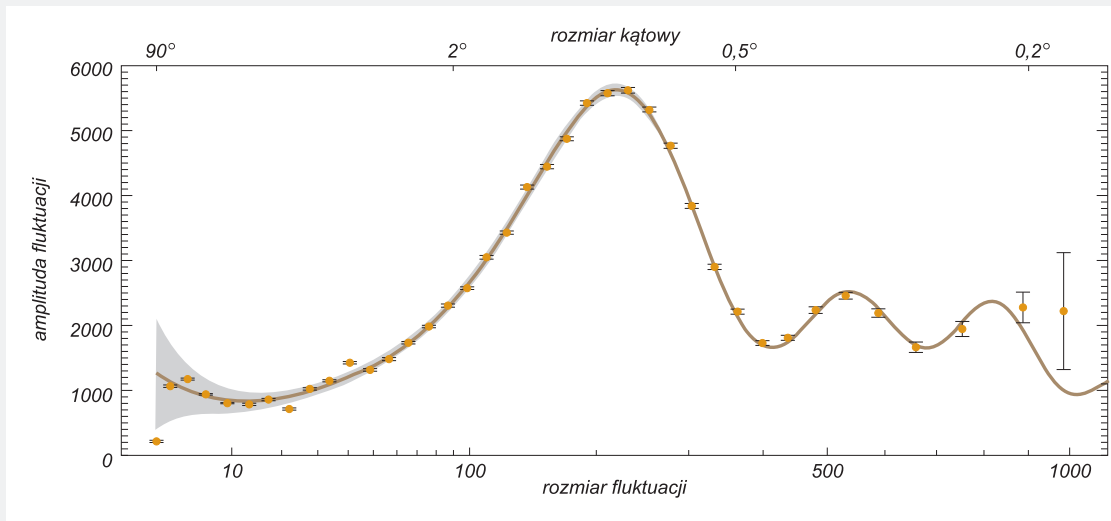
a potem znika. Człot ten miałby pochodzić od *ad hoc* wprowadzonego pola fizycznego wypełniającego próżnię, które przy pewnej temperaturze (czyli w określonym momencie ewolucji Wszechświata) dokonuje przejścia fazowego z wydzieleniem dużej ilości energii, ale z ujemnym ciśnieniem rozdymającym Wszechświat (normalne — dodatnie ciśnienie działa jak masa i jest źródłem grawitacji). O ile weryfikacja istnienia i własności takiego pola jest obecnie poza naszym zasięgiem, o tyle wyznaczenie średniej gęstości Wszechświata wydaje się zadaniem nietrudnym. Wchodzą do niej wszelkie formy masy i energii. Zaczniemy zatem od tych form materii, które bezpośrednio obserwujemy. Nazwiemy je materią świecąca. Oceniając, w dostatecznie dużej objętości, masę wszystkich gwiazd i pozostałości po nich, materii międzygwiazdowej i międzygalaktycznej oraz znanych form promieniowania możemy obliczyć jej gęstość. Okazuje się, że jest bardzo mała, rzędu 5% wartości krytycznej. Z taką gęstością Wszechświat byłby otwarty, o ujemnej krzywiznie i ekspandowałby w nieskończoność. Ale czy to wszystko? Okazuje się,

że nie. Badanie dynamiki obiektów krążących na peryferiach galaktyk (np. obłoków materii międzygwiazdowej) wskazuje, że ich prędkości są o wiele za duże w stosunku do tych, jakie wynikają z przyciągania grawitacyjnego produkowanego przez samą materię świecąca. Ponieważ obłoki nie odrywają się od swoich galaktyk, musi na nie w istocie działać kilkakrotnie większa siła przyciągania. Stąd wynika, że większość masy galaktyki zawarta jest w nieznanej nam formie materii, nazwanej ciemną materią, która manifestuje swoją obecność wyłącznie przez oddziaływanie grawitacyjne. Ilościowa analiza dynamiki badanych obiektów oraz całych galaktyk poruszających się w układach wielokrotnych pozwala wyznaczyć dość dokładnie ilość ciemnej materii i policzyć odpowiadającą jej gęstość. Dostajemy wartość rzędu 25% wartości krytycznej. Łącznie mamy zatem około 30%, czyli wciąż ponad 3 razy mniej niż wymagane przez inflację.

W 90-tych latach XX wieku podjęto próbę rozszerzenia diagramu Hubble'a na znacznie większe odległości — sięgające wielu miliardów lat świetlnych. Wyznaczono je przez obserwację wybranego typu gwiazd supernowych, służących jako świece standardowe, w odległych galaktykach. Niektóre z nich wybuchły, gdy Wszechświat był dwukrotnie mniejszy niż obecnie. Okazało się, że ówczesne tempo rozszerzania się Wszechświata było mniejsze niż dzisiaj. To bardzo zaskakujący fakt, gdyż siły grawitacji spowalniają ekspansję, a więc dawniejsze tempo ekspansji powinno być większe niż obecne. Tymczasem obserwacje wskazują, że istnieje czynnik działający przeciwnie niż grawitacja i rozpychający Wszechświat. A zatem, już po raz trzeci, dochodzimy do konieczności wprowadzenia niezerowej stałej kosmologicznej! Możliwie dokładne odtworzenie obserwowanego diagramu Hubble'a przez modele z różnymi wartościami parametrów wymaga, by wartość stałej kosmologicznej była taka, że odpowiadająca jej gęstość energii wynosi około 70% wartości krytycznej (rys. 2). Jest to energia ukryta w próżni fizycznej, a nazwano ją ciemną energią. W sumie dostajemy wymaganą przez inflację wartość. Kosmologowie odczuli z ulgą, tym bardziej że silne, niezależne potwierdzenie takiego modelu przyszło od obserwatorów mikrofalowego promieniowania tła. Teoretyczne modele przewidują, że powinniśmy obserwować na niebie fluktuacje mikrofalowego promieniowania tła przejawiające się jako niewielkie (rzędu 10^{-5}) odchylenia temperatury od średniej. Innymi słowy, niebo powinno być pokryte plamkami gorętszego i chłodniejszego mikro-



Rysunek 2. Współczesna wersja diagramu Hubble'a. Najlepiej dopasowana linia opisuje model przedstawiony w tekście artykułu. Ω_M i Ω_Λ są stosunkiem gęstości materii (świecącej i ciemnej) oraz gęstości ciemnej energii do gęstości krytycznej. Wykres Hubble'a z poprzedniego rysunku miałby tu rozmiar kropki w początku układu współrzędnych

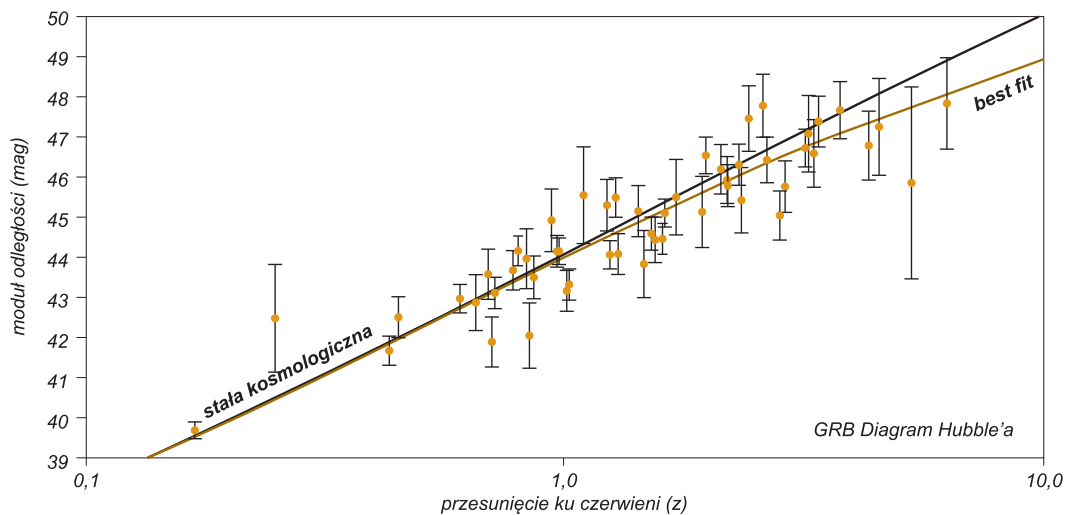


Rysunek 3. Amplituda fluktuacji temperatury mikrofalowego promieniowania tła w funkcji jej rozmiaru kąowego (dane z NASA/WMAP). Wielkość i położenie kolejnych maksimum zależy od parametrów Wszechświata. Linia ciągła odpowiada modelowi opisanemu w tekście artykułu

falowego promieniowania tła. Rozmiary kątowe plamek i amplituda odchytek zależą od podstawowych parametrów Wszechświata – krzywizny jego przestrzeni, względnej zawartości materii świecącej do ciemnej itd. Wieloletnie obserwacje tych fluktuacji potwierdzi-

ły model otrzymany na podstawie diagramu Hubble'a (rys. 3). Kosmologowie mogli więc podać stosunkowo dokładne wartości parametrów opisujących Wszechświat wraz z ich niepewnościami. Oto one:

– stała Hubble'a – 71 ± 4 km/sMpc (Megaparsek),



Rysunek 4. Opublikowany przez B. Schaefera w styczniu 2006 roku diagram Hubble'a. Zdaniem autora, model ze stałą kosmologiczną źle opisuje obserwacje

- średnia gęstość Wszechświata (w jednostkach gęstości krytycznej) $- 1,02 \pm 0,02$,
w tym wkład:
 - materii baronowej (świeczącej) $- 0,044 \pm 0,004$,
 - materii ciemnej $- 0,22 \pm 0,04$,
 - energii ciemnej $- 0,74 \pm 0,04$,
- obecny wiek Wszechświata $- 13,7 \pm 0,2$ mld lat,
- wiek w momencie oddzielenia się mikrofalowego promieniowania tła od materii $- 379 \pm 8$ tys. lat,
- wiek w momencie powstania pierwszych gwiazd $-$ około 200 mln lat.

Wydawało się, że wiemy wreszcie, w jakim Wszechświecie żyjemy – ma płaską (euklidesową) geometrię, jest nieskończony i będzie ekspandował nieskończenie. Dalsze badania, jak sądzono, zawężą jedynie granice niepewności tych parametrów. Tymczasem w 2006 roku ukazała się praca z diagramem Hubble’a sięgającym do o wiele większych odległości, niż ten otrzymany za pomocą supernowych. Jej autor wykorzystał źródła rozbłysków promieniowania gamma jako świece standardowe (rys. 4). Wyniki wskazują, że stała kosmologiczna nie jest stała. Gdyby to była prawda, zamiast stałej należałoby wprowadzić do równań zmienną zależną od czasu, opisującą tzw. kwintesencję, a wynika-

jący stąd model Wszechświata różniłby się od obecnie akceptowanego. Praca została przyjęta krytycznie, gdyż kalibracja źródeł rozbłysków jako świec standardowych jest kontrowersyjna i wymaga weryfikacji. Tym niemniej praca spowodowała, że aktualne pozostaje nadal pytanie, w jakim Wszechświecie żyjemy.

Abstract

Cosmology, as an exact science, was developed first in the XXth century. Discovery of the redshift-distance relation for galaxies by Edwin Hubble and General Relativity Theory by Albert Einstein resulted in a model of expanding Universe with Big Bang at the beginning. The recent observations and models consistently point towards the Universe which went through the inflation phase early in its history, has a flat geometry with average density equal (within uncertainties) to the critical value and consists of dark energy, dark matter and the baryon matter in proportions 0.74:0.22:0.04. The dark energy, described by the cosmological constant, makes the expansion accelerate. Very recently there appeared a paper arguing that the cosmological constant may actually not be constant.

Profesor Kazimierz Stępień polski astronom, wieloletni pracownik i dyrektor Obserwatorium Astronomicznego Uniwersytetu Warszawskiego. Wyznaczył, m.in. temperatury efektywne wielu gwiazd typu Ap, odkrył zależność między wiekiem, szybkością obrotu i poziomem aktywności chłodnych gwiazd ciągu głównego, a ostatnio przedstawił nowy model ewolucji chłodnych podwójnych układów kontaktowych. Jest autorem około 150

publikacji naukowych poświęconych głównie gwiazdom zmiennym (zwłaszcza posiadającym pola magnetyczne), wielu artykułów popularno-naukowych i haseł encyklopedycznych oraz podręcznika *Fizyka atmosfer gwiazd*. Sprawował liczne funkcje w organizacjach krajowych i zagranicznych. Jest członkiem m.in. Polskiego Towarzystwa Astronomicznego, Międzynarodowej Unii Astronomicznej oraz Europejskiego Towarzystwa Astronomicznego.



Wieloskalowe modelowanie molekularne białek

Na podstawie odczytu wygłoszonego w dniu 16 listopada 2006 roku

Andrzej Koliński

Wydział Chemii, Uniwersytet Warszawski

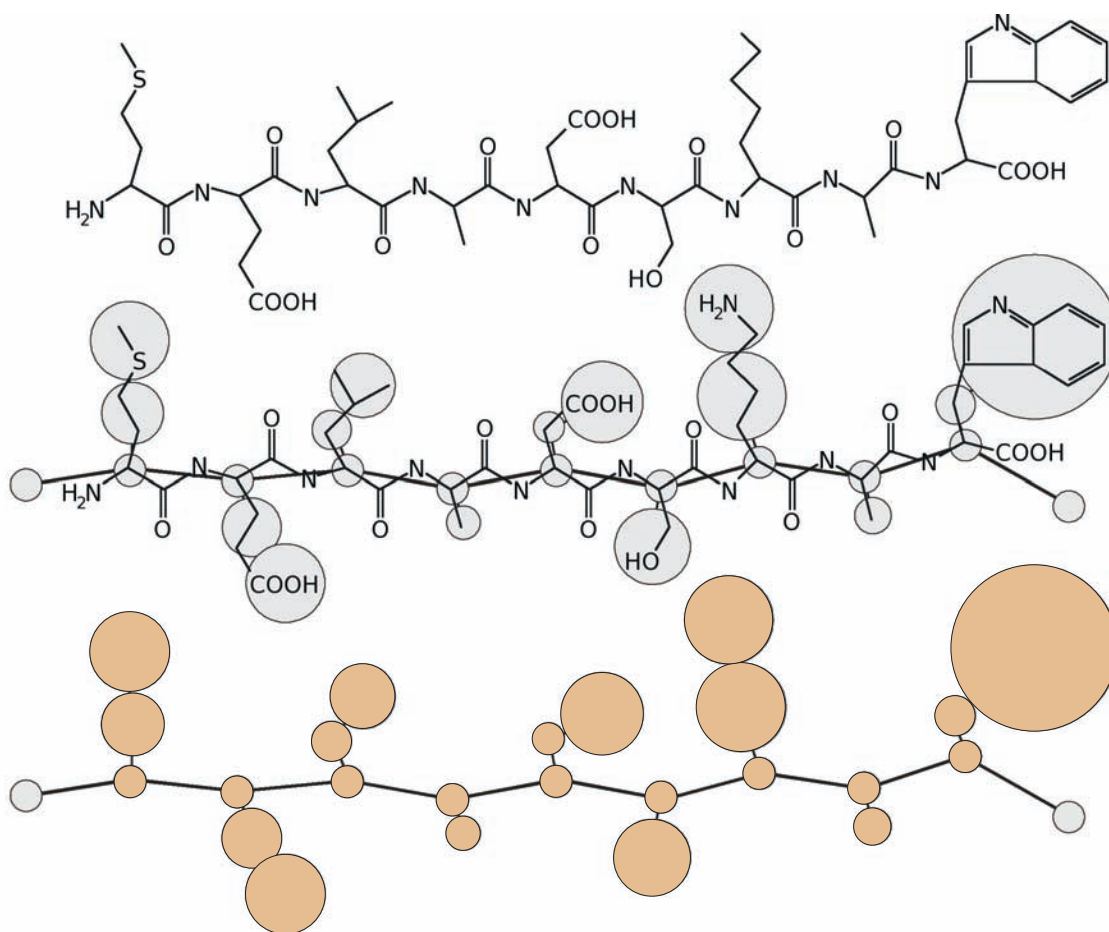
87

Znamy obecnie około 30 milionów sekwencji aminokwasowych białek. Struktury przestrzenne udało się wyznaczyć doświadczalnie jedynie dla około 30 tysięcy z nich. Ta dysproporcja stale się powiększa. Sekwencjonowanie materiału genetycznego i tłumaczenie wyników na sekwencje aminokwasów jest tanie i wykonywane na ogół przez automaty, natomiast wyznaczenie ich struktury – czasochłonne, drogie i wymaga udziału wysokiej klasy specjalistów. Znajomość struktur jest niezbędna dla wielu celów [1], np. projektowania nowych leków i biotechnologii, zrozumienia molekularnych mechanizmów procesów chorobotwórczych itp., stąd olbrzymie znaczenie teoretycznych metod przewidywania struktury białek.

Procesy molekularne w organizmach żywych są zwykle związane z bardzo skomplikowanym przegrupowaniem olbrzymiej liczby atomów i molekuł, a charakterystyczne czasy tych przemian (czy reakcji) mogą różnić się między sobą o rzędy wielkości. Dla przykładu, proces spontanicznego zwijania się białek globularnych od losowej struktury zdenaturowanej do mniej lub bardziej jednoznacznej globularnej struktury natywnej trwa od milisekund do minut, zależnie od wielkości molekule, typu struktury, warunków zewnętrznych itd. Znanne są jednak dość liczne wyjątki dużo szybszych i dużo wolniejszych procesów denaturacji-reanaturacji. Tylko bardzo szybkie (od pikosekund do mikrosekund), a co z tym na ogół się wiąże, bardzo lokalne procesy, można dziś modelować metodami klasycznej mechaniki molekularnej. W ciągu ostatnich 10–15 lat pokazano,

że dobrze zaprojektowane zredukowane modele makromolekuł mogą być bardzo przydatnymi narzędziami modelowania molekularnego dużej (w odniesieniu do liczby atomów i czasu trwania procesu) skali [2–3]. Jeden ze starszych algorytmów autora, służących do mezoskopowego modelowania białek metodą Monte Carlo, został zintegrowany z pakietem dynamiki molekularnej CHARMM (*Chemistry at Harvard Molecular Mechanics*) i jest publicznie dostępny na stronach internetowych The Scripps Research Institute (MMTSB – *Multiscale Modeling Tools for Structural Biology*. <http://mmtsb.scripps.edu/software/mmtsbToolSet.html>). Jest to pierwszy przykład ogólnie dostępnej i w pełni zautomatyzowanej metody do modelowania dynamiki białek na różnych poziomach rozdzielczości. W najbardziej ogólnym sformułowaniu zasada działania MMTSB w zastosowaniu do dynamiki białek polega na wykonaniu symulacji za pomocą przybliżonego (ale za to bardzo szybkiego) modelu niskiej rozdzielczości (na poziomie zjednoczonych reszt), odbudowaniu detali atomowych dla wybranych klatek trajektorii niskiej rozdzielczości i wykonaniu szczegółowej analizy dynamiki pełnoatomowej CHARMM dla krótkich przedziałów czasu.

Zredukowane modele molekularne białek projektowane są w różnorodny sposób [2], ale te najbardziej sprawdzone w praktyce zakładają zwykle uproszczoną reprezentację łańcucha głównego w postaci łańcucha pseudowiązań pomiędzy atomami węgla alfa. Łańcuchy boczne aminokwasów zastępowane są zazwyczaj jednym lub dwoma zjednoczonymi atomami, repre-



Rysunek 1. Schemat redukcji reprezentacji geometrycznej krótkiego fragmentu łańcucha polipeptydowego. Górny panel pokazuje reprezentację pełnoatomową, w której w celu zwiększenia czytelności pominięto atomy wodoru i zaznaczono odpowiednimi literami nazwy niektórych atomów. Środkowy panel pokazuje sposób łączenia grup atomów w formie „zjednoczonych atomów”, dolny panel – otrzymaną pseudostrukturę – zredukowany model fragmentu polipeptydowego

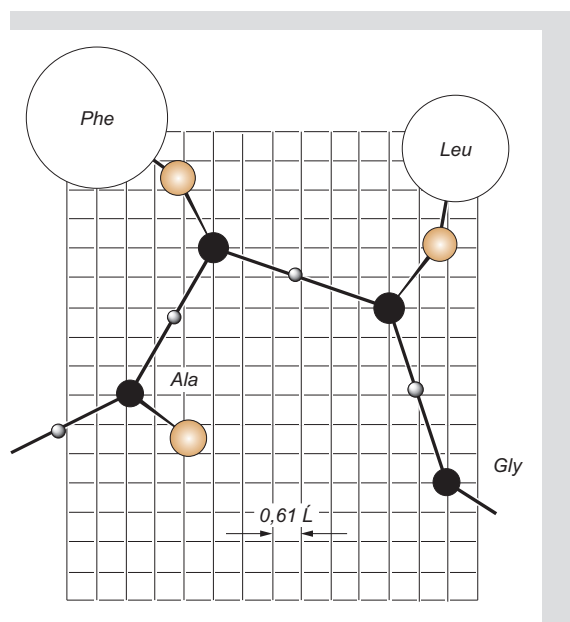
zentuującymi fragmenty grup bocznych. Przykładowy sposób redukcji geometrycznej reprezentacji łańcucha polipeptydowego pokazano na rysunku 1.

Model zredukowany CABS rozwijany przez autora [3] zakłada, podobny do naszkicowanego powyżej, sposób reprezentacji polipeptydów. Łańcuch główny ograniczony jest do szkieletu węgla alfa. Dodatkowo położenia węgla alfa ograniczone są do węzłów prostej sieci kubicznej o skoku $0,61 \text{ \AA}$, co odpowiada około $1/2$ średniej długości wiązań atomowych. Wykorzystanie reprezentacji siatkowej ma istotne znaczenie praktyczne,

pozwalając wielokrotnie przyspieszyć procesy obliczeniowe w porównaniu z równoważnymi modelami w ciągłej przestrzeni stanów. Grupy boczne reprezentowane są za pomocą dwóch zjednoczonych atomów – jednego centrowanego na węglu beta, a drugiego w środku masy pozostałej części grupy bocznej, o ile aminokwas takie posiada. Model zawiera dodatkowy pseudoatom umieszczony na środku wirtualnego wiązania między węglami alfa. Ten pseudoatom wykorzystywany jest do definicji kierunkowych oddziaływań naśladujących wiązania wodorowe między grupami

peptydowymi łańcucha głównego. Na rysunku 2 przedstawiono w sposób schematyczny geometrię modelu CABS.

Oddziaływania molekularne w modelu CABS reprezentowane są przez szereg potencjałów średniej siły wyprowadzonych na podstawie statystycznej analizy regularności strukturalnych obserwowanych w już znanych strukturach białek. Potencjały te odzwierciedlają tendencje odpowiednich sekwencji aminokwasów do przyjmowania określonej lokalnej geometrii łańcucha, którą z kolei określają odpowiednie preferencje co do struktury drugorzędowej (helis, beta-kartek itd.). Podobnie potencjały kontaktowe dla grup bocznych odzwierciedlają preferencje do charakterystycznego dla białek upakowania tych grup w stanie natywnym. Model oddziaływań CABS różni się w sposób jakościowy od in-



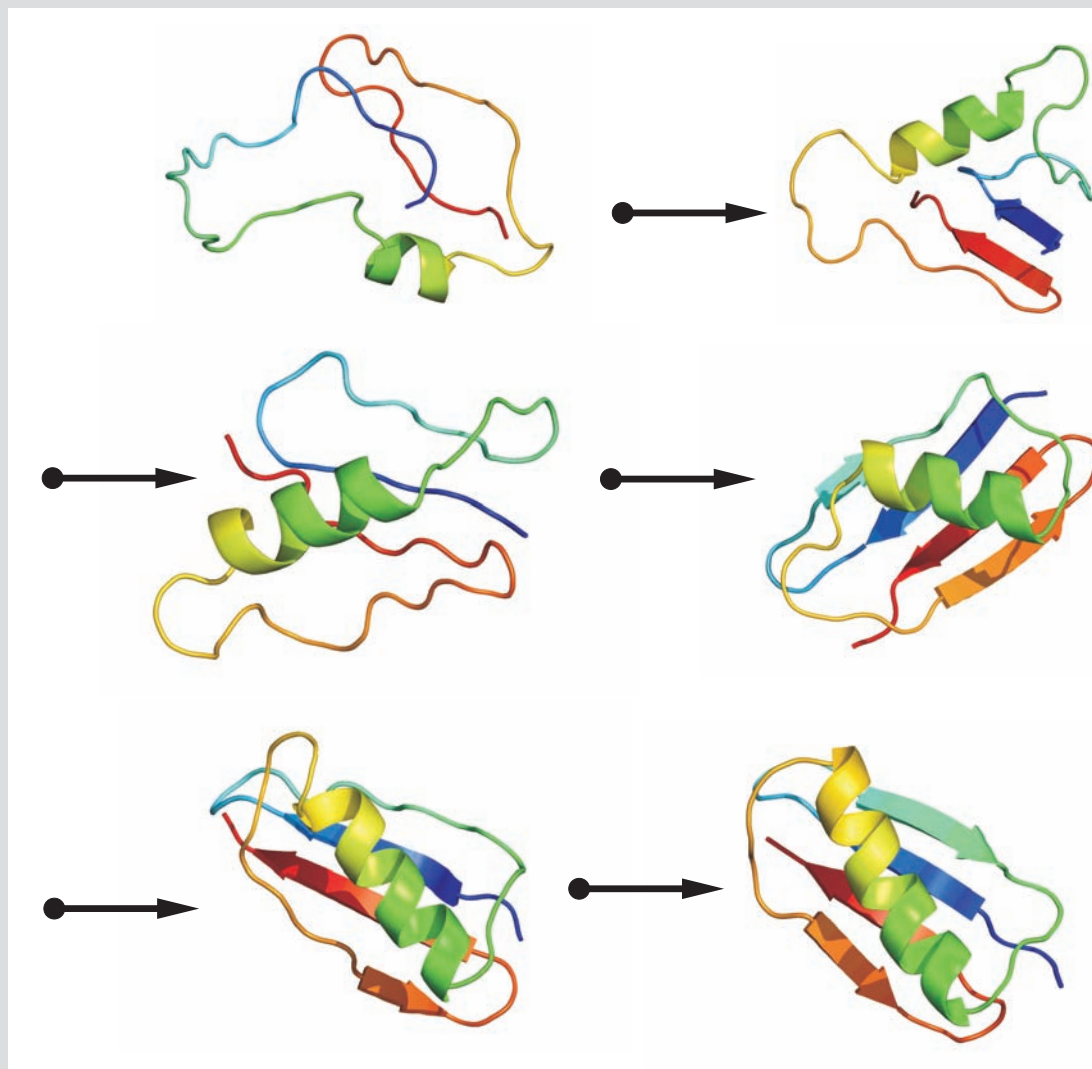
Rysunek 2. Geometria modelu CABS (skrót od C-Alpha, Beta, Side-group). Czarne kulki symbolizują położenie węgla alfa, ograniczone do siatki. Liczba możliwych pseudowiązań (wektorów siatkowych) Ca-Ca wynosi 800, eliminując w ten sposób ewentualne artefakty sieci. Na środkach wiązań Ca-Ca zaznaczono pseudoatomy używane do obliczania oddziaływań nasładowujących wiązania wodorowe. Szkielet Ca definiuje także dogodny układ współrzędnych lokalnych do obliczania położenia węgli beta (jaśniejsze kulki) zjednoczonych atomów zastępujących pozostałe części grup bocznych (duże białe kulki). Rozmiary kulek na rysunku nie odpowiadają rzeczywistym rozmiarom odpowiednich grup atomów

nych mezoskopowych modeli białek. Między innymi wprowadzono tu jawną zależność różnych oddziaływań od lokalnej geometrii łańcucha, uwzględniając w ten sposób skomplikowane efekty wielociałowe, szczególnie ważne dla modeli zredukowanych. Szczegółowy opis pola sił modelu CABS można znaleźć w niedawno opublikowanym artykule [3] oraz na stronach internetowych autora (<http://www.biocomp.chem.uw.edu.pl>).

Dyskretna przestrzeń konformacyjna łańcuchów polipeptydowych w reprezentacji CABS próbkowana jest za pomocą metody Monte Carlo. Proces obliczeniowy polega na wykonaniu olbrzymiej liczby małych losowych zmian konformacji łańcucha w jego losowo wybranych punktach [1–2]. Zmiany takie akceptowane są zgodnie ze znanym kryterium Metropolis'a. Lokalne mikromodyfikacje są zaprojektowane w sposób pozwalający interpretować ich długą sekwencję jako model dynamiki stochastycznej łańcucha w roztworze. Wykorzystywane są różne schematy symulacji: izotermiczne, tzw. symulowane schładzanie i metoda wymiany replik Monte Carlo [2].

Zredukowana reprezentacja przestrzeni konformacyjnej CABS i szybkie metody próbkowania pozwalają symulować *de novo* (tzn. wychodząc z sekwencji aminokwasów jako jedynej informacji specyficznej dla rozważanego białka) cały proces związania się małych białek globularnych. Na rysunku 3 pokazano 6 fotografii wybranych z trajektorii symulowanego schładzania małego białka globularnego – domeny B1 białka G (symbol 2GB1 w Protein Data Bank, PDB). Kończąca klatka przedstawia strukturę bardzo podobną do struktury natywnej – 1,9 Å RMSD (*Root-Mean-Square Deviation* – średnia kwadratowa odległość odpowiadających sobie atomów po najlepszym wzajemnym nałożeniu obu struktur, modelowej i doświadczalnej, z PDB). Wykonanie takiej symulacji zajmuje kilkanaście minut do godziny czasu pracy pojedynczego procesora PC.

CABS jest elementem centralnym szeregu wieloskalowych procedur do przewidywania struktury białek i asocjacji białkowych, modelowania molekularnego białek na podstawie fragmentarycznych danych doświadczalnych czy też obliczeniowych badań termodynamiki i dynamiki białek. Najbardziej ogólnie, wieloskalowość modelowania polega na wykonaniu odpowiednich symulacji za pomocą algorytmu CABS i wybraniu odpowiednich struktur (czy sekwencji struktur) w reprezentacji zredukowanej, które następnie stanowią punkt wyjścia dla bardziej szczegółowego modelowania na poziomie atomowym.



Rysunek 3. Wybrane klatki trajektorii z typowej symulacji procesu zwijania struktury małego białka globularnego. Pierwsza klatka odpowiada startowej strukturze kłębaka losowego, a ostatnia przedstawia strukturę końcową – bardzo podobną do struktury natywnej. Dla klarowności rysunku wybrano reprezentację wstęgową łańcucha polipeptydowego, pomijając szczegóły atomowe

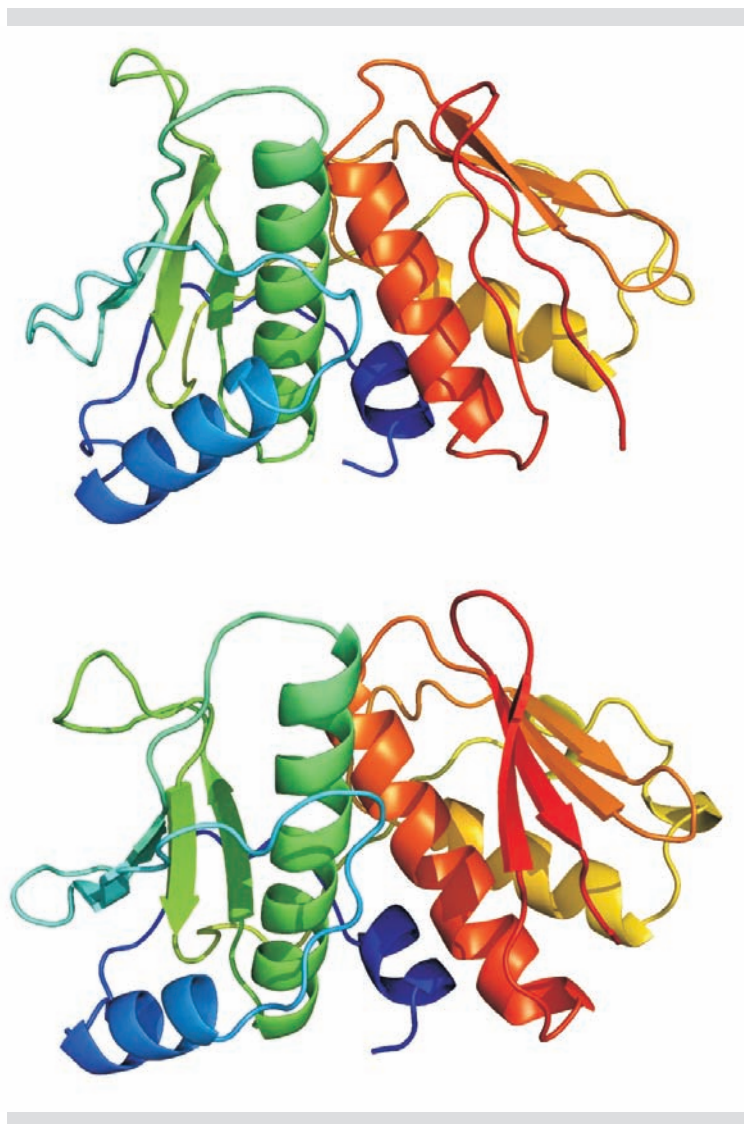
Nieco inne podejście wieloskalowe oparte na modelu CABS zostało zastosowane podczas eksperymentu CASP6 (6th *Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction* – CASP6), w którym wzięto udział ponad 200 grup z całego świata. Eksperyment CASP polega na komputerowym modelowaniu struktur białek, które właśnie mają być wyznaczone doświadczalnie w bliskiej przyszłości. Po opublikowaniu struktur doświadczalnych ocenia się zgodność, wcześniej zdeponowanych

w bezpiecznych komputerach organizatorów CASP, modeli obliczeniowych (<http://predictioncenter.org/casp6>). Procedura modelowania zastosowana podczas CASP6 przez grupę Koliński-Bujnicki [4] polegała na zbudowaniu szeregu, często bardzo przybliżonych i niepełnych, modeli na podstawie ogólnodostępnych serwerów bioinformatycznych. Modele te służyły do zgromadzenia dużej liczby więzów odległości pomiędzy atomami modelowanej struktury. Więzy użyto w algorytmie CABS do ograniczenia, w przybliżony sposób,

przestrzeni konformacyjnej modelu. Dużą liczbę otrzymanych modeli w zredukowanej przestrzeni konformacyjnej CABS poddano następnie analizie skupień [5-6], odbudowie detali atomowych [7] dla reprezentatywnych struktur, a wreszcie ich ocenie i porządkowaniu na poziomie reprezentacji petnoatomowej. Taka hierarchiczna procedura okazała się bardzo wydajna – grupa Koliński-Bujnicki została sklasyfikowana jako druga pod względem poprawności przewidzianych modeli molekularnych. Przykład struktury teoretycznie przewidzianej podczas CASP6 [4] pokazano na rysunku 4.

Podejście wieloskalowe pozwala również modelować układy składające się z większej liczby biomakro-

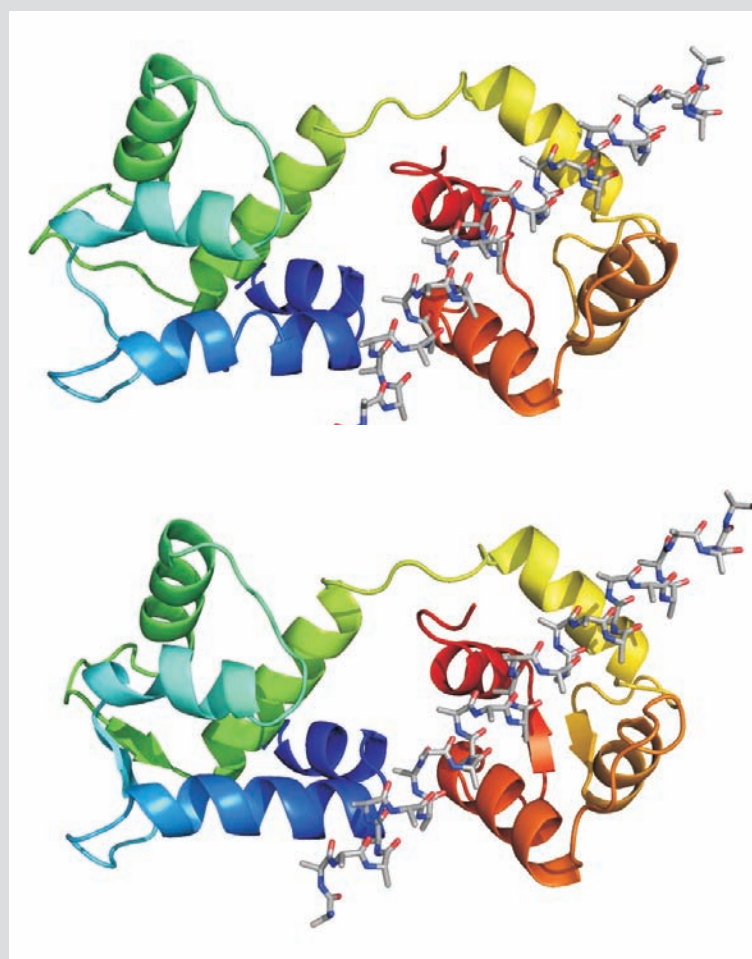
molekuł. Ma to znaczenie dla wspomaganego komputerowo racjonalnego projektowania nowych leków [8], zrozumienia procesów regulacyjnych w żywych komórkach, czy wyjaśnienia molekularnych podstaw niektórych procesów chorobowych [9]. Bardzo ważnym zagadnieniem molekularnej biologii obliczeniowej jest tzw. dokowanie. Dokowanie polega na znalezieniu sposobu łączenia się dwóch lub więcej molekuł i określeniu struktury powstałego kompleksu. Typowe zadanie obliczeniowe często wykonywane podczas projektowania leków [8] polega na zadokowaniu związku małocząsteczkowego (liganda) do znanej struktury białka (receptora). O ile znana jest struktura receptora, trady-



Rysunek 4. Przewidziana teoretycznie (góra rysunku) i doświadczalna (dół rysunku) struktura celu T0223. W celu schematycznego przedstawienia struktur użyto modelu wstęgowego opartego na węglach alfa. Kolor niebieski oznacza N-końce, a czerwony C-końce łańcuchów polipeptydowych. Model różni się od struktury krystalograficznej o 3 Å RMSD. Głównym źródłem błędów są dwa krótkie kawałki pętli na powierzchni N-końcowej domeny i między domenami (obie w odcieniach zieleni). Poprawnie przewidziano również wzajemną orientację domen

cyjne metody mechaniki molekularnej pozwalają czasami przewidzieć konformację liganda i miejsce jego dokowania. Słaba strona tego podejścia wynika z faktu, że struktura receptora w kompleksie może jakościowo różnić się od jego struktury w izolacji lub w innym kompleksie. Tzw. „giętkie dokowanie” w tradycyjnej (pełnoatomowej) mechanice molekularnej ogranicza się do optymalizacji jedynie małych fragmentów receptora. Zastosowanie zredukowanego modelu CABS umożliwia traktowanie w sposób giętki całego kompleksu [10]. Obecnie pole siłowe CABS ograniczone jest do białek i peptydów. Planowane jest rozszerzenie umożliwiające modelowanie również ligandów niepeptydowych. W pokazanych tu przykładach dokowania wieloskalowego założono ograniczoną swobodę konformacyjną całego receptora i nieograniczoną swobodę liganda (peptydu lub drugiej molekuly białka). Startowa struktura

receptora może pochodzić z badań doświadczalnych lub być wynikiem modelowania porównawczego (homologicznego). Symulacje CABS prowadzono metodą wymiany replik Monte Carlo, a końcowe struktury optymalizowano za pomocą pełnoatomowej mechaniki molekularnej [10], po uprzednim odbudowaniu reprezentacji atomowej. Struktury modelowych kompleksów porównano z ich strukturami wyznaczonymi doświadczalnie. Modelowano bardzo różne struktury, gdzie większa molekula składała się z 31–179 aminokwasów, a mniejsza z 5–63 aminokwasów. We wszystkich badanych przypadkach udało się poprawnie przewidzieć miejsce dokowania ligandów i ich własną konformację w kompleksie (przykład pokazano na rysunku 5). Tak więc, zaproponowana metoda może być wykorzystywana do przewidywania nieznanych jeszcze struktur kompleksów białko–peptyd i białko–białko.



Rysunek 5. Kompleks Troponin C – Troponin I. Receptor przedstawiono za pomocą modelu wstęgowego, a ligand – szkieletowego. Model teoretyczny (góra rysunku) i struktura krystalograficzna (dół) są bardzo podobne (1,76 Å RMSD po najlepszym natożeniu)

Dalsze prace nad przedstawionym tu wieloskalowymi sposobami modelowania biomakromolekuł będą zmierzały w kierunku pełnego zautomatyzowania procedur obliczeniowych do przewidywania struktur białek dla całych genomów, opracowania mezoskopowej reprezentacji innych niż białka molekuł oddziałujących z białkami (związki maćcząsteczkowe, kwasy nukleinowe, fosfolipidy itd.), a także mezoskopowego modelowania wielkich układów biomakromolekularnych.

Literatura uzupełniająca

- [1] D. Baker, A. Sali, "Protein Structure Prediction and Structural Genomics", *Science*, **294**:93–6 (2001).
- [2] A. Kolinski and J. Skolnick, "Reduced Models of Proteins and Their Applications", *Polymer*, **45**:511–524 (2004).
- [3] A. Kolinski, "Protein Modeling and Structure Prediction with a Reduced Representation", *Acta Biochimica Polonica*, **51**:349–371 (2004).
- [4] A. Kolinski and J.M. Bujnicki, "Generalized Protein Structure Prediction Based on Combination of Fold-recognition with *de novo* Folding and Evaluation of Models", *Proteins*, **61**(S7):84–90 (2005).
- [5] D. Gront and A. Kolinski, "HCPM – Program for Hierarchical Clustering of Protein Models", *Bioinformatics*, **21**:3179–3180 (2005).
- [6] D. Gront and A. Kolinski, "BioShell – A Package of Tools for Structural Biology Computations", *Bioinformatics*, **22**:621–622 (2006).
- [7] D. Gront, S. Kmiecik and A. Kolinski, "BBQ – Backbone Building from Quadrilaterals. A Fast and Accurate Algo-

rithm for Protein Backbone Reconstruction from Alpha Carbon Coordinates", *J. Comput. Chem.* (w druku).

- [8] R. Sicinski, A. Kolinski, P. Rotkiewicz, W. Sicinska, J.M. Prahl, C.M. Smith and H.F. De Luca, "2-Ethyl and 2-Ethylidene Analogs of 1 α , 25-Dihydroxy-19-norvitamin D₃: Synthesis, Conformational Analysis, Biological Activities, and Docking to the Modeled rVDR Ligand Binding Domain", *J. Medicinal Chem.*, **45**:3366–3380 (2002).
- [9] E. Malolepsza, M. Boniecki, A. Kolinski and L. Piel, "Theoretical Model of Prion Propagation: a Misfolded Protein Induces Misfolding", *Proc. Natl. Acad. Sci. USA*, **102**:7835–7840 (2005).
- [10] M. Kurcinski and A. Kolinski, "Steps Towards Flexible Docking: Modeling of Three-dimensional Structures of the Nuclear Receptors Bound with Peptide Ligands Mimicking Co-activators' Sequences", *J. Steroid Biochem. and Mol. Biol.* (w druku).

Abstract

Reduced computer modeling of proteins has now about 30 years history. In spite of enormous increase of computing abilities the reduced models are still very important tools for theoretical studies of proteins. It is shown that the reduced-space modeling can be integrated with a detailed all-atom simulations. Such multiscale approach is crucial for high-resolution protein structure predictions, predictions of protein interactions, computer-aided drug design and study of protein dynamics and thermodynamics.

Profesor Andrzej Koliński chemik, profesor na Wydziale Chemii Uniwersytetu Warszawskiego. Specjalizuje się w biofizyce teoretycznej i termodynamice statystycznej. Odbił liczne zagraniczne pobyty naukowe, m.in. w Washington University w Saint Louis, w The Scripps Research Institute w Kalifornii, w Donald Danforth Plant Science Center oraz w Center of Excellence in Bioinformatics na

Uniwersytecie w Buffalo. Od 2004 roku jest profesorem stowarzyszonym z L.H. Baker Center for Bioinformatics and Biological Statistics na Uniwersytecie Stanowym w Iowa. W latach 1995–1999 był badaczem międzynarodowym Instytutu Howarda Hughesa. W 2008 roku został odznaczony Medalem Jana Zawidzkiego Polskiego Towarzystwa Chemicznego, a w 2009 otrzymał Nagrodę Fundacji na Rzecz Nauki Polskiej.



Dynamika regularna i chaotyczna w układach technicznych z tarciami i uderzeniami

Na podstawie odczytu wygłoszonego w dniu 14 grudnia 2006 roku

Jan Awrejcewicz i Grzegorz Kudra

Katedra Automatyki i Biomechaniki Politechniki Łódzkiej

95

W fizyce i technice spotykamy wiele układów, które można opisać jako pracujące w różnych procesach, przy czym czas przejścia z jednego reżimu do innego jest bardzo krótki w porównaniu z czasem pracy układu w poszczególnych procesach. W takich przypadkach zwykle możliwe jest przyjęcie, że przejście z jednego reżimu do drugiego jest natychmiastowe i dyskretne. W ten sposób obiekt taki modeluje się jako układ dynamiczny kawałkami gładki (PWS) z pewnymi rodzajami nieciągłości, którego dynamikę można opisać równaniem

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$$

gdzie $\mathbf{x} = \mathbf{x}(t) \in \mathbb{R}^n$ reprezentuje stan układu w chwili t , natomiast odwzorowanie $\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ jest funkcją kawałkami gładką, czyli przestrzeń fazowa $D \in \mathbb{R}^n$ jest podzielona na skończoną ilość podobszarów V_i (we wnętrzu których funkcja \mathbf{f} jest gładka) rozdzielonych $(n-1)$ -wymiarowymi hiperpowierzchniami $\Sigma_{i,j}$, również co najmniej kawałkami gładkimi.

Gdy ruch układu odbywa się we wnętrzu jednego z obszarów V_i , to układ zachowuje się jak układ gładki. Natomiast gdy orbita przecina jedną z powierzchni $\Sigma_{i,j}$, układ wykazuje zachowanie nieciągłe. W zależności od stopnia nieciągłości, układy dynamiczne kawałkami gładkie (PWS) możemy podzielić na trzy grupy:

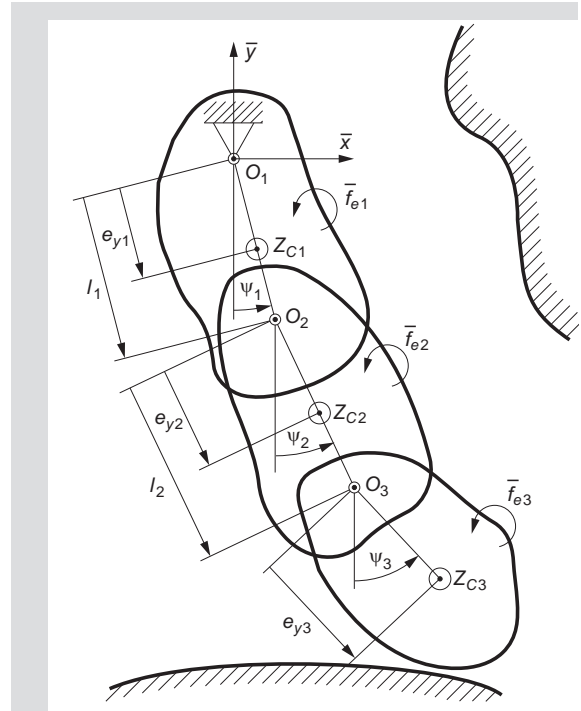
- I. Układy z nieciągłym Jakobianem $D\mathbf{f}$, z ciągłym lecz niegładkim polem wektorowym \mathbf{f} oraz z gładkim stanem układu \mathbf{x} (funkcja \mathbf{f} jest klasy C^0).
- II. Układy z nieciągłym polem wektorowym \mathbf{f} oraz z ciągłym lecz niegładkim wektorem stanu \mathbf{x} .

III. Układy z nieciągłym wektorem stanu \mathbf{x} . W tym przypadku za każdym razem, gdy układ wchodzi w kontakt z jedną z powierzchni nieciągłości $\Sigma_{i,j}$, jego stan doznaje skoku $\mathbf{x}^+ = \mathbf{g}(\mathbf{x}^-)$, gdzie \mathbf{x}^- oznacza stan układu bezpośrednio przed kontaktem z powierzchnią $\Sigma_{i,j}$, \mathbf{x}^+ jest stanem układu bezpośrednio po kontakcie, natomiast $\mathbf{g}(\mathbf{x})$ jest funkcją co najmniej kawałkami gładką na powierzchniach $\Sigma_{i,j}$.

Układy dynamiczne kawałkami gładkie (PWS) grupy II (również grupy I jako podklasy) są też nazywane **układami Filippowa**, dla których istnieje osobna teoria. Układy elektroniczne z elementami posiadającymi charakterystyki, które mogą być traktowane jako nieciągłe (diody, tranzystory itp.) są często modelowane jako układy dynamiczne kawałkami gładkie (PWS). Wśród układów mechanicznych spotyka się układy z tarciami suchym, które można modelować jako układy o skokowej charakterystyce tłumienia (grupa II). Poza tym układ mechaniczny może posiadać także nieciągłą charakterystykę sztywności. W szczególności układy mechaniczne z uderzeniami mogą być modelowane jako układy o gwałtownie zmieniającej się sztywności (grupa II) — jeśli przyjmiemy model ciał podatnych. Zwykle jednak przyjmuje się model sztywnych zderzających się ciał. W tym przypadku układ jest układem dynamicznym kawałkami gładkim (PWS) grupy III, gdyż podczas kontaktu układu z powierzchnią $\Sigma_{i,j}$ następuje skokowa zmiana prędkości odpowiednio do prawa Newtona opartego na współczynniku restytucji. Układ taki jest faktycznie układem o więzach

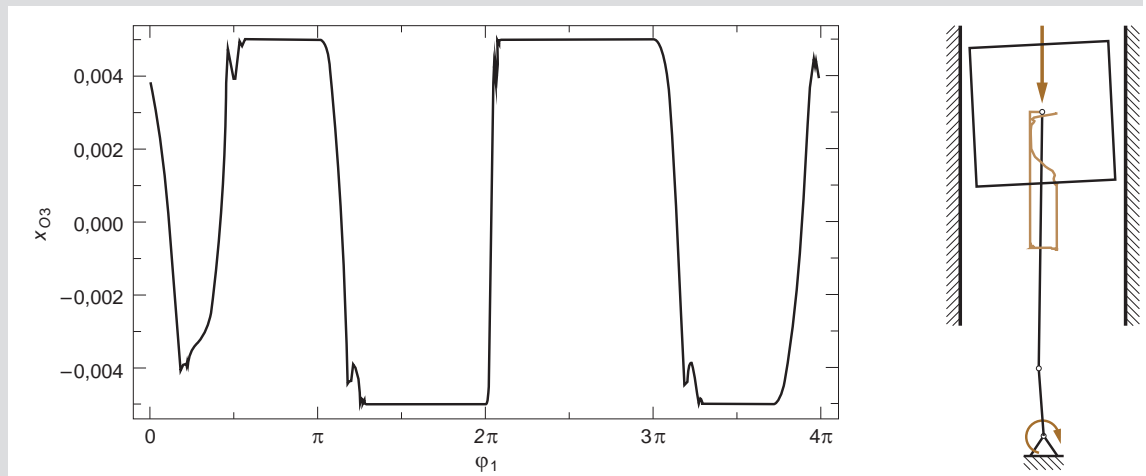
jednostronnych opisanych nierównościami algebraicznymi. Należy zauważyć, że jest to równoważne impulsowi funkcji \mathbf{f} typu delty Diraca po jednej ze stron powierzchni $\Sigma_{i,j}$. Jednak, jak już wielokrotnie zostało pokazane, model podatnych zderzających się ciał jest zbieżny do modelu zderzeń opartego na współczynniku restytucji, gdy sztywność ciał rośnie do nieskończoności. Ponadto w pewnych układach mogą występować jednocześnie różne rodzaje nieciągłości. Przykładem może być układ mechaniczny, w którym występują jednocześnie uderzenia i tarcie suche. Oba te zjawiska mogą być od siebie niezależne lub model uderzeń może uwzględniać zjawiska tarcia suchego na powierzchni zderzających się ciał.

Po zdefiniowaniu modelu matematycznego, następnym problemem do rozwiązania jest uzyskanie rozwiązania numerycznego tak określonych równań PWS, czyli symulacja układu. Zauważmy, że można to zrobić „sklejając” rozwiązania gładkie, otrzymane metodami klasycznymi pomiędzy kolejnymi punktami nieciągłości i wykrywając te punkty nieciągłości. W każdym z tych punktów możliwe jest też wprowadzenie odwzorowania stanu układu do innego stanu poprzez funkcję \mathbf{g} . Jeśli ilość punktów nieciągłości jest skończona w skończonym czasie, to taka aproksymacja jest wystarczająca. Może jednak się zdarzyć, że rozwiązanie w pewnym czasie nie opuszcza określonej powierzchni $\Sigma_{i,j}$ i wtedy konieczna jest redukcja formalnego rzędu problemu w tym przedziale lub też problem re-



Rysunek 1. Model potrójnego wahadła fizycznego

dukuje się do zagadnienia opisanego za pomocą równań algebraiczno-różniczkowych (DAEs). Zanim pojawi się stan trwałego pokrycia trajektorii i powierzchni

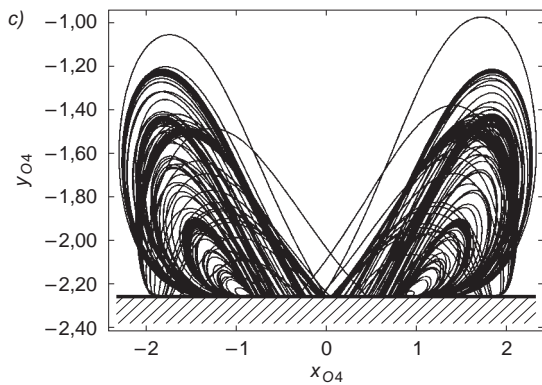
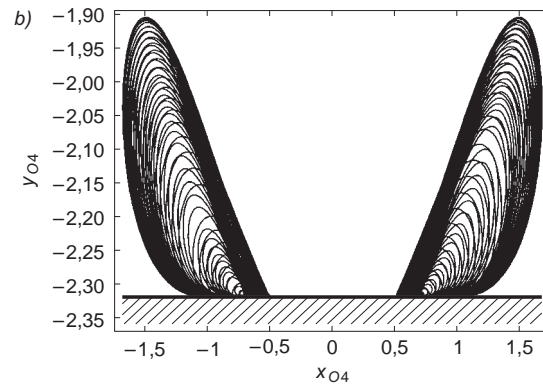
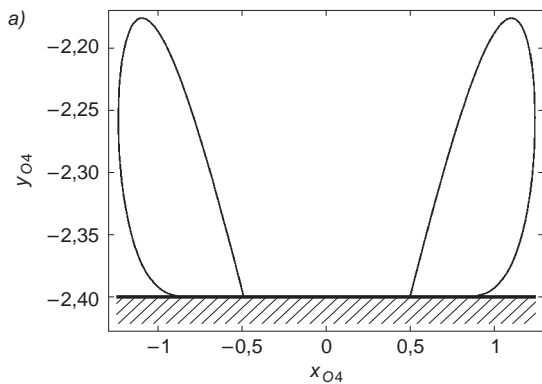


Rysunek 2. Odpowiedź układu korbowego dla parametrów współczynnika restytucji $e = 0,5$ (x_{O3} jest współrzędną położenia osi sworznia tłokowego wzdłuż osi prostopadłej do osi cylindra, φ_1 jest położeniem kątowym wału korbowego)

nieciągłości $\Sigma_{i,j}$, może dojść także do nieskończenie wielu przebiec tej powierzchni przez trajektorię. W celu aktywnego rozwiązania tego zagadnienia numerycznie można „ustalić sztucznie” pewien najkrótszy czas pomiędzy kolejnymi przebieciami, po przekroczeniu którego następuje trwały kontakt trajektorii i powierzchni $\Sigma_{i,j}$.

Opisane podejście autorzy zastosowali przy modelowaniu i analizie numerycznej płaskiego potrójnego wahadła ze sztywnymi ogranicznikami ruchu [1–3], pokazanego na rysunku 1. Układ został „zamodelowany” jako układ III typu. Przeszkody stanowią powierzchnie niętgładkości w przestrzeni stanów, a przebiecie jednej z nich przez trajektorię jest wykrywane i stan układu doznaje skoku realizowanego poprzez funkcję \mathbf{g} . Funkcja \mathbf{g} określa prawo uderzenia układu o przeszkodę o zadanej powierzchni, przy czym przyjęto, że powierzchnia przeszkody jest gładka, czyli nie występuje tarcie suche. Oznacza to, że impuls uderzenia jest prostopadły do powierzchni uderzenia. Wspomniany warunek oraz dodatkowe równanie określające

zmianę wektora prędkości w kierunku prostopadłym do powierzchni uderzenia pozwalają, zgodnie z definicją współczynnika restytucji, określić prędkości po uderzeniu o pojedynczą przeszkodę. Uwzględniono również możliwość pojawienia się trwałego kontaktu układu z jedną lub większą liczbą przeszkód – bardzo często poprzedzonego „zagęszczeniem” w czasie do nieskończoności ilości zderzeń. Problem ten został rozwiązany w sposób opisany powyżej. Kontakt trwały układu z przeszkodą wymagał także określenia takiej reakcji przeszkody, która zapewnia ten kontakt. Zgodnie z przedstawionymi założeniami, autorzy zbudowali model układu korbowego silnika spalinowego czterosuwowego (będącego szczególnym przypadkiem potrójnego wahadła), w którym tłok w cylindrze porusza się z luzem i dochodzi do uderzeń [3]. Rysunek 2 przedstawia wynik symulacji takiego układu dla współczynnika restytucji 0,5. Jest to rozwiązanie okresowe opisujące typowe zachowanie się tłoka, polegające na jego sześciokrotnym przechodzeniu z jednej strony cylindra na drugą w czasie jednego cyklu pracy silnika.



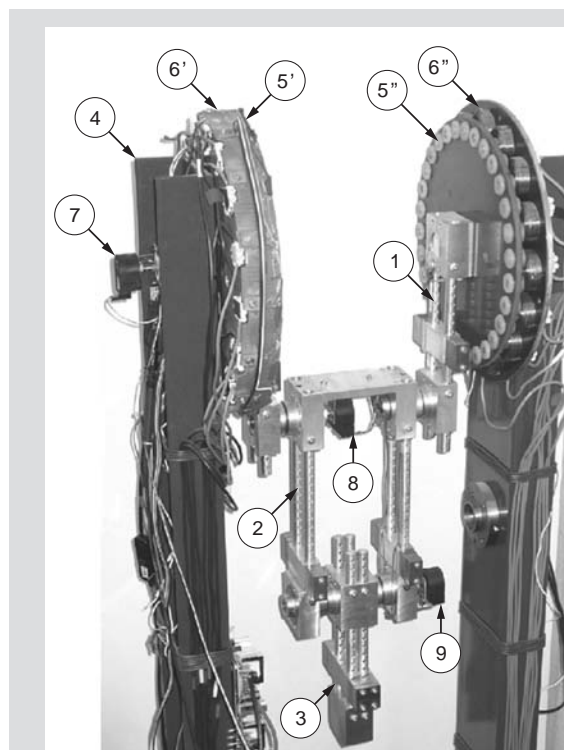
Rysunek 3. Rozwiązanie okresowe (a), quasi-okresowe (b) i chaotyczne (c) w układzie potrójnego wahadła z wymuszeniem okresowym i poziomą przeszkodą (x_{O4} i y_{O4} są współrzędnymi końcówki trzeciego wahadła w prostokątnym układzie współrzędnych umieszczonym w płaszczyźnie ruchu wahadła). Wykładniki Lapunowa: (b) – 0,00, 0, –0,10, –1,05, –2,0, $-\infty$, $-\infty$; (c) – 0,01, 0, –0,12, –0,74, –1,92, $-\infty$, $-\infty$. Współczynnik restytucji $e = 0$

Kolejnym krokiem było badanie stabilności orbit w tego typu układach. Ponieważ są to układy kawałkami gładkie, to na każdym odcinku, na którym trajektoria jest gładka, również zaburzenia rozchodzą się według typowych równań otrzymanych z linearyzacji równań układu wokół badanej orbity. Natomiast w punktach niegładkości należy odpowiednio przekształcić stan zaburzenia [1–3, 5] w zależności od Jakobianów Dg i Dh (gdzie $h = 0$ wyznacza powierzchnię przeszkody). Podejście takie pozwala na stosowanie klasycznych metod do obliczania wykładników Lapunowa, badania stabilności orbit okresowych i badania klasycznych bifurkacji orbit okresowych w układzie mechanicznym ze sztywnymi ogranicznikami ruchu. Na rysunku 3 zostały przedstawione przykładowe rozwiązania okresowe, quasi-okresowe i chaotyczne w układzie potrójnego wahadła z harmonicznym wymuszeniem pierwszego wahadła i przy wprowadzeniu tylko z poziomej przeszkody. Zauważmy, że ze względu na istnienie segmentów orbity o trwałym kontakcie z przeszkodą, po dwa wykładniki Lapunowa są równe minus nieskończoności.

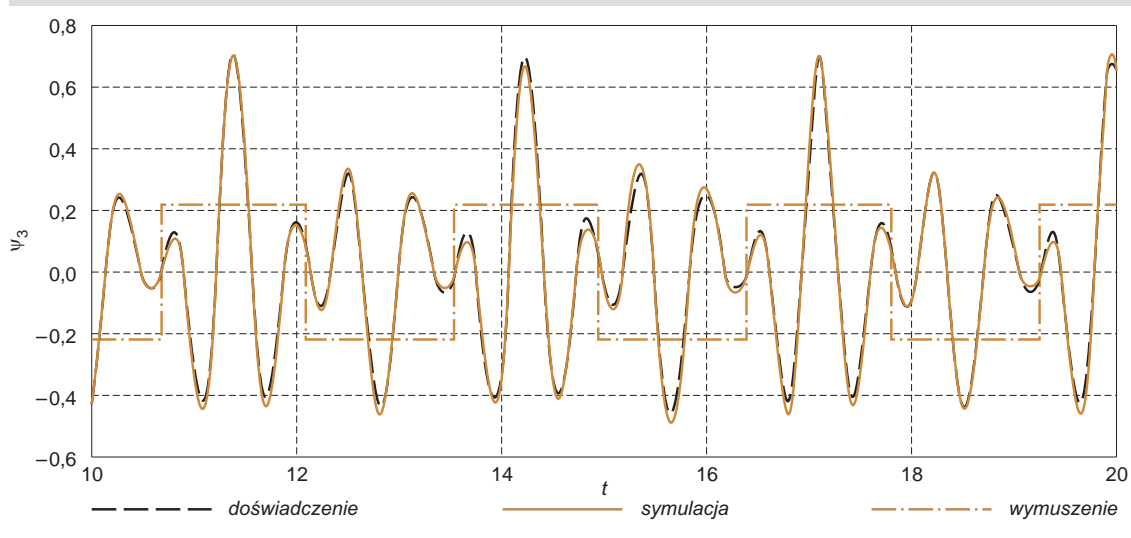
W układach dynamicznych kawałkami gładkich (PWS) mogą występować wszystkie bifurkacje typowe dla układów gładkich i można je badać klasycznymi metodami z pewnymi, wcześniej omówionymi modyfikacjami. Oprócz wspomnianych rozwiązań bifurkacyjnych mogą się pojawić jeszcze inne bifurkacje – charakterystyczne tylko dla układów niegładkich. Podstawowym typem bifurkacji charakterystycznym dla układów dynamicznych kawałkami gładkich (PWS) jest bifurkacja znana w literaturze jako **bifurkacja typu grazing**. Bifurkacja ta pojawia się wówczas, gdy część trajektorii (np. orbity okresowej, quasi-okresowej lub innej) staje się styczna do jednej z powierzchni nieciągłości (gdy powolnej zmianie ulega parametr bifurkacyjny). Istnieje bardzo szczególna cecha bifurkacji typu *grazing* odróżniająca ją od klasycznych bifurkacji lokalnych gładkich pól wektorowych. Nie można jej przewidzieć obserwując Jakobian bezpośrednio przed bifurkacją, ponieważ trajektoria zachowuje się tak jak orbita układu gładkiego i nie ma żadnej informacji o tym, kiedy nastąpi kontakt z powierzchnią nieciągłości, co może prowadzić nieraz do dramatycznych zmian jakościowych zachowania się układu.

Oprócz badań numerycznych potrójnego wahadła, autorzy prowadzą równoległe badania eksperymentalne [4]. W lutym 2005 roku w Katedrze Automatyki i Biomechaniki zakończono budowę i uruchomiono

stanowisko potrójnego wahadła fizycznego (rysunek 4). Pierwsze wahadło jest wymuszane okresowym momentem siły o przebiegu prostokątnym w czasie. Dane pomiarowe są zbierane za pośrednictwem precyzyjnych potencjometrów obrotowych i systemu LabView, przekazywane do komputera i tu dalej przetwarzane. Jako model matematyczny został wykorzystany zbudowany wcześniej model potrójnego wahadła. W celu uzyskania lepszej zgodności modelu z układem rzeczywistym opory w łożyskach „zamodelowano” jako złożenie tłumienia wiskotycznego i tarcia suchego opisanego przez charakterystykę Stribeck. Parametry modelu zostały oszacowane przez porównanie sygnałów wyjściowych z modelu i układu rzeczywistego (przy tym samym sygnale wejściowym), a jako kryterium dopasowania przyjęto sumę kwadratów odchyłek między sygnałami w kolejnych punktach próbkowania. Do minimalizacji kryterium (dla kilku rozwiązań okresowych) została użyta metoda sympleksów, z tym, że



Rysunek 4. Potrójne wahadło fizyczne – model doświadczalny: 1, 2, 3 – ogniw; 4 – statyw; 5', 5'' – tarcze wirników (tarcza 5' niewidoczna); 6', 6'' – tarcze stojanów (tarcza 6' z optycznym komutatorem); 7, 8, 9 – przetworniki położenia kąтового ogniw

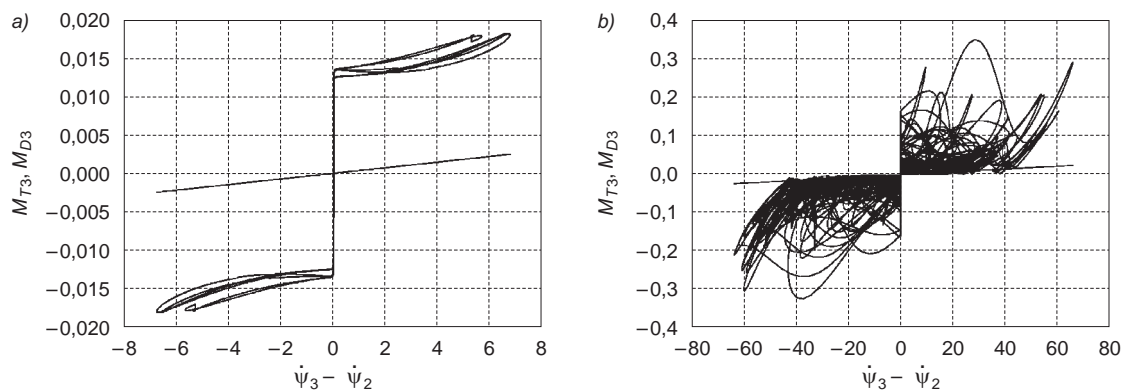


Rysunek 5. Rozwiązanie okresowe dla częstotliwości wymuszenia $f = 0,35$ Hz otrzymane z eksperymentu i numerycznie

wraz z parametrami oszacowaniu podlegaty również warunki początkowe rozwiązania numerycznego.

Na rysunku 5 przedstawiono porównanie rozwiązań otrzymanych z eksperymentu i z modelu matematycznego, a na rysunku 6 – charakterystyki momentów tarcia suchego i tłumienia wiskotycznego w trzecim przegubie dla przykładowych rozwiązań – okresowego i chaotycznego. Z zaprezentowanych rysunków wynika, że otrzymano bardzo dobrą zgodność ekspe-

rymentu i wyników otrzymanych z modelu numerycznego [4]. Wykryte zostały numerycznie różne obszary chaosu, których istnienie zostało następnie potwierdzone eksperymentalnie. Potwierdzenie to dotyczy zarówno granic poszczególnych obszarów, jak i pewnych jakościowych własności rozwiązań chaotycznych – np. pełnych obrotów wykonywanych przez poszczególne wahadła. Prowadzi to do wniosku, że zastosowany model matematyczny wahadła wraz z parametrami



Rysunek 6. Charakterystyki momentów tarcia suchego i tłumienia wiskotycznego w trzecim przegubie: (a) dla rozwiązania okresowego przy częstotliwości wymuszenia $f = 0,6$ Hz; (b) dla rozwiązania chaotycznego przy $f = 0,73$ Hz

może służyć jako narzędzie do szybkiego wyszukiwania i wyjaśniania różnorodnych zjawisk występujących w układzie rzeczywistym.

Na koniec tego krótkiego streszczenia do wykładu wygłoszonego w ramach „Konwersatorium” należy podkreślić, że prezentowane tu wyniki uzyskano dzięki wsparciu finansowemu Ministerstwa Edukacji i Szkolnictwa Wyższego w latach 2005–2008 w ramach projektu badawczego 4 T07A 031 28.

Bibliografia

- [1] Awrejcewicz J., Kudra G., Lamarque C.-H.: *Investigation of Triple Physical Pendulum with Impacts Using Fundamental Solution Matrices*. Int. J. Bifurcation and Chaos **14** (2), 4191–4213 (2004).
- [2] Awrejcewicz J., Kudra G.: *The Piston-connecting Rod-crankshaft System as a Triple Physical Pendulum with Impacts*. Int. J. Bifurcation and Chaos **15** (7), 2207–2226 (2005).
- [3] Awrejcewicz J., Kudra G.: *Stability Analysis and Lyapunov Exponents of a Multi-body Mechanical System with Rigid Unilateral Constraints*. Nonlinear Analysis **63** (5–7), 909–918 (2005).
- [4] Awrejcewicz J., Supet B., Kudra G., Wasilewski G., Olejnik P.: *Numerical and Experimental Study of Regular and*

Chaotic Behaviour of Triple Physical Pendulum. Fifth EU-ROMECH Nonlinear Dynamics Conferencem, ENOC2005, Eindhoven, The Netherlands, 1817–1824 (2005).

- [5] Müller P.C.: *Calculation of Lyapunov Exponents for Dynamic Systems with Discontinuities*. Chaos, Solitons and Fractals **5**, 1671–1691 (1995).

Abstract

The work is devoted to modeling and numerical analysis of technical systems with impacts and dry friction. The model of a mechanical system subjected to unilateral constraints, together with the model of stability, are developed. Examples of analysis are presented for plane triple physical pendulum with rigid limiters motions. The model of the piston-connecting rod-crankshaft system of the combustion engine, as a special case of triple pendulum, is also built and gives results conforming experiments. Then an experimental rig of the triple physical pendulum with the first body periodically forced is built. A mathematical model of the real pendulum is created, where friction in joints is modeled as a composition of dry friction and damping. Good agreement between model and real system is obtained.

Profesor Jan Awrejcewicz specjalista w obszarze mechaniki nieliniowej obejmującej nieklasyczne zagadnienia dynamiki bifurkacyjnej i chaotycznej w dynamice maszyn i konstrukcji. Kierownik Katedry Automatyki i Biomechaniki Politechniki Łódzkiej. Otrzymał liczne nagrody, stypendia i wyróżnienia, jak np.: Fundacji Fulbrighta, Fundacji im. Tadeusza Kościuszki (Nowy Jork), Japońskiej Fundacji Promocji Nauki (Tokio), Centrum Badawczego Nauki i Inżynierii (Uniwersytet Tokijski i Mitsubishi), NATO, rządu francuskiego i wiele innych. Uhonoro-

wany w 2006 roku prestiżową nagrodą „Złota Lampa” przyznaną przez PGNiG, w 2009 roku stypendium Mistrz otrzymanym od Fundacji na Rzecz Nauki Polskiej, a w 2011 roku został uhonorowany prestiżową nagrodą badawczą przyznaną przez niemiecką Fundację im. Alexandra von Humboldta.

Doktor Grzegorz Kudra specjalista w dziedzinie mechaniki i dynamiki nieliniowej. Pracuje w Katedrze Automatyki i Biomechaniki Politechniki Łódzkiej.

Einstein po stu latach

Na podstawie odczytu wygłoszonego w dniu 22 marca 2007 roku

Andrzej Kajetan Wróblewski

Wydział Fizyki Uniwersytetu Warszawskiego

101

O dokonaniach Alberta Einsteina większość ludzi wie tyle, że stworzył on teorie, które podobno zrewolucjonizowały fizykę. Panuje też przekonanie, zwłaszcza wśród części humanistów, że dzieła Einsteina nie da się w ogóle zrozumieć, ponieważ poglądy głoszone przez niego na temat czasu, przestrzeni i kwantów są sprzeczne ze zdrowym rozsądkiem.

W tym wykładzie spróbujemy zastanowić się nad rolą „zdrowego rozsądku” w poznawaniu świata. Nie trudno będzie wykazać na przykładach, że pojęcie i zakres zdrowego rozsądku w ciągu wieków ulegały daleko idącym zmianom. Analiza twórczości Einsteina pomoże nam także wyjaśnić współczesne znaczenie słowa „rozumieć”.

Przypomnijmy, że przez długi czas uważano – zgodnie ze zdrowym rozsądkiem – że Ziemia jest płaska i leży w środku świata. O tym, że świat jest płaski, zdaje się nas przecież przekonywać codzienne doświadczenie. Wprawdzie w starożytnej Grecji pitagorejczycy pierwsi rozważali możliwość, że Ziemia ma kształt kuli, ale były to poglądy właściwe tylko nielicznej warstwie uczonych filozofów, podczas gdy zwykli ludzie byli przekonani, że żyją w świecie płaskim. Zresztą poglądy filozofów greckich zostały szybko odrzucone i wyśmiane u schyłku starożytności i we wczesnym średniowieczu.

W IV wieku Laktancjusz napisał znane dzieło *Divinarum institutionum libri VII* (*Boskie ustanowienie w siedmiu księgach*). W trzeciej księdze *De falsa sapientia philosophorum* (*O fałszywej mądrości filozofów*) udowodnił jak niebezpieczne jest logiczne argumen-

owanie oparte na fałszywych przesłankach. Niektórzy filozofowie z faktu, że ciała niebieskie wschodzą na wschodzie a zachodzą na zachodzie – pisał Laktancjusz – wyciągnęli absurdalny wniosek, iż Ziemia jest kulista i że istnieją antypody. W rozdziale 24 tej książki Laktancjusz wyrażał się więc z ironią:

Czy jest ktoś tak nierozsądny, żeby uwierzył, że są ludzie, którzy mają stopy nad głowami, ... że zboża i drzewa rosną w dół, że deszcz, śnieg i grad padają w górę na ziemię...

Postęp nauk geograficznych i astronomii doprowadził jednak w końcu do uznania, że Ziemia ma kształt kulisty. Dziś uczy się o tym dzieci w szkole i niemal wszyscy są o tym przekonani. Niemal wszyscy, gdyż zdarzają się jednak wyjątki. Od dawna istniały różne stowarzyszenia, których celem było udowadnianie, że uczeni się mylą, a Ziemia jest w rzeczywistości płaska. Nawet postęp astronautyki i rozpowszechnienie widoków Ziemi z kosmosu nie zakończyły tej sprawy. Nadal istnieje Towarzystwo Płaskiej Ziemi (*Flat Earth Society*), które ma swą siedzibę w Lancaster (w Kalifornii) i ogłasza się w internecie. Za niedużą składkę członkowską można otrzymywać broszury propagandowe udowadniające płaskość Ziemi.

Kolejnym przekonaniem „zdroworozsądkowym” było to, że żyjemy w środku świata. Na tej zasadzie oparty był cały system Arystotelesa, który trwał przez wiele stuleci. Przypomnijmy więc, że kiedy Mikołaj Kopernik zaproponował inne rozwiązanie, to naraził się na ostrą krytykę. Astronom jezuita z Rzymu –

Christopher Clavius wyraził się w 1570 roku, że Kopernik stawia hipotezy:

niedorzeczne, absurdalne, sprzeczne ze zdrowym mniemaniem i zdrowym rozsądkiem i rzec można zuchwałe.

Znów mieliśmy tu nawiązanie do zdrowego rozsądku, który po raz kolejny okazał się ztym doradcą.

Isaac Newton na początku swego wielkiego dzieła *Zasady matematyczne filozofii naturalnej* podał definicje czasu i przestrzeni:

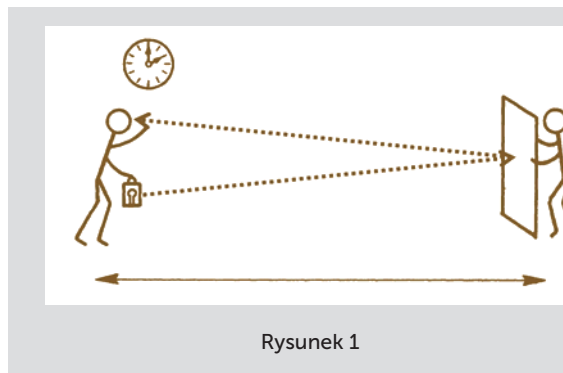
Czas absolutny, prawdziwy i matematyczny, sam z siebie i przez swą naturę upływa równomiernie bez związku z czymkolwiek zewnętrznym i inaczej nazywa się trwaniem...

Przestrzeń absolutna, przez swą naturę, bez związku z czymkolwiek zewnętrznym, pozostaje zawsze taka sama i nieruchoma...

Te definicje, nazwane przez Newtona aksjomatami, były zgodne z powszechnym zdrowym rozsądkiem. Przestrzeń miała być tylko sceną, na której rozgrywają się zjawiska fizyczne i astronomiczne. Czas miał także nieubłaganie odmierzanie i nie można było mieć wpływu na jego bieg. Nic dziwnego, że aksjomaty Newtona dotyczące czasu i przestrzeni weszły do potocznej świadomości i do dziś przyjmuje je większość ludzi.

Tymczasem postęp fizyki zmusił nas do zmiany tych od dawna utrwalonych wyobrażeń o czasie i przestrzeni. To właśnie jest pierwsze z osiągnięć Einsteina, które tu omawiam. Zaczęto się to wszystko dawno temu od prędkości światła.

Zasadę pomiaru prędkości światła podał już Galileusz w swoich *Rozmowach i dowodzeniach matematycznych* (1638). Jego zdaniem do pomiaru potrzebne były dwie osoby, z których każda trzymała zapaloną latarkę lub inne źródło światła, w taki sposób by mogła ręką ją zastaniać i odstaniać przed wzrokiem drugiej. Te osoby, usytuowane naprzeciwko siebie w pewnej odległości, na przemian odstaniały i zastaniały światło — gdy jedna zobaczyła odstaniające się światło u drugiej, zaraz odstaniała swoje. Po nabraniu wprawy przez eksperymentatorów, po odstąpieniu jednego światła miało zaraz następować odstąpienie drugiego. Znając odległość między obserwatorami oraz czas, po którym do pierwszego obserwatora dochodziło światło od drugiego, można było w zasadzie obliczyć prędkość świa-



ła. Zamiast drugiego obserwatora można było też zastosować zwierciadło, od którego mogło odbijać się światło (rysunek 1). Przepis podany przez Galileusza był poprawny, ale prędkość światła jest tak ogromna, że oczekiwane przez niego ewentualne opóźnienie było niemierzalnie małe.

Isaac Newton zastosował Metodę Galileusza, chcąc zmierzyć prędkość dźwięku. Osoby zwiedzające piękny gmach Trinity College w Cambridge mogą nadal oglądać długą galerię, w której były wykonywane te pomiary. Galeria ta służy jako dobry falowód dla fal głosowych, dźwięk rozchodzi się tam doskonale i można słyszeć nawet wielokrotne echa. Stojąc na jednym końcu, Newton klaskał w dłonie i za pomocą wahadła mierzył czas do usłyszenia echa — dwu-, trzy-, a nawet czterokrotnego.

Metodę Galileusza udoskonalono, przystosowując ją do pomiaru niezmiernie krótkich odstępów czasu. Po raz pierwszy w połowie XIX wieku można było już zmierzyć prędkość światła z dokładnością lepszą niż jeden procent. Następnie uzyskano jeszcze większą dokładność — drobnego ułamka procenta i wtedy stwierdzono **zadziwiający fakt**. Otóż ze wszystkich doświadczeń, w których wykonywano pomiary prędkości światła wynikało, że prędkość światła (w próżni) c jest zawsze taka sama — niezmienna, niezależna od kierunku, barwy światła, prędkości ruchu źródła względem obserwatora itd.

Ten **fakt doświadczalny** jest oczywiście sprzeczny ze zdrowym rozsądkiem, ponieważ w doświadczeniach mechanicznych (na przykład przy rzucaniu piłką) obserwujemy proste dodawanie się prędkości — piłka rzucona przez osobę jadącą na wózku ma względem nieruchomego obserwatora większą prędkość niż względem osoby rzucającej.

Doświadczenia nie można jednak ignorować. Najwidoczniej światło ma szczególne właściwości, inne niż piłki i podobne przedmioty. W 1905 roku Albert Einstein postawił więc kropkę nad „i”. Niezmiennność prędkości światła przyjął jako postulat – podstawę swojej teorii, którą nazywamy dziś **szczególną teorią względności**. Przyjął jeszcze jeden postulat – nazwany zasadą względności – mówiący, że:

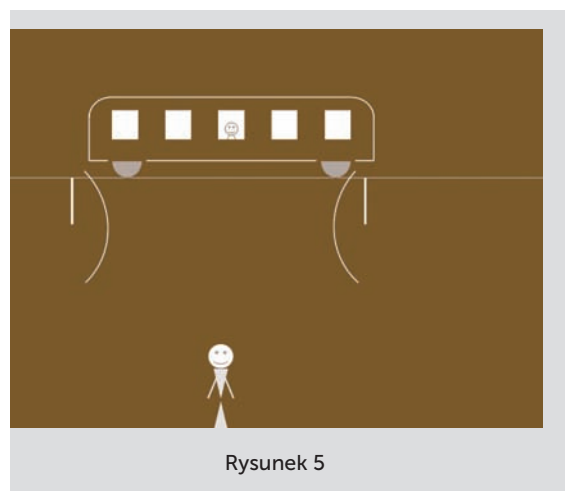
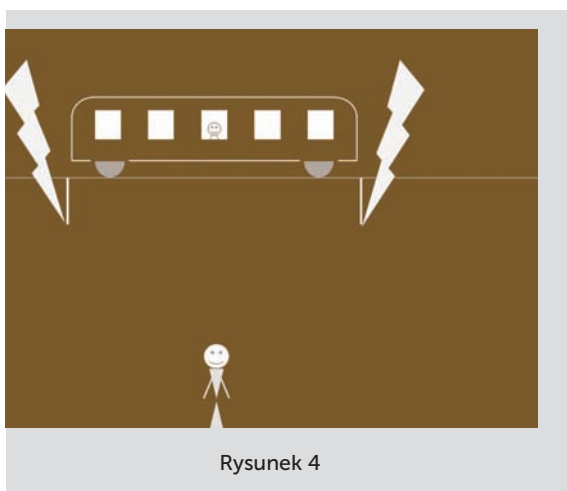
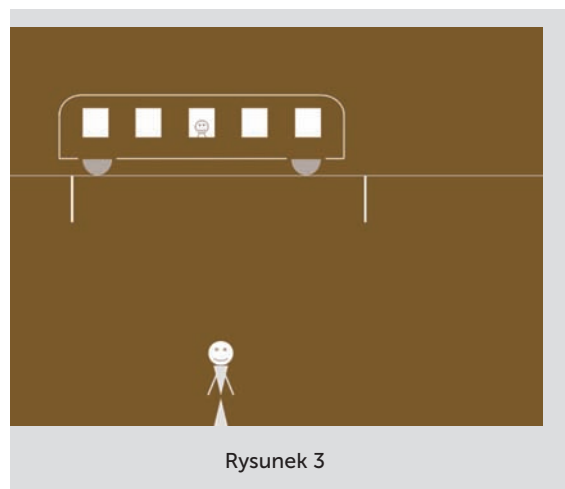
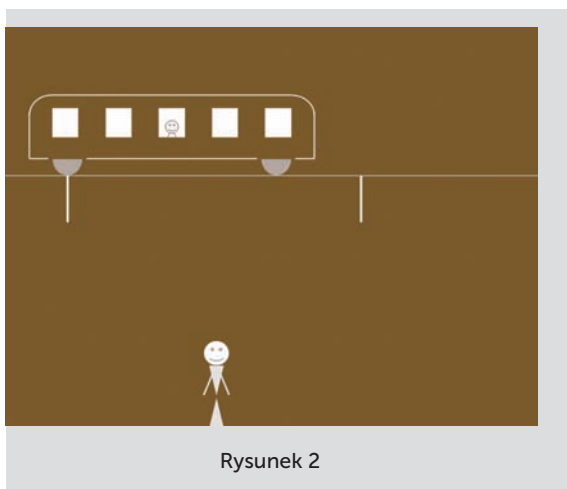
prawa fizyki są takie same dla wszystkich obserwatorów, którzy poruszają się względem siebie jednostajnie po prostej.

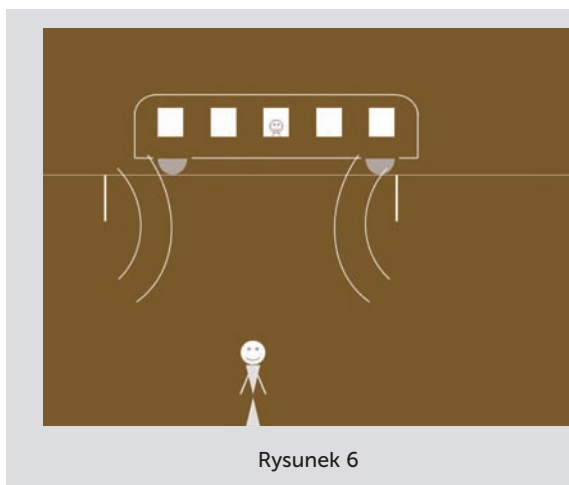
Ten postulat jest zgodny ze zdrowym rozsądkiem, bo przecież skąd miałyby się brać różnice między obserwatorami będącymi względem siebie w ruchu.

Teraz jednak zaczynają się niespodzianki. Na podstawie powyższych postulatów można łatwo udo-

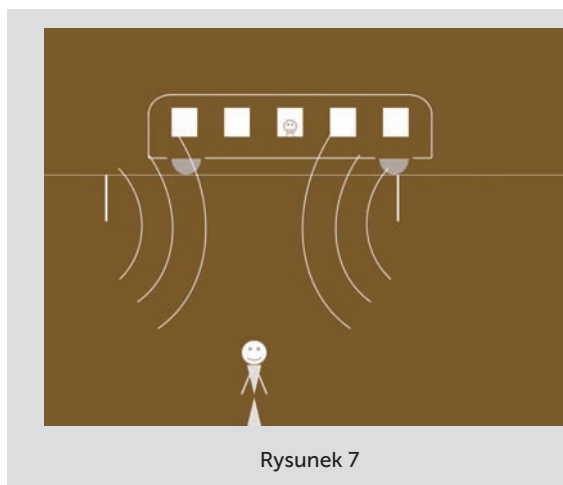
wodnić, że równoczesność zjawisk jest pojęciem względnym, a to oznacza, że czas płynie różnie dla różnych obserwatorów!

Względność równoczesności można wykazać bardzo łatwo w doświadczeniu myślowym, rozpatrując dwie osoby – jedną w wagonie jadącego pociągu, a drugą siedzącą obok torów. Ilustruje to seria ośmiu rysunków (rysunki 2–9). Wagon porusza się jednostajnie z lewa na prawo. Obok szyn są zaznaczone dwa wskaźniki odpowiadające długości wagonu. W chwili gdy wagon znajduje się przy tych wskaźnikach, uderzają w nie dwa pioruny (rysunek 4). Światło błyskawic biegnie w kierunku obserwatora stojącego przy torze i obserwatora w środku wagonu. Kolejne rysunki 5–7 obrazują rozchodzenie fali światła błyskawic. W pewnej chwili (rysunek 8) do obserwatora w wagonie

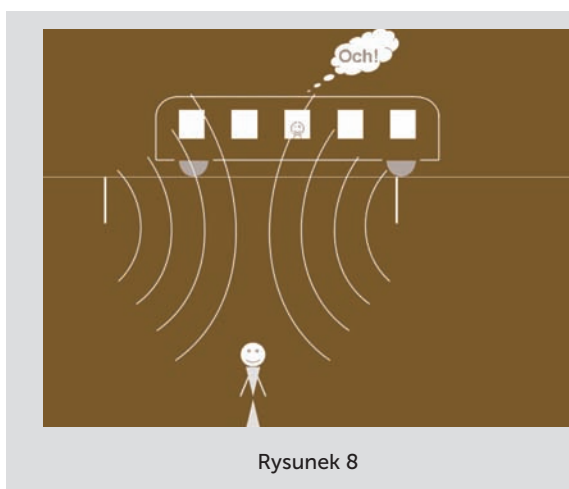




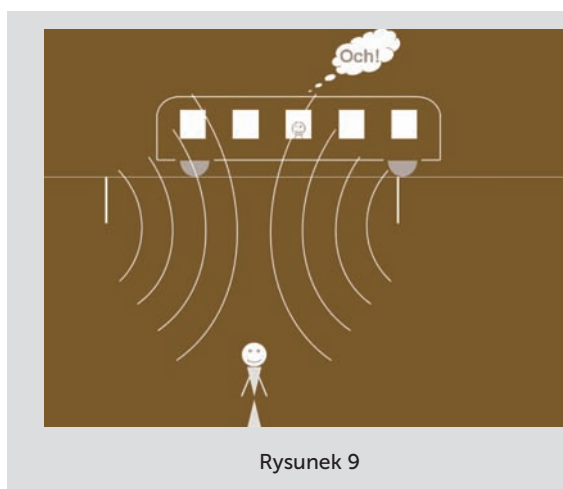
Rysunek 6



Rysunek 7



Rysunek 8



Rysunek 9

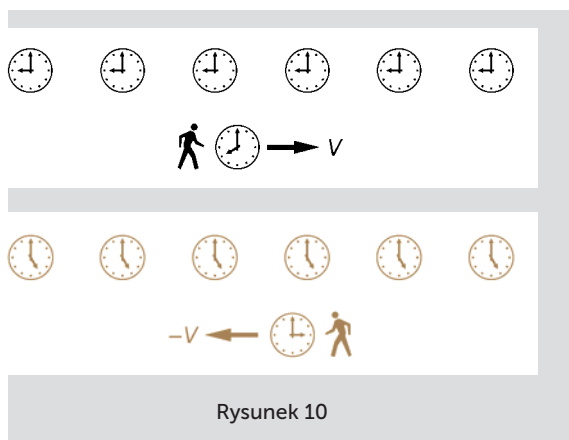
dochodzi błysk światła z prawej strony, ale o błyskawicy z lewej jeszcze wtedy się nie dowiaduje. Po chwili do obserwatora stojącego przy torze dojdą błyski światła od obu błyskawic (rysunek 9) i będzie on traktował uderzenia obu piorunów za zdarzenia jednoczesne. Tymczasem do obserwatora w wagonie błysk światła z lewej dojdzie dopiero później, więc on potraktuje uderzenia piorunów jako zjawiska niejednoczesne!

Na podstawie prostych rozważań Einstein udowodnił, że zegar w ruchu idzie wolniej – to zjawisko nazywamy wydłużeniem (dylatacją) czasu. Warto podkreślić, że jest to efekt czysto kinematyczny i wzajemny. Wyjaśnia to rysunek 10. Kiedy obserwator „brązowy” porównuje swój zegar z ciągiem zegarów zsynchronizowanych w układzie, względem którego jest w ruchu z prędkością V , to stwierdza, że **jego zegar się późni**.

To samo stwierdza obserwator „czarny”, porównując chód swojego zegara z ciągiem zegarów zsynchronizowanych w układzie, względem którego jest w ruchu z prędkością $-V$. W obu przypadkach porównuje się chód **jednego zegara** w jednym układzie z chodem **ciągu zegarów** zsynchronizowanych w drugim układzie.

W życiu codziennym zjawisko dylatacji czasu jest na ogół poniżej dokładności pomiarów. Na przykład statki „Apollo” w drodze na Księżyc miały prędkość około 10 000 razy mniejszą od prędkości światła, czyli $v/c \approx 0,0001$. To daje różnicę w chodzie zegarów zaledwie około 0,00000001.

Przewidziany przez Einsteina efekt można było potwierdzić w bardzo dokładnych eksperymentach z zegarami atomowymi, które przeprowadzili w 1972 roku



Rysunek 10

amerykańscy fizycy J.C. Hafele i R.E. Keating („Science”, vol. 177, 1972, s. 168). Skorzystali oni z czterech niezwykle dokładnych zegarów atomowych, które przewozili samolotami na wschód i na zachód i porównywali ich wskazania ze wskazaniem zegarów pozostawionych w laboratorium. Tak więc, przy locie na zachód przewidywane opóźnienie zegarów miało wynosić 275 (\pm 21) miliardowych części sekundy, podczas gdy eksperyment dał wynik 273 (\pm 7) w tych jednostkach. Była to średnia ze wskazań czterech zegarów, które pokazały opóźnienia odpowiednio 277, 266, 284 i 266 miliardowych części sekundy. Wyniki były więc bardzo spójne. Zgodne z przewidywaniami teorii były również wskazania zegarów przewożonych na wschód.

Musimy zatem przyjąć do wiadomości – jakkolwiek jest to sprzeczne ze zdrowym rozsądkiem – że czas płynie w różny sposób dla różnych obserwatorów, a nie tak jak to sobie wyobrażał Newton.

Kolejny przykład, który przytoczę, dotyczy oporu przeciw koncepcji grawitacji wysuniętej przez Newtona. Nawet wielcy uczeni nie mogli uwierzyć w istnienie siły działającej na odległość. W liście do Gottfrieda Wilhelma Leibniza pisany 18 listopada 1690 roku wielki Christiaan Huygens wyraził się o Newtonie:

Nie jestem przekonany przez jego teorie budowane na zasadzie przyciągania, która wydaje mi się absurdem. Dziwię się często, jak mógł on zadać sobie taki trud wykonania licznych badań i trudnych rachunków, nie mających innej podstawy niż ta zasada...

Kolejny raz okazało się, że zdrowy rozsądek nie jest dobrym doradcą. Dziś przecież mało kto nie jest przekonany o istnieniu siły ciężkości.

Jak Einstein doszedł do swojej koncepcji? Wyda się, że najlepsze wyjaśnienie znajdujemy w jego wspomnieniach autobiograficznych, w których pisał, że w czasach kiedy był studentem najbardziej fascynującym przedmiotem była dla niego teoria Maxwella. Ta teoria opisywała matematycznie pole elektromagnetyczne i przewidywała istnienie fal elektromagnetycznych. Einstein pisał dalej, że gdy miał szesnaście lat natknął się na paradoks:

...Jeżeli podążę za promieniem światła z prędkością c (prędkość światła w próżni), to taki promień powinienem widzieć jako pole elektromagnetyczne w spoczynku, ale oscylujące w przestrzeni. Wydaje się jednak, że coś takiego nie może istnieć, co wynika zarówno z doświadczenia, jak z równań Maxwella. Od samego początku wydawało mi się intuicyjnie jasne, że z punktu widzenia takiego obserwatora wszystko musi się dziać zgodnie z tymi samymi prawami, co dla obserwatora pozostającego w spoczynku względem Ziemi. W jaki bowiem sposób pierwszy obserwator mógłby wiedzieć lub stwierdzić, że jest w szybkim ruchu jednostajnym...

Rozumując w podobny sposób Einstein uznał, że obserwator nie może osiągnąć prędkości światła w próżni, a to oznacza, że jest to pewna szczególna prędkość graniczna. Na tej podstawie w 1905 roku zbudował **szczególną teorię względności**, w której (przypomnijmy) przyjął, że prawa fizyki są takie same dla wszystkich obserwatorów, którzy poruszają się względem siebie jednostajnie po prostej.

Analizując fakty dotyczące grawitacji, Einstein w 1916 roku poszedł jeszcze dalej w swych rozważaniach i uogólnił zasadę względności, odrzucając ograniczenie jej stosowalności tylko do ruchu jednostajnego po linii prostej. W nowym ujęciu zasada ta głosi, że prawa fizyki są takie same dla wszystkich obserwatorów. Teoria Einsteina z 1916 roku nazywa się **ogólną teorią względności**. Opisuje ona właściwości przestrzeni i wpływ, jaki ma na nie materia powodująca zakrzywienie przestrzeni. Jednym z efektów przewidzianych przez tę teorię jest zakrzywanie toru promieni świetlnych w pobliżu masywnych ciał niebieskich. Efekt jest bardzo niewielki i wynosi zaledwie 1,75 sekundy łuku dla światła biegnącego tuż przy powierzchni Słońca. Przewidywane przez Einsteina zjawisko zostało potwierdzone po raz pierwszy przez obserwacje podczas

całkowitego zaćmienia Słońca w 1919 roku, a potem jeszcze wielokrotnie, z coraz większą dokładnością.

Kończąc ten wątek przypomnę, że najbardziej znany osiągnięciem Alberta Einsteina było stwierdzenie równoważności masy i energii, wyrażone równością $E = mc^2$. Ten najstynniejszy wzór fizyki został wielokrotnie sprawdzony doświadczalnie. Ostatni pomiar (wykonany w 2005 roku) potwierdził tę równość z dokładnością 0,00004% („Nature”, vol. 438, 2005, s. 1096).

Ostatnim zagadnieniem, które omówię, są kwanty energii. Dzięki ogromnemu autorytetowi Newtona już na początku XVIII wieku przyjęto wyrażony przez niego punkt widzenia, że światło ma naturę korpuskularną. Tak na przykład w pierwszym wydaniu słynnej *Encyclopaedia Britannica* (1771) czytamy:

Światło składa się z niewyobrażalnie wielkiej liczby cząstek wylatujących we wszystkie strony ze świecącego ciała. Te cząstki są tak małe, że wykraczają poza ludzkie wyobrażenia... Z jednej świecy wylatuje w każdej sekundzie ponad 6 000 000 000 000 razy tyle cząstek światła, ile jest ziarenek piasku w całej Ziemi, zakładając, że każdy cal sześcienny zawiera tych ziarenek 1 000 000. Te cząstki światła, wpadając do naszych oczu, wywołują w naszych umysłach wrażenie światła...

Tymczasem od dawna znane były zjawiska, które trudno było wyjaśnić, przyjmując pogląd Newtona. Mowa o interferencji światła. Zjawisko interferencji – nakładania się fal na wodzie – było znane od bardzo dawna. Już kilkaset lat temu było przedstawiane nawet w malarstwie (np. w obrazie *Chrzest Chrystusa* flamandzkiego malarza Gerarda Davida z 1508 roku). W przypadku interferencji światła występują barwy interferencyjne, które można obserwować, patrząc na cienką warstwę oleju na chodniku albo patrząc pod kątem na płytę kompaktową. Te barwy były opisane już w tekstach klinowych ze starożytnej Mezopotamii. Jednak pierwsze próby ich wyjaśnienia podjęto dopiero w drugiej połowie XVII wieku (Robert Hooke, Isaac Newton). Te wyjaśnienia nie były jednak przekonujące i dopiero na początku XIX wieku Anglik Thomas Young i Francuz Augustin Fresnel rozwinęli teorię faliową światła, w której wspomniane barwy wyjaśnia się jako wynik interferencji fal poprzecznych rozchodzących się w eterze. Kiedy liczne doświadczenia okazywały się zgodne z teorią Younga-Fresnela, uznano ją za

ostatecznie potwierdzoną, a sławny francuski uczony Henri Poincaré wyraził się, że:

Teoria światła oparta na pracach Fresnela i jego następców jest najdoskonalszą ze wszystkich teorii fizycznych (Théorie mathématique de la lumiere, Paris 1889).

Na tym historia się nie kończy, ponieważ w XX wieku uzyskano niezbita dowody na to, że światło zachowuje się także tak, jakby było strumieniem porcji energii – kwantów, które nazywamy fotonami. Pierwszy krok w tym kierunku był dziełem Alberta Einsteina, który wprowadził pojęcie kwantów światła w pracy z 1905 roku (*O pewnym heurystycznym punkcie widzenia na wytwarzanie i przemiany światła*, „Annalen der Physik”, vol. 17, s. 132–148) i w ramach tej teorii podał przekonujące wyjaśnienie wielu zjawisk optycznych – w tym zjawiska fotoelektrycznego. Cel tej pracy Einstein wyjaśnił w początkowych jej zdaniach:

Istnieje głęboka różnica formalna między pojęciami teoretycznymi, które fizycy uformowali na temat gazów i innych ciał ważkich oraz teorią Maxwella procesów elektromagnetycznych w tak zwanej pustej przestrzeni. Podczas gdy uważamy, że stan ciała jest całkowicie wyznaczony przez położenia i prędkości bardzo wielkiej, ale skończonej liczby atomów i elektronów, dla opisu stanu elektromagnetycznego elementu przestrzeni używamy ciągłych funkcji przestrzennych, tak że skończonej liczby wielkości nie można uznać za wystarczającą dla całkowitego opisu stanu elektromagnetycznego przestrzeni.

Dopiero w końcowej części pracy Einstein nawiązał do zjawiska fotoelektrycznego:

Wydaje mi się, że obserwacje „promieniowania ciała czarnego”, fotoluminescencji, wytwarzania promieni katodowych przez światło ultrafioletowe i inne zjawiska związane z emisją i przemianą światła, są łatwiej zrozumiałe jeśli się założy, że energia światła jest w przestrzeni rozłożona nierównomiernie. Zgodnie z rozważanym tu założeniem, przy rozchodzeniu się promienia świetlnego wystanego ze źródła punktowego, jego energia nie jest rozłożona w sposób ciągły w stale zwiększającej się objętości przestrzeni, lecz składa się ze skończonej liczby kwantów energii, które są zlokalizowane

w punktach przestrzeni, poruszają się bez po-
działu i mogą być pochłaniane lub wytwarzane
tylko jako kompletne całości.

Nieliczni fizycy, którzy zainteresowali się wtedy ar-
tykułem Einsteina, wyobrażali sobie, że kwanty światła
to pewnego rodzaju „kule światła”. Przykładem mogą
być spekulacje Lorentza na temat kwantów (1910).
Zwracał on uwagę, że eksperymenty interferencyjne
Lummera i Gehrckego, w których różnica dróg optycz-
nych dochodziła do 80 cm dają dolną granicę **roz-
ciągłości podłużnej** kwantu światła. Podobnie, na co
zwracał uwagę Lorentz, dolną granicę **rozciągłości po-
przecznej** kwantu światła daje średnica 150 cm (śred-
nica największego wówczas teleskopu na Mt. Wilson).

*Jak to jest możliwe, że tak monstrualnie wielki
kwant przechodzi przez źrenicę oka nie ulega-
jąc podziałowi?*

– pytał Lorentz.

Od tego czasu nauczyliśmy się bardzo wiele o kwan-
tach światła — fotonach (samą nazwę „foton” zapro-
ponował w 1926 roku Gilbert N. Lewis). Właściwości tego
niezwykłego obiektu fizycznego można podsumować
w następujących punktach.

Foton:

- nie jest „kulką” światła,
- niesie własny moment pędu,
- bardzo „lubi tćok” (opisuje to kwantowa statystyka
Bosego-Einsteina),
- ma zdolność dyfrakcji i interferencji,
- ma bogatą i skomplikowaną strukturę,
- w próżni ma nieskończony czas życia,
- może być „splątany” z innym fotonem (co jest wy-
korzystywane w teleportacji).

Trudno się dziwić, że starając się pojąć naturę tak
niezwykłego obiektu zdesperowany Einstein napisał
w 1921 roku do Paula Ehrenfesta:

*Problem kwantów wystarczy żeby mnie zapro-
wadzić do domu wariatów.*

Dziś wiemy, że wyobrażanie sobie fotonu w ra-
mach fizyki klasycznej jest skazane na niepowodze-
nie. Fotonem nie jest ani bardzo krótki impuls światła
laserowego, ani bardzo krótki odcinek sinusoidy, którą
tak chętnie wykorzystujemy do popularnego wyjaśnie-
nia falowej natury światła.

Dzisiaj fotony opisujemy, korzystając z elektrody-
namiki kwantowej — najdokładniejszej teorii fizycznej.
Wynika z niej, że pole elektromagnetyczne jest skwan-
towane, a jego energia może się zmieniać tylko skoko-

wo, w porcjach wynoszących $h\nu$, co można wyrazić
wzorem:

$$\int \frac{1}{2} (\epsilon_0 \vec{E}^2 + \mu_0 \vec{H}^2) dV = \left(n + \frac{1}{2} \right) h\nu$$

W ujęciu elektrodynamiki kwantowej fotony są to
mody pola elektromagnetycznego.

Dziś jesteśmy przekonani, że **światło jest zarówno
falą jak i strumieniem fotonów**. Taka dwoistość prze-
czy oczywiście „zdrowemu rozsądkowi”. Czyż jednak
nie przekonaliśmy się już, jak często on nas zawodził
w przeszłości i nadal zawodzi?

Wybitny fizyk niemiecki, jeden z twórców mecha-
niki kwantowej — Werner Heisenberg wspominał, że
w 1922 roku, jako bardzo młody fizyk, słuchał w Getyn-
dze wykładów Nielsa Bohra na temat fizyki atomów.
Pewnego dnia zapytał Bohra:

Czy w ogóle zrozumiemy kiedyś atomy?

Bohr zwlekał przez chwilę i odpowiedział:

*Tak. Ale jednocześnie dopiero wtedy dowiemy
się, co znaczy słowo „rozumieć”...*

Istotnie, we współczesnej fizyce słowo „rozu-
mieć” ma znaczenie nieco inne niż w życiu codzien-
nym, w którym staramy się zawsze wyobrazić sobie
zjawiska i korzystać ze zdrowego rozsądku.

Bardzo ładnie wyraził to wybitny fizyk brytyjski Free-
man Dyson:

*Elektrodynamika kwantowa zajmuje pozycję
wyjątkową we współczesnej fizyce. Jest to jedy-
na część naszej nauki, która została całkowicie
zredukowana do układu precyzyjnych równań.
Jest to jedyny obszar, w którym możemy wy-
brać hipotetyczny eksperyment i przewidzieć
jego wynik z dokładnością pięciu miejsc po
przecinku i mieć pewność, iż teoria bierze pod
uwagę wszystkie istotne czynniki. Elektrodyna-
mika kwantowa daje nam kompletny opis za-
chowania się elektronu; zatem w pewnym sen-
sie pozwala nam zrozumieć, czym jest elektron...*

Richard Feynman, jeden z twórców elektrodynami-
ki kwantowej, w jednym ze swych popularnych wykła-
dów w 1965 roku użył sformułowania, że:

...nikt nie rozumie mechaniki kwantowej.

Miał oczywiście na myśli potoczne znaczenie słowa
„rozumieć”. Inne znaczenie tego słowa miał na myśli
Einstein, kiedy wyraził się, że:

*...najbardziej niezrozumiałe jest to, że wszech-
świat można zrozumieć.*

Bibliografia

A. Einstein, *5 prac, które zmieniły oblicze fizyki*, tłum. z ang. Piotr Amsterdamski, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa 2005.

A.K. Wróblewski, *Historia fizyki od czasów najdawniejszych do współczesności*, Wydawnictwo Naukowe PWN, Warszawa 2006.

Abstract

In 1905 Albert Einstein published four papers which revolutionized physics. Einstein's ideas concerning energy quanta and electrodynamics of moving bodies were received with scepticism, which only very slowly

went away in spite of their solid experimental confirmation. For many people Einstein's results were contrary to common sense. However, history of physics provides numerous convincing examples of important discoveries, which were first rejected as contrary to common sense, and accepted only after a long time.

Słowa kluczowe

szczególna i ogólna teoria względności, względność równoczesności, dylatacja czasu, kwanty energii, fotony.

Profesor Andrzej Kajetan Wróblewski fizyk, dziekan Wydziału Fizyki Uniwersytetu Warszawskiego (1986–1989), rektor Uniwersytetu Warszawskiego (1989–1993), przewodniczący Rady Naukowej Instytutu Historii Nauki PAN. Pracuje w Zakładzie Cząstek i Oddziaływań Fundamentalnych Instytutu Fizyki Doświadczalnej na Wydziale Fizyki Uniwersytetu Warszawskiego. Jego specjalnościami są fizyka cząstek elementarnych (fizyka wysokich energii) oraz historia fizyki. Prowadzi prace

dotyczące mechanizmu produkcji hadronów (znalazł m.in. związek między średnią krotnością produkowanych cząstek i dyspersją ich rozkładu). Zajmuje się też historią i metodologią fizyki (m.in. książki *Prawda i mity w fizyce*, 1982; *Historia fizyki*, 2007). Z J.A. Zakrzewskim jest współautorem podręcznika *Wstęp do fizyki*. Został odznaczony Medalem Mariana Smoluchowskiego, wielokrotnie uhonorowany tytułem Doktor *honoris causa*, jest również laureatem Nagrody Marii Skłodowskiej-Curie.

Czarne dziury: obiekty odkryte w przyrodzie czy wymyślone przez człowieka?

Na podstawie odczytu wygłoszonego w dniu 24 maja 2007 roku

Jerzy Kijowski

Centrum Fizyki Teoretycznej Polskiej Akademii Nauk

109

Rola spekulacji teoretycznych w fizyce

Fizyka jest nauką empiryczną. Tę prawdę wbijano mi do głowy przez całe dzieciństwo i młodość, ostro przeciwstawiając „postępowe” nauki doświadczalne „scholastycznym spekulacjom”, które jedynie hamowały rozwój cywilizacji. Taka postawa filozoficzna nie jest nowa. Wyraziście była eksponowana np. w dziele Sykstusa Empiryka napisanym około 200 r.n.e., a zatytułowanym bardzo wymownie *Przeciw Matematykom* (Προϋ μαθηματικουϋ).

Tymczasem u podstaw nowożytnej fizyki leżą mocno spekulatywne zasady dynamiki Galileusza-Newtona! Rzeczywiście — czy ktoś z nas miał kiedykolwiek do czynienia w potocznym doświadczeniu z pojazdem, który porusza się ruchem jednostajnym, mimo że nie działa nań żadna siła (bo np. wyczerpała się benzyna)?

Nieraz więc bywało tak, że **nauka rozwijała się właśnie dzięki czysto teoretycznym „spekulacjom”**, a przewidywane w ten sposób obiekty czy zjawiska fizyczne musiały potem długo jeszcze czekać na „prawdziwe” — to znaczy eksperymentalne — odkrycie, stanowiące ostateczne potwierdzenie słuszności rozumowania, które doprowadziło do ich pierwszego — teoretycznego — odkrycia.

Tak się zdarzyło np. przewidzianym przez Maxwella falom elektromagnetycznym. Pojawiły się one jako matematycznie konieczne rozwiązania równań elektrodynamiki, gdy ich twórca poprawił równanie Ampere’a a czysto „spekulatywnych” powodów, wprowadzając

doń tzw. **prąd przesunięcia**, by stało się zadość prawu zachowania ładunku. Podnosząc to prawo do rangi podstawowej zasady uniwersalnej, Maxwell odkrył „po raz pierwszy” fale radiowe, bez których trudno sobie wyobrazić nasze współczesne życie. A przecież na drugie — doświadczalne — odkrycie musiały potem czekać jeszcze prawie 40 lat, kiedy to zostały „wprężgnięte w służbę ludzkości” przez Heinricha Hertza, Guglielmo Marconiego lub Aleksandra Popowa (to, którego z nich uznajemy za twórcę radia zależy od tego w jakim kraju, czy nawet ustroju politycznym, kończyliśmy szkoły).

Tak zwane „czarne dziury”, podobnie zresztą jak fale grawitacyjne, są matematycznie konsekwencją nowoczesnej teorii grawitacji sformułowanej w 1915 roku przez Alberta Einsteina. Nie dysponujemy do tej pory wynikami obserwacji, które w sposób niewątpliwy dowodziłyby ich istnienia w Kosmosie, choć mamy wiele pośrednich argumentów za tym, by pewne obiekty astronomiczne interpretować właśnie jako czarne dziury. Status epistemiologiczny „czarnych dziur” jest więc obecnie analogiczny do statusu fal radiowych w roku, powiedzmy, 1870.

Czy grawitacja da się sprowadzić do czystej geometrii?

Historia einsteinowskiej teorii grawitacji, popularnie zwanej „ogólną teorią względności”, też zaczęta się

od czystej spekulacji. Wspaniale rozwinięta w XIX wieku mechanika nieba, oparta na teorii newtonowskiej, znakomicie tłumaczyła obserwowane zjawiska astronomiczne i z czysto eksperymentalnego punktu widzenia nie było żadnych powodów, by tę teorię zmieniać. Jednak w 1905 roku Einstein zaproponował nową teorię ruchu i nowy opis pola elektromagnetycznego. Chodziło m.in. o to, by wytłumaczyć zaobserwowany fakt niezależności prędkości światła od prędkości mierzącego ją obserwatora. Ta nowa teoria, zwana „szczególną teorią względności”, odniosła ogromny sukces, dostarczając znakomitego narzędzia do opisu zjawisk fizycznych polegających na oddziaływaniu elektromagnetycznym – i to opisu zgodnego z doświadczeniem Michelsona-Morleya. Ponieważ miała ona charakter uniwersalnej teorii czasu i przestrzeni, wielki myśliciel – Einstein, spróbował opisać siły grawitacyjne w swoim nowym, czterowymiarowym, „relatywistycznym formalizmie”. Okazało się to niemożliwe, a w każdym razie nie było na to żadnej prostej i naturalnej recepty. Po kilku latach rozmyślań Einstein doszedł do wniosku, że trajektorie swobodnie spadających ciał (to znaczy ciał poddanych jedynie działaniu sił grawitacyjnych) są po prostu „najbardziej prostymi ze wszystkich możliwych” liniami w, prawdopodobnie, krzywej czasoprzestrzeni. Punktem wyjścia do tej hipotezy była powszechnie obserwowana równość masy inercyjnej i masy grawitacyjnej różnych ciał, począwszy od skali słynnego newtonowskiego jabłka do skali obserwowanych ciał niebieskich. Łatwo zrozumieć o co chodzi, przyglądając się np. równaniu Newtona opisującemu ruch ciała o masie m i ładunku elektrycznym e , poruszającego się pod wpływem pola elektrycznego \vec{E}

$$m\vec{a} = e\vec{E} \quad (1)$$

Ciała o równych masach, ale różnych ładunkach poruszają się zupełnie inaczej! Tymczasem dla sił grawitacyjnych analogiczne równanie Newtona wygląda następująco

$$m\vec{a} = m\vec{G} \quad (2)$$

gdzie \vec{G} oznacza „pole sił grawitacyjnych” (równe np. gradientowi potencjału grawitacyjnego z ujemnym znakiem). Występujący po lewej stronie czynnik m określa bezwładność ciała – podobnie jak to miało miejsce w przypadku równania (1). Natomiast występująca po prawej stronie wielkość m odgrywa rolę „ładunku grawitacyjnego” i mierzy „wrażliwość ciała”

na oddziaływanie pola grawitacyjnego. Z punktu widzenia teorii Newtona to m po prawej stronie mogłoby być różne od tego m po lewej stronie i nie prowadziłoby to do żadnych komplikacji filozoficznych. Tymczasem jednak wszelkie pomiary (począwszy od Galileusza, rzucającego ponoć różne przedmioty z Krzywej Wieży w Pizie) pokazują, że obie masy są sobie **zawsze równe**. Oznacza to, że równanie (2) można podzielić przez m . Wobec tego trajektorie ciał spadających swobodnie są takie same dla różnych ciał. Jabłko powinno poruszać się po takiej samej orbicie okotłoczonej jak planeta. Powinniśmy zatem traktować tę orbitę jako jakąś uniwersalną własność geometryczną czasoprzestrzeni, w której „żyje” pole grawitacyjne \vec{G} , niezależną od tego, jakim ciałem (jabłkiem czy planetą) postugujemy się w celu jej zbadania. O jaką własność może tutaj chodzić?

Einstein znalazł odpowiedź w powstałej wcześniej geometrii nieeuklidesowej i zaproponował, by odejść od obrazu czasoprzestrzeni jako obiektu „idealnie płaskiego”, do którego stosują się pojęcia geometrii euklidesowej, znane nam z kursu szkolnego, na przykład możliwość absolutnego, niezależnego od drogi, transportu równoległego na dowolne odległości. To, co w małej skali jest banalnie proste, np. przyłożenie ekerki do linijki i przesunięcie jej dożądanego punktu, w wielkiej skali jest, być może, niewykonalne. Czasoprzestrzeń, w której żyjemy, jest prawdopodobnie zakrzywiona, a ciała spadające swobodnie poruszają się po liniach „najbardziej prostych z możliwych”. Do linii tych nie stosują się twierdzenia geometrii euklidesowej, jak np. aksjomat Euklidesa o prostych równoległych czy też twierdzenie o sumie kątów w trójkącie. Grawitacja to po prostu odstępstwo geometrii czasoprzestrzennej od „płaskości”. W dalszym ciągu tego wykładu pokażę, jak najprościej można opisać taką strukturę.

Co to są linie „proste” w krzywej przestrzeni?

Najwcześniej, bo już w dobie wielkich odkrywców geograficznych, poznaną „krzywą przestrzenią”, dla której opis oparty na geometrii euklidesowej przestał wystarczać, była powierzchnia globu ziemskiego. Żeglując po oceanie, najłatwiej jest trzymać się tzw. loksodromy, czyli linii przecinającej siatkę geograficzną, złożoną z południków i równoleżników, pod stałym kątem.

W takiej żegludze sternik musi po prostu trzymać ciągle ten sam kurs kompasowy. We współrzędnych geograficznych (θ, φ) równanie takiej trajektorii ruchu wygląda rzeczywiście tak, jak równanie linii prostej na płaszczyźnie wyrażone we współrzędnych kartezjańskich

$$\theta(t) = \theta_0 + at \quad \varphi(t) = \varphi_0 + bt \quad (3)$$

gdzie t jest dowolnym parametrem afinicznym (na przykład czasem, gdy statek płynie ze stałą prędkością). Równania te można zapisać równoważnie w postaci warunku na znikanie drugich pochodnych obu funkcji

$$\ddot{\theta}(t) = 0 \quad \ddot{\varphi}(t) = 0 \quad (4)$$

co gwarantuje liniową zależność obu funkcji od parametru.

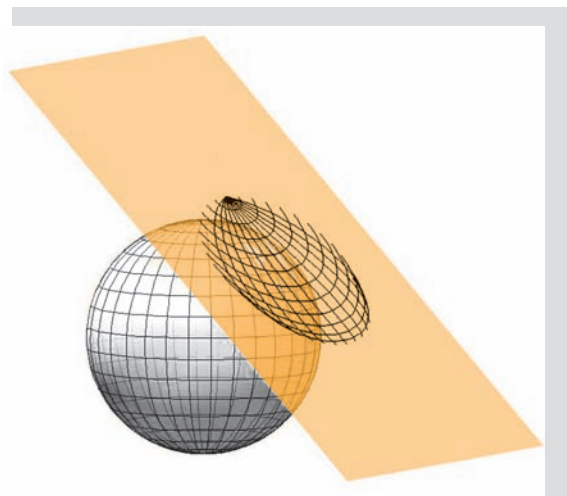
Nazwa **loksodroma** pochodzi od greckich słów *loksos* (ukośny, krzywa) oraz *dromós* (bieg, linia). Ale cóż to za linia „skośna”? W żegludze po Morzu Bałtyckim może być nawet przydatna, ale wystarczy popatrzeć na globus, by zauważyć, że żegluga po loksodromie z Plymouth do Nowego Yorku to ogromna strata czasu! A w pobliżu biegunów loksodroma coraz bardziej zbliża się do spirali Archimedesesa i poruszanie się po niej przypominałoby raczej taniec św. Wita niż jakiegokolwiek sensowne zmierzanie do celu.

Gdyby na globusie naciągnąć gumkę umocowaną na końcach w dwóch miastach leżących po różnych stronach oceanu, to wyznaczyłaby ona zupełnie inną trasę, zwaną „ortodromą” (gr. *orto*, *orthós* – prosty), czyli linią prostą. Jest to łuk wielkiego koła na kuli i czujemy instynktownie, że wyznacza najkrótszą drogę między punktem startu a punktem docelowym.

Wyznaczanie ortodromy to najbardziej typowe zadanie, jakie rozwiązują studenci wydziału nawigacji w szkole morskiej. Trzeba się tu wyzybyć przywiązania do funkcji liniowych typu (3), bowiem równanie różniczkowe opisujące ortodromę we współrzędnych geograficznych (θ, φ) jest dużo bardziej skomplikowane niż warunek (4) na znikanie „przyspieszenia”, czyli drugich pochodnych. Aby je tutaj wyprowadzić zauważmy, że w każdym punkcie globu (θ_0, φ_0) można wybrać takie **lokalne współrzędne** (x, y) , żeby **przynajmniej w tym jednym punkcie** równanie ortodromy wyglądało tak jak (4)

$$\ddot{x}(t) = 0 \quad \ddot{y}(t) = 0 \quad (5)$$

W tym celu rozważymy płaszczyznę styczną do globusa w naszym „miejscu postoju” (θ_0, φ_0) – niech będzie to na przykład kartka papieru. Jak to widać na



Rysunek 1. Rzut siatki geograficznej na płaszczyznę styczną do kuli

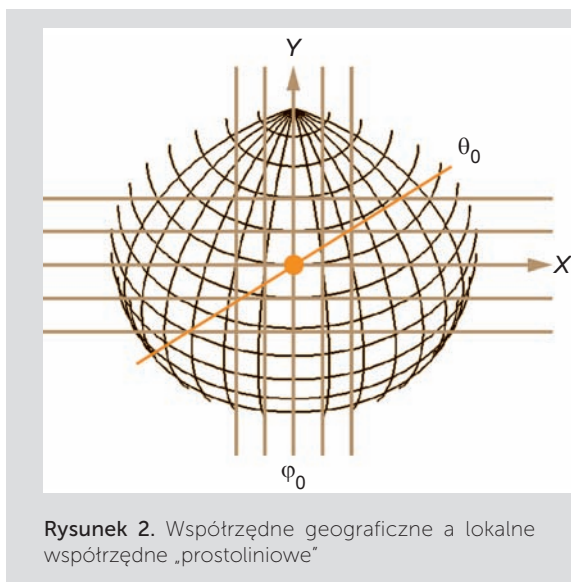
rysunku 1, rzut siatki geograficznej z globusa na naszą kartkę daje wysoce nieprostoliniowy układ współrzędnych. Gdyby nasze miejsce postoju znajdowało się na lądzie, to siatka ta zupełnie nie nadawałaby się jako lokalna mapa do celów geodezyjnych, takich jak wytyczanie sieci równoległych ulic czy obrysów działek w naszej miejscowości.

Do tych celów najlepiej używać siatki współrzędnych kartezjańskich na kartce. Gdy zrzućmy je, za pomocą rzutu prostokątnego, z naszej kartki papieru na globus, otrzymamy właśnie lokalnie najlepsze, „wyprostowane” współrzędne. Wybierzmy na przykład oś X w kierunku wschodnim, zaś oś Y – prostopadłe (w kierunku północnym), przy czym, dla uproszczenia rachunków, niech obie będą zaczepione właśnie w naszym „miejscu postoju” (θ_0, φ_0) (zob. rys. 2). Odrobina znajomości trygonometrii pozwoli nam stwierdzić, że zależność między współrzędnymi geograficznymi a naszymi „lokalnie prostoliniowymi” współrzędnymi (x, y) , skopiowanymi z płaskiej kartki, wyraża się następującymi wzorami

$$\theta = \theta_0 - y + Ax^2 + \text{człony wyższego rzędu} \quad (6)$$

$$\varphi = \varphi_0 + x(1 + By) + \text{człony wyższego rzędu} \quad (7)$$

Pozwoliłem sobie na pewną nonszalancję w zapisie powyższych wzorów, bowiem chcę zwrócić uwagę na jego strukturę, a do tego ani wartość stałych A i B , ani szczegółowa informacja o członach rzędu wyż-



Rysunek 2. Współrzędne geograficzne a lokalne współrzędne „prostoliniowe”

szego niż kwadratowy, nie jest potrzebna. Otóż człony rzędu zerowego zostały wprowadzone tylko po to, by odpowiednio „scentrować” nasze współrzędne, czyli by zero układu (x, y) przypadło właśnie w punkcie (θ_0, φ_0) . Wybór ten jest nieistotny z punktu widzenia naszego celu, jakim jest „wyprostowanie” równania ortodromy do najprostszej postaci (5). Człony rzędu pierwszego zostały wybrane tak, by osie X i Y były odpowiednio skierowane. Znak „minus” przed zmienną y w pierwszym równaniu pochodzi stąd, że jako matematyk liczę „szerokość geograficzną” θ od bieguna północnego w dół (na południe). Tymczasem jako nawigator wolę liczyć współrzędną y na mapie w górę (na północ). Ale to również jest **nieistotne** z punktu widzenia naszego celu i dowolna transformacja liniowa współrzędnych (x, y) będzie równie dobra. Także człony wyższego rzędu są nieistotne, bowiem w punkcie $(x, y) = (0, 0)$ nie dają wkładu do równania (5). To co istotne, to człony kwadratowe, reprezentowane tutaj przez dwie stałe A i B . Odrobina znajomości trygonometrii wystarczy, by się przekonać, że ich wartości wynoszą

$$A = \frac{1}{2} \sin \theta_0 \cos \theta_0 \quad B = \operatorname{ctg} \theta_0 \quad (8)$$

Gdybyśmy mieli osiąść na stałe w miejscowości (θ_0, φ_0) , to już nigdy inna mapa nie byłaby nam potrzebna – ta jest najlepsza ze wszystkich możliwych. Jeśli jednak punkt (θ_0, φ_0) znajduje się na oceanie, a my jesteśmy nawigatorami w podróży, to niestety nasza doskona-

ła (w otoczeniu punktu (θ_0, φ_0)) mapa wkrótce się zdeaktualizuje i trzeba będzie ją szybko wymienić na inną, dostosowaną do innego punktu. Nie możemy jednak wozić ze sobą tak ogromnej, po jednej dla każdego małego otoczenia kolejno mijanych punktów globu, liczby map! Przepraszamy się zatem z „bardziej globalnymi” współrzędnymi geograficznymi i przeliczmy równanie ortodromy (5) (oznaczonej kolorem pomarańczowym na rys. 2) do tych współrzędnych. Różniczkując stronami wzory (6) i (7) otrzymujemy

$$\dot{\theta} = -\dot{y} + 2Ax\dot{x} \quad (9)$$

$$\ddot{\theta} = -\ddot{y} + 2A\dot{x}\dot{x} + 2Ax\ddot{x} \quad (10)$$

$$\dot{\varphi} = \dot{x} + Bx\dot{y} + B\dot{x}y \quad (11)$$

$$\ddot{\varphi} = \ddot{x} + 2B\dot{x}\dot{y} + Bx\ddot{y} + B\dot{x}\dot{y} \quad (12)$$

Ale w zmiennych „wyprostowanych” równanie ortodromy to znikanie drugich pochodnych (5). Zatem w naszym punkcie postoju $(x, y) = (0, 0)$ równanie ortodromy wygląda następująco

$$\begin{aligned} \ddot{\theta} &= 2A\dot{x}\dot{x} = 2A\dot{\varphi}^2 \\ \ddot{\varphi} &= 2B\dot{x}\dot{y} = -2B\dot{\varphi}\dot{\theta} \end{aligned}$$

(uwzględniono związki $\dot{\theta} = -\dot{y}$ oraz $\dot{\varphi} = \dot{x}$, wynikające z (9) i (11)). Po wstawieniu wartości stałych (8) otrzymujemy następujące równanie ortodromy

$$\ddot{\theta} = \dot{\varphi}^2 \operatorname{si} r \theta \cos \theta \quad (13)$$

$$\ddot{\varphi} = -2\dot{\varphi}\dot{\theta} \operatorname{ctg} \theta \quad (14)$$

obowiązujące już uniwersalnie, w każdym punkcie (θ, φ) , czyli na całym globie.

Powyższe, prościutkie opowiadanie można teraz sformalizować w postaci następującej „wycieczki” istotnych z naszego punktu widzenia własności, jakie wykorzystaliśmy do stworzenia wygodnego modelu matematycznego powierzchni globu ziemskiego M jako tzw. **przestrzeni ze strukturą powiązania**. Przestrzeń ta może być krzywa, ale lokalnie przypomina przestrzeń afiniczną, w której jest dobrze określone pojęcie linii prostej. Oto te własności:

- Przestrzeń M jest różniczkowalną. Oznacza to, że lokalnie można parametryzować punkty zbioru M układem współrzędnych $(x^a) = (x^1, x^2, \dots, x^n)$, gdzie liczba naturalna n nazywa się **wymiarem przestrzeni** (dla powierzchni globu ziemskiego $n = 2$, ale to nie ma żadnego znaczenia). Jeśli (y^a) , $a = 1, \dots, n$, jest innym układem

współrzędnych (mapą), to (tam gdzie to możliwe) transformacja $y^a = y^a(x^k)$ jest odpowiednio gładka (np. różniczkowalna klasy C^∞).

- W każdym punkcie $m \in M$ można wprowadzić relację równoważności \sim_m między układami współrzędnych wokół m . Powiemy, że układy (x^k) oraz (y^a) są równoważne w m , jeśli ich wzajemne drugie pochodne znikają w tym punkcie

$$((x^k) \sim_m (y^a)) \Leftrightarrow \left(\frac{\partial^2 y^a}{\partial x^k \partial x^l}(m) = 0 \right)$$

Łatwo sprawdzić, że jest to relacja równoważności (zwrotna, symetryczna, przechodnia). Zatem zbiór wszystkich map wokół m rozpada się na wzajemnie rozłączne klasy równoważności.

- W każdym punkcie m wyróżniona jest pewna klasa równoważności I_m , którą nazwiemy „lokalnym układem prostoliniowym” lub też – żeby podkreślić inspirację fizyczną – „lokalnym układem inercyjnym”. Współrzędne odpowiadające mapie należącej do tej klasy będziemy nazywali „lokalnie prostoliniowymi” albo „lokalnie inercyjnymi” w punkcie m . Zakładamy, że I_m zależy gładko (różniczkowalnie) od punktu m . Na razie nie będę precyzował, co to oznacza, ale w dalszym ciągu wykładu stanie się to jasne.
- Sparametryzowaną linię w M będziemy nazywali ortodromą (czy może nawet – nadużywając nieco terminologii – linią „prostą”, a w każdym razie „najprostszą z możliwych”), jeśli w każdym punkcie M spełnia ona warunek

$$\ddot{y}^a(m) = 0 \quad (15)$$

gdy jako współrzędne weźmiemy dowolne współrzędne lokalnie inercjalne w m , czyli należące do klasy I_m .

- Współrzędne inercjalne w jednym punkcie wcale nie muszą być inercjalne w punktach sąsiednich. Gdyby jednak istniał układ współrzędnych **globalnie inercjalny**, to taką przestrzeń nazwalibyśmy **płaską**.

Gdybyśmy mieli zamieszkać na stałe w miejscowości m i badać jedynie ruchy przejeżdżających przez nią pojazdów, to nie warto byłoby używać innych układów współrzędnych, niż tylko inercjalne w punkcie m . Tak postępowali fizycy przez ostatnich kilka tysięcy lat. Jeśli jednak interesują nas zjawiska zachodzące daleko od

nas, jeśli chcemy badać ruchy odległych planet, komet czy może nawet gwiazd albo galaktyk, to – podobnie jak to miało miejsce w przypadku nawigacji po ziemskim globie – zmuszeni będziemy używać układów nieinercjalnych, ale za to opisujących dużo większe obszary. Przeliczmy zatem równania ortodromy z układu inercjalnego (y^a) do jakiegokolwiek układu (x^k) . Prawo transformacji prędkości z jednego układu do drugiego jest oczywiste

$$\dot{y}^a(m) = \sum_{k=1}^n \frac{\partial y^a}{\partial x^k} \dot{x}^k = \frac{\partial y^a}{\partial x^k} \dot{x}^k \quad (16)$$

W ostatnim wyrażeniu opuściliśmy znak sumy, korzystając z tak zwanej **konwencji sumacyjnej Einsteina**, która bardzo ułatwia pisanie formuł opisujących struktury geometrii różniczkowej. Głosi ona, że powtórzony wskaźnik – w naszym przypadku „ k ” – raz na dole, raz na górze oznacza sumowanie po wszystkich jego możliwych wartościach. Różniczkując jeszcze raz po parametrze, otrzymujemy następujące równanie ortodromy wyrażone w nieinercjalnym układzie współrzędnych (x^k) (pamiętamy o konwencji sumacyjnej!)

$$0 = \ddot{y}^a = \frac{\partial y^a}{\partial x^k} \ddot{x}^k + \frac{\partial^2 y^a}{\partial x^k \partial x^l} \dot{x}^k \dot{x}^l \quad (17)$$

Powyższy układ n równań można zapisać w równoważnej postaci, jeśli podzielimy na obie strony macierz $(\partial x^n / \partial y^a)$, odwrotną do macierzy $(\partial y^a / \partial x^k)$ (ich iloczyn daje macierz jednostkową (δ_k^n)). W rezultacie otrzymamy następujące, uniwersalne równanie ortodromy, będące naturalnym uogólnieniem równań nawigatora (13) oraz (14)

$$\ddot{x}^n + \Gamma_{kl}^n \dot{x}^k \dot{x}^l = 0 \quad (18)$$

Symbolem Γ_{kl}^n oznaczyliśmy tutaj następującą kombinację pochodnych

$$\Gamma_{kl}^n(m) := \frac{\partial x^n}{\partial y^a}(m) \cdot \frac{\partial^2 y^a}{\partial x^k \partial x^l}(m) \quad (19)$$

We wzorze tym (y^a) są **dowolnymi współrzędnymi inercjalnymi** w punkcie m , w którym liczymy wartość współczynników Γ . Łatwo zauważyć, że wynik nie zależy od wyboru układu współrzędnych (y^a) , byle tylko należały do klasy I_m . Wielkości te charakteryzują jednoznacznie układ inercjalny I_m względem naszych (dowolnych) „roboczych” współrzędnych (x^k) . Można zatem powiedzieć, że cała informacja o polu lokalnych układów inercjalnych jest zawarta we „współczynnikach powiązania” Γ_{kl}^n . W szczególności układ jest iner-

cyjny w punkcie m jeśli, współczynniki te znikają w tym punkcie

$$((x^k) \in I_m) \Leftrightarrow (\Gamma_{kl}^n(m) = 0)$$

Założenie o „gładkiej zależności” klasy I_m od punktu m oznacza teraz, że elementy tablicy $\Gamma_{kl}^n(m)$ są odpowiednio gładkimi funkcjami na przestrzeni M .

To opowiadanie można ciągnąć dalej – np. łatwo wyprowadzić wzory na transformację obiektu Γ między dwoma dowolnymi (na ogół nieinercjalnymi) układami współrzędnych. Zainteresowanych odsyłam do bogatej literatury z geometrii różniczkowej. Jednak główną własnością przestrzeni z powiązaniem jest istnienie pola lokalnych układów inercjalnych, co pozwala na wyróżnienie szczególnych linii „prostych”, dla których „przyspieszenie” (liczone w układzie inercjalnym) znika.

Grawitacja jako pole lokalnych układów inercjalnych

Właśnie taka struktura geometryczna czasoprzestrzeni odpowiada polu grawitacyjnemu według Einsteina. Równania ruchu ciała spadającego swobodnie, to równania ortodromy (18) w czterowymiarowej czasoprzestrzeni, które można przepisać jako

$$m \ddot{x}^n = -m \Gamma_{kl}^n \dot{x}^k \dot{x}^l \quad (20)$$

Interpretację tradycyjną, w której prawa strona jest siłą grawitacyjną działającą na ciało o masie m , zgodnie z drugą zasadą Newtona, Einstein proponuje zastąpić interpretacją geometryczną – prawa strona tylko dlatego nie znika, że do opisu ruchu użyliśmy układu odniesienia, który nie jest inercjalny! Siły grawitacyjne dadzą się (lokalnie, w punkcie m) wyeliminować, jeśli tylko przejdziemy do **układu inercjalnego** w punkcie m . I nie jest to żadna *science fiction*, lecz realna możliwość. Sto lat temu uważano, że polega ona na zamknięciu się w swobodnie spadającej windzie, gdzie (zanim nastąpi katastrofa) znajdziemy się w stanie całkowitej nieważkości. W takiej sytuacji będziemy mogli uważać, że sił grawitacyjnych nie ma! W dzisiejszych czasach stan nieważkości stał się doświadczeniem wielu ludzi – astronautów i pracowników stacji kosmicznych.

Siły grawitacyjne uzyskały zatem status „sił pozornych” z mechaniki klasycznej, takich jak siła odśrodkowa czy siła Coriolisa. Są one proporcjonalne do masy

ciała. Można je wyeliminować przez wybór odpowiedniego (inercjalnego!) układu odniesienia. Możliwość sprawdzenia, czy „nasz układ odniesienia” jest inercjalny „względem odległych gwiazd” bardzo absorbowwała myśl Ernesta Macha, którego idee wywarły duży wpływ na Einsteina, czemu dał on wielokrotnie wyraz w swych pismach. Pasażer obracającej się karuzeli – argumentował Mach – jeśli nie pozwolić mu na obserwację gwiazd, nie potrafi stwierdzić, czy działające nań siły, to siły pozorne, czy też „prawdziwe” siły grawitacyjne. Ze wstydem wyznam, że z wielkim trudem podążałem za niektórymi myślami Macha, spisany mi w języku **globalnych układów odniesienia**. Wydaje mi się, że cała sprawa staje się niezwykle prosta i klarowna, jeśli zrezygnujemy z globalnych układów odniesienia i całe rozważania ograniczymy do przytoczonej w niniejszym wykładzie dyskusji **lokalnej struktury czasoprzestrzeni**. W tej teorii dylemat Macha – czy działają na mnie siły pozorne (odśrodkowa i Coriolisa), czy grawitacyjne – nie ma żadnego znaczenia. Niezależnie od nazwy, siły te mogą zostać wyeliminowane, ale tylko **lokalnie**.

Jak mierzyć krzywiznę czasoprzestrzeni

Jeśli w zwykłej, płaskiej przestrzeni wybierzemy krzywoliniowy układ współrzędnych, to równania linii prostych staną się bardzo skomplikowane. Można się o tym łatwo przekonać, próbując napisać np. równanie prostej w układzie współrzędnych sferycznych, w poczciwej, znanej ze szkolnej „stereometrii” – trójwymiarowej przestrzeni euklidesowej. Wystąpią wtedy wysoce nietrywialne współczynniki powiązania Γ_{kl}^n . Część z nich – te, które odpowiadają współrzędnym kątowym – pojawiła się już zresztą w niniejszym artykule pod postacią współczynników (8). Pokażemy teraz, w jaki sposób odróżnić sytuację, gdy przestrzeń jest naprawdę prosta (a jedynie prezentuje się jakby była krzywa, bo widać używamy skomplikowanego, krzywoliniowego układu) od sytuacji, kiedy przestrzeń jest **naprawdę krzywa**. Rozpocznijmy od „wyprostowania” krzywych współrzędnych (x^k) w jednym punkcie m za pomocą wzoru

$$y^k := x^k + \frac{1}{2} \Gamma_{ij}^n(m) x^i x^j + \text{poprawki trzeciego i wyższych rzędów} \quad (21)$$

Aby się przekonać, że tak poprawione współrzędne (y^k) są inercjalne w punkcie m wystarczy zauważyć, że zachodzi równość (19)

$$\begin{aligned}\Gamma_{ij}^n(m) &:= \frac{\partial^2 y^k}{\partial x^i \partial x^j}(m) = \\ &= \frac{\partial x^k}{\partial y^a}(m) \cdot \frac{\partial^2 y^a}{\partial x^i \partial x^j}(m)\end{aligned}\quad (22)$$

bowiem macierz pierwszych pochodnych jest macierzą jednostkową. Czy nie dałoby się — do tej pory nieistotnych — poprawek wyższych rzędów dobrać tak, by wyprostować te współrzędne zarówno w samym punkcie, jak i wokół niego? Na początku przynajmniej tak, by wyzerować w tym punkcie nie tylko współczynniki powiązania (co już się stanie, gdy tylko przejdziemy do zmiennych (y^a)), ale również ich pierwsze pochodne

$$\Gamma_{bcd}^a := \partial_d \Gamma_{bc}^a$$

W tym celu znów zaczniemy poprawiać już raz poprawione współrzędne, przy czym istotne teraz będą poprawki trzeciego rzędu, bowiem tylko one dadzą wkład do pochodnych liczonych w zerze układu współrzędnych. Ogólna taka poprawka mogłaby wyglądać następująco

$$\tilde{y}^a := y^a + \frac{1}{2} P_{bc}^a y^b y^c + \frac{1}{6} Q_{bcd}^a y^b y^c y^d \quad (23)$$

Po to jednak, by coś naprawić, ale przy tym nie popsuć już znikających współczynników Γ_{bc}^a , musimy położyć $P_{bc}^a = 0$. Wtedy współczynniki te zmienią się o drugie pochodne wyrazu trzeciego rzędu (czyli o człon liniowe, znikające w zerze), natomiast ich pochodne zmienią się o trzecie pochodne, czyli o człon zerowego rzędu, wyznaczony przez tablice współczynników Q_{bcd}^a . Z symetrii trzecich pochodnych wynika, że będzie to część całkowicie symetryczna tej tablicy. Gdyby jednak Q_{bcd}^a zawierała cokolwiek poza tą częścią symetryczną, to owe „cokolwiek” (np. część antysymetryczna w jakichś dwu wskaźnikach) i tak nie wpłynę na poprawkę (23). Aby zatem uprościć nasze rachunki, możemy założyć, że tablica Q_{bcd}^a jest całkowicie symetryczna

$$Q_{bcd}^a = Q_{(bcd)}^a \quad (24)$$

$$\begin{aligned} &:= \frac{1}{6} \{ Q_{bcd}^a + Q_{cdb}^a + Q_{dbc}^a + \\ &\quad + Q_{dcb}^a + Q_{cbd}^a + Q_{bdc}^a \} \end{aligned}\quad (25)$$

Wtedy pochodne współczynników powiązania Γ w nowym układzie współrzędnych zmienią się po prostu o Q

$$\tilde{\Gamma}_{bcd}^a = \Gamma_{bcd}^a + Q_{bcd}^a \quad (26)$$

które są całkowicie do naszej dyspozycji. Jak widać, po-trafilibyśmy w ten sposób „zabić” całą tablicę pochodnych Γ_{bcd}^a jedynie wtedy, gdyby była ona całkowicie symetryczna — wystarczyłoby położyć $Q_{bcd}^a = -\Gamma_{bcd}^a$. W ogólnym przypadku, gdy Γ_{bcd}^a nie jest całkowicie symetryczna, zdołamy jedynie „zabić” jej część całkowicie symetryczną. Wynika stąd, że cała reszta — to, co zostaje po odjęciu części całkowicie symetrycznej

$$K_{bcd}^a = \Gamma_{bcd}^a - \Gamma_{(bcd)}^a \quad (27)$$

nie zależy już od żadnych poprawek, czyli jest pewną wielkością charakteryzującą badaną geometrię, niezależną od tego w jakim układzie współrzędnych (byle był inercjalny!) dokonujemy obliczeń. Rzeczywiście, po dokonaniu poprawek (23) otrzymujemy

$$\begin{aligned} K_{bcd}^a &= \tilde{\Gamma}_{bcd}^a - \tilde{\Gamma}_{(bcd)}^a = \\ &= \Gamma_{bcd}^a + Q_{bcd}^a - \Gamma_{(bcd)}^a - Q_{(bcd)}^a = K_{bcd}^a \end{aligned}\quad (28)$$

na mocy równania (24). Wielkość K nazywa się **tenso-rem krzywizny** przestrzeni M . Gdy jest ona różna od zera, to — jak widać z definicji — nie ma szans na istnienie globalnego układu inercjalnego, który „zabiłby” współczynniki Γ globalnie, bowiem „zabiłby” on również ich pochodne oraz (w szczególności) ich kombinacje (27). Zatem znikanie tensora krzywizny jest warunkiem koniecznym na płaskość przestrzeni. Czy jest to również warunek dostateczny? Czy po wyzerowaniu pierwszych pochodnych będziemy musieli zająć się zerowaniem kolejnych i tak w nieskończoność?

Okazuje się, że nie będzie to potrzebne, bowiem zerowanie się tensora krzywizny K w „grubym” zbiorze pozwala już skonstruować współrzędne globalnie inercjalne w tym zbiorze. Zatem znikanie K jest warunkiem koniecznym i dostatecznym na płaskość. **Tensor krzywizny mierzy rzeczywistość krzywiznę**, rozumianą jako niemożność skonstruowania globalnie inercjalnych współrzędnych. Mistyfikacja polegająca na zapisaniu współczynników powiązania **płaskiej przestrzeni** w krzywoliniowym układzie współrzędnych, przez co będą one wyglądały bardzo skomplikowanie, zostanie natychmiast odkryta, gdy obliczymy tensor krzywizny, bowiem mimo pozornie skomplikowanej postaci funkcji $\Gamma_{bcd}^a(m)$, ostateczny rezultat takich rachunków będzie trywialny: $K_{bcd}^a \equiv 0$.

Materia powodem zakrzywienia czasoprzestrzeni

W teorii grawitacji Einsteina pole grawitacyjne jest zatem polem lokalnych układów inercjalnych, które można opisać w ustalonym układzie współrzędnych jako pole zależnych od punktu czasoprzestrzeni współczynników powiązania $\Gamma_{bc}^a(m)$. Ale czym zastąpić równania pola, które w teorii Newtona wyglądały następująco

$$\vec{G} = -\gamma \frac{\mu \vec{x}}{\|\vec{x}\|^3} \quad (29)$$

gdzie μ jest masą ciała będącego źródłem pola (np. słońca w układzie planetarnym), \vec{x} – wektorem wiodącym ciała próbnego względem źródła, zaś γ – stałą grawitacyjną. Gdy źródłem pola nie jest masa punktowa, ale ciągły rozkład mas, opisany gęstością ρ , to pole \vec{G} jest superpozycją wkładów od poszczególnych punktów. Łatwo zauważyć, że odpowiada to następującym warunkom, które spełnia to pole:

1. Pole \vec{G} jest bezwirowe, czyli $\text{rot} \vec{G} = 0$.
2. Spełnione jest równanie wiążące rozbieżność pola ze źródłami, mianowicie $\text{div} \vec{G} = 4\pi\gamma\rho$.

Okazuje się, że w teorii Einsteina powyższe warunki należy zastąpić następującymi:

1. Warunek zgodności z resztą fizyki, opisaną przez strukturę metryczną czasoprzestrzeni, bowiem inne działy fizyki (np. elektrodynamika) wykorzystują pole tensora metrycznego g_{kl} , opisujące „odległości” czterowymiarowe między różnymi zdarzeniami przestrzennymi. Strukturą tego tensora zajmuje się tzw. szczególna teoria względności. Zgodność obu struktur polega na tym, że tensor metryczny ma być w jakimś sensie „stały” w czasoprzestrzeni. Cóż to jednak ma znaczyć, skoro przechodząc od jednego układu współrzędnych do innego, można współrzędne g_{kl} znacznie zmienić i np. ze stałych uczynić zmienne? Otóż warunek ten oznacza, że pochodne tensora metrycznego mają zerować się **w układzie inercjalnym**. Mówi się wówczas, że metryka jest **kowariantnie stała** na M . Warunek ten zastępuje jednorodny warunek $\text{rot} \vec{G} = 0$ z teorii newtonowskiej.
2. Warunek niejednorodny zastąpiony jest **równaniem Einsteina**, wiążącym krzywiznę ze źródłami pola

$$G_{kl} = 8\pi\gamma T_{kl}$$

W równaniu tym

$$G_{kl} := R_{kl} - \frac{1}{2} g_{kl} R$$

oznacza tzw. tensor Einsteina, $R_{kl} := \frac{3}{2} K_{kl}^n$ jest tzw. tensorem Ricciego, powstałym z tensora krzywizny przez zwężenie po pierwszym i ostatnim wskaźniku, zaś $R = g^{kl} R_{kl}$ jest śladem tensora Ricciego, zwanym **skalarem krzywizny**. W każdym przypadku po lewej stronie stoi (na miejscu dywergencji pola \vec{G} z teorii Newtona) jakaś informacja o krzywiznie, wyrażona pochodnymi pola Γ . Natomiast po prawej stronie stoi tensor energii-pędu materii T , niosący informację o źródłach pola grawitacyjnego.

Jeśli czasoprzestrzeń nie jest pusta, czyli gdy stojąco po prawej stronie równań Einsteina tensor energii-pędu materii nie jest równy zeru, to również lewa strona nie może zniknąć, zatem czasoprzestrzeń nie może być płaska. Można skrótkowo powiedzieć, że obecność materii, której konfiguracja jest mierzona wielkością tensora energii-pędu T , „wykrzywia” czasoprzestrzeń.

Nie jest natomiast prawdą, że przestrzeń pusta musi być płaska. Z równań Einsteina wynika jedynie znikanie tensora Einsteina G lub – co jest równoważne – tensora Ricciego R , ale jest to jedynie część informacji zawartych w pełnym tensorze krzywizny K . Dziś wiemy już, że wcale nie musi to pociągać znikania całego tensora krzywizny. Istnieją bowiem rozwiązania równań Einsteina opisujące przestrzeń pustą, ale mimo to krzywą. Jej krzywizna propaguje się w sposób nieco zbliżony (choć dużo trudniejszy, ze względu na nieliniowość równań Einsteina) do propagacji fal elektromagnetycznych i opisuje zjawiska, które nazywamy **falami grawitacyjnymi**. Ten fakt nie był oczywisty od początku badań teorii. Sam Einstein miał poważne wątpliwości na ten temat i dość długo sadził, że fal grawitacyjnych nie ma. W odkryciu (teoretycznym) tych fal wybitny udział ma wielki polski fizyk-teoretyk – profesor Andrzej Trautman, absolwent Politechniki Warszawskiej. Na ich powtórne – tym razem obserwacyjne – odkrycie czekamy do tej pory.

Krótką historia odkrywania „czarnych dziur”

Już kilka miesięcy po ogłoszeniu Ogólnej Teorii Względności – na przełomie 1915 i 1916 roku – wybitny niemiecki astronom, fizyk i matematyk, Karl Schwarzschild znalazł pierwsze ścisłe rozwiązanie równań Einsteina,

opisujące sferycznie symetryczne pole grawitacyjne wokół ciężkiego ciała o danej masie. Rozwiązanie to jest relatywistycznym odpowiednikiem pola (29). Scenerię tego odkrycia znamy dobrze z książki Jaroslava Haška o dobrym wojaku Szwejku – okolice Przemyśla, błoto i śnieg, okopy pierwszej wojny światowej. Karl Schwarzschild został zmobilizowany na wojnę w charakterze oficera artylerii, jednak nawet w tak trudnych warunkach nie poniechał aktywności intelektualnej. Zresztą kilka miesięcy po swoim wspaniałym odkryciu, w maju 1916 roku, zmarł na skutek rzadko spotykanej choroby skórnej – pęcherzycy, której nabawił się na wojnie.

Podobnie jak pole (29) w punkcie $\vec{x} = 0$, rozwiązanie Schwarzschilda cechuje się osobliwością, gdy opisujące je formuły tracą sens, bowiem zawierają dzielenie przez zero. Jednak w przypadku pola Schwarzschilda ta osobliwość występuje wcześniej, zanim zbliżymy się do „centrum” punktowej masy, którą chcielibyśmy opisywać. W przypadku masy równej masie Ziemi ten krytyczny „promień Schwarzschilda” wynosi około centymetra, a dla masy Słońca – około 3 km. Można byłoby pocieszać się, że w przyrodzie nigdy taka osobliwość nie wystąpi, bo nie potrafimy przecieżyć „sprasować” masy Ziemi do rozmiarów centymetra! Dziś wiemy jednak, że materia w tzw. gwiazdach neutronowych może być upakowana **dużo gęściej** niż to, z czym mamy do czynienia w potocznym doświadczeniu i istnieją, być może, sytuacje astrofizyczne, w których cała masa będąca źródłem geometrii Schwarzschilda jest ukryta wewnątrz tej „osobliwości”. Zatem problem pozostaje. Co się dzieje w „osobliwości” rozwiązania Schwarzschilda? Czy przestaje obowiązywać znana nam fizyka? Czy wewnątrz obszaru o silnej krzywiznie dzieje jakaś otchłań, czy jest tam po prostu „czarna dziura”? Okazuje się, że sama osobliwość występująca we wzorach na rozwiązanie Schwarzschilda, która tak niepokoiła fizyków od początku Ogólnej Teorii Względności, jest pozorna i opisuje jedynie własności szczególnego układu współrzędnych, których użył jego twórca.

Sytuacja jest podobna do tej, w której chcielibyśmy punkty paraboloidy obrotowej w trójwymiarowej przestrzeni euklidesowej, danej równaniem

$$\sqrt{x^2+y^2} = 1 + z^2$$

parametryzować za pomocą dwu zmiennych (x, y) , traktując współrzędną z jako ich funkcję

$$z = +\sqrt{\sqrt{x^2+y^2}-1}$$

Widać, że na kole danym równaniem $x^2 + y^2 = 1$, taki opis zawodzi, bowiem występuje „czarna dziura” odpowiadająca wartościom $x^2 + y^2 < 1$, a przecież nie ma tu żadnej osobliwości, należy jedynie do zbioru rozwiązań powyższego równania, opisującego górną połowę paraboloidy, dokleić jej dolną połowę, odpowiadającą znakowi „minus” przed pierwiastkiem. Bardzo łatwo znaleźć współrzędne, które zachowują się nieosobliwie w pobliżu „wąskiego gardła”, wzdłuż którego sklejono obie połowki tej powierzchni.

Takim właśnie „wąskim gardłem” trójwymiarowego cięcia $\{t = \text{const}\}$ jest pozorna osobliwość Schwarzschilda (zob. rys. 3). Z dala od niej to cięcie coraz bardziej przypomina naszą płaską, euklidesową przestrzeń trójwymiarową. Zbliżając się do „wąskiego gardła” przestrzeń staje się coraz bardziej zakrzywiona, a po jego przekroczeniu znów zaczyna się wypłaszczać – być może aż do jakiegoś „drugiego końca” przestrzeni, jak na rysunku 3, ale być może aż do schowanej wewnątrz tego gardła materii, będącej źródłem pola grawitacyjnego, tak jak na rysunku 4.

Tym niemniej, obszar czasoprzestrzeni odpowiadający punktom leżącym wewnątrz tego „gardła” ma dość szczególne właściwości, bowiem nie może się zeń wydostać „na zewnątrz” żadna informacja. Co więcej, **tam naprawdę jest osobliwość** (ale nie przestrzenna, którą można byłoby pokazać na rysunku 4, lecz **czasoprzestrzenna**) nie dająca się usunąć przez manipulację współrzędnymi. Obecnie „czarną dziurą” nazywa się taką właśnie sytuację.

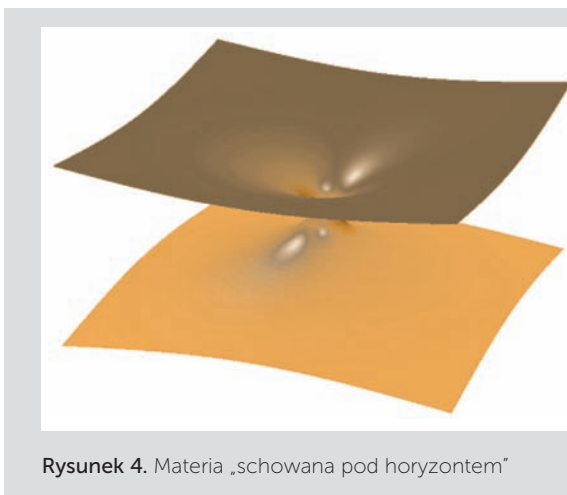
Teorii czarnych dziur nadał impet wielki astrofizyk hinduski Subrahmanyan Chandrasekhar (1910–1995), uhonorowany za swe prace nagrodą Nobla w 1983 roku. W 1930 roku, wiążąc dwie bardzo młode teorie fizyczne – Ogólną Teorię Względności i Mechanikę Kwantową – zauważył, że ewolucja gwiazd do stacjonarnego stanu Białego Karła, w którym ciśnienie gazu elektronowego wewnątrz gwiazdy powinno zrównoważyć siły grawitacyjne, będzie zachodziła zupełnie inaczej dla gwiazd cięższych niż pewna wartość krytyczna, znana dziś jako masa Chandrasekhara i wynosząca 1,44 masy Słońca. Otóż dla ciężkiej gwiazdy ciśnienie wewnętrzne nie zdoła zrównoważyć przyciągania grawitacyjnego i w końcu cała masa zapadnie się „pod horyzont” jak na rysunku 4. Uformuje się więc czarna dziura, a w każdym razie widziane z zewnątrz pole grawitacyjne będzie miało wiele cech sytuacji opisywanej przez geometrię Schwarzschilda. Wydaje się, że jest to sytuacja dość powszechna we Wszechświecie



Rysunek 3. Cięcie $\{t = \text{const}\}$ czasoprzestrzeni Schwarzschilda

i powinniśmy na niebie widzieć wiele takich obiektów. Aby je odróżnić od zwykłych gwiazd czy galaktyk, potrzebne są dokładniejsze obserwacje niż te, których można dokonywać za pomocą tradycyjnych teleskopów. W tym celu buduje się obecnie obserwatoria umieszczone w Kosmosie. Takie dwa orbitalne obserwatoria – Spitzer oraz Chandra – dostarczyły ogromnej ilości danych potwierdzających istnienie czarnych dziur. Astrofizycy mówią ostatnio o „setkach czarnych dziur, które zostały ostatnio odkryte w odległych galaktykach pyłowych”.

Bardzo istotna metoda testowania wyników tych obserwacji, pomagająca odróżnić czarną dziurę od „banalnej” gwiazdy lub galaktyki, polega na wykorzystaniu zjawiska soczewkowania grawitacyjnego. Wykorzystuje ona zjawisko uginania promieni świetlnych w polu grawitacyjnym. Jednym z animatorów i współautorów tej metody był niedawno zmarły, wybitny polski



Rysunek 4. Materia „schowana pod horyzontem”

astrofizyk Bogdan Paczyński (ur. 1940 w Wilnie, zm. 2007 w Princeton).

Dzięki tym wszystkim obserwacjom i pomiarom astrofizycy coraz mocniej utwierdzają się w przekonaniu, że matematyczna konsekwencja teoretycznych idei Alberta Einsteina, jaką jest struktura czarnej dziury, istnieje jako fakt realnie obserwowany w przyrodzie.

Bibliografia

- [1] Kocznyński W., Trautman A.: *Czasoprzestrzeń i grawitacja*. PWN, Warszawa 1981.
- [2] Misner C.W., Thorne K.S., Wheeler J.A.: *Gravitation*. N.H. Freeman and Co, San Francisco, Cal. 1973.
- [3] Wald R.: *General Relativity*. University of Chicago Press, Chicago 1984.

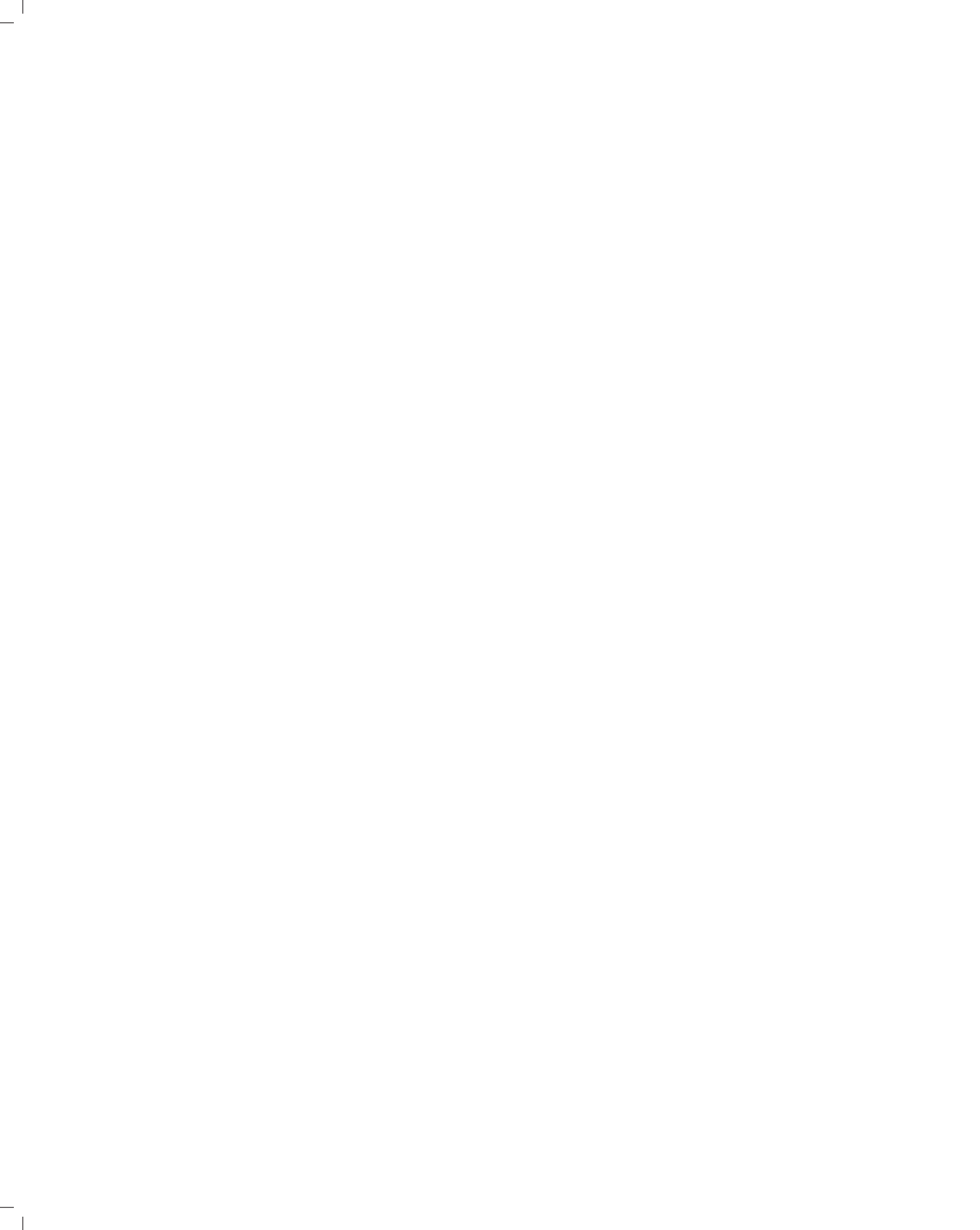
Abstract

Short history of the General Relativity Theory (together with theory of “black holes”) is presented. A mathematical structure which enables us to describe gravitational field as a field of local inertial frames is proposed in a simple form. The notion of the curvature tensor is derived as an obstruction against existence of linear coordinates.

Słowa kluczowe:

czarne dziury, ogólna teoria względności, geometria nieeuklidesowa, nieinercjalne układy współrzędnych.

Profesor Jerzy Kijowski specjalista z zakresu analizy matematycznej, algebry, geometrii, mechaniki kwantowej, geometrii różniczkowej i jej zastosowań w fizyce teoretycznej. Pełnił funkcje dyrektora Centrum Fizyki Teoretycznej Polskiej Akademii Nauk i Instytutu Fizyki Matematycznej Uniwersytetu Warszawskiego. Wyróżniony m.in.: nagrodą Polskiego Towarzystwa Matematycznego dla Młodego Matematyka, nagrodą im. Stanisława Zaremby, nagrodą „Mistrz” Fundacji Nauki Polskiej oraz nagrodą Polskiej Akademii Nauk za wyniki uzyskane w obszarze fizyki teoretycznej. Członek zespołu redakcyjnego czasopism „Reports on Mathematical Physics”, „Acta Fisica Polonica” oraz „Journal of Geometry and Physics”.



Czy krzem może chodzić — czyli o mikrosystemach, które myślą, czują i pracują

Na podstawie odczytu wygłoszonego w dniu 5 listopada 2007 roku

Ryszard Sławomir Jachowicz

Wydział Elektroniki i Technik Informatycznych Politechniki Warszawskiej

121

Gdy słyszymy o krzemie, o materialnych elementach wykonanych w krzemie, to natychmiast przychodzi na myśl układ scalony o wielkiej skali integracji z wieloma tysiącami tranzystorów w jednej strukturze. Jeśli pytamy, co się może w nim dzieć w czasie jego pracy, to fizycy powiedzą, że jest w nim nieustanny, kontrolowany ruch elektronów, elektronicy zapewnią, że jest w nim kontrolowany przepływ prądów i kontrolowana zmiana potencjałów, a informatycy stwierdzą, że jest w nim nieustanny „przepływ bitów” (co jest słuszne dla układów cyfrowych).

To wszystko co powiedziano powyżej jest prawdziwe. Należy jednak postawić pytanie, czy możliwości techniczne krzemu, a może bardziej właściwe byłoby spytać, czy możliwości technologii półprzewodnikowej, ograniczają się tylko do tak rozumianych układów scalonych. Czy technologia półprzewodnikowa pozwala produkować tylko mikroprocesory, pamięci masowe i inne układy logiczne, układy analogowe i elementy optyczne?

Otóż nie. Może znacznie więcej. Będę starał się przekonać szanownych Czytelników, że dzisiejsza technologia półprzewodnikowa pozwala zbudować prawie wszystko to co znamy ze skali makro w mechanice, chemii, optyce itd., ale w skali mini-, mikro-, a czasami nawet nano-. Uzyskując te same cechy funkcjonalne, a często znacznie lepsze.

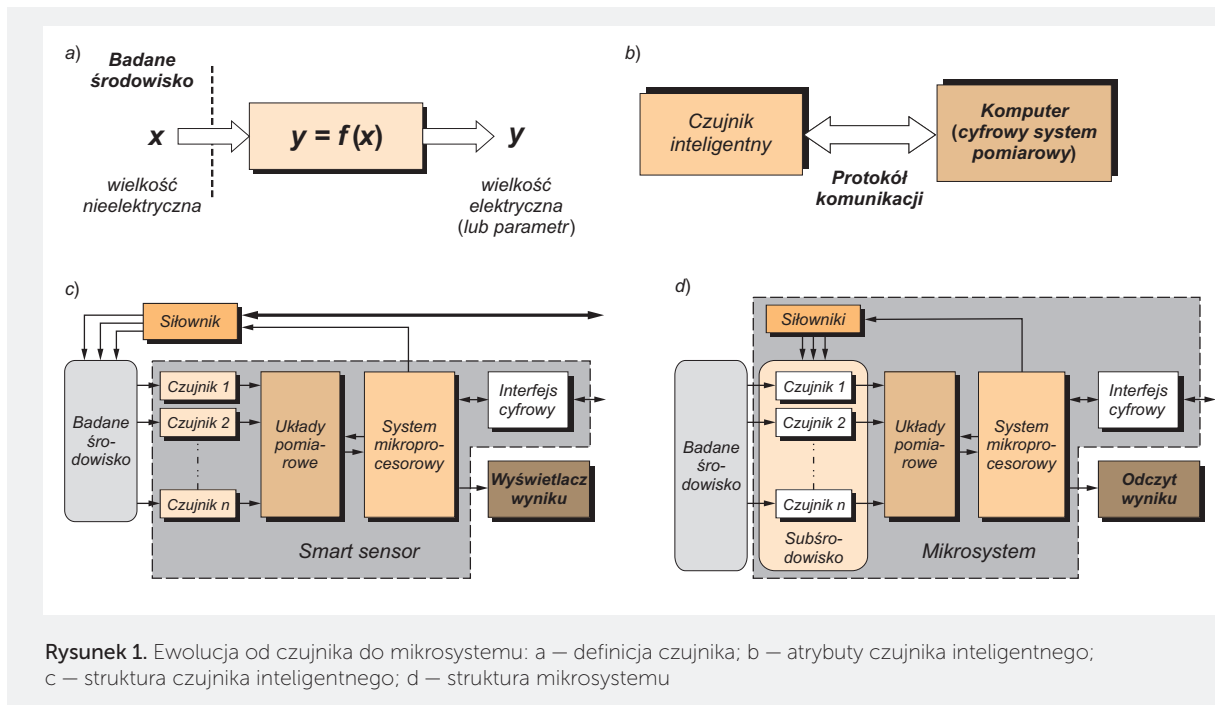
Stąd więc ten zaskakujący tytuł, który miał z jednej strony zwrócić uwagę potencjalnych słuchaczy referatu i Czytelników niniejszego streszczenia na niezwykle możliwości konstrukcyjne krzemu i technologii pół-

przewodnikowej, a z drugiej strony zasygnalizować nadzwyczajne możliwości wykonywanych w tej technologii nowych zintegrowanych elementów, które noszą nazwę mikrosystemów lub MEMS'ów (ang. *Micro Electro Mechanical Systems*).

Co to jest mikrosystem? Czym różni się od czujnika i czujnika inteligentnego?

Na rysunku 1 pokazano w zarysie ewolucję systemowo-funkcjonalną od czujnika do mikrosystemu. Czujnik to blok pomiarowy (rys. 1a), który potrafi jednoznacznie przetworzyć informację o stanie wielkości nieelektrycznej x (np. ciśnieniu, przyspieszeniu, temperaturze, przepływie itd.) na sygnał elektryczny y (np. prąd, napięcie, częstotliwość, pojemność, rezystancję itd.).

Czujnik inteligentny (rys. 1b) to taki czujnik, którego sygnał wyjściowy jest w postaci cyfrowej i posiada dodatkowo interfejs cyfrowy oraz określony „język komunikacji”, czyli tzw. protokół komunikacji. Oznacza to, że musi w swoim „wnętrzu” zawierać blok elektroniczny pomiaru parametrów czujnika (a czasami wielu czujników), układy przetwarzania cyfrowego (np. odpowiedni mikroprocesor z pamięcią i z przetwornikiem analog–cyfra) i rzeczony interfejs cyfrowy. Te wszystkie bloki mogą być zintegrowane ze sobą w jednej strukturze półprzewodnikowej (krzemowej) albo, w najgorszym razie, w jednej obudowie. Poziom inteligencji tego



Rysunek 1. Ewolucja od czujnika do mikrosystemu: a – definicja czujnika; b – atrybuty czujnika inteligentnego; c – struktura czujnika inteligentnego; d – struktura mikrosystemu

czujnika zależy od rodzaju mikroprocesora i jego oprogramowania, a czasami również od obecności innych czujników. Taki inteligentny czujnik (rys. 1c) może również sterować bezpośrednio pracą zewnętrznego elementu wykonawczego, czyli siłownika (np. silnika, elektromagnesu, grzałki itd.).

Natomiast **mikrosystem** to taki czujnik inteligentny, który w tej samej strukturze półprzewodnikowej integruje dodatkowo mikrośiownik (rys. 1d). Mikrośiowniki na ogół zawierają elementy mechanicznie ruchome i stąd ich bardziej popularna nazwa **MEMS** (ang. *Micro Electro Mechanical Systems*).

Krzem – doskonały materiał na mikrosystemy

Kryształ krzemu ma doskonałe parametry mechaniczne (moduł sprężystości podobny do stali, wytrzymałość mechaniczną prawie dwukrotnie wyższą od stali) i parametry cieplne też lepsze (przewodnictwo cieplne wyższe o około 60% od stali a rozszerzalność cieplną około 5-krotnie niższą od stali). Oznacza to, że krzem znakomicie nadaje się do wykonywania różnego rodzaju elementów mechanicznych. Natomiast niektóre związki krzemu, np. SiC lub Si₃N₄, popularne w tech-

nologii półprzewodnikowej, mają wiele z tych parametrów jeszcze lepszych. Co więcej, koszty wytwarzania nawet bardzo skomplikowanych mikrosystemów w masowej produkcji są bardzo niskie. Czy są zatem jakieś wady mikrosystemów półprzewodnikowych? Tak, są! Opracowanie każdego produktu jest bardzo kosztowne, a często bardzo trudne i czasochłonne. Ale po opracowaniu są już same korzyści – w jednym cyklu technologicznym wytwarza się od kilku do kilkuset tysięcy struktur w prawie takiej samej cenie jak wytworzenie pojedynczej sztuki. Parametry techniczne każdej sztuki są prawie identyczne zarówno w jednym cyklu produkcyjnym, jak i w kolejnych cyklach, a tzw. „uzysk produkcyjny” jest rzędu 90–95%.

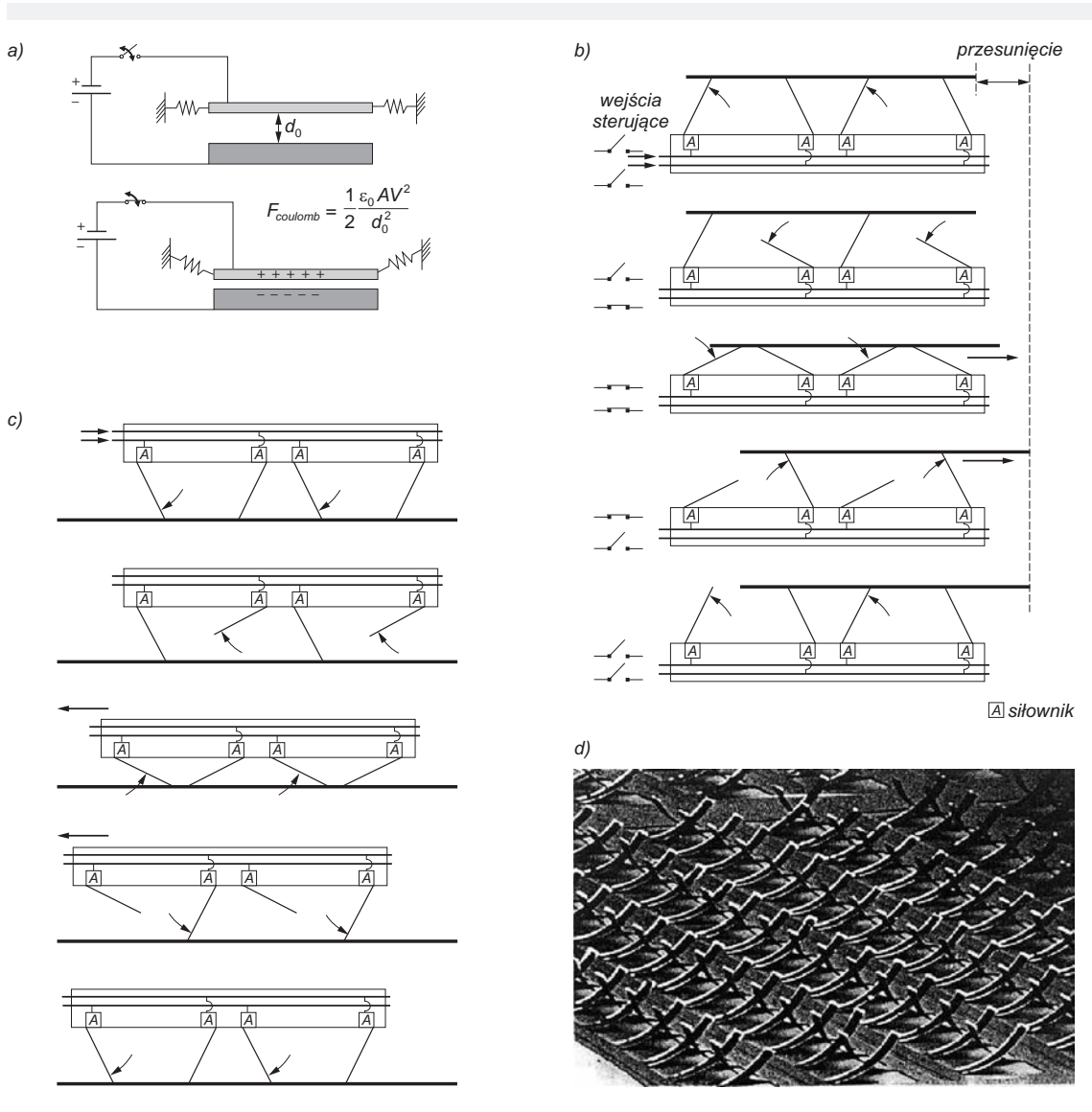
Czy można zrobić siłowniki w mikroskali?

Na wstępie trzeba powiedzieć jakie siłowniki, czyli elementy wykonawcze (ang. *actuators*), można wykonać w mikroskali. Otóż wszystkie, które są znane w makroskali – mikrośiowniki liniowe i obrotowe, mikropompy, mikrozawory, mikroprzetłaczniki oraz wiele innych elementów ruchomych (mikroprzekładnie obrotowe, mikrochwytki, ruchome mikrozwierciadła itd.). Ich praca

oparta jest na wykorzystaniu energii elektrostatycznej, elektromagnetycznej, termicznej lub piezoelektrycznej. Należy zauważyć, że zarówno poziom energii potrzebny do ich pracy, jak i czasy zadziałania, podlegają skalowaniu proporcjonalnie do skalowania ich wymiarów geometrycznych. Mamy więc do czynienia z mocami rzędu mW lub ułamków mW i czasami działania od mikro- do milisekund. Wiele z nich osiąga większą

sprawność, niezawodność i funkcjonalność niż ich odpowiedniki w makroskali, a są i takie, które w makroskali w ogóle nie mają odpowiedników, np. silniki elektrostatyczne.

Na rysunku 2 pokazano, jak w oparciu o oddziaływanie elektrostatyczne – siły Coulomba (rys. 1a) można zrealizować mikrotransporter krzemowy (rys. 1b). Odwrócona struktura o 180° i położona na podłożu



Rysunek 2. „Krzem, który chodzi” – mikrositownik elektrostatyczny: a – koncepcja działania pojedynczego sitownika; b – koncepcja transportu przez zespół sitowników; c – koncepcja struktury „samo chodzącej”; d – rzeczywista struktura realizująca funkcje z rysunku b i c

(rys. 1c) opiera się na matrycy mikroświatłowodów. Sterowana identycznie jak w przypadku mikrotransportera „chodzi” po podłożu. Na rysunku 3 pokazano jeszcze inne zastosowania mikroświatłowodów elektrostatycznych – do sterowania pociągami mikroświatłowodów.

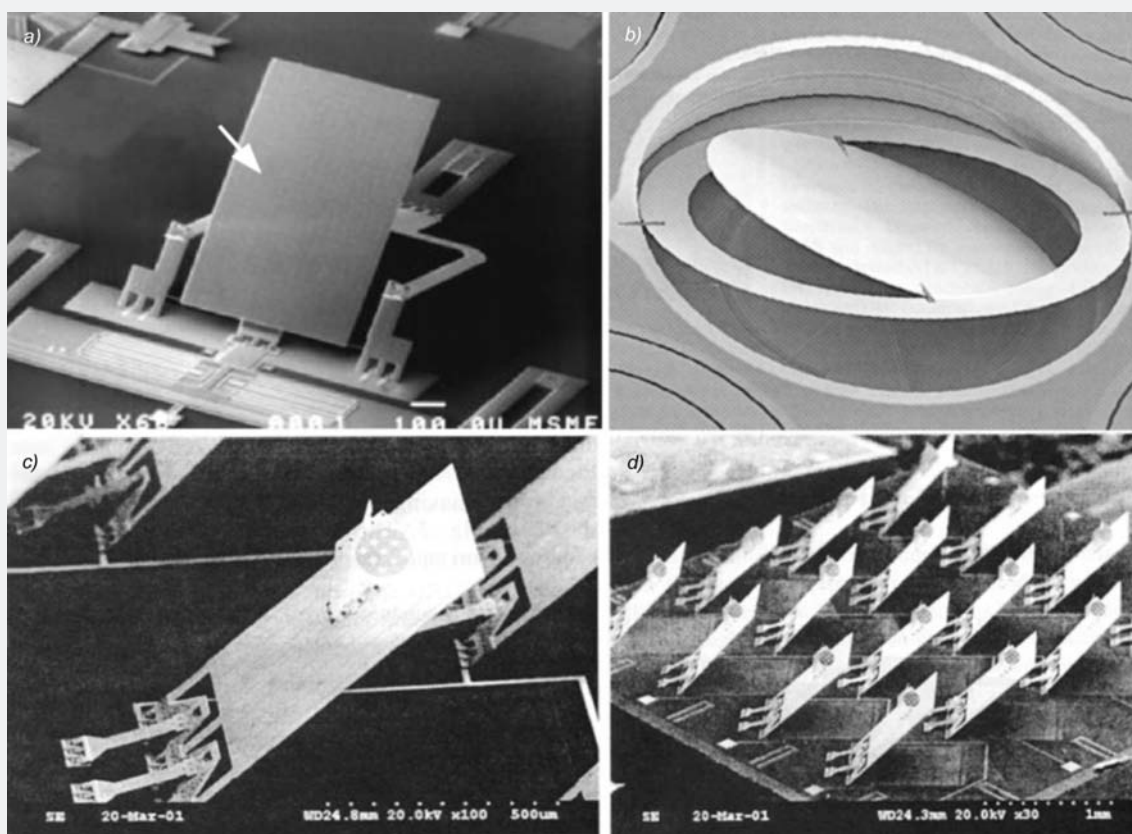
Przykładowe konstrukcje mikrosystemów

Z każdym rokiem liczba nowych mikrosystemów jest coraz większa. Część z nich trafia na rynek i jest masowo produkowana, a reszta jest dalej doskonalona. Do największych sukcesów rynkowych w dziedzinie mikrosystemów należy zaliczyć mikrosystem drukarski w drukarkach atramentowych (roczna sprzedaż około 3 mld USD), matryca mikroświatłowodów dla projektorów multimedialnych, tzw. DMD (ang. *Digital Mirror*

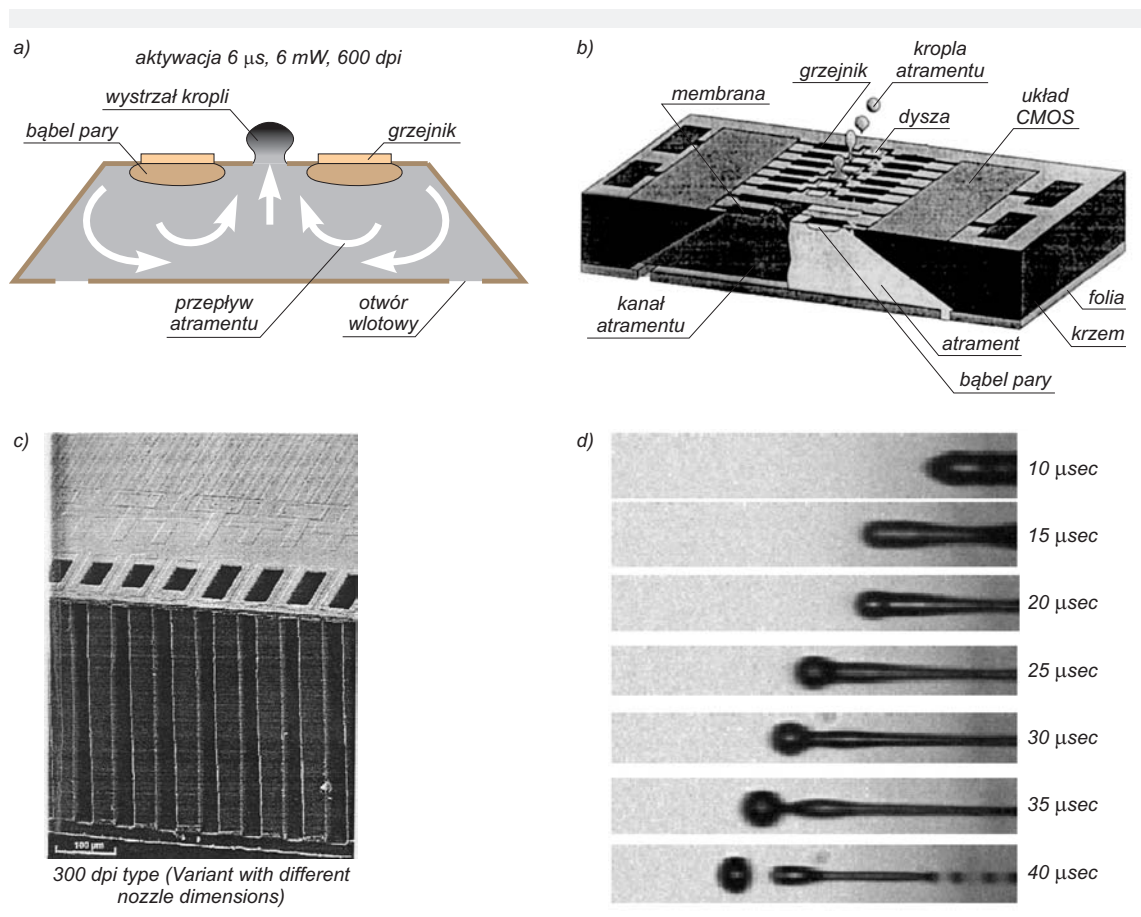
Devices, 1,5 mld USD/rok), żyroskopy z silnikiem liniowym i najnowsze z silnikiem obrotowym (0,8 mld USD na rok łącznie), precyzyjna głowica magnetyczna do odczytu i zapisu CD ROM (2,5 mld USD na rok) i wiele innych.

Przyjrzyjmy się niektórym z nich:

- **Mikrosystem drukarski** jest matrycą mikroświatłowodów (od 20 do 60 i więcej) umieszczonych w dnie pojemnika z atramentem. Działanie każdego z nich opiera się na prostej zasadzie (rys. 4): strzela kroplę atramentu o objętości rzędu 2 nanolitrow na odległość około 2 mm w czasie około 6 μ s. Stosuje się w nich siłowniki termiczne (grzejniki umieszczone w mikrownęce) lub piezoelektryczne (z membraną w dnie wnęki).
- **Mikrożyroskop** jest obecnie najbardziej kompleksowym mikrosystemem jaki występuje na rynku. Na rysunku 5 przedstawiono dwa różne typy żyroskopów.



Rysunek 3. Konstrukcje mikroświatłowodów sterowanych: a – elektrostatycznie liniowym mikroświatłowodem; b – elektrostatycznie 2D; c – elektrostatycznie hybrydowo 2D; d – matryca zwierciadeł z rysunku c



Rysunek 4. Mikrodziąto drukarki atramentowej: a – zasada działania; b – wylot rzeczywistego „dziąta”; c – matryca „dziąt”; d – kropla wystrzelonego tuszu w różnych fazach czasowych

Pierwszy z silnikiem elektrostatycznym obrotowym (rys. 5 a), który jest w stanie detekować prędkość kątową w dwóch ortogonalnych kierunkach z czułością 1 obr./12 godz. i przyspieszenie w trzech kierunkach z czułością 0,01 g (rys. 5 b). Jego praca opiera się na zasadzie Coriolisa i wykorzystaniu mikrosilnika elektrostatycznego o zawrotnej prędkości obrotowej 74 000 obr./min, gdzie wirnik o grubości 50 μm lawituje (aby wyeliminować tarcie) w pionowym polu elektrostatycznym 3 μm nad powierzchnią, a odległość między elektrodami rotora i stojana wynosi 2,2 μm . Żyroskop ten posiada niezwykle rozbudowany system elektrod sterujących, które służą nadaniu prędkości obrotowej wirnika, jego lewitowaniu i utrzymaniu wirnika w stanie równowagi niezależnie od sił Coriolisa i od poszczególnych składowych przyspieszenia (rys. 5 c).

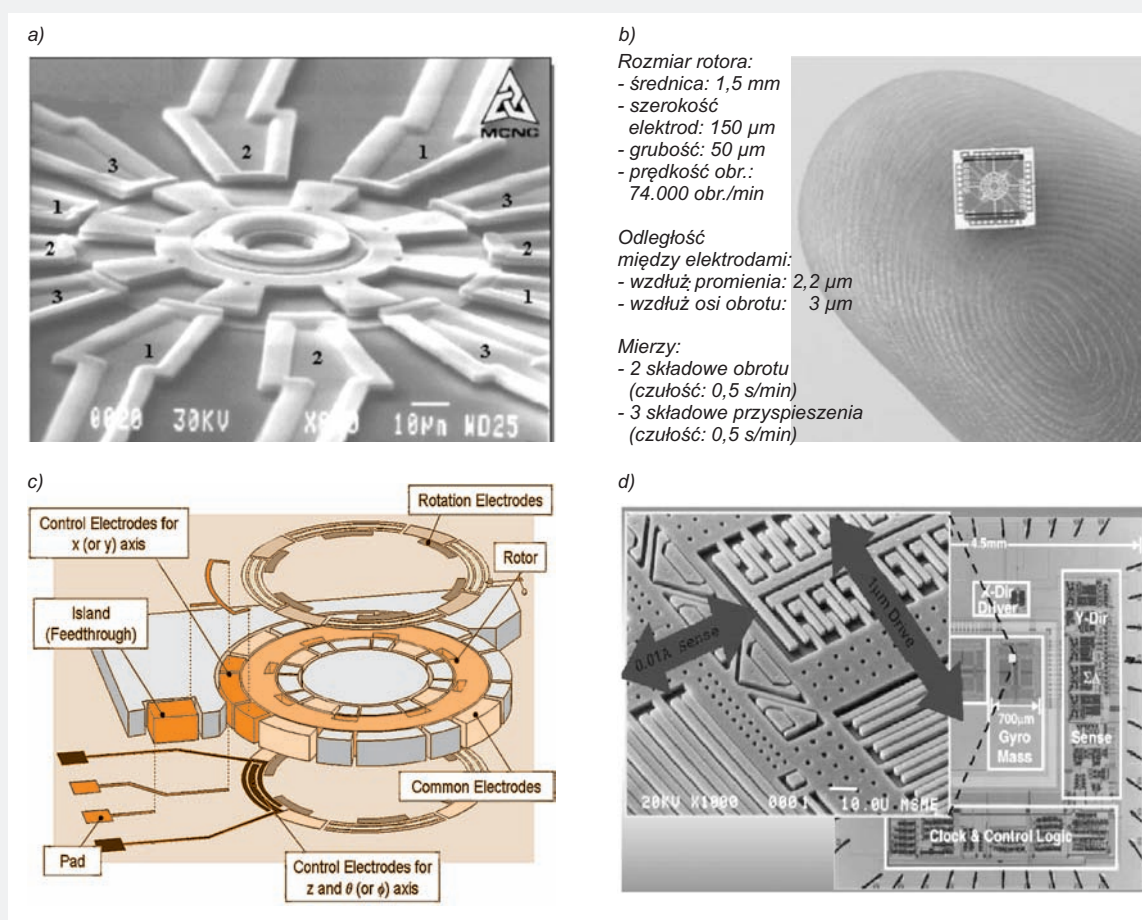
Na rysunku 5 d pokazany jest mikrożyroskop z silnikiem liniowym (o amplitudzie ruchu masy sejsmicznej 1 μm) o parametrach pomiarowych o klasę gorszych od żyroskopu obrotowego.

— **Sztuczne oko** z przeznaczeniem dla osoby widzącej jest już dopracowane do perfekcji (odbiór obrazu odbywa się przez *display*). Przy realizacji sztucznego oka dla osoby niewidzącej pojawia się w zasadzie tylko jeden, ale niestety bardzo poważny, problem – konieczność stworzenia interfejsu podłączonego do systemu nerwowego człowieka. Na rysunku 6 a przedstawiono koncepcję rozwiązania tego problemu – minikamera umieszczona jest w okularach, a sygnał elektryczny przesyłany jest przez zintegrowany z nią nadajnik do odbiornika wszczepionego do oka. Odbie-

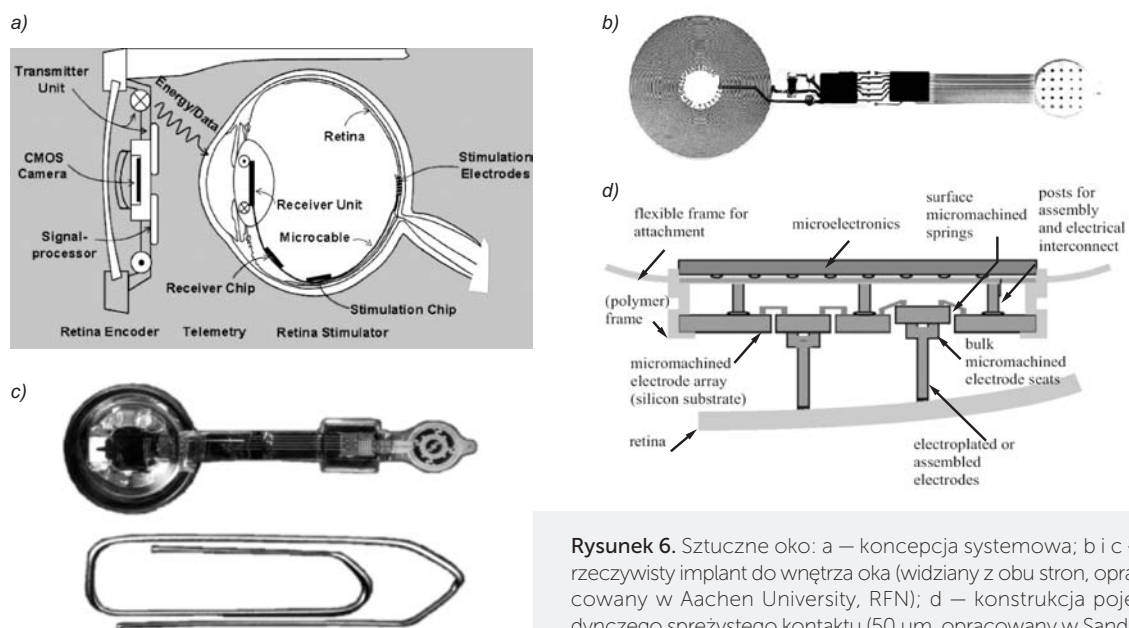
rany sygnał jest przetwarzany na matrycę potencjałów (jak w kamerze CCD) i przekazywany na dwuwymiarową matrycę kontaktów. Jednym z trudniejszych do rozwiązania problemów jest zapewnienie dopasowania się poszczególnych pixeli (kontaktów) do krzywiny siatkówki (ang. *retina*), tak aby zapewnić w każdym punkcie siatkówki dobry kontakt elektryczny, umożliwiający potencjatowe pobudzenie odpowiednich axonów.

– **Mikrosystem do dynamicznych pomiarów wilgotności** opracowano w Zakładzie Mikrosystemów i Systemów Pomiarowych (ISE, WEiTI, PW; patrz rys. 7). Umożliwia on pomiar wilgotności gazu od 3–5 razy na sekundę (prędkość nieosiągalna w innych czujnikach

wilgotności). Konstrukcja i uzyskane parametry pomiarowe pozwoliły na zastosowanie omawianego systemu w dermatologii do **pomiaru przelnaskórkowej utraty wody w organizmie człowieka** (tzw. TEWL – ang. *TransEpidermal Water Loss factor*). Jest to ważny pomiar diagnostyczny przy schorzeniach dermatologicznych i zabiegach kosmetycznych (rys. 7c). Charakteryzuje stopień wysuszenia skóry i umożliwia ocenę stanu skóry przed i po zastosowaniu kuracji (np. smarowaniu maścią lub kremem). Drugie ważne zastosowanie dotyczy **pomiaru wilgotności w nosie i gardle pacjenta w czasie oddychania** (rys. 7d). Ten system pomiarowy jest niezwykle przydatny do poznania mechanizmów funkcjonowania nosa i gardła w kondycjonowaniu powietrza trafiającego do płuc w procesie



Rysunek 5. Mikrozyroskopy: a – zdjęcie pierwszego na świecie mikrosilnika elektrostatycznego o średnicy wirnika 100 μm; b – mikrozyroskop z elektrostatycznym mikrosilnikiem obrotowym japońskiej firmy Tokimec; c – uproszczony schemat konstrukcyjny; d) mikrozyroskop z silnikiem liniowym – [5]



Rysunek 6. Sztuczne oko: a – koncepcja systemowa; b i c – rzeczywisty implant do wnętrza oka (widziany z obu stron, opracowany w Aachen University, RFN); d – konstrukcja pojedynczego sprężystego kontaktu (50 μm , opracowany w Sandia Laboratory, USA) – [6]

oddychania. Umożliwia on wykonanie i zarejestrowanie od kilkunastu do kilkudziesięciu pomiarów w czasie jednego cyklu oddechowego. Opracowany mikrosystem i zbudowane na jego bazie systemy pomiarowe wydają się dobrze wypełniać niszę rynkową w dziedzinie aparatury diagnostycznej dla medycyny i są wysoko oceniane przez naukowców-laryngologów.

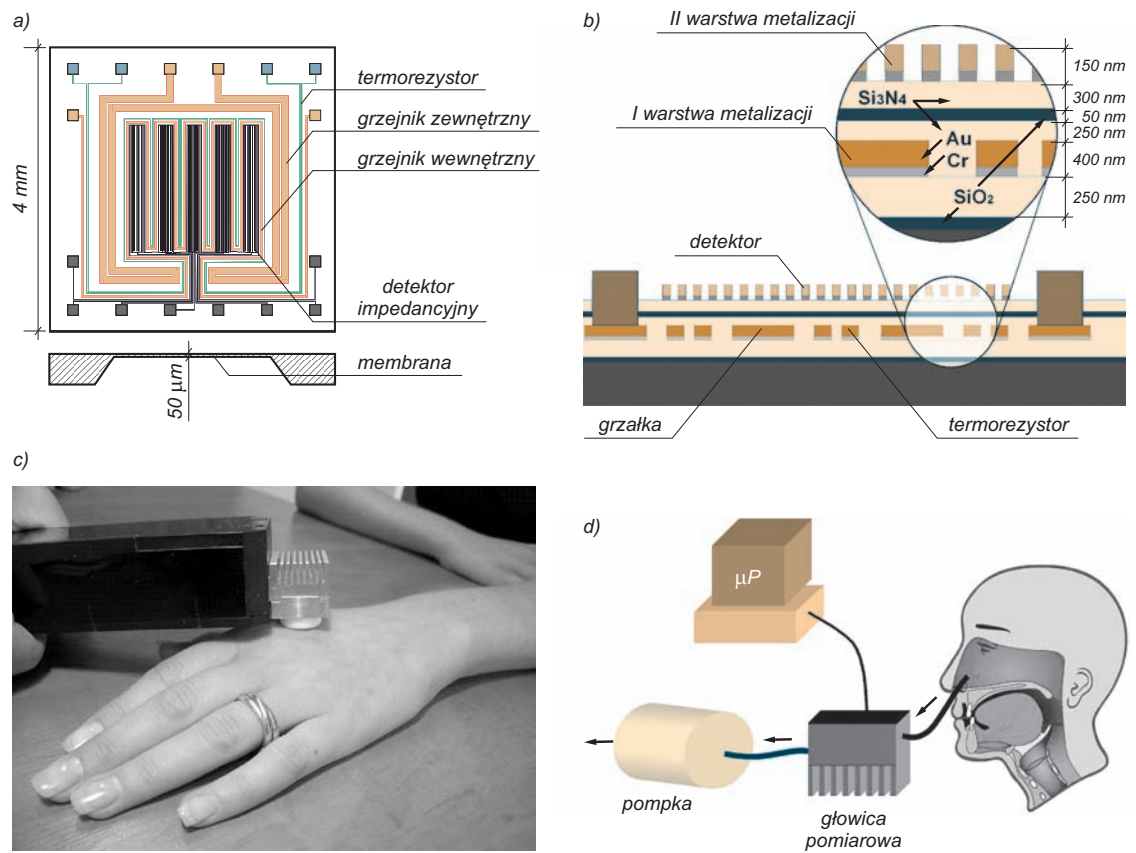
Czy warto inwestować w rozwój mikrosystemów?

Jak już wcześniej wspomniano, cechą charakterystyczną mikrosystemów jest ich niewielki koszt jednostkowy (po uruchomieniu produkcji masowej), przy jednoczesnym wysokim zaawansowaniu technologicznym i wysokiej funkcjonalności. Stwarza to nieskończone wręcz możliwości opracowywania i wprowadzania na rynek zaawansowanych urządzeń pomiarowych i wykonawczych oraz elementów czy systemów (mikrosystemów) do jednorazowego, masowego wykorzystania, zwłaszcza w zastosowaniach medycznych, w sprzęcie militarnym, w przemyśle samochodowym i w przemyśle w ogóle. Masowo produkowana i powszechnie stosowana drukarka atramentowa nie istniałaby, gdyby nie była wyposażona w mikrosystem. Należy spodziewać

się, że w najbliższym czasie pojawi się wiele nowych produktów, które „zmienią świat”. Trzeba również powiedzieć, że wiele mikrosystemów można wytwarzać przy wykorzystaniu technologii o znacznie niższym standardzie niż jest wymagany do produkcji mikroprocesorów, masowej pamięci półprzewodnikowej czy innych układów scalonych. To właśnie z tego powodu sukces – najpierw badawczy, a później rynkowy – w dziedzinie mikrosystemów zależy głównie od pomysłu, talentu naukowego i pasji badawczej, a w mniejszej mierze od wielkości środków finansowych zainwestowanych w nowoczesną linię technologiczną (oczywiście przy zachowaniu pewnego niezbędnego minimum).

Bibliografia

- [1] *Digest of Technical Papers „TRANSDUCERS'05 – The 13th International Conference on Solid-State Sensors, Actuators and Microsystems*. Seoul, Korea, czerwiec 2005, s. 2159.
- [2] Rai-Choudhury P.: *MEMS and MOEMS Technology and Applications*. „SPIE Press”, 2000, s. 520.
- [3] Jachowicz R., Weremczuk J., Tarapata G.: *Transepidermal Water Loss Sensor Based on Fast Dew Point Hygrometer*. „Sensors & Actuators A”, Elsevier Seq., 2005, vol. 123–124, ss. 7–11.



Rysunek 7. Mikrosystem do dynamicznych pomiarów wilgotności i jego medyczne aplikacje: a – struktura mikrosystemu w widoku z góry; b – jego przekrój (różna skala w kierunku x i y); c – tewametr do pomiaru przeznaskórkowej utraty wody; d – system do pomiaru wilgotności w nosie i gardle człowieka w czasie oddychania

- [4] Weremczuk J., Tarapata G., Paczesny D., Jachowicz R.: *Fast Dew Point Hygrometer with Silicon Integrated Detector – Optimization of Dynamic Properties*. „Sensors & Actuators A”, 2006, vol. 132, ss. 195–198.
- [5] Nakamura S.: *MEMS Inertial Sensor toward Higher Accuracy & Multi-axis Sensing*. Proc. 4th IEEE International Conf. on SENSORS, CA., USA, 2005; TOKIMEC, Japan.
- [6] Prof. W. Mokwa, Aachen University, RFN – prywatne konsultacje autora.

Abstract

A short review of microsensors and microsystems has been presented in the report. After some introduction, where differences between smart sensors and microsystems have been defined, few real constructions of

microactuators and microsystems have been discussed and illustrated in the paper. Typical positive features of semiconductor microsensors and microsystems as well have been commented in the aspect of their reliability, quality of technical parameters and cost effectiveness.

Słowa kluczowe:

mikrosystemy, mikroczipy, czujniki inteligentne, mikroświatniki

Profesor Ryszard Sławomir Jachowicz

– specjalista z zakresu badań, projektowania i modelowania czujników pomiarowych i mikrosystemów. Kierownik Zakładu Mikrosystemów i Systemów Pomiarowych na Wydziale Elektroniki i Technik Informatycznych Politechniki Warszawskiej. Odbił staże podoktorskie w Stanach Zjednoczonych, między innymi w National Bureau of Standards (obecnie National Institute of Standards and Technology) i Massachusetts Institute of Technology. Był profesorem wizytującym w Technical University of Vienna (Austria) oraz w Lehigh University (Bethlehem, Stany Zjednoczone). Członek (*Senior Member*) stowarzyszenia IEEE (przewodniczący Polskiej Sekcji IEEE w latach 1997–2002). Fellow Member Institute of Physics (Wielka Brytania). Kierownik około 20 projektów krajowych (1990–2012), 2 projektów badawczych prowadzonych we współpracy z naukowymi ośrodkami w Stanach Zjednoczonych oraz 4 projektów realizowanych w kooperacji z ośrodkami europejskimi. Prezes Polskiego Towarzystwa Techniki Sensorowej (PTTS), członek Komitetu Metrologii i Aparatury Pomiarowej PAN, członek wielu komitetów redakcyjnych czasopism, w tym „Elektroniki”, „Measurement Science and Technology” (IoP, Wielka Brytania) w latach 2002–2009 oraz „Journal of Measurement Science and Instrumentation” (Chiny). Recenzent wielu czasopism międzynarodowych i krajowych, recenzent ponad 250 projektów badawczych. Autor lub współautor 15 patentów oraz ponad 190 artykułów i referatów konferencyjnych (z czego więcej niż połowa została opublikowana za granicą), w tym 5 wykładów inauguracyjnych konferencji międzynarodowych i 16 referatów zaproszonych.



Dokąd zmierza Świat i Polska

Na podstawie odczytu wygłoszonego w dniu 23 kwietnia 2009 roku

Grzegorz W. Kołodko

Akademia Leona Koźmińskiego

131

Kultura a rozwój*

Kiedy spojrzeć na dzieje rozwoju i zastój gospodarczego, to widać, że historia jednego uczy nas nader wyraziście: decyduje kultura. Podkreślał to już Max Weber (1864–1920), a miniony wiek jeszcze dobitniej wykazał słuszność takiej obserwacji. Nie mam żadnych wątpliwości, że po upływie kolejnych, wieku XXII i następnych, będzie ona jeszcze bardziej bezsporna.

Ekonomiści nie mogą lubić takiej konstatacji. Nie tylko dlatego, że kultura to pojęcie miękkie, niedające się precyzyjnie mierzyć, a zarazem wielce pojemne, w które sporo można włożyć. Przede wszystkim dlatego, że sugerować to może zbyt daleko posunięty determinizm. Kultura co prawda nie jest dana raz na zawsze, ale jednak na długo. Skoro to kultura przesądzać miała o rozwoju, cóż zatem można zrobić? Niewiele. Ograniczeni jesteśmy gorsetem spuścizny z epok poprzednich. Myślimy i działamy pod wpływem obciążeń religijnych, rasowych, narodowościowych, mentalnych. I na dodatek śliskie to tematy, o których nie zawsze wygodnie jest mówić, zwłaszcza gdy trzeba zachować polityczną poprawność, o którą troszczą się całe armie obłudników.

Ekonomia to nauka o zmianach. Nawet skrajnie liberalni ekonomiści lubią przy tych zmianach manipulować. Na krótką metę poprzez zmiany rozmaitych para-

metrów ekonomicznych — a to stopy procentowej, a to podatków, a to kursów walutowych — i na długą, gdy poprzez reformowanie instytucji pożądanym procesom chcą sprzyjać, niechciane usiłują blokować. A kulturą jako produktem historii manipulować łatwo się nie da, gdyż ewoluuje ona powoli, w wymiarach pokoleniowych, i doraźnie niewiele można zmienić.

To bardziej kultura manipuluje nami niż my nią. Podczas gdy niektórzy antropolodzy uważają, że pewnych kulturowych cech w ogóle zmienić się nie da, inni wskazują na cechy wysrane ponoć wraz z mlekiem matki, przenoszone z pokolenia na pokolenie. Może akurat nie tą drogą, ale bez wątpienia ciągną się one przez całe generacje i niekiedy na tym polu więcej jest kontynuacji aniżeli zmiany. Gdy są to cechy pozytywne, to dobrze, gdy negatywne, źle.

Skoro tak jest, skoro jesteśmy spętani uwarunkowaniami kulturowymi, to jak wpływać na rozwój gospodarczy? Może idzie on po prostu koleinami przez kulturę wyznaczonymi i nic tu już więcej zrobić nie można? Może zaiste to kultura, mocno osadzona w historii i tradycji, wyznacza ścieżki rozwojowe i narzuca narodom ramy, poza które nie za bardzo mogą wykroczyć? Bynajmniej.

Łatwiej pojąć znaczenie kultury w rozwoju gospodarki, gdy przyrównać ją do roli charakteru w rozwoju człowieka. Inny miał Kain, inny Abel; inny Dr Jekyll,

* Fragment VIII rozdziału („Instytucje, polityka i kultura a zastój i rozwój”) książki pt. „Wędrujący świat”, Prószyński i S-ka, Warszawa 2008, s. 440 (www.wedrujacyswiat.pl)

a inny Mr Hyde. Inny charakter (czyli kulturę) ma nordycka społeczna gospodarka rynkowa, inny anglosaska gospodarka neoliberalnego kapitalizmu (i odpowiadające jej społeczeństwa oraz ich elity).

O narodach — albo o społeczeństwach — mówi się podobnie jak o ludziach. A ludzie są różni. Podczas gdy są tacy, co spędzają mnóstwo czasu przed telewizorem albo bezmyślnie wertują gazety, inni w tym czasie wolą szydełkować albo ogródek uprawiać. Jedni śpią długo, inni długo pracują. Ktoś woli majsterkować, ktoś inny plotkować. Wszystkie te zachowania mają swoje kulturowe uwarunkowania i ekonomiczne konsekwencje. Niekoniecznie są one ujmowane w statystykach dochodu narodowego, aczkolwiek wpływają nie tylko na rzeczywisty poziom konsumpcji, ale i na satysfakcję z życia. Kluczowe są tu pragmatyczność i kreatywność. Wtedy też, mając pieniądze, można mieć i fantazję. Gdy zaś ma się tylko wiele fantazji, trudno mieć pieniądze.

Także jest i z narodami. One też mają na swój sposób do wyboru, czy — powiedzmy — doksztatcać się wieczorami, czy pić wódkę. Społeczeństwom, jak ludzким charakterom, przypisuje się cechy. Jednym razem głośno się, że są pracowite i zapobiegliwe, innym, że się lenią i brakuje im przeczności. Jednym przypisuje się zmysł przedsiębiorczości, podczas gdy innym rozlazłość. Są takie, co słyną z porządku, ale i takie, którym zarzuca się bałaganiarstwo. Dostownie i w przenośni. Jedne bywają gościnne, inne przez lata obcego nie wpuszczają za swoje progi. Bywają narody i grupy etniczne słynące z wojowniczości i takie, które zawsze bardziej spokój i pokój miłowały. Szkotów kojarzy się ze skąpstwem, Gruzynów z rozrzutnością. Niemcom przypisuje się wielkie poczucie dyscypliny, ale za to zupełny brak poczucia humoru, dokładnie odwrotnie niż Polakom, co może w jakiejś mierze wyjaśniać, dlaczego u tych pierwszych drogi są znakomite, ale za to u tych drugich jest dużo weselej. Zrozumiałe, że jest tu też wiele stereotypów i często różne narody — zwłaszcza sąsiedzkie — oskarżają się o te same przywary.

Człowiek od najdawniejszych czasów jest uwikłany w relacje międzykulturowe. Gdy wychodzi się za oplotki własnej zagrody, jest to wydarzenie kulturowe, spotyka się bowiem innych. Tak rozumiana interkulturowość ma swój wymiar mikro, poczynając dostownie od tej zagrody, jak i makro, a więc na szczeblu społeczeństwa, narodu, państwa. Ma też swój wymiar transnarodowy i globalny. Nic tak nie zmienia charakterów — ludzi, grup społecznych, narodów — jak stosunki międzykulturowe.

Na charakter zawsze składają się cechy pozytywne i negatywne w swoich przeróżnych odcieniach. Podczas gdy jedne są wrodzone, innych się nabywa. Wiadomo, niedaleko pada jabłko od jabłoni. Ale wiadomo też, że czym skorupka za młodu nasiąknie, tym na starość trąci. Gdzieś tak mniej więcej w połowie nasze cechy są zdeterminowane genetycznie, a więc niejako przetransportowane z pokolenia na pokolenie (spadające jabłko), w połowie kształtują się w trakcie życia, zwłaszcza w jego wcześniejszych fazach (nasiąkająca skorupka).

Jesteśmy wychowywani w rodzinnych domach, obracamy się w rozmaitych środowiskach sąsiedzkich, szkolnych, uniwersyteckich, zawodowych, politycznych, towarzyskich. Nabieramy nawyków sprzyjających ewolucji i dojrzewaniu osobowości, ale — niestety — i takich, którą ją psują. Wciąż jeszcze zaledwie w przedszkolach i szkołach są laureaci Nagrody Nobla, które będą przyznawane w drugiej połowie XXI wieku. Także ekonomiści, którzy nawet jeszcze sami nie wiedzą, że ekonomistami zostaną. Tamże są też najwięksi zbrodniarze, którzy się w tym samym czasie objawiają. Teraz jeszcze mogą być nawet grzecznymi dziećmi i przykładowymi uczniami. I o ile chyba nie da się z tych drugich procesem wychowawczym odpowiednio ukierunkowanym uczynić tych pierwszych, to można wiele zrobić, aby więcej było ludzi światłych i uczciwych niż głupków i szubrawców. Przecież poprzez proces edukacji i wychowania — a także coraz bardziej w wyniku społecznych interakcji, w których niebywale wzrosła pozycja środków masowego przekazu i internetu — można znakomicie na charaktery wpływać. I próbują to czynić rodzice, nauczyciele, religie, media, rozmaite organizacje. Z rozmaitymi skutkami.

Trochę podobnie jest z gospodarką. Ona też ma różne „charaktery”, cechy „wrodzone” i „nabyte”, choć nie potrafimy powiedzieć, jakie dokładnie są ich wzajemne proporcje. Uważam, że współcześnie zdecydowanie większą wagę mają te nabyte, ale jest z tym różnie w różnych społeczeństwach. Społeczeństwa zaiste są jak ludzie. Jedne uczą się szybciej, inne wolniej. Jedne przyswajają cechy bardziej postępowe, inne nabierają znamion bardziej konserwatywnych. Bywają i tak patologiczne, że przez pokolenia albo i wieki całe niewiele dobrego potrafią się nauczyć.

Wiele paraleli można by nakreślić. Weźmy Chile i Czechy. Ich porównywanie lat temu dwadzieścia pięć od razu wskazywało istotne różnice. Teraz widać więcej podobieństw i cech wspólnych, bo oba kraje

nauczyły się sporo na temat prowadzenia sensownej działalności ekonomicznej w zglobalizowanej gospodarce. Podobnie z Argentyną i Węgrami. Starczyło jedno pokolenie. Zaledwie albo aż. Ćwierć wieku temu były to istotnie odmienne systemy i odmienna kultura ekonomiczna, dzisiaj można by wskazać sporo analogii strukturalnych i instytucjonalnych. Już nie tylko zewnętrznych podobieństw na ulicach i polach, zbiegiem okoliczności powstałych wcześniej — co znakomicie wykorzystał Alan Parker, kręcąc *Evitę* zarówno w Buenos Aires, jak i w Budapeszcie, na argentyńskich pampasach i w węgierskiej puszczy — ale i wewnętrznych, tych objaśniających sposób funkcjonowania i rozwoju gospodarki.

Cechy wrodzone, będące funkcją siły kulturowych tradycji oraz mechanizmu instytucjonalnej inercji i kroczenia wspomnianą wcześniej ścieżką uzależnienia, mają tendencję do trwania. Czasami zbyt długo, bo szkodzą, innym razem akuratnie, bo sprzyjają postępowi naukowo-technicznemu i przedsiębiorczości. Notabene, tęsknota za „nową *Evitą*” w Argentynie jest niepomernie większa aniżeli za nowym królem na Węgrzech. Do tego stopnia, że Cristina Fernández de Kirchner bez trudu wygrała wybory prezydenckie jesienią 2007 roku już w pierwszej turze. Społeczeństwa bowiem mają — jak ludzie — swoje sentymenty i resentymenty. Z niektórych nie potrafią wydobyć się przez lata całe, nawet gdy już wiedzą, że nie sprzyjają one rozwojowi. Trochę podobnie jak z obżarstwem. Powinno się zaniechać i rzucić, ale się nie potrafi albo najprościej nie chce.

Nowych cech nabywa się, ucząc się podczas procesu budowania nowych instytucji oraz uprawiania aktywnej strategii i polityki rozwoju. I tu właśnie jest szczególne pole do popisu. Podobnie jak w przypadku człowieka, u którego można wykreować pożądane społecznie cechy i kształtować jego dobry indywidualny charakter, w gospodarce również można tworzyć pozytywne cechy, nadając jej prorozwojowy charakter. Jeśli obciążenia genetyczne przeszkadzają, to trzeba się jeszcze bardziej wysilać w edukacji i wychowywaniu, a nie narzekać na dopust boży albo, bliżej, na rodziców. Jeśli spuścizna kulturowa gospodarczej ekspansji nie pomaga, to też trzeba więcej wysiłku w stymulowaniu postępu ekonomicznego, zamiast psioczyć na dopust historii albo, bliżej, na poprzednie ekipy gospodarcze.

Przy szczęśliwym zbiegu okoliczności nie trzeba na to całych pokoleń. I, niestety, odwrotnie. Można bo -

wiem szanse na wykreowanie takiego „dobrego charakteru” gospodarki, a więc jej prorozwojowości, łatwo zaprzepaścić. Wystarczy tylko porównać po sąsiedzku Chile z Argentyną, Kostarykę z Hondurasem, Dominikanę z Haiti, Botswanę z Zimbabwe, Senegal z Gwineą, Polskę z Ukrainą, Słowenię z Chorwacją, Katar z Bahrajnem, Malezję z Filipinami, Samoa z Fidżi. To nie warunki naturalne (podobne) ani nie położenie geograficzne (też podobne) spowodowały, że pierwsze w tych parach kraje potrafiły podczas ostatniego ćwierćwiecza wyraźnie zdystansować w rozwoju te drugie.

Kultura — podobnie jak charaktery — bardzo silnie wiąże się z inspiracją i motywacją. Jest takie stare rosyjskie powiedzonko: z kawatka drewna można równie dobrze zrobić ikonę, jak i patkę. Otóż to. Wychodząc z takich samych czy podobnych środków, można mieć różne inspiracje i odmienne motywacje. To one mogą decydować o tym, jaki jest produkt końcowy zaangażowania w aktywność produkcyjną. Oczywiście, przyjmując, że takie same kwalifikacje są potrzebne do namalowania ikony i do wyciosania patki, co już prawdą nie jest. Ale w ludziach drzemią różne umiejętności, a w społeczeństwie wiele z nich występuje obok siebie. Można więc robić albo patki, albo ikony. Albo też jedno i drugie w rozmaitych proporcjach. Może też być i tak, że ci z kwalifikacjami do robienia patek dorywają się do władzy (niekoniecznie siłą, za pomocą tych patek, bo czasami również poprzez demokratyczne wybory) i tłamszą tych zdolnych do dostarczania ikon. Wtedy na rynku jest za dużo patek i za mało ikon. Ekonomista powiedziałby, że występuje nierównowaga rynkowa. Szybko prowadzić musi ona także do nierównowagi społecznej i politycznej, a wszystkie trzy typy nierównowagi obracają się przeciwko wzrostowi gospodarczemu.

Kluczowa jest tutaj rola motywacji. Aby coś robić — wymyślać, projektować, organizować, zarządzać, produkować, dzielić, transportować, magazynować, sprzedawać, konsumować — trzeba chcieć. Oddolne pragnienie podejmowania jakiegokolwiek ekonomicznej działalności jest dużo lepszym podłożem motywacji niż przymus narzucany z zewnątrz. W gospodarce mamy do czynienia z przymusem ekonomicznym, ale tylko w określonych granicach. Warto jednak tu podkreślić, że to właśnie liberalna teoria kapitalizmu (liberalna, nie jej neoliberalna dewiacja) najtrafniej formułuje zasady ekonomicznej gry, gdyż lepiej niż jakkolwiek inna współgra z ludzką naturą, w tym także z właściwymi jej psychologicznymi mechanizmami motywacji. I je eksploatuje. Kapitalizm wolnego rynku nie poszedł ani na

manowce prób tworzenia „nowego człowieka”, jak różne utopie od komunistycznej poprzez islamską do społecznej nauki Kościoła katolickiego, lecz w skuteczny niezwykle (a może jak najbardziej zwykle?) sposób wykorzystuje naturalne ludzkie cechy. Oczywiście łącznie z tymi najgorszymi, jak egocentryzm, chciwość, zawiść, zachłanność, krótkowzroczność, a nie tylko tymi pożądanymi, jak inicjatywa, pracowitość, zaradność, zapobiegliwość, przezorność, grupowa solidarność.

Największe osiągnięcia i sukcesy w zarządzaniu mikroekonomicznym (na szczeblu firm) i w strategii rozwoju społeczno-gospodarczego (na szczeblu państw) nie brały się z klasycznego przymusu ekonomicznego. Ani Henry Ford, rewolucjonizując przemysł i umasawiając produkcję samochodów na początku XX wieku, ani Mohamad Mahathir, przywódca Malezji, w jego końcu nie byli przymuszeni przez konkurencję do tego, co potrafili zrobić. Mieli motywację, która płynęła z osobistych charakterów i otaczającej ich społecznej kultury. Musiało się dla nich szczęśliwie nalożyć na to jeszcze wiele innych czynników, ale bez odpowiedniej motywacji wszystkie one razem wzięte nie wystarczyłyby.

Bez motywacji bowiem nie ma przedsiębiorczości i sensownej działalności. Nie tylko gospodarczej. O ile jednak w sztuce i nauce może ona płynąć z subiektywnego poczucia wartości czy też z samoistnego dążenia do piękna i prawdy, o tyle w gospodarce powinna być skorelowana z racjonalnością i zgodna z prakseologicznymi zasadami gospodarowania. Do tego sama kultura nie starcza. Historia pokazała nam przecież, że przy określonej kulturze trzeba jeszcze odpowiednio ustawić instytucje albo poprawić politykę, aby z fazy застоju przejść do etapu rozwoju. Albo też istotnie przyspieszyć jego tempo lub po prostu tylko je utrzymać. Widzimy to ostatnio najlepiej w krajach Azji Południowo-Wschodniej. Także w gospodarkach posocjalistycznej transformacji w Europie Środkowo-Wschodniej. Ale już na mniejszą skalę — również ze względów kulturowych — w poradzieckich republikach Azji Środkowej, w Ameryce Łacińskiej i na Karaibach. W najmniejszym stopniu na Bliskim Wschodzie, w Afryce i w Oceanii. Przedsiębiorczość, poza indywidualnymi inklinacjami ludzi nią się parających, jest funkcją nauki, techniki i polityki. Ta ostatnia może mieć szczególne znaczenie dla pobudzania przedsiębiorczości. Z jednej strony, jak widzieliśmy to w naszej krótkiej wędrówce poprzez dzieje, może ona przedsiębiorczości szkodzić i ją tłumić. Wtedy mamy zastój. Ale, z drugiej strony, nauka i tech-

nika może nie być w stanie samodzielnie przebić się do sfery produkcji. Wtedy trzeba jej w tym pomóc. Nie ma przy tym lepszego sposobu stymulacji przedsiębiorczości jak oddziaływanie, zwłaszcza pośrednie, na motywacje poprzez wykorzystywanie instrumentów polityki gospodarczej. Gdy przy ich odpowiednim ustawieniu jakieś zadanie będzie opłacalne, to i będzie przedsięwzięte.

Przedsiębiorczość jest zarówno ludzką cechą (aczkolwiek nie wszystkim daną), jak i umiejętnością. Wrodzone talenty zawsze się tu przydają, ale umiejętności można się uczyć i nauczyć. W jakimś stopniu także przedsiębiorczości. Jest to jednak aż tak specyficzna umiejętność, że by ją osiągnąć, nie wystarczą studia w najlepszych nawet szkołach biznesu. Niezbędna jest nauka poprzez działanie, a więc poprzez doświadczenie.

Problem sprowadza się zatem do wzajemnego oddziaływania na siebie kultury, instytucji i polityki. Gdy patrzymy na wyzwania rozwojowe w bardzo długiej perspektywie czasowej, to widzimy, że właśnie na polu sprzężeń pomiędzy tymi trzema wielkimi kategoriami rozgrywają się batalie o przyszłość. I choć bezsprzecznie protestancka kultura mieszkańców Beneluksu czy nordyckie cechy mieszkańców Skandynawii sprzyjać będą rozwojowi bardziej niż islamska kultura Arabii czy Sahelu — podobnie jak to było przez ostatnie kilkadziesiąt lat — to również w tym drugim przypadku można wyobrazić sobie takie ustawienie instytucji i polityki, i starać się tak je wykorzystywać, aby ludność tych regionów też potrafiła skutecznie rozwijać swoje gospodarki. Z wszystkimi tego pozytywnymi następstwami dla światowej gospodarki i społeczności. Nie jest to przecież li tylko ich sprawa, skoro konsekwencje — dla migracji, dla bezpieczeństwa, dla wzrostu — są także nasze.

Jakże wiele przecież potrafiły osiągnąć niektóre niearabskie kraje muzułmańskie, jak na przykład Indonezja, a zwłaszcza te, gdzie występuje wielokulturowe bogactwo, jak na przykład Malezja. Tam współistnieją różne narodowości i religie. Obok Malajów wielu jest Chińczyków i Hindusów, trochę też Europejczyków. Obok islamu kultywowane jest chrześcijaństwo, a na szerszą skalę taoizm i buddyzm. A filozofia tych dwu ostatnich religii szczególnie sprzyja rozwojowi. Ale już nieopodal, w indonezyjskiej prowincji Aceh, taka mieszanina się nie sprawdza. Ludzie tam sami żartują, że Aceh to skrót od pierwszych liter Asian – Chinese – European – Hindu. Jednakże tam taka kulturowa kompozycja nie tylko nie starcza do rozwoju, lecz wręcz jest

zarzewiem wciąż niezażegnanych konfliktów. A to dlatego, że brakuje innych czynników niezbędnych do przetoczenia się pozytywnej masy krytycznej, zwłaszcza trwałego pokoju i sensownego — z wizją i bez iluzji — przywództwa politycznego. Niektórzy czasami mają do szczęścia, że takie właśnie im się przytrafia. W dużym stopniu wskutek szczęśliwego zbiegu okoliczności.

Tak oto — abstrahując na chwilę od konfliktowych interesów — jeśli wskutek jakichś doktrynalnych zahamowań o kulturowym podłożu blokuje się budowanie prorozwojowo zorientowanych instytucji i uniemożliwia realizację sensownej polityki, wtedy taka kultura skazuje narody na brnięcie przez gospodarczą stagnację. Można tę konstatację przyjąć do wiadomości, ale to bynajmniej nie zwalnia z poszukiwania teorii rozwoju, która może okazać się pomocna w przezwyciężaniu takiego syndromu niemocy i w wyrwaniu się z zastoju na gruncie pragmatycznym.

Najogólniej biorąc, kultura oznacza system wartości i płynących z nich ludzkich zachowań. O losach rozwoju gospodarczego decyduje kultura nie sama w sobie, ale w konkretnym naturalnym, społecznym i technicznym otoczeniu, z czasem zwrotnie na nie oddziałując. Bardziej na to społeczne i techniczne, mniej na to naturalne. Nie sama z siebie, ale w powiązaniu z wieloma innymi czynnikami, które poprzez ułożenie się w konkretną wiązkę sprzyjają albo rozwojowi, albo zastojowi. Jednak to kultura przesądza na długą metę, kto jest wygrany, a kto przegrywa, kto się wzbogaca, a kto nędziej, która gospodarka rozkwita, a która więdnie, jakie narody prosperują, a jakie wegetują. A to dlatego, że w jej ramach zbiegają się inne okoliczności o współdecydującym dla ludzkiej przedsiębiorczości i wydajności znaczeniu. Raz kultura im sprzyja, kiedy indziej przeszkadza, różne bowiem sama ma oblicza i na różne dodatkowe czynniki się natyka.

Gdy jednak rozważa się przyszłość i sposoby pożądanego z gospodarczego punktu widzenia jej kształtowania, o ogólnych kulturowych uwarunkowaniach trzeba pamiętać, ale zejść należy na grunt zdecydowanie bardziej konkretny. Zawsze działa się w określonym otoczeniu kulturowym, ale też i zawsze decyduje podej-

muje się w konkretnym otoczeniu fizycznym i ekonomicznym, co modele teoretyczne także konkretnie muszą brać pod uwagę. Jednakże pomimo nieustannych studiów i wielu kroków naprzód, żadna z dotychczasowych teorii wzrostu i rozwoju nie jest zadowalająca. Nie wyjaśniają one bowiem wszystkich aspektów zastoju i rozwoju i, analizując poszczególne przypadki, nieustannie napotykały trudności interpretacyjne lub też widzimy, że rzeczywistość nie mieści się w modelowych ujęciach bądź niekiedy stoi wobec nich w sprzeczności. Teoretycznie coś powinno rosnąć, a tu akurat spada. Coś zgodnie z teoretycznym modelem powinno przepływać z krajów bogatych do biednych, a dzieje się odwrotnie. Kryzys powinien się pojawić, a jakoś mu nie spieszo. Albo, co częstsze, nie powinien nas zaskoczyć, a właśnie atakuje.

Dlaczego tak się dzieje? I co można na to poradzić? Odpowiedź znajdziemy w prezentowanym tu podejściu do teoretycznej interpretacji procesów wzrostu i rozwoju społeczno-gospodarczego. Najwyższy czas porzucić zgraną już jak stara płyta utopijną teorię neoliberalizmu i zaproponować nowy paradygmat. Z jednej strony, jest to koincydencji teoria rozwoju. Odnosi się ona do warstwy opisowej (deskryptywnej) ekonomii. Z drugiej strony, jest to opierający się na tej teorii nowy pragmatyzm, który odnosi się do warstwy postulatywnej (normatywnej). Nawiązując do wcześniejszych uwag metodologicznych, teoria ta w ujęciu deskryptywnym wyjaśnia nam, jak sprawy się mają i jakie mechanizmy rządzą biegiem procesów gospodarczych, a w ujęciu normatywnym mówi nam, co i jak czynić, aby miały się lepiej.

Otóż przede wszystkim trzeba odejść od wszelkiego dogmatyzmu, gdyż ten ciąży na twórczym myśleniu nie mniej niż ignorancja. A czasami nawet bardziej. Dogmatyzm — powtórzmy — dobry jest w religii, ale nie w nauce. Tu niezbędny jest nonkonformizm, nieustanny pęd do przodu, zmysł godny reformacji i reform, a nie tradycjonalizmu i zachowawczości.

Pieniądze kultury nie dają. To kultura, sama wiele warta, daje pieniądze.

Profesor Grzegorz Kołodko specjalista z zakresu nauk ekonomicznych, dyrektor Centrum Badawczego Transformacji, Integracji i Globalizacji TIGER w Akademii Leona Koźmińskiego w Warszawie. Wicepremier i minister finansów w latach 1994–1997 i 2002–2003. Wykładał na prestiżowych uniwersytetach w Stanach Zjednoczonych: Yale, UCLA i Rochester, New York. Był ekspertem w Organizacji Narodów Zjednoczonych, Banku Światowym czy też Międzynarodowym Funduszu Walutowym. Członek Europejskiej Akademii Nauki, Sztuki i Literatury. Obszary zainteresowań naukowych to: polityka rozwoju, długofalowe zmiany systemowe oraz teoretyczne i praktyczne problemy globalizacji i transformacji posocjalistycznej. Autor i redaktor naukowy 40 książek oraz ponad 400 artykułów i referatów naukowych. Odznaczony przez Prezydenta RP Krzyżem Komandorskim Orderu Odrodzenia Polski.

Counting Single Electrons Using a Carbon Nanotube Field-Effect Transistor

Na podstawie odczytu wygłoszonego w dniu 16 października 2008 roku

Adrian Bachtold

Catalau Institute of Technology, Barcelona

A novel scheme based on a carbon nanotube field-effect transistor allows the detection of single-electron events at relatively high temperature in a reliable way, demonstrating the ultimate capabilities of the carbon nanotubes as single charge sensors.

The detection and manipulation of individual electron charges are one of the hot topics in nanoscale electronics due to the promising applications in ultra low dissipative devices and information processing in molecular circuits. Single-electron detectors consisting of micro-fabricated devices in metal or semiconducting material have been already reported in literature but with the drawback of being operable at millikelvin temperatures [1]. On the other hand, single electron detection carbon nanotubes has been resolved for electrons hopping onto defects randomly trapped in a silicon oxide layer [2]. However, these single-electron processes are poorly controlled and the controlled detection by a nanotube transistor of single electrons on a nearby nanosystem still needed to be demonstrated.

The present work shows for the first time a real-time detection of the transfer of single electrons between a gold nanoparticle and the carbon nanotube with an operation temperature close to 150 K, three orders of magnitude higher than the one of previous detectors. This work also demonstrates that nanotube transistors can probe electrons on other molecular systems, which cannot do the other microfabricated detectors. Moreover, the single-electron detection measurements allow

for the full electron characterization of the system circuit (Fig. 1 b).

The carbon nanotube transistors were fabricated by means of standard nanofabrication techniques with nanotubes grown under chemical vapor deposition on a Si/SiO₂ substrate. Gold nanoparticles were deposited onto the wafer from an aqueous suspension and positioned on top of the nanotube by atomic force microscopy (AFM) manipulation (Fig. 1 a).

The transfer of single electrons into the Au nanoparticle can be detected by measuring the conductance G_{tube} of the nanotube while sweeping the gate voltage V_G (Fig. 2 a), as the tube conductance is extremely sensitive to the presence of electric charges. As V_G is swept, the conductance is turned off as for typical *p*-doped semiconducting SWNTs. However, we additionally observe abrupt conductance jumps (vertical blue bars) that indicate discrete electron transfers from the nanotube into the particle. Each extra electron in the particle changes the electrostatic potential in the particle and, in turn, the charge density in the nanotube, which shifts the conductance G_{tube} horizontally in V_G .

In order to unequivocally confirm that these discrete jumps correspond to single electron events, we have performed repetitive scans in V_G (Fig. 2 b). A collection of curves is obtained which are periodically spaced in gate voltage with a period of about $\Delta V_G^{shift} = 60$ mV. This periodicity suggests that adjacent curves differ by one

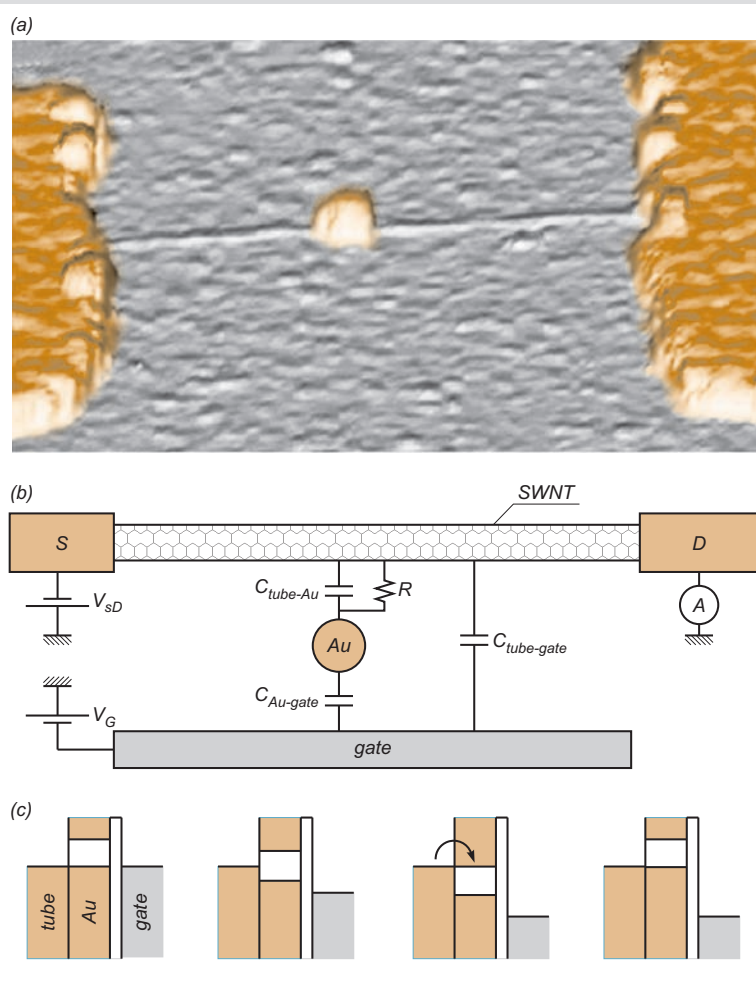


Figure 1. (a) Atomic force microscopy image of the device geometry. (b) Schematic of the measurement setup. (c) Schematics of the potentials in the nanotube and the particle as the gate potential is swept down. Each time an empty energy level of the particle matches the electrochemical potential of the tube, an electron is transferred onto the particle, which is detected by the nanotube transistor.

Reprinted with permission [3]; Copyright (2007) American Chemical Society

electron in the Au particle and in turn that the observed jumps correspond to transfers of single electrons.

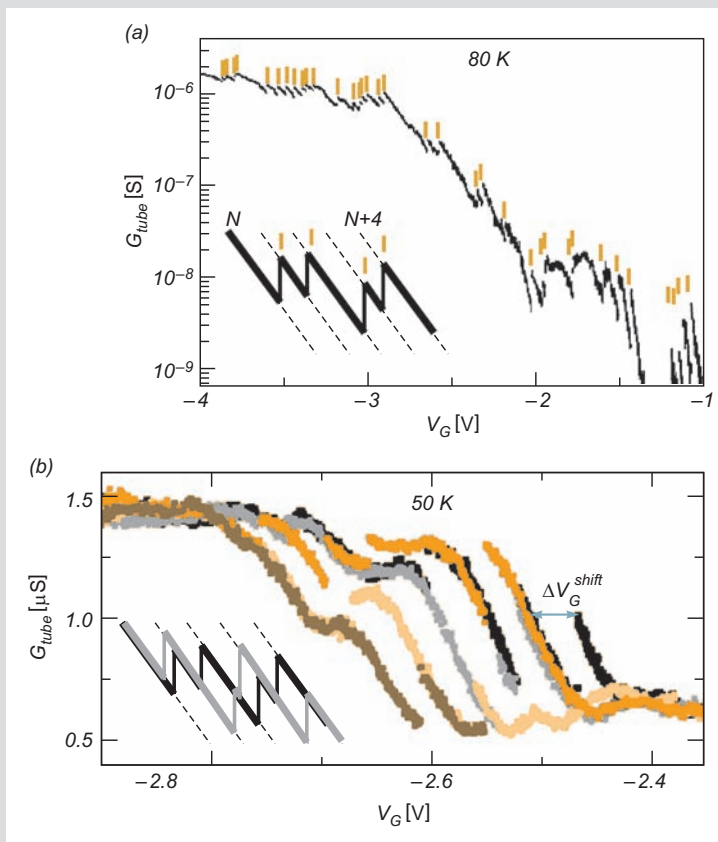
The detection of single electrons allows the characterization of Au nanoparticle electrical properties by looking at the time dependence of the electron transfer versus temperature. For instance, if the gate voltage is set at a fixed value while measuring the tube conductance at 50 K (Fig. 3), the tube conductance fluctuates between two values on a time scale of several hundred seconds corresponding to an electron going back and forth into the Au particle due to thermal excitation and changing the number of electrons between N and $N + 1$. This kind of measurements is very useful, since the fluctuations of N due to thermal excitation can provide information on the energy separation be-

tween electron states of the Au particle. For further information, see reference [3].

In conclusion we have demonstrated well-controlled single electron detection in a simple, well-defined highly resistive molecular circuit consisting of a carbon nanotube transistor and a gold nanoparticle. The nanotube transistors are shown to be excellent detectors of single electrons at high operation temperatures. We have exploited such single electron counting and its low transfer rate to electrically characterize the Au particle. Single electron counting with nanotubes offers great promise for the future studies on organic molecules, biomolecules, or semiconducting particles which more often are highly resistive and it is not possible to pass a current measurable with conventional electronics.

Figure 2. Detection of single electrons. **(a)** Tube conductance as the gate voltage is swept from -4 to -1 V. Vertical blue bars indicate conductance jumps. The inset shows the relation between $G_{tube}(V_G)$ and the number of electrons in the Au particle. **(b)** Tube conductance as a function of V_G in a smaller range. Each color corresponds to a different scan. The inset shows two traces of $G_{tube}(V_G)$ in black and gray; jumps appear at different V_G values as a result of the stochastic nature of the electron transfer.

Reprinted with permission [3]; Copyright (2007) American Chemical Society



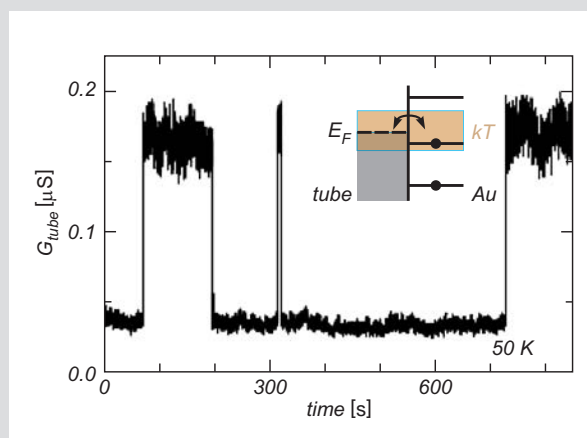
For example, single electron photoelectric effects can be investigated in CdSe particles as well as charge transfer in biomolecules involved in photosynthesis and respiration activities.

References

- [1] T. Fujisawa, T. Hayashi, R. Tomita and Y. Hirayama, *Bidirectional Counting of Single Electrons*, *Science* **312** (5780), pp. 1634–1636, 2006. doi: 10.1126/science.1126788

Figure 3. Fluctuations of the electron number due to thermal excitation. Tube conductance as a function of time at 50 K for $V_G = -1.35$ V. The conductance experiences two levels at 50 K. We attribute the extra level at 50 K to the electrochemical potential of the tube that matches the center of the Coulomb gap. The insets show the energy levels in the tube and in the Au particle for different numbers N of electrons. The thermal energy is shown in blue.

Reprinted with permission [3]; Copyright (2007) American Chemical Society



- [2] T. Durkop, B.M. Kim, M.S. Fuhrer, *Properties and Applications of High-mobility Semiconducting Nanotubes*, **J. Phys.: Condens. Matter.**, **16** (18), pp. R553–R580, 2004. doi: 10.1088/0953-8984/16/18/R01
- [3] A. Gruneis, M.J. Esplandiú, D. García-Sánchez and A. Bachtold, *Detecting Individual Electrons Using a Carbon Nanotube Field-Effect Transistor*, **Nano Lett.**, **7** (12), pp 3766–3769, 2007. doi:10.1021/nl072243w

Profesor Adrian Bachtold pracownik naukowy i szef grupy badawczej Quantum Nano-Electronics w Katalońskim Instytucie Nanotechnologii, Barcelona, Hiszpania. Pracował w Narodowym Centrum Badań Naukowych w Ecole Normale Supérieure w Paryżu. Absolwent Uniwersytetu Bazylejskiego (Szwajcaria), Uniwersytetu Kalifornijskiego w Berkley (Stany Zjednoczone) oraz Uniwersytetu Technicznego w Delft (Holandia). Posiada prawie 5000 cytowań. Laureat brązowego medalu z 2004 roku przyznanego przez Narodowe Centrum Badań Naukowych (CNRS), stypendysta w ramach programu EURYI (Europejska Nagroda dla Młodego Naukowca) oraz laureat konkursu na stypendium naukowe przyznawane przez Europejską Radę ds. Badań (ERC).

Wiara, technika i medycyna

Na podstawie odczytu wygłoszonego w dniu 11 grudnia 2008 roku

Abp Henryk Hoser SAC

141

Trynomiczne sformułowanie tytułu niniejszego wykładu sygnalizuje w założeniu trzy rzeczywistości i relacje między nimi zachodzące. Dla niektórych te terminy mogą wydawać się jasne i kluczowe, kiedy inni odczytują je jako dwuznaczne lub wieloznaczne, o niejasnym zakresie znaczeniowym. A skoro żyjemy w czasach zawirowań semantycznych, nierzadko podbarwionych ideologicznie, będzie rzeczą użyteczną sprecyzować, co mam na myśli odwołując się do tytułowych rzeczowników.

Wiarę możemy określić jako jeden z biegunów relacji między dwoma podmiotami – człowiekiem i Bogiem. Relacja ta powstaje i utrwała się dzięki temu, że człowiek – *capax Dei* – jest zdolny do poszukiwania i poznania Boga, co należy do jego specyficznej ludzkiej natury.

Historia nie zna żadnej cywilizacji i pierwotnej kultury wyzbytych z religijności i wierzeń oraz ich przejawów: modlitw, ofiar, kultów, medytacji, a często pism sakralnych i sakralnej ikonografii. Człowiek jest więc z natury istotą religijną, tak jak do jego natury należy zdolność tworzenia kultur. Innymi słowy, religia i kultura, często ze sobą powiązane, są wyznacznikami ludzkiej natury jako jej powszechnie spotykane formy wyrazu. W tej religijnej naturze jest zakodowane poszukiwanie Boga również w sposób racjonalny, przez badanie i pogłębianie istoty otaczającego nas świata w aspektach filozoficznych, jak: ruch i stawanie się, przygodność zjawisk, porządek świata wyrażony w prawach naukowych. Znajduje tu miejsce również estety-

ka z odkrywaniem i ukazywaniem piękna. Do ludzkiej natury należy również zdolność do tzw. uczuciowości wyższej i uogólniającego myślenia. Ta zdolność determinuje jego działania wolitywne, np. z motywów przekonania i dążenia do ideałów.

W świat i w kosmos, będące również przedmiotem badań nauk przyrodniczych i ścisłych, Bóg jest niejako wpisany, zakodowany – jako początek wszystkiego i podstawowy *Logos* istnienia. Odnajdujemy ów *Logos* we wszystkich „logiach” szeroko pojętych nauk.

Również sam człowiek jest drogą racjonalną, prowadzącą do Boga. Biologicznie podobny do zwierząt, biologicznie będący ssakiem, jest jednocześnie od świata zwierząt radykalnie różny i nieporównywalny. W swej dynamice życiowej niesformatowany somatycznie, lecz ideologicznie, w tym znaczeniu, że żyje ideałami i ideałami, prowadzi życie duchowe, do życia potrzebuje sensu. Istnienie bezsensowne nie jest w pełni ludzkie. Człowiek myślący abstrakcyjnie ma też świadomość samego siebie – ma refleksyjną tożsamość.

Ta wyjątkowość człowieka streszcza się w pojęciu osoby. Jako osoba człowiek ma cechy boskie i dlatego wiemy, że jest stworzony na obraz i podobieństwo Boże. Nawet starożytny Olimp był zamieszkały przez bogów człekokształtnych, mających określoną osobowość.

Pośrednie i domyślne poznanie Boga poprzez fenomen człowieka i rzeczywistość otaczającego go mikro i makro kosmosu, w poznaniu naturalnym i czysto racjonalnym jest niedokładne, rozmyte i mgliste. Jest

określany jako praprzyczyna, praenergia, demiurg, najwyższy byt. Istnieją jego różne interpretacje w światowych religiach i wierzeniach.

Tymczasem drugim terminem, czy biegunem relacji wiary, jest Bóg **osobowy**. Wiara odkrywa, że poszukiwany Bóg objawia się. **Objawienie** — pojęcie teologiczne — to samoukazywanie się Boga, komunikowanie Jego obecności, ukazywanie nie tylko Jego istoty, ale i istoty świata i człowieka.

Objawienie to dokonuje się w czasie i historii. Jego jakby kroniką jest **Biblia**, zwana też **Pismem Świętym**. Biblia, to mała biblioteka, składająca się z kilkudziesięciu ksiąg (45 ST + 27 NT), powstających sukcesywnie, a ich datacja, dzięki pracy biblistów, jest coraz lepiej znana. „Kronikarzy” nazywa się **autorami biblijnymi**, którzy posługiwali się całą gamą **rodzajów literackich**. Biblia opisuje w części retrospekcyjnej historię świata i człowieka od ich początku, a w swej warstwie prospektywnej wizję końca świata w znanej nam jego postaci. Głównym jednak nurtem biblijnego przekazu jest obecność — stopniowe ujawnianie się Boga w czasie, przetrzeni i w ponadczasowym Absolucie.

To dzięki przekazowi biblijnemu i towarzyszącej mu żywej tradycji wiemy, że Bóg jest bytem osobowym, aczkolwiek natury duchowej. Ukazuje się również chrześcijański rys objawienia nieistniejący w innych religiach. Dotyczy on, od strony bytowej, prawdy o Bogu jako jedności Trzech Osób. Drugą prawdą chrześcijańską jest **dogmat wcielenia**. Bóg wkroczył w historię ludzkości, przyjmując w drugiej Osobie naturę człowieka.

Wiara zawiera i zakłada precyzyjną treść. Ujęta ona została w zbiorze zwanym **Credo** — czyli wyznanie wiary. Kolejne wersje *Credo* od pierwszej, zwanej Symbolem albo Składem Apostolskim mają tę samą strukturę i zawartość treściową. Najpierw jest mowa o Bogu, następnie o Kościele, wreszcie o celu ostatecznym człowieka. Kolejne wersje *Credo*, jak nicejsko-konstantynopolska i *Credo* mszalne aż do Wyznania Wiary Pawła VI, są bardziej eksplikatywne, ale treściowo tożsame.

Naukową refleksją nad treścią wiary zajmuje się **teologia**. Wszystkie najstarsze uniwersytety europejskie od początku posiadały wydział teologiczny, będący zresztą warunkiem zatwierdzenia uczelni.

Również w Polsce istnieją liczne wydziały teologiczne i kadra naukowa wszystkich szczebli, od asystentów po belwederskich profesorów. To ich praca i dorobek wykorzystywany jest w oficjalnych orzeczeniach Kościoła — tzw. Magisterium, czyli Urzędu Nauczycielskiego.

Refleksja teologiczna pozwala też badać relacje zachodzące między zawartością wiary a obszarami ludzkiej działalności.

Dzisiaj przyjrzymy się tym relacjom i spojrzeniem na dwie z nich, zresztą coraz bardziej ze sobą powiązane, czyli na technikę — przedmiot nauczania goszczącej nas uczelni — i na medycynę — zawód mówiącego te słowa.

Powiedzenie, że **technika** towarzyszy człowiekowi od zarania jego egzystencji na ziemi jest truizmem. Posługiwanie się narzędziami i znajomością fizycznego otoczenia człowieka wykazuje początkowo niezwykle powolny ale stale postępujący proces o charakterze linearnym, ale niekiedy przerywanym. Można cytować zatrzymanie i zaprzestanie rozwoju imponujących osiągnięć inżynierskich w starożytnym Egipcie, czy cywilizacji Indian w Ameryce Środkowej i Południowej.

Prowadzenie rachunków i obliczeń, mierzenie przetrzeni i czasu, umiejętności czysto praktyczne „z praktyki zrodzone i praktyce służące”, były dla przyszłej nauki materiałem wyjściowym, który wypadło przetworzyć w teorię i opracować naukowo, sięgając do przyczyn zjawisk i koniecznych uogólnień. W ten sposób technice towarzyszy **technologia** rozumiana jako nauka o technice i jej procesach.

To właśnie technika wprowadziła do naszego słownictwa ideę postępu, początkowo i w terminach historycznych trudno dostrzegalnego. Od niepamiętnych czasów ludzkość żyła w kręgu **kultury rolniczej** z jej cechami charakterystycznymi, jak zależność od natury, pogody, pór roku oraz suwerenność pracy w społeczeństwie ograniczonym kręgu. Była to kultura oralna języka opisowego, ujednoliconych postaw i obyczajów oraz typowej strukturze społecznej.

Rola tradycji była w obszarze tej kultury decydująca. Zwrócone do przeszłości społeczeństwa tradycyjne utożsamiały się z wiedzą i mądrością przodków. Udoskonalanie i rozwój techniki w życiu były powolne i czasochłonne.

W takiej też kulturze agro-pastoralnej osadzona jest kultura biblijna, która jednak coraz wyraźniej zwraca się ku przyszłości. Chodzi tu o przejście z czasu cyklicznego, zamkniętego w czas otwarty i progresywny.

Przekroczenie progu **kultury przemysłowej** oznaczało znaczne przyspieszenie odkryć i zastosowań technicznych z wykorzystaniem nowych źródeł energii oraz metod produkcji. Były one synonimem postępu rzeczywistego, ale i podpartego ideologią. Urbanizacja, odkrycia geograficzne, powstanie i rozwój burżuazji

dokonywały się również w atmosferze mitu o postępie. Postępowi technicznemu nie zawsze towarzyszy postęp moralny i cywilizacyjny.

Nowa kultura przemysłowa oznaczała: racjonalne planowanie, produkcję zespotową, mechanizację, nowy język (bardziej syntetyczny i sformalizowany), powstanie państw narodowych na bazie nacjonalizmów, dopuszczenie różnorodności postaw i zachowań. Tym zmianom towarzyszyły już stale innowacje techniczne i niemal rewolucyjny rozwój komunikacji i transportu.

Kolejną fazą linearnego rozwoju o coraz większym przyspieszeniu jest obecnie panująca **kultura elektro-niczna**. Praca indywidualna i zespołowa korzysta z komputerowego wsparcia. Robotnika taśmowego zastąpiła robotyzacja. Programy produkcyjne osiągnęły niezwykłą precyzyjność. Pojawił się wykwinicie rozwinięty syndrom wydajności i zysku. Język syntetyczny zero-jedynkowy (tak – nie) zubożył język mówiony i pisany, a jednocześnie ma niemal nieograniczoną pojemność.

Postępuje standaryzacja postaw, zachowań i obyczajów w skali globalnej – „American style of life”, „global English”. Globalna jest komunikacja i transport, obywatelstwo staje się kontynentalne i światowe.

Elektroniczne media o światowym zasięgu przedstawiają rzeczywistość społeczną, kulturalną i polityczną, jednocześnie je deformując i propagując uniformistyczne sądy, ideologiczne poglądy oraz modne trendy. W ten sposób zmienia się mentalność ludzi ich **etnopsychologia** i wzajemne relacje. Jednocześnie, paradoksalnie, narasta samotność jednostek i ich zagubienie w życiu.

Wszystkie te, powierzchownie opisane zjawiska, są wynikiem ogromnie poszerzonych możliwości technicznych. Analitycy społeczeństwa zauważają, że technika zawładnęła już prawie wszystkimi dziedzinami życia: medycyną z jej laboratoriami analitycznymi i diagnostyką obrazową, finansami wraz z technikami obliczeniowymi, rolnictwem maksymalnie zmechanizowanym i nasyconym środkami chemicznymi. Rytm życia ulega ciągłemu przyspieszeniu, a to przyspieszenie znajduje podstawę w postępie technicznym, który modeluje i przekształca kulturę.

Niektórzy badacze (G. Devreux) rozwój kultur upatrują w sumie wszystkich technik, które nie są przekazywane biologicznie, tym bardziej, że owe techniki nie nakładają się przypadkowo na siebie, lecz tworzą ściśle powiązaną całość, utworzoną według określonych osi koncepcyjnych stanowiących „szkielet” czy „model” kultury.

Inni, jak J. Ellul, za technikę uważają wszystkie wynalazki przekazywane z pokolenia na pokolenie i odnoszące się do czynników, które utatwiając życie, modyfikują postawy i zachowania ludzi.

Co na to wiara i przesłanie biblijne? Czy kultura tak uformowana jest kompatybilna z naszym etosem, czy też od niego się oddala?

Charles-Daniel Maire, świecki teolog protestancki uważa, iż Biblia pokazuje, jak Bóg wkraczał w historię, by wyrwać swój lud z alienującej go władzy. Uczy, jak rozróżnić fałszywe uzasadnienia, pełne nadużyć uproszczenia i ujawnia motywy, które ideologia usiłuje ukryć.

Wśród radykalnych pytań, jakie towarzyszą ludziom jest i takie o granice oraz relacje między przyrodą, techniką i moralnością. Są to problemy, *które zdecydowanie odwołują się do odpowiedzialności osobistej i zbiorowej na płaszczyźnie postaw, jakie należy przyjmując w odniesieniu do tego kim jest człowiek, co może uczynić i kim powinien być* [Kompendium nauki społecznej Kościoła, nr 16].

Omawiane zagadnienia wpływu techniki na ludzkie życie i kształtowanie przez nią nowego paradygmatu ludzkiej egzystencji jest tak obszerne, że przekraczają ramy niniejszego wykładu.

Zacytuję tylko, konkludując, ten fragment refleksji inspirowanej przez wystąpienie Jana Pawła II w UNESCO 2 czerwca 1980 roku:

Zagubienie metafizycznej perspektywy, utrata tęsknoty za Bogiem w wyniku narcystycznego samouwielbienia i obfitości środków oferowanych przez konsumpcyjny styl życia; pierwszeństwo przyznawane technologii i badaniom naukowym, które stały się celami samymi w sobie, przesadne przywiązywanie wagi do powierzchowności, kreowanie wizerunku, technik komunikacji; wszystkie te zjawiska należy postrzegać w aspektach kulturowych i analizować w zestawieniu z centralnym tematem, jakim jest osoba ludzka, jej integralny rozwój, zdolność do komunikowania się i nawiązywania relacji z innymi ludźmi, ciągłe szukanie odpowiedzi na ważne egzystencjalne pytania.

Jeśli technika zmienia radykalnie ludzkie życie, jeśli przekształca nie tylko kulturę, ale i ludzką mentalność i osobowość, to łatwo dociec, iż wprowadziła rewolucyjne zmiany w praktykowaniu i kształcie innej dziedziny od niepamiętnych czasów, towarzyszącej ludzkości, czyli **medycyny**.

Za ojca medycyny europejskiej przyjmuje się Hipokratesa, lekarza greckiego. Już wtedy kładł on pod-

waliny pod medycynę racjonalną, przyczynowo-skutkową, poznającą zewnętrzne i wewnętrzne mechanizmy fizjologii patologii. On to jest do dzisiaj twórcą deontologii lekarskiej, czyli etyki zawodowej.

Medycyna, określana jako sztuka i nauka leczenia, implikuje dwie osoby, wchodzące w bezpośredni kontakt – człowieka chorego i lekarza. Kontakt ten służył najpierw ustaleniu diagnozy, a następnie śledził etapy terapii. Lekarz miał współpracować z naturą – z fizjologią człowieka. Nie wszyscy uświadamiają sobie, w jakim stopniu osoba pacjenta, jako rzeczywistość psychosomatyczna, jest narzędziem diagnostycznym. Zbieranie wywiadu, czyli anamneza, ma na celu odtworzenie z pamięci chorego historii choroby, jej początku i dalszego rozwoju. To pacjent sygnalizuje zaobserwowane objawy, samopoczucie, zachowanie czynności życiowych. W tym etapie odgrywa on rolę podmiotu. Jego ciało jest znakomitym „aparatem” samodiagnozy, który znamy z urządzeń mechanicznych i elektronicznych.

W kolejnym etapie badania chorego, rolę podmiotową odgrywa lekarz i jego działania, a przedmiotową obecność pacjenta. Lekarz ogląda wzrokiem, bada dotykiem, posługuje się słuchem w opukiwaniu i ostuchiwaniu chorego. To jego, lekarza, ciało jest „aparatem” diagnostycznym w tej fazie, którą nazywamy badaniem fizykalnym.

I znów, jak w życiu codziennym, technika stała się przedłużeniem ludzkich zmysłów i czynności. Lupa i mikroskop zastąpiły wzrok, stetoskop i fonendoskop wzmocniły słuch, zastosowanie promieni rentgenowskich i ultradźwięków, endoskopii, czyli cała tzw. diagnostyka obrazowa umożliwiły zajrzenie do wnętrza organizmu przyżyciowo, a nie tylko w prosektorium. Podobnie też narzędzia i urządzenia stały się domeną terapii, nie mówiąc o postępkach chemii stosowanej w farmakoterapii czy nowych materiałach w protetyce.

Lawinowo wzrastająca obecność techniki w medycynie zmienia jej naturę, uwarunkowania i rozsadza relację normatywną między zasadami i wartościami.

Rozwój nowoczesnej medycyny i jej strefy zwiększającego się wptywu na ludzkie ciało są niewątpliwie jednym z decydujących elementów nowej kultury osobistej coraz bardziej zainteresowanej badaniami mogącymi poprawić parametry organizmu (sport wyczynowy, *body building*, chirurgia plastyczna).

Inną stroną dzisiejszej praktyki medycznej jest wręcz redefinicja medycyny, dystrybucji wiedzy w jej dziedzinie, technik biomedycznych, aż do zmiany sen-

su i znaczenia przypisywanego chorobie i choremu. Zmieni się tu, i to radykalnie, kliniczny model relacji lekarza i chorego.

Medycyna staje się rzeczywistością wielopostaciową, w której funkcjonują instytucja medyczna (szpital, kliniki itp.), finanse służby zdrowia, polityka zdrowotna, biotechnologia i wreszcie bioetyka.

Ośrodki decyzyjne stają się rozproszone, dzieląc między sobą, to co nazwał Michel Foucault biowładzą – władzę polityczną, jurydyczną i technologiczną – mającą za przedmiot ludzkie życie.

W sytuacji, gdy medycyna bywa nazywana „totalnym faktem społecznym”, słuszną wydaje się sugestia, by w centrum debaty umieścić na nowo relację między dwiema osobami – pacjentem i lekarzem.

To właśnie masowe nasycenie medycyny techniką zepchnęło na dalszy plan rolę i znaczenie fizykalnego badania pacjenta (wielu młodych lekarzy tego nie potrafi), słownego i pozawerbalnego z nim kontaktu, ekranizując w przenośni i dostownie wzajemne widzenie. Coraz częściej lekarz przyjmujący w swoim gabinecie jest w interaktywnym kontakcie z ekranem komputera, pacjenta widzi poprzez listy wyników badań szczegółowych, a jego obecność w pomieszczeniu, ma znaczenie drugorzędne i nie jest konieczna.

Opisany model ma dwie inne, poważne wady. Eliminuje z medycyny tak ważną intuicję lekarską z jednej strony, doprowadzając ten szlachetny zawód do odarcia go z elementu sztuki. Lekarz przekształcony w inżyniera biologicznego, jest jakby obok swojej misji dostrzegania pacjenta jako osoby obdarzonej nie tylko biologicznym ciałem, ale też specyficzną ludzką psychiką i duchowością.

Poważnym wyzwaniem współczesnej medycyny jest uporanie się z coraz to innymi problemami bioetycznymi. Wyeliminowanie etyki z techniki i medycyny jest śmiertelnym zagrożeniem dla poszczególnych osób i społeczeństw.

Problemy bioetyczne pojawiły się z co najmniej dwóch powodów. Pierwszym jest omawiany już lawinowy rozwój biomedycyny z zastosowaniem biotechnologii. Znamy początek tego procesu, nie wiemy natomiast dokąd on prowadzi. Drugim powodem jest zmienny *ethos* kulturowy. O ile tradycyjne kultury mieszcza się w dość jednolitym nurcie podstawowych odniesień, to nowe, tworzone w oparciu o radykalne pomysły, kultury od tego nurtu odbiegają.

Wobec zmienności wymienionych powodów narastania problemów bioetycznych dochodzi do kon-

fliktów i napięć między tym, co może się zmieniać i tym, co jest niezmiennie i takie pozostać musi pod groźbą zatraty samej idei humanizmu. Potrzebna jest zatem regulacja etyczna, czyli szukanie zgodności zasad i wartości.

Kościół postępujący się poszerzonym polem poznania – naukami pozytywnymi, humanistycznymi i ściślymi – oraz treścią Bożego Objawienia, czyli treścią wiary, jest doskonale przygotowany do dzisiejszej debaty etycznej i nie zajmuje postawy wątpienia w szczegółowych rozważaniach, właśnie dzięki solidnym podstawom **aksjologicznym**.

Podstawową statą wartością jest definicja człowieka i jego życia biologicznego.

Ciało ludzkie nie może być uważane tylko za zespół tkanek, narządów i funkcji; nie może być oceniane na równi z ciałem zwierzęcym, jest bowiem istotną częścią osoby, która przez to ciało objawia się i wyraża [Donum Vitae, 74].

Każda osoba ludzka w swojej niepowtarzalnej wyjątkowości nie jest złożona tylko z ducha, lecz także z ciała, i dlatego w ciele i przez ciało dociera się do samej osoby w jej konkretnej rzeczywistości [JPiI, Przemówienie WHO z dn. 29.10.1983].

Ciało, będące wyrażeniem osoby, w swoim kształcie i dynamizmie biologicznym, jest podstawą i źródłem wymogów moralnych [JPiI, Audjencja Ogólna, z dnia 09.01.1984 i 20.02.1984].

Są to elementarne określenia antropologii chrześcijańskiej o powszechnym pozawyznaniowym charakterze, ujętym w tzw. **prawo naturalne**, przystugujące każdej ludzkiej osobie i w całkowitym jej istnieniu od początku do naturalnej śmierci.

W konsekwencji, ludzkie życie cieszy się wymogiem nienaruszalności i nierozporządności.

Życie ludzkie jest podwójnie sakralne – jako pochodzące od Boga, którego jest obrazem i podobieństwem i sakralnością naturalną, którą może uznać każdy prawy rozum, także abstrahując od wiary religijnej [JPiI, Przemówienie z dn. 24.10.1986].

Zmierzając do konkluzji, z konieczności pobieżnej refleksji, warto sięgnąć do kluczowego w omawianej materii tekstu Jana Pawła II *Fides et Ratio*. Przypomina on, że łącznikiem między wiarą a rozumem, między teologią a naukami szczegółowymi, jest **filozofia**. W okresie powstawania pierwszych uniwersytetów, te dwie dziedziny szeroko pojętej wiedzy pozostawały w ścisłym kontakcie i dialogu. Jednak od późnego średnio-

wiecza metodyczne rozgraniczenie tych obszarów poznania przekształciło się w szkodliwy rozdział.

Zacytuję dwa fragmenty:

Przesadny racjonalizm niektórych myślicieli doprowadził do radykalizacji stanowisk i do powstania filozofii praktycznie oderwanej i całkowicie autonomicznej w stosunku do treści wiary. Jedną z konsekwencji tego rozdziału była także narastająca podejrzliwość wobec samego rozumu. Niektórzy przyjęli postawę całkowitej nieufności, sceptycyzmu i agnostycyzmu, bądź to aby rozszerzyć przestrzeń wiary, bądź też aby pozbawić ją wszelkich racjonalnych odniesień.

W sferze badań przyrodniczych rozpowszechniła się stopniowo mentalność pozytywistyczna, która nie tylko zerwała wszelkie powiązania z chrześcijańską wizją świata, ale – co ważniejsze – zrezygnowała też z wszelkich odniesień do wizji metafizycznej i moralnej. W wyniku tego zaistniało niebezpieczeństwo, że niektórzy ludzie nauki, rezygnując z jakichkolwiek odniesień etycznych, nie stawiają już w centrum swej uwagi osoby ludzkiej i całości jej życia. Co więcej, część z nich, świadoma możliwości otwartych przez rozwój techniki, wydaje się ulegać nie tylko logice rynku, ale także pokusie zdobycia demiurgicznej władzy nad przyrodą, a nawet nad samym bytem ludzkim.

Konsekwencje kryzysu racjonalistycznego przyjęły w końcu postać nihilizmu. Dla naszych współczesnych ma on swoisty urok jako filozofia nicości. Według teorii jego zwolenników poszukiwanie stanowi cel sam w sobie, nie istnieje bowiem nadzieja ani możliwość osiągnięcia celu, jakim jest prawda. W interpretacji nihilistycznej życie jest jedynie sposobnością do poszukiwania doznań i doświadczeń, wśród których na pierwszy plan wysuwa się to, co przemijające. Nihilizm jest źródłem rozpowszechnionego dziś poglądu, że nie należy podejmować żadnych trwałych zobowiązań, ponieważ wszystko jest ulotne i tymczasowe [FR, 45].

Należę do tych, którzy uważają, iż każda wyższa uczelnia w poszukiwaniu swej nobilitacji powinna dochodzić do ponaddiscyplinarnych poszukiwań w zakresie jej właściwej kompetencji. Czy to będzie filozofia techniki, czy medycyny, obydwie odwołują się nie tylko do inteligencji, wiedzy i praktycznej sprawności, ale stanowią podstawę mądrościową, chroniącą uprawiane dyscypliny od ciągłe i coraz bardziej aktualnej perspektywy zwrócenia się przeciwko człowiekowi.

Widniejący w tym gmachu cytat z Jana Zamoyskiego, tę konkluzję uprawomocnia.

Abp Henryk Hoser polski biskup rzymskokatolicki, pallotyn, misjonarz, lekarz, biskup diecezjalny warszawsko-praski od 2008 roku. W Episkopacie Polski pełni funkcję przewodniczącego Zespołu Ekspertów KEP ds. Bioetycznych, członka Komisji Duszpasterstwa, członka Rady ds. Rodziny, członka Komisji ds. Misji. Był współzałożycielem Afrykańskiej Federacji Akcji Rodzinnej. Jest członkiem Rady Dzieła Nowego Tysiąclecia. Jako rektor Pallotyńskiej Prokury Misyjnej w Brukseli zajmował się duszpasterstwem w strukturach Wspólnoty Europejskiej. W latach 2005–2008 był Sekretarzem Pomocniczym Kongregacji Ewangelizacji Narodów i Przewodniczącym Papieskich Dzieł Misyjnych. Od 2009 roku jest konsultorem Kongregacji ds. Ewangelizacji Narodów. Odznaczony Orderem Zasługi (Portugalia, 2008), Honorowy Obywatel Warszawy (2011).

Science, Society and Sustainability

Artykuł, zapowiadający odczyt profesora Harolda Kroto, laureata Nagrody Nobla w dziedzinie chemii w 1996 roku.

Odczyt Profesora, pod tym samym tytułem, został wygłoszony w dniu 14 lipca 2010 roku.

147

Harold Kroto i fullereny — rys historyczny

Harold Kroto z Univeristy of Sussex (Wielka Brytania) oraz Robert Curl i Richard Smalley z Rice University w Houston (USA) otrzymali w 1996 roku nagrodę Nobla w dziedzinie chemii za odkrycie fullerenów.

Prześledźmy historię tego odkrycia poprzez prace Harry'ego Kroto, uważanego za głównego inspiratora badań, które doprowadziły zespół do tego sukcesu.

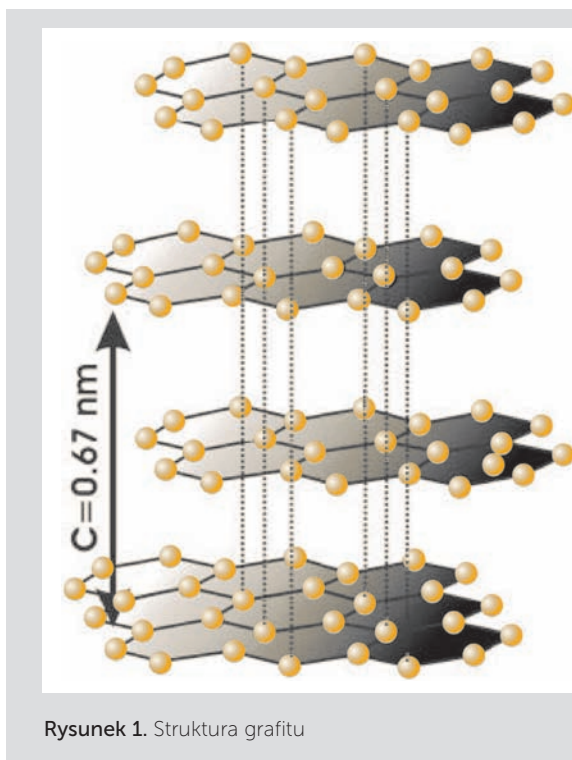
Harold Kroto związał swoją pracę naukową z University of Sussex na przełomie lat sześćdziesiątych i siedemdziesiątych XX wieku. Zajmował się m.in. identyfikacją liniowych cząsteczek węglowych poprzez ich oddziaływanie z promieniowaniem mikrofalowym z zastosowaniem spektroskopii rotacyjnej. Ważnym momentem w trakcie tych prac była synteza cząsteczki cyjanodwuacetyleny HC_5N i określenie zbioru częstotliwości absorpcji mikrofal przez te cząsteczki. Miało to związek z pracami radioastronomów nad poszukiwaniem międzygwiazdowego HC_5N . Kroto zastana wiał się nad mechanizmami powstawania międzygwiazdowych molekuł, zawierających w swojej strukturze liniowe tańcuchy węglowe. Według niego miejscem powstawania długołańcuchowych związków chemicznych była atmosfera zimnych gwiazd węglowych. Powstało pytanie czy tego typu cząsteczki nie stanowią pośrednich struktur pomiędzy krótkimi cząsteczkami C_n , powstającymi podczas sublimacji węgla, a cząsteczkami grafitu obserwowanymi wokół czerwonego

olbrzyma. Ta hipoteza stała się inspiracją do poszukiwania techniki otrzymywania wielkocząsteczkowych struktur węglowych. Kluczowym surowcem do wytwarzania takich struktur mógłby być grafit — jedna z podstawowych (obok diamentu) odmian alotropowych węgla. Struktura grafitu (rys. 1), składa się z warstw, w których występują sprzężone, sześciocząsteczkowe układy cykliczne.

Podobnie jak w cząsteczce benzenu, każde wiązanie C-C w warstwie ma charakter zdelokalizowanego $3/2$ -krotnego wiązania aromatycznego, tworzącego obszary zdelokalizowanych orbitali π , umożliwiających swobodny ruch elektronów równoległe do warstw pierścieniowych. Między warstwami występują słabe oddziaływania van der Waalsa. Podstawowe elementy takiej struktury mogłyby być wygodnym budulcem form wielkocząsteczkowych, zawierających duże ilości atomów węgla.

Jak stworzyć warunki do wykonania takich zlepów?

W latach osiemdziesiątych XX wieku Richard Smalley w Rice University zajmował się wykorzystaniem technik wiązek molekularnych do badania klastków atomowych. Aparat o nazwie AP2, który skonstruował i stosował do tych badań, umożliwiał tworzenie klastków z wybranego materiału. Zasada działania polegała na napromienianiu impulsową wiązką lasera tarczy materiału, który odparowywał do strumienia sprzężonego helu. Odpowiednie warunki aerodynamiczne układu i czasy przebywania umożliwiały łączenie odparowywanych elementów tarczy w klastery



Rysunek 1. Struktura grafitu

odpowiedniej wielkości, analizowane za pomocą spektrometru masowego.

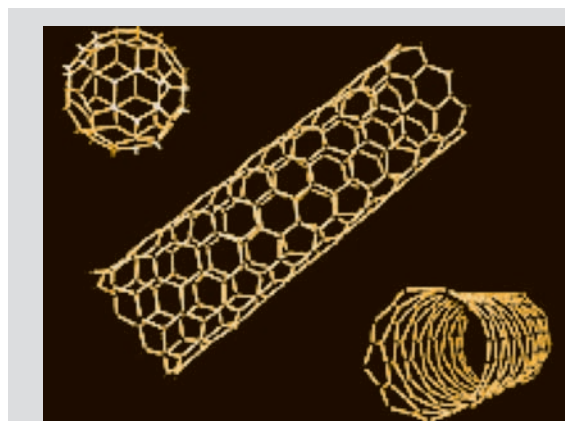
Harold Kroto, który na zaproszenie Roberta Curla, współpracującego ze Smalleyem, wizytował w 1985 roku Rice University, przedstawił swoją koncepcję i cel badań, proponując aby do eksperymentu jako materiału tarczy użyć grafitu.

Zespół, w skład którego wchodził jeszcze doktoranci, zmieniając warunki operacyjne procesu otrzymywał klaster węglowe różnej wielkości. Analiza spektralna wykazała, że w produkcie silny pik widma wykazują struktury zawierające 60 atomów. Pik ten był kilkadziesiąt razy silniejszy od pików odpowiadających sąsiednim klasterom. Badacze doszli do wniosku, że otrzymali stabilną i trwałą strukturę wielcząsteczkową węgla C_{60} . Jaka była jej budowa przestrzenna? Jak połączone są w niej atomy węgla? Kolejne hipotetyczne struktury C_{60} nie spełniały warunków spójności wszystkich elementów definiujących molekułę. Na pewnym etapie rozważań Harry Kroto zaproponował wielowarstwową strukturę sześciokątów grafitu — czteropiętrowy układ, w którym pomiędzy dwoma sześciokątnymi pierścieniami z sześciu atomów węgla ściśnięte są dwie sieciowate „blaszki” z 24 atomami węgla każda. W sumie w takiej

strukturze jest 60 atomów węgla. Ale jak zagospodarowane są 32 wiązania brzegowe? Energia tych wiązań typu delta jest duża. Jak mogą się one potoczyć? Czy możliwe byłoby przestrzenne wygięcie wiązania chemicznego żeby to uzyskać?

Przez kolejne przybliżenia oraz fakt, że szczególnie stabilne są struktury C_{60} i C_{70} , twórcy doszli do wniosku, że molekuly te mają zwartą strukturę przestrzenną zamykającą się do kształtu kuli.

Pierwsze pomiary modeli przestrzennych, wykonywanych z papieru, wykazały, że same sześciokąty nie wystarczą do zbudowania zamkniętej struktury. Inspiracją do ostatniego modelu zawierającego sześciokąty obok pięciokątów, były obserwacje konstrukcji architektonicznych pokrycia hal w postaci tzw. kopuł geodezyjnych, zaproponowanych przez R. Buckminstera Fullera — amerykańskiego architekta, którego pracami fascynował się młody Kroto. Konstrukcje połączenia foremnych sześciokątów z pięciokątami utworzyły drogę do uzyskania krzywizny zamkniętej bryły, złożonej z dwudziestu ścianek sześciokątnych i dwunastu pięciokątnych oraz sześćdziesięciu wierzchołków. Wierzchołki są sobie równoważne i każdy leży we wspólnym narożniku jednego pięciokąta i dwóch sześciokątów. Dokładna analiza struktury potwierdziła, że taka konstrukcja odpowiada sytuacji, w której wiązania pomiędzy atomami węgla, znajdującymi się w narożach, wysycają się jak w klasycznych cząsteczkach związków aromatycznych. Powstała nowa molekula C_{60} , nazwana przez twórców początkowo żartobliwie „jajami Buckiego”, na cześć architekta Fullera, a później bardziej poważnie „Buckminster fullerenem”, potocznie fullerenem (rys. 2).



Rysunek 2. Model fullerenu i nanorurek

Było to wielkie i inspirujące odkrycie. Należy tu nadmienić, że wcześniej, nad podobnymi strukturami pracował zespół badawczy Exxona oraz badacze japońscy i niemieccy. Pełny i udokumentowany opis fullerenów podali jednak Curl, Kroto i Smalley. Ich odkrycie stało się inspiracją do tworzenia struktur pochodnych, które zaliczamy do grupy fullerenów. Szczególnymi izomerami strukturalnymi fullerenów są np. nanorurki. Są to walce powstające ze zwinięcia pojedynczej płaszczyzny grafitowej, domknięte z obu stron potówkami fullerenów odpowiednimi do średnicy cylindra.

Można wyróżnić następujące rodziny fullerenów:

- fullereny wtaściwe (C_{60} , C_{70}),
- fullereny cebulkowe (wielowarstwowe),
- fullereny olbrzymie (C_x , $x > 500$),
- nanorurki.

W trakcie badań okazało się, że fullereny mogą mieć szereg zastosowań praktycznych. Można je łączyć z innymi pierwiastkami lub związkami chemicznymi (tzw. fullereny egzohedralne), umieszczać w ich wnętrzu atomy różnych pierwiastków lub związki nadające takim strukturom własności funkcjonalne (fullereny endohedralne). Fullereny otworzyły nowy obszar badawczy i aplikacyjny, nazwany nanotechnologiami.

Opracował — prof. dr hab. Leon Gradoń; materiały źródłowe — prace H. Kroto i E. Osawy; zdjęcia — Wikipedia; <http://www.comunicacion.amc.edu.mx/fotos/reconocen-premios-nobel-la-calidad-de-la-ciencia-mexicana-2/>

Abstract wykładu profesora Harolda Kroto pt. *Science, Society and Sustainability*

The fact that our modern world is so completely and precariously balanced on Science, Engineering and Technology (SET) makes an understanding of these disciplines by all in positions of significant responsibility vital. *Although wise decision-making may not be guaranteed by knowledge, common sense suggests that wisdom is not a likely consequence of ignorance.* So far SET have truly revolutionised our lives and there is no doubt that the humanitarian contributions have improved the quality of life in the developed world immeasurably. It is worth reminding ourselves that, for instance, in the 18th Century, half of all children died by the age of 8. It is also worth reflecting on the fact that this improvement was brought about by scientific/technological advances based on doubt and ques-

tioning — evidence dependent philosophies totally at variance with the belief-based concepts that underpin all mystical societal attitudes. The conflict between these two orthogonal philosophies is nowhere more obvious than in the recent avalanche of attempts to undermine the validity of Darwin's Theory of Evolution and in particular corrupt the intellectual integrity of our educational system. The fact that this is taken at all seriously indicates a level of basic public gullibility which is being wilfully and maliciously exploited by intellectually corrupt organisations focused primarily on undermining the rational principles that led to Democracy and the Enlightenment. The mindless acceptance of the claims is hard to credit when one considers that almost everyone in the western world has benefited from the socio-economic advances of SET — in particular medical advances based on a deep understanding of the evolutionary process.

Society has the power to use technology so that it can be of benefit or be detrimental. On the latter score one should note that political decisions have resulted in the existence of some 28,000 nuclear weapons worldwide. In addition, it now appears that our technologies may well have also catalysed a mindless mass-production driven plundering of the Planet's resources. We may be hurtling towards disaster — we may not need an asteroid. For a 50:50 chance of surviving into the next century every segment of society, from industrialists, engineers and scientists to politicians, farmers and fishermen must now recognise that these issues are the most serious that the human race has ever confronted. Our only hope for survival rests on the shoulders of those who take survival and sustainability issues seriously — and do something about it.

I see a key role for "Nanoscience and Nanotechnology" which is arguably — and I would argue it — just a new name for Chemistry where this discipline overlaps Condensed Matter Physics, Molecular Biology and Materials Engineering. I also see improved SET education, as vital. We have manifestly failed in this endeavour but there may be one last hope: *The Internet is a major new communications technology which we must exploit to educate people on a global scale in the rational attitudes to decision-making that are now vital to our very survival.* With the Vega Science Trust (www.vega.org.uk), we have made over 200 TV and Internet programmes (almost half shown on the BBC). It is a highly successful platform for scientists to communicate directly and

improve the public awareness and understanding of SET. With an exciting new Global Educational Outreach for Science Engineering and Technology (GeoSet, www.geoset.info) initiative we are working with other universities to make outstanding educational material available

in any part of the world. The major aim is to empower teachers, worldwide, by giving them access to the best teaching material, packaged for direct use in the classroom, together with expert examples of how the material might best be presented.

Profesor Harold Kroto urodził się w miejscowości Wisbech w hrabstwie Cambridge w Anglii w 1939 roku. Po ukończeniu szkół w Bolton, w 1958 roku rozpoczął studia na wydziale chemicznym University of Sheffield, które ukończył z wyróżnieniem w 1961 roku. Prace nad doktoratem w zakresie spektroskopii molekularnej prowadził również na University of Sheffield, uzyskując stopień doktora (Ph.D.) w 1964 roku. Staż podoktorski odbył w latach 1964–1966 w National Research Council w Ottawie, a następnie kontynuował swoje badania w Bell Telephone Laboratories (1966–1967). W 1967 roku wrócił do University of Sussex (Brighton), gdzie po kolejnych awansach akademickich w 1985 roku został profesorem. Od 2004 roku Harold Kroto jest profesorem na Wydziale Chemii, Florida State University.

Na początku kariery naukowej zajmował się badaniami własności wolnych rodników i oddziaływań molekularnych w fazie gazowej i ciekłej z zastosowaniem metod spektroskopowych. Następnie skierował swoje zainteresowania na tworzenie nowych molekuł z wielokrotnymi wiązaniami między węglem a innymi pierwiastkami, w szczególności siar-

ką, selenem i fosforem. Udowodnił możliwość istnienia takich związków. Dalsze badania Harolda Kroto dotyczyły molekuł w postaci długołańcuchowych liniowych struktur węglowych, których istnienie stwierdzono w gwiazdach i przestrzeni międzygwiazdowej. Zajął się więc możliwością otrzymywania takich struktur w warunkach laboratoryjnych. Doprowadziło to m.in. do odkrycia struktury fullereny, za które wraz z Robertem Curlem i Richardem Smalleyem otrzymał w 1996 roku Nagrodę Nobla.

Najnowsze badania Harolda Kroto dotyczą chemii fullerenów, otrzymywania i badania własności materiałów nanostrukturalnych, w tym nanorurek.

Za swoją działalność profesor Harold Kroto został uhonorowany szeregiem wyróżnień, w których najważniejszą jest wspomniana Nagroda Nobla. Jest członkiem wielu towarzystw naukowych, m.in. Towarzystwa Królewskiego (1990), Narodowej Akademii Nauk USA, Królewskiego Towarzystwa Chemicznego (1993), którego prezydentem był w latach 2002–2004. Został wyróżniony doktoratami honorowymi 29 uniwersytetów na całym świecie. Ponadto otrzymał szereg medali i nagród naukowych i edukacyjnych.

Nauka w świecie islamu

Na podstawie odczytu wygłoszonego w dniu 14 stycznia 2010 roku

Janusz Danecki

Katedra Arabistyki, Uniwersytet Warszawski

151

Od najwcześniejszych lat szkolnych dowiadujemy się o wkładzie nauki arabskiej do dziedzictwa cywilizacji światowej. Przewijają się nazwy, które stają się codziennością — algebra, sinus, nadir, zenit. O niektórych szybko zapominamy, inne wchodzą na trwałe do naszej pamięci. Rzadko jednak pamiętamy o ich arabskim pochodzeniu, a prawie wcale nie wiemy, jak powstały i jak do nas dotarły.

Choć nie brak interesujących książek o nauce arabskiej, to jednak poza specjalistami rzadko kto do nich zagląda. A szkoda, bo wiedza o niej jest przydatna i ciekawa. Pokazuje bowiem, jak cywilizacje na siebie wpływają, a zarazem, jak upadają i przemijają. Nigdy jednak bez śladu.

W tym omówieniu postawiłem sobie dwa cele. Po pierwsze, chciałbym odpowiedzieć na pytanie: dlaczego w kulturze arabskiej powstała tak wielka nauka, że oddziałuje na nas do dziś. Przyglądając się bowiem zamieszaniu w świecie muzułmańskim, zdarza się nam sceptycznie myśleć również o przeszłości kultury islamu.

Po drugie, na kilku przykładach chcę pokazać, jak cywilizacja arabska rozwijała i tworzyła nową wiedzę.

Od czasów niemieckiego arabisty C.H. Beckera (zm. 1933) dzieje klasycznej cywilizacji muzułmańskiej dzieli się na dwa okresy — arabizację i islamizację, które obrazowo określa się jako ruch odśrodkowy i przeciwstawiany mu ruch dośrodkowy.

Arabizacja przypada na początek islamu i trwa mniej więcej jeden wiek, czyli od śmierci proroka Mahometa w 632 roku do upadku pierwszej dynastii muzułmańskiej Umajjadów w 750 roku. Wtedy to imperium kalifów roz-

rosło się ze skromnych, choć rozległych, siedzib plemiennych na Półwyspie Arabskim i objęło obszar od południowej Francji na zachodzie po granice Chin na wschodzie. Na tym obszarze językiem administracji i kultury stał się język arabski, stąd nazwa tego okresu — arabizacja. Ludy podbite przez Arabów zaczęły się posługiwać językiem arabskim jako wspólnym dla wszystkich. Wielką rolę w tym procesie odegrał islam, którego językiem był i pozostawał arabski. Arabizacja była procesem odśrodkowym — nastąpiła ekspansja Arabów we wszystkich kierunkach i narzucania języka arabskiego.

Drugi okres — islamizacja, rozpoczął się od przejścia władzy przez dynastię Abbasydów w 750 roku i trwał przez kilka następnych stuleci, zwanych złotym wiekiem lub renesansem islamu. W tym okresie nastąpiło utrwalenie panowania arabskiego na obszarze kalifatu. Towarzystwo mu nowe zjawisko, jakim była integracja wielu różnych kultur w nową cywilizację, zwaną muzułmańską. Z tego powodu islamizacja oznacza nie tyle islamizowanie (nawracanie na islam), co budowanie nowej cywilizacji z kultur, które znalazły się na obszarze islamu. Jest to więc ruch dośrodkowy, w którym dziedzictwo wielu kultur zostało włączone w nową spójną cywilizację, umownie nazywaną muzułmańską. Nie budowali jej wyłącznie Arabowie, ani nawet muzułmanie, lecz przede wszystkim przedstawiciele wszystkich tradycji kulturowych, które znalazły się w kalifacie.

Jest wiele przyczyn przejmowanie nauki z poprzednich kultur. Po pierwsze, Arabowie z Półwyspu nie mieli rozwiniętej ani tradycji naukowej, ani kultury miejskiej, ponieważ życie koczownicze rozproszonych plemion

przedmuzułmańskich nie sprzyjało rozwojowi zintegrowanej nauki. Budowanie imperium wymagało jednak dostosowania się do kultury miejskiej i stąd wynikała potrzeba przejmowania wzorców kulturowych od poprzedników. Po drugie, dopiero powstanie organizmu państwowego, jakim był kalifat, umożliwiło integrację dziedzictwa różnych kultur. I wreszcie, po trzecie, Arabowie, którzy sami nie mieli wielkich tradycji, okazali się niezwykle otwarci na dziedzictwo innych. Dotyczyło to wszelkich sfer życia – od filozofii i religii, przez medycynę i botanikę po nauki ścisłe. Ciekawe, że nie przejęto literatury pięknej. Ani poezja, ani dramat starożytny nie były przedmiotem zainteresowania w kulturze islamu.

Przede wszystkim niemal w całości przejęto wzory uprawiania nauki na Bliskim Wschodzie. Dawne akademie hellenistyczne nie tylko się rozwijały, ale na ich wzór powstawały nowe, podobne instytucje, takie jak na przykład Dom Mądrości w Bagdadzie (założonym w 767 roku). W Domu Mądrości gromadzili się wybitni uczeni, prawie wyłącznie reprezentujący inne narodowości i wyznania – chrześcijanie, Żydzi, Syryjczycy (Aramejczycy), Mandejczycy. Jednocześnie przybywali tu głodni wiedzy uczniowie, którzy mogli swobodnie zdobywać wiedzę zgodnie z utrwaloną w islamie maksymą (przypisaną z czasem Prorokowi) – *Wiedzy szukajcie choćby i w Chinach*.

Rozwój nowej muzulmańskiej nauki oparty był na istnieniu wspólnego języka wszystkich ludów, czyli arabskiego. Cały znany dorobek nauki zaczęto przekładać na język arabski. Ruch tłumaczeniowy stał się podstawą rozwoju nowej nauki, którą poprawnie należy określać jako arabsko-muzulmańską, ponieważ powstawała w języku arabskim, ale tworzyli ją przede wszystkim nie-Arabowie wciągnięci w krąg oddziaływania systemu politycznego islamu – kalifatu.

Tłumaczami byli specjaliści – lekarze, matematycy, astronomowie, botanicy, chemicy, fizycy (by użyć współczesnych określeń). Nic dziwnego, że tłumacząc dzieła kultury hellenistycznej, indyjskiej czy aramejskiej, od razu przystąpili do sprawdzania zawartej w nich wiedzy i jej uzupełniania. Zatem nie tylko tłumaczyli, ale również tworzyli, rozwijając dotychczasową wiedzę. Przy czym głównym asumptem do rozwijania nauki zawsze były cele praktyczne – jej zastosowania w różnych dziedzinach życia. Prowadziło to do poszerzenia

zastosowań dotychczasowej nauki. Rodząca się klasa uczonych służyła nowej cywilizacji i ją budowała. Rozwijający się ruch dośrodkowy, islamizacja, przyczynił się do budowania tego, co nazywamy kulturą klasycznego islamu.

Ten złożony mechanizm najlepiej można przedstawić na przykładach.

Zacznijmy od najprostszego – papieru. Na Bliskim Wschodzie, podobnie jak w Europie, papier nie był znany. Używano drogich, a przez to ograniczonych w zastosowaniu materiałów, przede wszystkim pergaminu i papirusu. Gdy w połowie VIII wieku podbito Azję Środkową, nowi władcy zetknęli się z chińskim wynalazkiem – papierem wytwarzanym z drewna. Był on tani i łatwy w produkcji. Chińska produkcja papieru i jego zastosowanie było jednak dość ograniczone. Początkowo – w II i I wieku p.n.e. – wyrabiano z niego chusteczki bądź używano do pakowania, np. podarunków¹. Znacznie później, bo dopiero za panowania dynastii Tang (618–907), dzięki temu, że udoskonalono technikę produkcji papieru, zaczęto go używać wyłącznie jako materiału piśmiennego.

Arabowie, kiedy już poznali technikę wytwarzania papieru, stanęli przed poważnym problemem – z czego go wytwarzać. Chińczycy używali tyka różnych powszechnych w Chinach roślin, czyli konopi, rattanu, bambusa, morwy. Tych roślin nie było w świecie islamu. Zastosowano więc inną technikę i inne materiały, a mianowicie zaczęto wytwarzać papier z odpadów powszechnie używanych tkanin bawełnianych i lnianych.

Istnienie jednego arabsko-muzulmańskiego imperium stworzyło warunki do upowszechnienia papieru na wielkim obszarze, od granic Chin po arabską Hiszpanię. Już w X wieku w arabskiej Asz-Szabatibie (dzisiejsza Jativa koło Almerii) działały młyny papierowe produkujące papier. Stąd stała otworem droga na północ – do Europy. Arabski wynalazek szybko się upowszechnił i już w połowie XII wieku produkowano „papier europejski”. Wraz z produkcją przyszła arabska terminologia, np. nasz wyraz **ryza** (papieru) pochodzi od arabskiego słowa *rizma* – wiązka, plik².

Upowszechnienie papieru jako materiału piśmiennego w kalifacie miało rewolucyjny wpływ na rozwój cywilizacji arabsko-muzulmańskiej. Zrodziła się kultura książki. Wynikało to przede wszystkim z potrzeb nauki.

¹ Jonathan M. Bloom: *Paper Before Print. The History and Impact of Paper In the Islamic World*. New Haven 2001, s. 34.

² Wacław Przemysław Turek: *Słownik zapożyczeń pochodzenia arabskiego w polszczyźnie*. Kraków 2001, s. 348.

Głód wiedzy w połączeniu z łatwym dostępem do materiałów piśmiennych, a także rozwijający się ruch tłumaczeniowy i naukowy spowodował, że książka stała się łatwo dostępna. Na mużulmańskich bazarach nie tylko sprzedawano książki, ale też masowo je produkowano — powstała klasa zawodowych przepisywaczy, kopistów, którzy na zamówienie mogli przygotowywać kopie każdego dzieła. Ciekawe, że przepisywaniem dzieł trudnili się dla zarobku także mużulmańscy uczeni. Szybko też przy akademiach zaczęły powstawać biblioteki, zwane skarbnicami mądrości.

Powstanie cywilizacji papieru przyczyniło się do rozwoju książki i nauki, natomiast nie miało dużego wpływu na rozwój dokumentalistyki. W świecie mużulmańskim w zasadzie nie powstawały wielkie archiwa

صنعة الإِسْزَا
 ناليفه
 الشيخ أبي العباس محمد القاقلشندي

(حقوق إعادة طبعه محفوظة لدار الكتب المصرية)

طبع
 بمطبعة دار الكتب المصرية بالقاهرة
 ١٩٢٢ - ١٣٤٠ م

Rysunek 1. Faksymile współczesnego wydania *Poranka krótkowidza* (Kair 1922). Strona tytułowa

	Arabskie	Indyjskie
0	٠	०
1	١	१
2	٢	२
3	٣	३
4	٤ ٤	४
5	٥ ٥	५
6	٦ ٦	६
7	٧	७
8	٨	८
9	٩	९

Rysunek 2. Cyfry indyjskie i arabskie³

dokumentów, za to wszystkie dokumenty były rejestrowane w książkach. Świadczy to o istotnej roli książki w kulturze mużulmańskiej. To książka była autorytatywna, a dokument — nie. Przykładem takiej księgi-archiwum może być wielka 13-tomowa encyklopedia sztuki pisania dokumentów dworskich z XIV wieku, zawierająca setki dokumentów od początku istnienia islamu. Jest nią dzieło o wymyślnym tytule *Poranek krótkowidza, czyli sztuka prowadzenia kancelarii* (*Subh al-asza fi sina'at al-insza*) Al-Kalkaszandiego (zm. 1418).

Mimo że papier nie był wynalazkiem Arabów, to jednak w imperium mużulmańskim udoskonalono jego produkcję, zmieniono przeznaczenie i nadano mu o wiele większą funkcjonalność niż w Chinach.

Podobnie było z cyframi, powszechnie nazywanymi arabskimi. Oczywiście cyfry zapożyczono z Indii, o czym świadczy stosowany do dziś w piśmie arab-

³ Cyfry 4, 5, 6 utrwały się w dwóch wariantach: wschodnim (perskim) — po prawej, i zachodnim (arabskim) — po lewej.

skim sposób zapisywania liczb – od lewej do prawej, choć teksty pisze się od prawej do lewej. Jednak już sama nazwa **cyfra** wywodzi się z arabskiego słowa *sifr* – zero. Zastosowanie cyfr do zapisywania liczb udoskonalono i, mimo że operacje wykonywane na liczbach nazywały się w średniowiecznym świecie muzutmańskim rachunkiem indyjskim, to metody obliczeń wynaleziono i udoskonalono tutaj. A dokonał tego wcale nie Arab, lecz prawdopodobnie Turek lub Irańczyk pocho-

dzący z Azji Środkowej, noszący przydomek Al-Chuwarizmi (zm. ok. 850) – Chorezmijczyk. Od jego nazwiska wywodzi się wyraz **algorytm**⁴, czyli sposób dokonywania operacji na liczbach. Algorytmem tacinicy nazywali opracowane przez niego sposoby wykonywania działań arytmetycznych – dodawania, mnożenia i dzielenia, przy czym pierwotnie dodawano, mnożono i dzielono liczby, zaczynając od strony lewej, a nie jak dzisiaj – od prawej.

Al-Chuwarizmiemu zawdzięczamy również stworzenie nowej dziedziny matematycznej – algebry. Elementy tego, co dziś nazywamy algebrą, istniały już w starożytności. Jednakże to właśnie Al-Chuwarizmi wyodrębnił algebrę, usystematyzował ją i opisał podstawowe działania prowadzące do znalezienia niewiadomej, zwanej po arabsku *szaj*, czyli *coś*. To *szaj* zapisywano po hiszpańsku jako *xei*, bo w średniowiecznej hiszpańskiej tacinie *x* czytano jako *sz*⁵. Z kolei od nazwy podstawowej w ówczesnej matematyce operacji algebraicznej – przenoszenia ujemnych liczb na przeciwną stronę równania ze znakiem dodatnim – pochodzi nazwa algebra, po arabsku [al-] *dżabr*⁶, czyli *nastawianie złamanych kości*. W ówczesnej matematyce nie umiano bowiem przeprowadzać działań na liczbach ujemnych i dlatego trzeba się było ich pozbyć z równania. Inną operacją, którą opisuje Al-Chuwarizmi, jest *mukabala* (zestawiania, porównywanie), która oznacza zestawianie tych samych elementów równania, czyli upraszczanie. Dzieło Al-Chuwarizmiego poświęcone algebrze nosi tytuł *Kitab al-dżabr wa-al-mukabala i porównywania*. Cel, jaki sobie w nim stawia Al-Chuwarizmi, jest na wskroś praktyczny – chodzi o obliczanie podziału spadku. Zgodnie z szariatem (prawem muzutmańskim) dzielenie spadku było bardzo skomplikowane i wymagało stosowania równań drugiego stopnia, które Al-Chuwarizmi opracował i objaśnił sposób ich rozwiązania.

Oto jak Al-Chuwarizmi rozpoczyna swoje dzieło: *Ułożyłem zwięzłą książkę o algebrze i al-mukabali obejmującą to, co szczegółowe i to co ogólne w arytmie-*

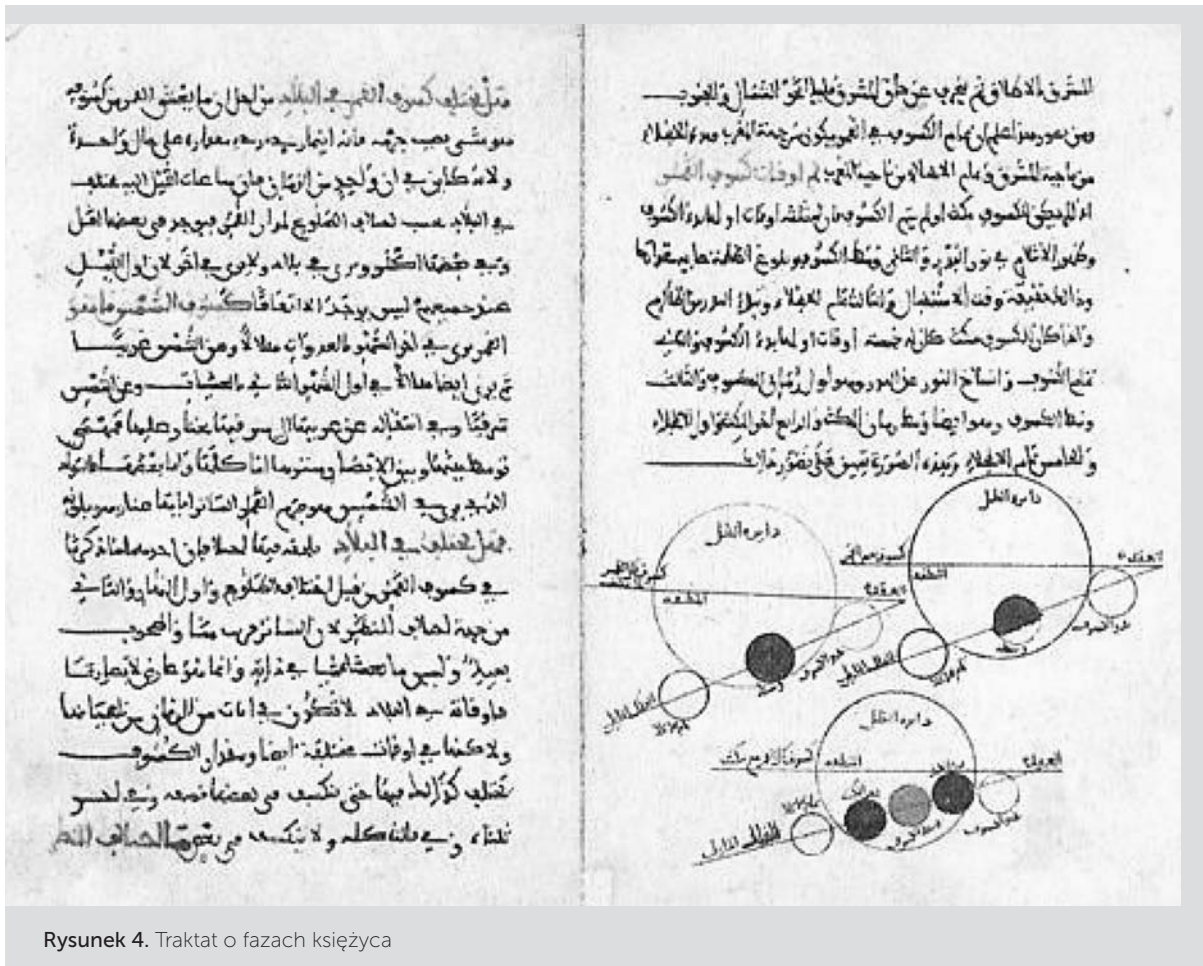


Rysunek 3. Strona tytułowa jedyne zachowanego arabskiego rękopisu *Kitab al-dżabr wa-al-mukabala* Al-Chuwarizmiego (1342, Bodlejana, Hunt. 214)

⁴ Podobieństwo między algorytmem a nazwiskiem Al-Chuwarizmi jest oczywiste. W tacinie arabski dźwięk *ch* oddano przez *g*, a nazwę *algorithmi* uznano za liczbę mnogą wyrazu *algorithmus*. Zapis za pomocą *th* powstał w wyniku kontaminacji z greckim wyrazem *arithmós* – *liczba*. Stąd pochodzi polski wyraz *algorytm* (początkowo *algorytmus*).

⁵ Na upowszechnienie się tego zapisu mogło mieć wpływ również i to, że po grecku od *x* zaczyna się wyraz *xénos* – *obcy, nieznan*. Al-Chuwarizmi używa innych terminów o indyjskiej proveniencji: *mal* (*majątek*) na oznaczenie kwadratu niewiadomej, *dżuzr* (*korzeń*) na pierwiastek drugiego stopnia oraz *dirham* (*drachma*) na oznaczenie konkretnych liczb naturalnych.

⁶ *al-* – to arabski rodzajnik określony, odpowiednik angielskiego *the*.



Rysunek 4. Traktat o fazach księżyca

tyce. Są to sprawy potrzebne ludziom w związku ze spadkami i testamentami, z podziałami, sprawami sądowymi i handlem oraz ze wszystkim, co dotyczy po miaru ziemi, wytyczania kanałów, miernictwa i innych temu podobnych spraw.⁷

Prawie na samym początku *Kitab al-džabr wa-al-mukabala* pojawia się wzorcowe równanie:

$$x^2 + 21 = 10x$$

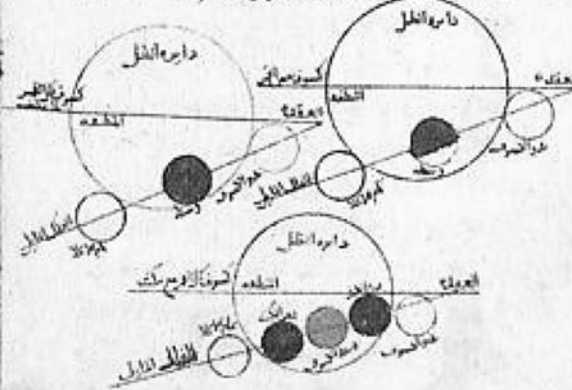
i jego opis:

To znaczy, że kwadrat liczby (majątek) powiększony o 21 dirhamów⁸ daje tążcnie 10 pierwiastków tego kwadratu.

⁷ *Kitab al-džabr wa-al-mukabala*, Kair 1968, s. 16.

⁸ *Dirhamami*, czyli *drachmami*, Al-Chuwarizmi nazywa liczby naturalne.

المشرق فلا حلق لهم فيرى عن حلق المشرق على نحو الشمال والجنوب
 من مودع من العلمان تمام الكسوف في اليوم يكون من جهة المغرب وهو الكسوف
 من ناحية المشرق في علم العلماء من ناحية المغرب لم اوقات كسوف الشمس
 اوله يمشي الكسوف في كل يوم سبع الكسوف من مختلفه اوقات او لما يروى الكسوف
 ويظهر الانعام في يوم اربعه واثني عشر الكسوف يولد في الكسوف على ما يمشي
 وهذا الكسوف في وقت الاستقبال والانتقال والقبول والاربعه من الكسوف
 والاربعه من الكسوف في كل يوم خمسة اوقات او لما يروى الكسوف في كل يوم
 ثلثه الكسوف واستراح النور عن مودع مودع الكسوف في كل يوم
 ونظ الكسوف ومودع ايضا ونظ زمان الكسوف والاربعه من الكسوف والاربعه
 والاربعه من الكسوف في كل يوم في كل يوم في كل يوم في كل يوم



Rozwiązanie tego typu zadań jest następujące:

- podziel liczbę pierwiastków na pół, a otrzymasz pięć,
- pomnóż to przez siebie, a otrzymasz dwadzieścia pięć,
- odejmij od tego dwadzieścia jeden, które jest przy kwadracie — zostanie cztery,
- weź od tego pierwiastek — otrzymasz dwa,
- odejmij je od połowy pierwiastków, czyli pięciu, a otrzymasz trzy.

To jest pierwiastek kwadratu, którego szukasz, a kwadrat wynosi dziewięć.

(Zamiast odejmować) możesz też dodać ten pierwiastek (czyli dwa) do połowy pierwiastków (czyli pięć),



Rysunek 5. Tablice astronomiczne z XIII wieku

a otrzymasz siedem. Wyjdzie siedem i to jest pierwiastek, którego szukasz, a jego kwadrat wynosi czterdzieści dziewięć.⁹

Podobnie jak algebrę rozwinięto również trygonometrię. Ponownie muzulmańskim matematykom przyświecały cele bardzo praktyczne, a nierzadko i religijne. Otóż muzulmanin powinien się modlić pięć razy dziennie zwrócony twarzą ku świątyni Al-Kaba w Mekce. Ten kierunek trzeba było określać w każdym rejonie świata muzulmańskiego, a do tego przydatna okazała się trygonometria.

Innym religijnym zastosowaniem był księżycowy kalendarz muzulmański, oparty na obserwacji faz księżyca, określający m.in. pory modlitwy oraz początek i koniec muzulmańskiego postu. I tu znów przydatna okazywała się trygonometria.

Już samo nazewnictwo wskazuje na źródło muzulmańskiej trygonometrii, którym znów są Indie. Po arabsku *sinus* nazywa się *dżajb*, co bezpośrednio wywodzi

się z sanskryckiego słowa *dźiva*, oznaczającego cięciwę. Ponieważ zupełnie przypadkowo arabskie słowo *dżajb* oznacza także kieszeń, więc po tacinie tłumacze użyli terminu *sinus* – czyli kieszeń.

Jak wszystkie inne dziedziny nauki, trygonometrię zaczęto poszerzać, uzupełniać dziedzictwo poprzedników, a przede wszystkim znajdować dla niego praktyczne zastosowania.

Niemniej trygonometria, jako nauka praktyczna, jest wynalazkiem czysto muzulmańskim, a jej zastosowanie nie ograniczało się do obliczania kierunku odprawiania modłów. Przede wszystkim znalazło zastosowanie w astronomii, geografii i nawigacji. Praktyczne znaczenie miały różnego typu tabele astronomiczne i trygonometryczne.

Jednym z najpiękniejszych, choć z pewnością anegdotycznych przykładów upowszechniania przez świat arabski i muzulmański wiedzy czerpanej ze wszystkich możliwych źródeł jest historia niepozornej cegły lepionej przez tysiąclecia z mułu nilowego i stomy.

W języku staroegipskim występuje słowo czytane jako *db.t*. Oznacza ono suszoną na słońcu cegłę z gliny pomieszczonej ze stomą. W ostatniej fazie rozwoju języka egipskiego, czyli w języku koptyjskim, wyraz ten brzmi *tobe*. Gdy w latach czterdziestych VII wieku Arabowie zajęli Egipt, zachowali istniejącą kulturę (i nieistniejącą Bibliotekę Aleksandryjską), a w tym techniki budowlane starożytnych. Specyficzna cegła egipska stosowana była przez Arabów i w języku arabskim używano jej starej nazwy – *tuba*. Ponieważ po arabsku zwykle dodawano rodzajnik określony *al-*, stąd zwykła arabska forma brzmiała *at-tuba* (z *l* zmienionym w *t* w wyniku asymilacji).

Arabowie nie skończyli swoich podbojów na Egipcie; ruszyli na zachód i już wkrótce znaleźli się w Hiszpanii i południowej Francji. Do Hiszpanii przenieśli swoją kulturę wzbogaconą o elementy, które poznali gdzieś indziej. Między innymi technikę wyrobu cegieł z suszonej gliny. Przyjęta się ona w arabskiej Hiszpanii. Nic dziwnego, że nazwa weszła do języka hiszpańskiego, w którym nazywa się *adobe*. W innych rejonach Europy taka technika nie miała szans utrzymania się ze względów klimatycznych.

Kiedy jednak po 1492 roku Hiszpanie wyruszyli w świat na podbój Ameryki, technika ta okazała się niezwykle przydatna na terenach dzisiejszego Meksyku

⁹ Tamże, s. 20.

i południowych stanów USA, gdzie z cegły *adobe* buduje się do dziś. W XIX wieku z kolei na podbój świata wyruszyli Amerykanie i po zajęciu południa dzisiejszego USA język angielski wzbogacił się o nowy wyraz pochodzenia hiszpańskiego *adobe*.

Kilkanaście lat temu tym wyspecjalizowanym terminem ozdobiła swoją nazwę firma komputerowa Adobe Systems Incorporated założona w południowej Kalifornii w 1982 roku¹⁰. Jej program Adobe Photoshop jest dzisiaj jednym z najpopularniejszych komputerowych narzędzi do obróbki obrazu¹¹.

Ta opowieść o wędrówce niepozornej cegły z Egiptu IV tysiąclecia p.n.e. do dwudziestowiecznych technik komputerowych wskazuje na rolę języka arabskiego na świecie, ale też na ciągłość kulturową naszych cywilizacji. Jak każdy język, arabski obstrugiwał pewną kulturę, a ponieważ była to kultura ruchliwa, jej znaczenie było olbrzymie i nie przestaje ona oddziaływać do dzisiaj. Podobnie jak nasza cywilizacja miała charakter odtwórczy i uzupełniający.

Ten wpływ wcale nie następował przez religię islamu, choć z pewnością również dzięki religii. Religia islamu była bowiem (i w dużej mierze jest nadal) otwarta na przyjmowanie nowych wartości, a zarazem w bardzo prosty sposób narzucała własne, tak jak to było z adobą. Wynikało to z faktu, że poddani muzulmanów chętnie przyjmowali ich pomysły, a nawet ich język i pismo.

Nauka arabska uczyniła coś, co dziś jest na porządku dziennym – teoretyczne marzenia badaczy i myślicieli przetwarzała w praktykę. Robiono to w sposób systematyczny. Najpierw gromadzono wiedzę ze wszystkich dostępnych źródeł, a następnie dokonywano syntezy. W ten sposób wiedza grecka spotkała się w islamie z indyjską, babilońską, chińską i perską. Wszystko rejestrowano w jednym wspólnym języku, a potem zaczęto te różnorodne elementy łączyć w jedną całość i porządkować.

Przede wszystkim jednak nauka arabska była otwarta dla wszystkich – nie tylko swoich, ale także dla obcych, bo przecież od obcych pochodziła. Nigdy i nikomu nie przyszło do głowy, by skąpić dostępu do wiedzy. Dzięki temu każdy mógł ją rozwijać i upowszechniać. Te z pozoru błahe stwierdzenia nie były kiedyś wcale

oczywistością. Częstsza bywała w nauce monopolizacja – odgradzania od nauki większości i skrytego ukrywania wynalazków.

Profesor Janusz Danecki wykładowca w Katedrze Arabistyki i Islamistyki Uniwersytetu Warszawskiego oraz w Centrum Bliskowschodnim Wyższej Szkoły Psychologii Społecznej, znawca problematyki świata arabskiego i muzulmańskiego, autor książek o islamie i języku arabskim, a także tłumacz klasycznej i współczesnej literatury arabskiej. Obszary zainteresowań to językoznawstwo, literaturoznawstwo, religioznawstwo. Członek Towarzystwa Naukowego Warszawskiego (Wydział I Językoznawstwa i Historii Literatury) oraz Komitetu Nauk Orientalistycznych Polskiej Akademii Nauk. Założyciel i redaktor naczelny międzynarodowego czasopisma „Studia Arabistyczne i Islamistyczne” oraz czasopisma „Bliski Wschód. Społeczeństwa–Polityka–Tradycje”.

¹⁰ Nazwa firmy nie pochodzi bezpośrednio od wyrazu *adobe*, lecz od nazwy strumienia Adobe Creek płynącego w Los Altos za domem założycieli firmy (Wikipedia angielska 10.05.2010).

¹¹ O historii wyrazu *adobe* zob. H. Wiesmann: *Adobe*, „Zeitschrift für Ägyptische Sprache und Altertumskunde” 52, 1916.



Semiconductors in 21st Century — the First Decade

Based on the lecture delivered on February 25, 2010*

Jerzy Rużyłto

Penn State University, University Park, PA, USA

159

Introduction

As a class of materials, semiconductors played, and continue to play, an undeniably pivotal role in the explosive growth of our technical civilization over the last six decades. The main driving force behind this growth was the unprecedented progress in digital integrated circuits (IC) technology as described by Moore's law [1].

During recent years, departure from the pattern noted above can be observed. It is no longer that progress in semiconductor technology is driven solely by technical breakthroughs needed to sustain the growth of digital electronics. During the last ten years, the impact of digital electronics is increasingly accompanied by the accelerated growth of distinct, readily identifiable semiconductor technical domains which are only partially related, or not related at all, to logic and memory IC technology.

The purpose of this overview is to identify and to briefly discuss, in layman terms, selected technologies which are perceived to contribute most to the recent expansion of priorities which define state-of-the-art semiconductor science and engineering.

To establish a foundation for the follow up discussion, a brief overview of the fundamental properties of semiconductors, key semiconductor materials as well as their uses is presented. Then, trends in semiconductor

science and engineering are discussed and emerging new directions are identified. Subsequently, a discussion of new generation technology drivers is presented, emphasizing the role of organic semiconductors, nano-ordered semiconductor material systems, carbon electronics, as well as photovoltaics, and MEMS/NEMS devices in defining emerging trends. The review is summarized by stressing the critical role, semiconductors continue to play in supporting high-tech endeavors of the 21st century.

Overview of Semiconductors

For the sake of clarity of the forthcoming discussion, it is appropriate to introduce key concepts related to its scope by answering the following three basic questions.

What Are Semiconductors?

The fundamental electrical property of any solid is its electrical conductivity, i.e. ability to conduct an electric current. As Figure 1 illustrates, insulators and conductors feature very low and very high conductivity respectively, regardless of external conditions such as temperature or illumination. The electrical conductivity of these materials, as predetermined by the nature of interatomic bonds which determine electron's freedom to move within a solid, cannot be altered. Metals, for instance, can only be very good conductors.

* This text is also available on www.semi1source.com/21stCentury

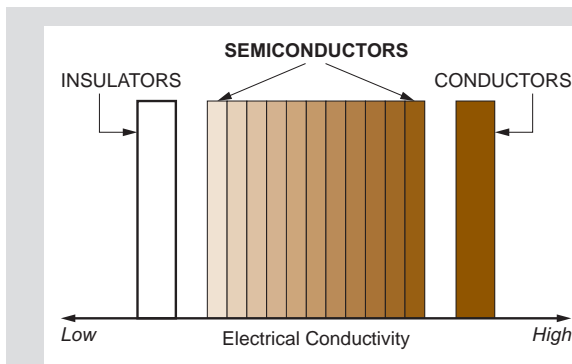


Figure 1. Electrical conductivity of solids.

What distinguishes electrical conductivity of semiconductors from metals and insulators is basically what defines semiconductors as a separate class of solids. First and foremost, in contrast to metals and insulators, elec-

trical conductivity of semiconductors can be controlled by orders of magnitude (Fig. 1) by introduction of alien elements (doping). Furthermore, conductivity of semiconductors can be controlled by two types of carrier: negative electrons or positive holes. In addition, it depends on temperature, illumination as well as electric and magnetic fields. Very importantly, and again in contrast to metals and insulators, when adequately processed, semiconductors can emit visible radiation. These outstanding characteristics allow a myriad of highly functional devices, both electronic and photonic to be made using semiconductors.

What Materials Display Semiconductor Properties?

In the Periodic Table of the Elements (Fig. 2a) a section singled out in Figure 2b is often referred to as a "semiconductor periodic table". All inorganic semiconductors either elemental in group IV, or compound synthesized

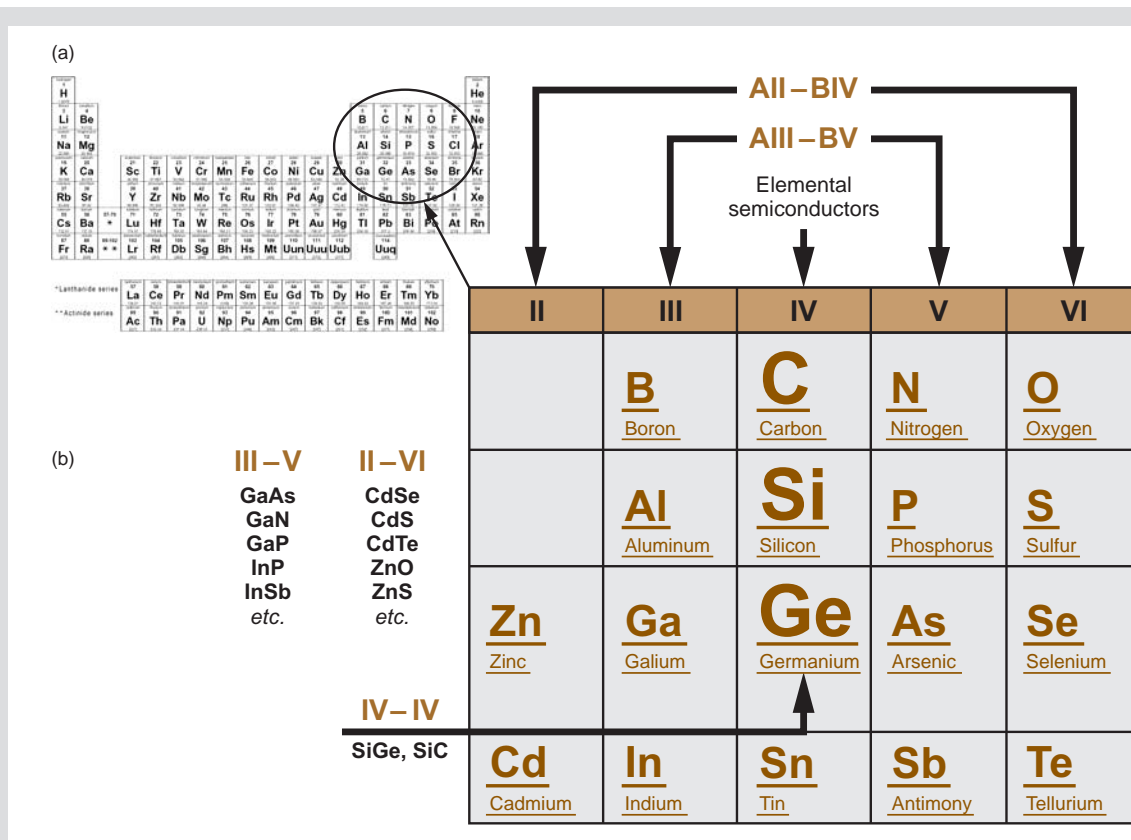


Figure 2. Materials displaying semiconductor properties.

from the elements from the group IV (e.g., silicon carbide, SiC), groups III and V (e.g., gallium arsenide, GaAs), as well as groups II and VI (e.g., cadmium selenide, CdSe) originate from this section of the periodic table.

In addition to a range of inorganic semiconductors identified in Figure 2b, selected organic compounds form a very distinct class of semiconductors featuring unique properties and offering a range of novel applications (see discussion later in this overview).

Semiconductors can be subdivided into classes featuring diversified properties based on the chemical composition, fundamental physical properties as well as the extent of an order in the three-dimensional arrangements of atoms. In this last case ordered (crystalline) and disordered (non-crystalline, or amorphous) semiconductors are distinguished. In addition, the nano-ordered semiconductors featuring a highly ordered structure within extremely confined geometries, such as nanowires, nanotubes, and quantum dots, can be singled out.

Silicon (see Fig. 2b) is by far the most important and the most widely used semiconductor material. Its use continues to grow not only through the needs of ever progressing microprocessor and memory integrated circuit technology, but also through the growing needs of solar cells (photovoltaics) and Micro-Electro-Mechanical Systems (MEMS) applications. Among compound semiconductors, gallium (Ga) compounds with group V elements (Fig. 2b), gallium nitride (GaN) and gallium arsenide (GaAs) in particular, are the most important due to their usefulness in a range of electronic and photonic applications.

How Are Semiconductors Used?

It can be safely assumed that with the exception of the most rudimentary ones, all instruments and any equipment which uses electricity to operate require semiconductor elements to be functional. From the most complex outer space instrumentation, weaponry, or information processing circuitry, to simple everyday tools and gadgets, semiconductors are the foundation upon which the operation of almost everything electronic and photonic is based.

In order to implement the desired operations, semiconductor material is processed into a device which can perform in a controlled and predetermined fashion electronic (e.g. diode, transistor, monolithic integrated

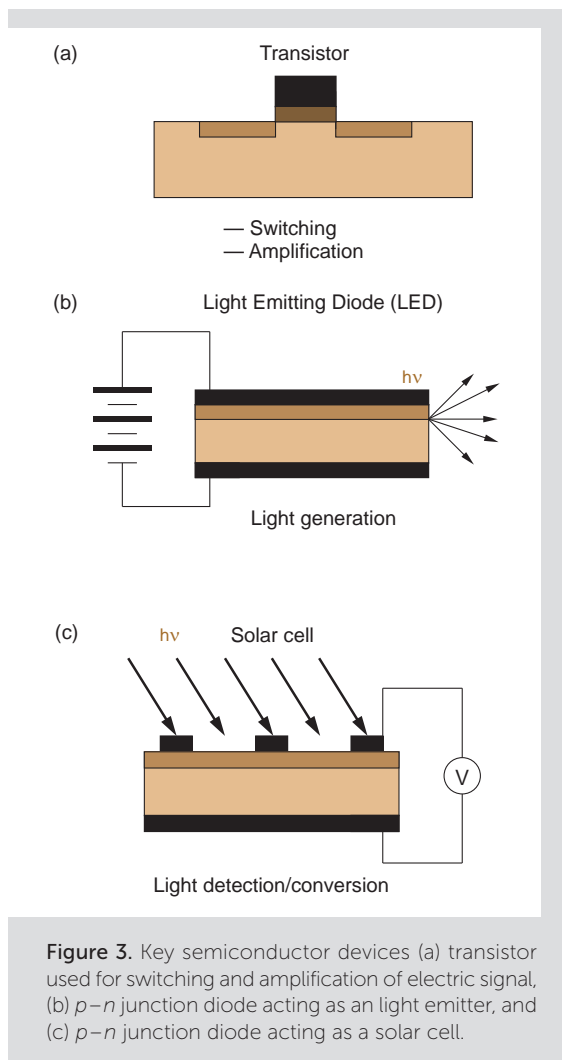


Figure 3. Key semiconductor devices (a) transistor used for switching and amplification of electric signal, (b) $p-n$ junction diode acting as a light emitter, and (c) $p-n$ junction diode acting as a solar cell.

circuit) or photonic (e.g. Light Emitting Diode, or LED, laser) functions. Among many classes of semiconductor devices the selected three, deemed the most important in terms of function, are shown in Figure 3. A transistor (Fig. 3a) is a device designed to amplify and/or switch an electrical signal and as such is a cornerstone of semiconductor device engineering. The transistor is a basic building block of all integrated circuits (IC), and hence, has been developed into by far the most complex and most important semiconductor device. Other than transistor, semiconductor diodes that can be used to either emit light of desired wavelengths (light-emitting diode in Fig. 3b) or convert light into electricity, as it is done in solar cells (Fig. 3c), are the other key semiconductor devices.

Technology Drivers in the 20th Century

Since the invention of the transistor in 1947 followed by the commercial introduction of integrated circuits (IC) some twenty years later, progress in semiconductor technology has been driven primarily by the need to process information faster and more efficiently. Processing information is all dependent upon devices which can be turned "on" and "off" fast enough so that corresponding sequence of "1"s and "0"s can be executed billions of times per second. Since the Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET) is the most effective device in carrying out these functions the progress in IC engineering was driven primarily by the improvements in MOS technology.

To be exact, it is not a MOSFET either *n*-channel or *p*-channel (Fig. 4a and b), but the combination of these two in the Complimentary MOS (CMOS) structure shown in Figure 4c that is the focus of attention. CMOS shows superior characteristics in switching applications including very low energy needed to switch it from "on" to "off" state, and hence, very limited power dissipation, as well as essentially no current in the "off"

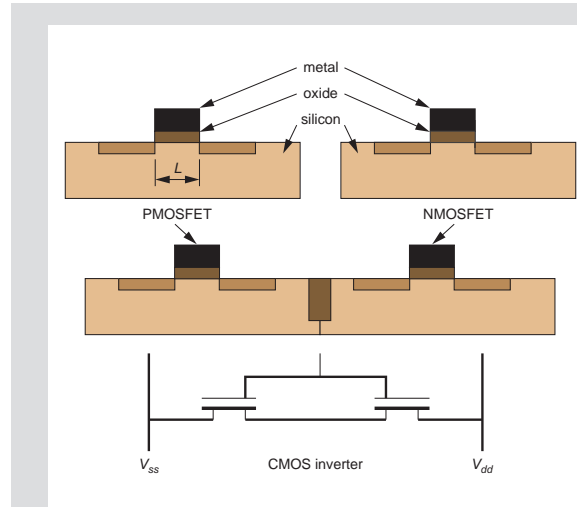


Figure 4. Complementary MOS structure, CMOS – an ultimate electric switch.

state. In more general terms, digital applications involving logic and memory ICs were driving forces behind the dramatic progress in semiconductor science and engineering over the last fifty years.

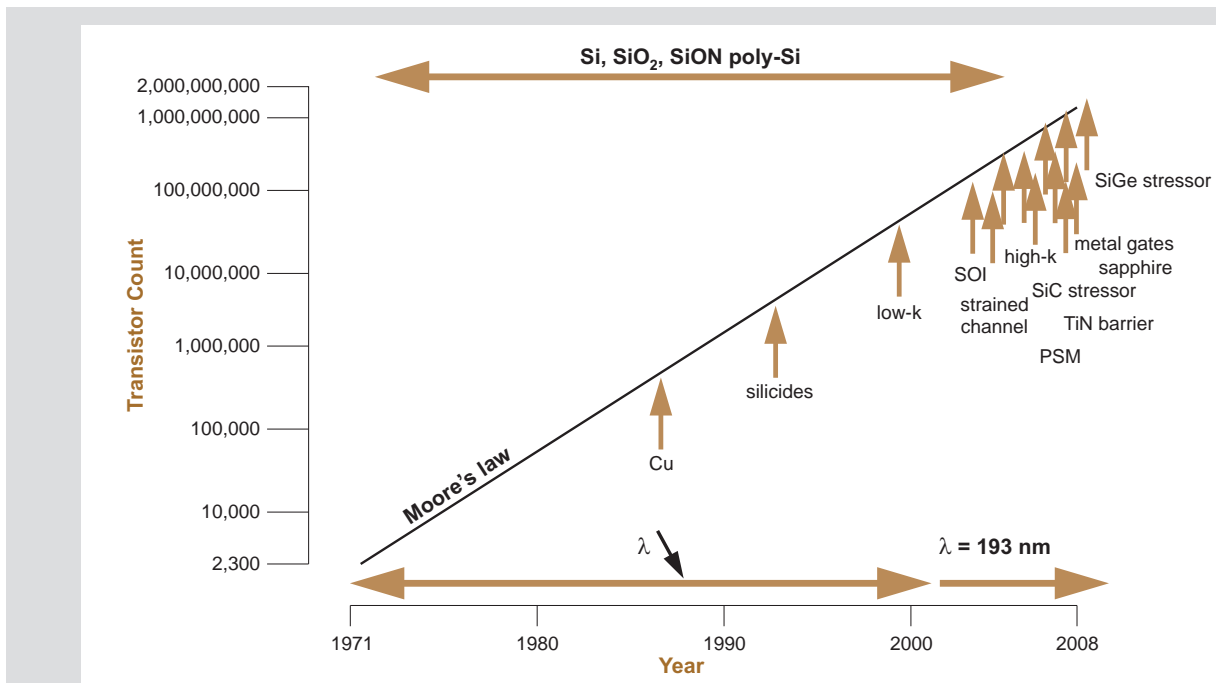


Figure 5. Recent progress in integrated circuit technology is driven primarily by material engineering solutions.

During those years, progress in IC technology expressed in terms of the number of transistors per chip by Moore's law (Fig. 5) was dependent primarily on the continued reduction of transistor's geometry in CMOS cells with scaling down of the channel length L (Fig. 4) being a lead target. Technically, the reduction of transistor geometry was possible because we had ways to keep on reducing wavelengths λ of radiation used for photoresist exposure in the photolithographic processes defining transistor geometry. That was until λ was reduced to the 193 nm emission wavelengths possible with ArF excimer laser. As further reduction of λ was not feasible without overcoming significant technical and cost-related barriers, the 193 nm exposure length remained unchanged (Fig. 5) and alternative ways of improving the performance of the CMOS were pursued.

While all this was happening, not much was changing in terms of materials used to fabricate advanced ICs (with the exception of copper replacing aluminum as an interconnect material, Fig. 5) and device configuration. It was all based on silicon and its derivatives (single-crystal Si, poly-Si, SiO₂, Si₃N₄, SiON), while planar CMOS remained a benchmark in terms of device layout.

In parallel to the rapid growth of the electronic component of semiconductor device technology, impressive progress has also been accomplished at the photonic end of the semiconductor device spectrum. Over several years the driving force in this domain was an effort to develop blue light emitting diodes needed in addition to the longer wavelength LEDs developed earlier. This effort was punctuated by the successful development of blue LEDs formed on InGaN [2].

Era of Materials

As discussed earlier, for over 30 years the progress in semiconductor technology was dependent mainly on the improvements in ICs manufacturing processes in particular photolithography. At the same time, developments that were highlighting the advancements were basically synonymous with pushing the boundaries of high-end logic and memory IC technology toward faster more potent circuits.

During the last decade or so the paradigm has shifted noticeably. With the exposure wavelength used in photolithography remaining unchanged, the progress in CMOS technology is currently accomplished primarily

through innovative materials engineering solutions combined with broader than ever before selection of materials used to process high-end devices. In conjunction with this renewed emphasis on materials, major modifications of the transistor's structure are aggressively pursued.

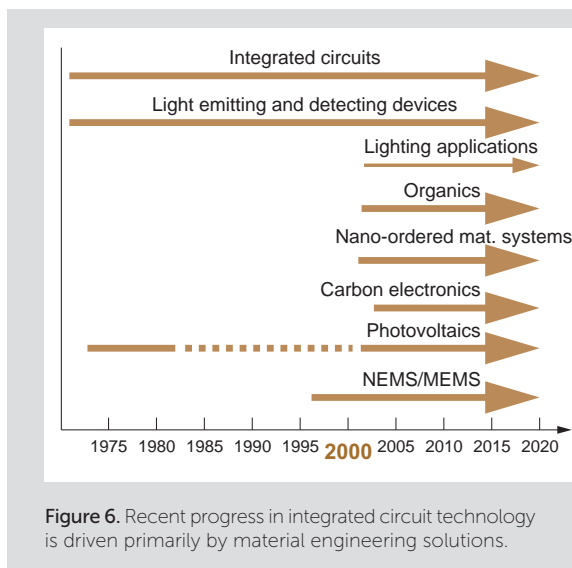
As a result, it is mostly all about the materials and the elaborately configured material systems these days. In fact, the latest process technology related breakthroughs are almost entirely related to materials engineering. Figure 5 shows examples of materials related developments that took place in advanced ICs engineering during the last few years. Introduction of high-k gate dielectrics, accompanied by the need to address a range of challenges (e.g. [3]), use of SOI substrates and the use of stressors in the MOSFET channels are just a few highlights underscoring this trend.

Technology Drivers in the Early 21st Century

Figure 6 is constructed in order to illustrate developmental trends currently observed in semiconductor engineering. As the horizontal bars in this figure indicate, integrated circuit technology continues to play pivotal role in a key technology driver. The selection of materials used, besides silicon, is growing continuously, device design evolves and the focus is shifting from sheer speed and power efficiency to more application specific solutions based on the range of materials and transistor layouts available.

At the photonics end of the spectrum, light emitting and detecting devices clearly continue as a self-contained technical domain with growing impact on an overall semiconductor business particularly through the increased role of semiconductor devices in display technology. The new emphasis in LED technology, which emerged during the last decade as a major force, is an urgent need to use LEDs in lighting applications as a replacement for highly inefficient incandescent and fluorescent bulbs. In this respect, the development effort spreads over a range of device solutions including inorganic and organic LEDs as well as quantum dot based light emitters.

During the last decade, several new or renewed, applications of semiconductors have emerged as major players (see Fig. 6), and hence can be seen as technology



drivers in the early 21st century. Those emerging areas of semiconductor science and engineering are briefly reviewed below. It needs to be emphasized that the selection of those areas is based entirely on the Author's assessment of the observed trends and as such is open for discussion. Another point which should be stressed is that several among new technologies are pursued to overcome roadblocks digital IC technology will be facing in the future and not necessarily as self-contained technical domains.

Organic Semiconductors

Until recently, all semiconductor materials used in commercial applications originated from the semiconductor series of the periodic table (Fig. 2) and were inorganic in nature. A new breed of cheap to process organic semiconductors, i.e. materials consisting primarily of carbon and hydrogen, show good promise in expanding applications of semiconductors into the areas in which conventionally thin-film inorganic semiconductors can not be used. An important factor distinguishing organic, or "plastic", semiconductors from their organic counterparts is that the former maintain their fundamental physical properties even if drastically bent. This characteristic opens up possibility of formation of low-cost functional semiconductor devices on flexible substrates.

From the point of view of physical properties, organic semiconductors differ in several ways from their con-

ventional, inorganic counterparts. Just like other semiconductors, however, they do allow control of charge distribution using external electric field and they do have the capability to emit visible radiation. Consequently, several semiconductor device structures both photonic (LEDs, solar cells) and electronic (Thin-Film Transistors, TFTs) can be implemented, using select organic compounds.

The fundamental difference between inorganic and organic semiconductors is in the charge transport mechanism. In the first case, electrons moving in wide bands as delocalized plane waves are subjected to very limited scattering, and hence feature relatively high mobility (e.g. $\sim 1.5 \times 10^3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ in Si at room temperature). In the case of organic semiconductors, the charge transport is based on carriers hopping between localized states associated with organic molecules. In the process electrons undergo significant scattering which result in very low electron mobility in organic semiconductors ($\sim 1\text{--}3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$).

The best promise for a breakthrough application that organic semiconductors show is in flexible displays as well as printed electronics and photonics. In the first case, organic LEDs (OLEDs) (e.g. [4]) and TFTs, both needed to engineer high quality active matrix full-color display, are the center of attention. In OLEDs, light generation is based on the decay to the ground state of excitons formed through interactions between electrons and "holes" rather than through the band-to-band recombination as is in the case of conventional LEDs. In an organic TFT, similarly to conventional TFTs, transistor action is based on the field effect in which potential applied to the gate contact changes the conductivity of organic semiconductor between source and drain.

Nano-Ordered Semiconductor Material Systems

As current technology allows us to manipulate a solid matter at the atomic and molecular levels, the formation and use of nano-scaled ($\text{nm} = 10^{-9} \text{ m}$; size of the atoms varies depending on the element from 0.2 nm to 0.5 nm) material systems has become entirely feasible. Due to their physical properties, being distinctly different from their bulk counterparts, the nano-ordered (see earlier discussion) semiconductor material systems offer a range of novel applications and have a potential for addressing challenges semiconductor technology, including digital ICs, will face in the future. As a result, nano-

-ordered semiconductors have emerged during the last decade as one among major technology drivers (Fig. 6), partially because of the role they are expected to play in resolving limitations of future generations digital electronics.

The starting point to understanding nano-ordered material systems is recognizing the fact that the fundamental properties of a solid confined to ultra-small geometries are different than the properties of the exact same solid in bulk form. Figure 7a shows a crystalline bulk material, silicon for instance, featuring well defined physical characteristics governed by the laws of classical physics. It is enough, however, that its one dimension is reduced to close to zero, effectively forming a few-atoms-thick silicon sheet (Fig. 7b) that its properties change significantly. This is because 2D-confinement alters the distribution of energy levels which can be occupied by electrons in the atom, and hence changes fundamental physical properties which now are defined by the laws of quantum physics. The same phenomena are further reinforced in the case of 1D confinement resulting in nanowire formation (Fig. 7c) and zeroD confinement leading to the formation of quantum nano-dots (Fig. 7d). Those last show dependence of the bandgap on the size of the dot allowing tunability of basic properties of the material by changing the size of the dot. This particular characteristic is very useful especially in photonic applications (e.g. [5]).

The ability to form nano-ordered semiconductor structures shown in Figure 7 opened up new areas of application of semiconductor technology and established this particular technical domain among the key tech-

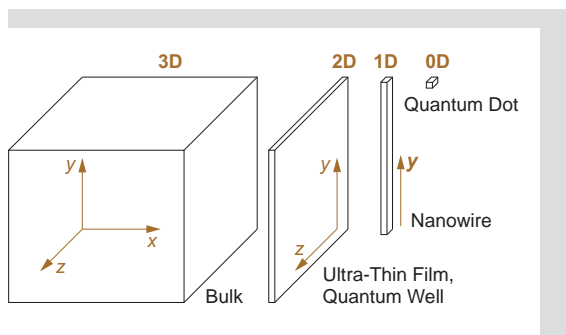


Figure 7. Physical properties of semiconductor material change as its geometry is scaled from 3D to 0D configuration.

nology drivers in semiconductor technology in the early 21st century.

Carbon Electronics

Due to manufacturability related issues, lack of substrates as well as lack of effective *n*-type dopants [6], bulk carbon in the highly ordered crystalline form (diamond) did not fulfill its promise of an excellent semiconductor. However, in the nano-ordered form, a range of carbon-specific material structures show great promise in a number of key applications.

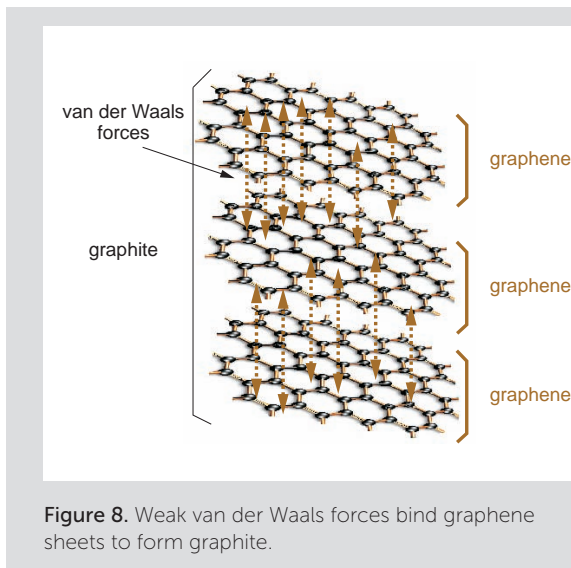
Among the most exciting forms of nano-ordered carbon, is its two dimensional version known as graphene. This unique structural configuration of pure carbon has the potential to have a real impact on how we will build active semiconductor devices in the future.

In terms of crystallographic structure, graphene is a two-dimensional part of the three-dimensional graphite, i.e. one-atom-thick planar sheet of carbon that is connected to adjacent sheets by relatively weak van der Waals forces (Fig. 8). Effectively then, graphite is a thick stack of graphene sheets each internally covalently bonded (a very strong bond) in the planar hexagonal configuration, but which are relatively weakly bonded between each other.

One of the many remarkable characteristics of this ultimate 2D material structure is that graphene displays excellent semiconductor properties. Particularly attractive is its very high carrier mobility in excess of 150,000 cm²/Vs. Interestingly, early observations based on conductance measurement suggests that in graphene electron and holes mobilities should be almost the same. This is in contrast to conventional 3D semiconductors (Fig. 7a) in which electrons feature significantly higher mobility than holes.

Regarding electronic applications, graphene can be used to form ultra-fast transistors. A graphene-based transistor featuring 100 GHz cut-off frequency has been recently demonstrated [7].

Graphene can be obtained by detaching, atomic layer by atomic layer, sufficiently large pieces of graphene from graphite and transferring them to other substrates or, by far more effectively, by growing it on silicon carbide substrates. Rolled into a cylinder, graphene will form a single-walled carbon nanotube which forms another unique carbon configuration with potentially important applications in semiconductor device manufacture.



Photovoltaics

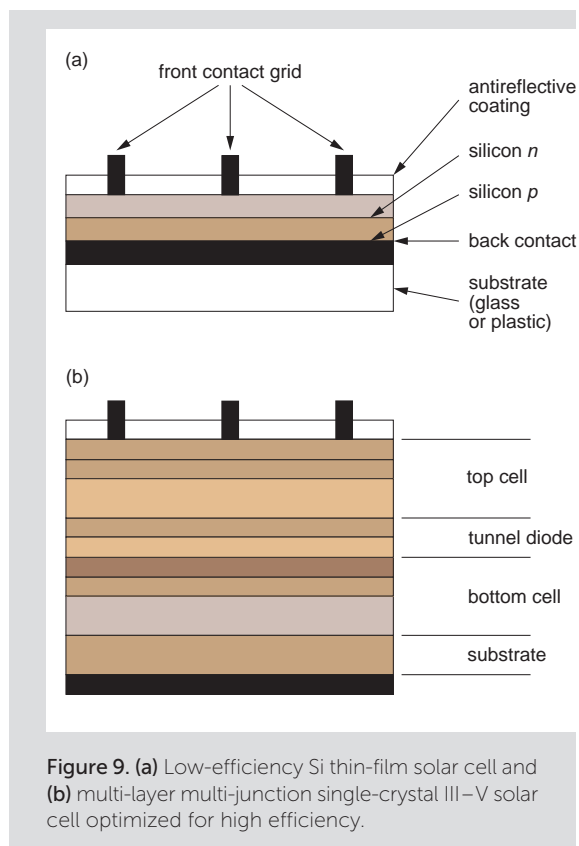
Photovoltaics constitutes yet another technology driver in state-of-the-art semiconductor science and engineering (Fig. 6). The photovoltaic effect is the effect based on what solid state devices convert light into electricity. This effect is by far the most efficient when such devices are made out of semiconductors. Hence, the term “photovoltaics”, which refers to the technical domain concerned with the conversion of solar energy into electricity using photovoltaic effect, is essentially synonymous with semiconductor solar cells.

The photovoltaic effect was experimentally observed some 150 years ago [8]. Since then, during the last fifty years in particular, commercial attempts to use it as a vehicle allowing direct conversion of solar light into electricity were highly dependent upon cost and balance between supply and demand for energy derived from fossil fuels. The interest was always growing during energy crises, e.g. in the early 1970’s (Fig. 6) then dissipated after the oil market regained its balance.

The above scheme was decidedly altered during the last decade when, sparked by yet another energy crisis, the world-wide push toward the alternative to fossil fuel sources of energy firmly established photovoltaics as a rapidly growing segment of semiconductor industry. Currently, a large number of commercial enterprises manufacturing solar cells are fully functional around the world.

It should be emphasized that, although shaped and processed differently, solar cell fabrication involves the same manufacturing methods and materials as most of other semiconductor devices. With the technology and knowledge available today the efficiency with which solar cell converts light into electricity is proportional to the cost of its fabrication, which in turn depends on the cost (quality) of materials used and complexity of manufacturing processes employed. Based on this paradigm, various classes of solar cells can be distinguished.

In terms of materials used, silicon dominates solar cells industry while the selection of technology roughly comes down to the choice between thin-film solar cells using non-crystalline silicon or CdTe for instance (solar cells based on organic semiconductors belong to this class), and crystalline semiconductor, again mainly silicon, in the form of very thin square substrate wafers. The former offers low-cost cells, but at the expense of efficiency. The latter produces more efficient cells, but at a higher cost. Figure 9 shows schematic diagram of the typical amorphous silicon thin-film solar cell



featuring efficiency in the 6–8% range (Fig. 9a) and single-crystal multi-layer III–V based solar cell featuring efficiencies approaching 40% (Fig. 9b). Both approaches are needed and the choice depends on the type of application. Regardless of technology/materials used, photovoltaics is certainly among key technologies determining how semiconductors are used now and will be used in the future.

MEMS/NEMS

Micro-electro-mechanical systems (MEMS) and follow-up generations of nano-electro-mechanical systems (NEMS) integrate mechanical and electrical functions using a somewhat modified, but otherwise standard CMOS IC manufacturing procedure. Such functional integration is possible only because silicon, besides discussed earlier advantageous electrical and cost/manufacturing related properties, also features outstanding mechanical properties. This combination is unique to silicon and cannot be reproduced using any other material.

As shown in Figure 6, the growth of MEMS technology took off several years ago and is expected to continue to grow in years to come, driven by the increasing complexity and functionality of MEMS/NEMS components of the complete Systems-on-Chip (SOC). Micro- and nano-sensors, including bio- and chemosensors, actuators, gyroscopes, accelerometers, etc. are just a few examples of MEMS devices.

While based on the Si IC manufacturing technology, the MEMS fabrication processes are governed by needs which in some aspects require specialized approaches. For instance, very deep etches performed on silicon in the course of MEMS device manufacture need specialized dry etching tools. Also, MEMS release processes, not used at all in conventional IC manufacturing, call for oxide etching in intricate lateral geometries and as such require development of innovative solutions specific to MEMS manufacturing [9].

In summary, the ultra-small mechanical devices integrated on the same chip with controlling them electronic circuitry, constitutes yet another technical domain which continues to drive progress in semiconductor science and engineering. Further growth of this segment of semiconductor technology is inevitable and is expected to have a growing impact both technically and commercially.

Concluding Remarks

This overview attempted to identify trends which emerged as forces driving progress in semiconductor science and engineering in the early 21st century. The discussion presented leads to the conclusion that semiconductors, now more than ever before, continue to have a major impact on the evolution of our technical civilization. It is also quite evident that, although several developments continue to be driven by the long-term needs of digital electronics, there is a range of new technologies which open up new areas of application for semiconductor materials well beyond today's uses. Some of them were briefly discussed in this overview.

Finally, it should be once again pointed out that this overview is concerned with new developments on semiconductor arena which are either commercialized already or are on the direct path to commercialization in the near future. Developments in several emerging domains such as spintronic or use of semiconductors in broadly understood "bio" applications, are left to the future discussion. Their impact will be judged on the basis of the future success of the process of conversion of the theoretical concepts into practical solutions.

Bibliography

- [1] G.E. Moore, "Cracking More Components Onto Integrated Circuits", *Electronics*, 38, 8, (1965).
- [2] S. Nakajima, M. Senoh, M. Iwasa, and S. Nagahama, "High-Brightness InGaN Blue, Green and Yellow Light Emitting Diodes with Quantum Well Structures", *Japan. J. Appl. Phys.*, 34, L797 (1995).
- [3] K. Chang, F.M. Chang, and J. Ruzyllo, "Charge Trapping in HfO₂ and HfSiO₄ MOS Gate Dielectrics", *Solid-St. Electronics*, 50, 1670–1672 (2006).
- [4] K. Shanmugasundaram, S.C. Price, W. Li, H. Jiang, Q.K. Wang and J. Ruzyllo, "Studies of Mist Deposition in the Fabrication of Blue Organic Semiconductor Light Emitting Diodes", *Semiconductor Science and Technology*, 23, 075036 (2008).
- [5] Z. Tan, F. Zhang, T. Zhu, J. Xu, A.Y. Wang, J.D. Dixon, L. Li, Q. Zhang, S.E. Mohny, and J. Ruzyllo, "Bright and Color-Saturated Emission from Blue Light Emitting Diodes Based on Solution-Processed Colloidal Nanocrystal Quantum Dots", *NanoLetters*, 7, (12), 3803–3807 (2007).
- [6] A. Badzian, "The Status of Diamond Electronics", www.semi1source.com
- [7] Y.-M. Lin, C. Dimitrakopoulos, K.A. Jenkins, D.B. Farmer, H.-Y. Chiu, A. Grill, Ph. Avouris, "100-GHz Transistors from Wafer-Scale Epitaxial Graphene", *Science*, 327, 5966, 662 (2010).

- [8] A.E. Becquerel, "Mémoire sur les effets électriques produits sous l'influence des rayons solaires". Comptes Rendus, 9: 561–567 (1839).
- [9] M. Erdamar, K. Shanmugasundaram, P. Roman, K. Chang, M. Klimkiewicz, and J. Ruzyllo, "Deep Lateral Anhydrous HF/Methanol Etching for MEMS Release Processes", Journal of Micro- and Nanolithography, MEMS and MOEMS, 7(3), 1 (Jul.–Sep. 2008).

Profesor Jerzy Rużyłło wybitny specjalista w dziedzinie mikro- i nanotechnologii materiałów i przyrządów półprzewodnikowych. Związany z Penn State University, USA, obecnie jako *Distinguished Professor* w Department of Electrical Engineering i Department of Materials Science and Engineering. Obszary badań naukowych to m.in. metody i narzędzia technologii zaawansowanych układów scalonych, osadzanie cienkich warstw nanokropek kwantowych, a także charakteryzacja materiałów elektronicznych i fotonicznych. Autor i współautor ponad 250 artykułów i referatów konferencyjnych, książki *Semiconductor Glossary* oraz sześciu patentów. Utworzył i prowadzi strony internetowe poświęcone inżynierii półprzewodników. Honorowy członek (*Fellow*) międzynarodowych organizacji naukowych IEEE (*Institute of Electrical and Electronics Engineers*) oraz Electrochemical Society.

Utrzymanie Życia jako podstawowa wartość przestrzeni Miast

Na podstawie odczytu wygłoszonego w dniu 28 października 2010 roku*

Marek Budzyński

Wydział Architektury Politechniki Warszawskiej

W tym wystąpieniu słowa *Kultura*, *Natura*, *przeciwieństwo*, *jedność*, *proces* są kluczowe.

Używam ich jako:

Kultura — całokształt utrwalonego dorobku człowieka i procesów go tworzących.

Natura — całokształt procesu *Stwarzania* bez udziału człowieka.

Kultura niewątpliwie wykreowana w *Naturze* stanowi jej przeciwieństwo.

Jedność przeciwieństw jest w kategoriach dalekowschodnich naturalną cechą procesu *Stwarzania* (*Ewolucji*).

Zacznę od „oczywistej oczywistości”:

Człowiek — obywatel miasta jest najwyższą wartością Miasta. Przestrzeń, tworząca warunki podtrzymujące życie obywatela, jest dla niego wartością podstawową, bo umożliwia mu życie, które dla każdego z nas jest bezcenne.

Pieniądże to genialny wynalazek ludzkości. Wynalazek, którego celem jest, wraz z rynkiem, ustalenie wartości rzeczy niewspółmiernych. Ale ani pieniądze, ani rynek nie są w stanie określić takich wartości jak Życie, a w nim Miłość, Wiara, Nadzieja. Nie kwestionowaną wartością dla każdego z nas jest nasze życie. Większość z nas ma Nadzieje lub Wierzy, że nie kończy się ono tu i teraz. Wierzymy również, że Miłość otwiera nowe życie, jak również nadaje sens istniejącemu. Staramy się poznać tajemnice Życia. Teologia, filozofia, fizyka i bioelektronika coraz częściej wracają do pojęcia Jedni, definiowanej już przez Demokryta, wywodzącej powstanie przyrody, a w niej życia, z pojęcia „boska cząstka”

jedni ducha, energii i materii. Pojęcia, które jest i go nie ma. Jest każdym z osobna i wszystkimi razem. Nie ma wymiaru ani potencjału, a jest.

Wielki Wybuch, choć znowu kwestionowany, ciągle jeszcze wyobrazeniowo otwiera niedefiniowalne pojęcia przestrzeni i czasu. W nich uruchamia celowe procesy stwarzania Życia w przestrzennej sieci przyczynowo-skutkowej, często zwane przypadkami. To otwiera pojęcie jedności przeciwieństw, ich wzajemnego definiowania się i tworzenia systemu wartości ukierunkowanego na tworzenie Życia i jego Celu. Procesy w przestrzeni i czasie ukierunkowują organizację boskich cząstek w systemy, tworzące Warunki powstawania życia, a następnie w systemie sprzężeń zwrotnych, tworzące coraz bardziej złożone organizmy żywe i odpowiednie im środowisko. Wzajemne oddziaływania kreują organizację inspirującą powstawanie coraz bardziej złożonych związków między życiem i środowiskiem, tworząc ścisłe związki między rybą a wodą, ptakiem a powietrzem itp. aż do powstania najbardziej złożonej

* oraz referatu wygłoszonego na Konferencji na Uniwersytecie Warszawskim w dniu 3 grudnia 2009 roku.

organizacji życia — człowieka, który świadomie zaczął kształtować swoje środowisko przestrzenne, dążąc do zabezpieczenia swojego życia i dominacji (innych, przyrody). Ten niezwykle syntetyczny wstęp służy wyłączenie do uwypuklenia, podkreślenia, że organizm żywy i jego środowisko są ściśle powiązane we wzajemnych ukierunkowaniach i zależnościach oraz, że w transcendentji są jednią (zbiorem, organizacją boskich cząstek).

Stwierdzenia, że człowiek jest bez ducha lub, że istnieje duch cełty, lub że przemknął promień światła, uświadamiają nam, że ten zbiór może się składać z materii i energii lub z materii i ducha, lub z czystej energii, czyli jego elementy niekoniecznie muszą występować w triadzie.

Przede wszystkim jednak pragnę uzasadnić, niemodne dziś w Polsce przekonanie, iż człowiek i jego środowisko jest i musi być swoistą Jednią. Człowiek uwarunkowuje środowisko, a ono uwarunkowuje człowieka (inspiruje lub deprymuje), i to zarówno środowisko ducha, energii, jak i materii, żywe czy nie ożywione.

Wierzmy lub Wiemy, że Życie powstało w nieskończonym procesie Stwarzania, zwanym również Ewolucją. Nauka uważa, że ten proces zaczął się 10 miliardów lat temu, trwa i nie potrafimy określić jego końca. Teologia stwierdza, że Proces Stwarzania trwa, nie znamy jego początku ani końca.

Wierzę, że proces ten był i jest celowy dla stworzenia życia i ukoronowania go Człowiekiem.

Problemem podstawowym w mojej ocenie jest fakt, że przez 100 000 lat egzystencji na Ziemi *Homo sapiens sapiens* przede wszystkim zawłaszcza przestrzeń, niszczy, często w obronie, i zużywa bez zastanowienia niezwykle subtelne procesy tworzące i podtrzymujące życie. Docieramy do punktu krytycznego.

Zagrożenia są globalne, wynikające z wewnętrznych przeciwieństw Kultury oraz z zaktócenia (niszczenia i zużycia) przez Kulturę naturalnego procesu tworzącego i podtrzymującego życie. Natura też w tym macza palce.

W ostatnich czasach za najdonioślejsze w Kulturze uważam zmuszenie przez cywilizację Północno-Atlantycką wielkich organizacji społecznych Chin, Indii i Rosji do przyjęcia ideologii niczym nie skrupowanej dominacji kumulowania i tworzenia wirtualnych pieniędzy i również wirtualnego rynku oraz zaktócenia, utajniania (ograniczania) przepływu informacji w oparciu o hasła wolności jednostki. Jednocześnie został stworzony najbardziej totalitarny, niedemokratyczny, system międzynarodowych finansów. Wytworzył on samo

kształtującą i samo zawężającą się elitę, głoszącą nadzędną systemu ekonomicznego nad demokracją i jedyną słuszność tego poglądu.

Ten niewątpliwie najważniejszy i najpotężniejszy proces wewnątrz Kultury, mimo wyraźnie już definiujących się ruchów proekologicznych, energooszczędnych czy Greenpeace i ich szerokiej akceptacji społecznej, pomija relację Życia i Przestrzeni. W warunkach polskich można powiedzieć, że i polityka, i biznes, i religia nie tylko nie zauważa tej relacji, ale traktują ją jako wrogą, a przecież:

— **Promieniowanie kosmiczne, słoneczne, ziemskie** — ich wzajemne proporcje i natężenie na powierzchni ziemi i w atmosferze to podstawa tworzenia i utrzymywania życia. Współczesna cywilizacja elektroniczna nasyci tę przestrzeń całkiem innymi proporcjami i natężeniami promieniowań. Człowiek jest najbardziej złożoną, wyrafinowaną istotą podlegającą oddziaływaniu kombinacji tych promieniowań.

Co wiemy o wzajemnych relacjach?

— **Atmosfera** — filtr tych promieniowań jest regulowany złożonym systemem procesów organicznych i jest częścią przekształcającej się biosfery.

Kultura i Natura ostro oddziałują na zmiany w atmosferze — nie kontrolujemy ich.

— **Biosfera** — organizacja bezpośrednio tworząca Życie i podtrzymująca je, powstała wskutek postępującej złożoności procesów submolekularnych, chemicznych i biologicznych. Człowiek, ludzkość, jego wiedza i niewiedza, ekonomia, produkcja ostro w nią ingerują — niszczą, przekształcają, zużywają, tworzą mutacje. Nie ma żadnej wiedzy syntetyzującej. Co gorsze nauka zajmująca się tymi relacjami Natury i Kultury została upolityczniona i skomercjalizowana.

Rytm, cykle, powtórzenia, ciągłości to podstawa procesów tworzących i podtrzymujących Życie. Przykładem to zaktócenia, modyfikujące stabilne cykle. Mogą być zarówno początkiem nowej, lepszej organizacji, jak i entropii. Przeciwnieństwo organizacji i entropii jest, z pozycji wartościowania przez człowieka, tym samym co przeciwieństwo dobra i zła. Dobra tworzy życie, złe niszczy.

Miasta to lokalne organizacje Kultury. Na ogół dynamicznie rozrastające się i stające się miejscem życia coraz większej liczby ludności. Obecnie w miastach mieszka około 50% ludzkości, a prognozuje się, że

w przeciągu kilkudziesięciu lat staną się one miejscem życia 80% ludzkości. Można przypuszczać, że z aktualnych 3,5 miliarda mieszkańców, ich liczba wzrośnie do 7 miliardów. Naturalny proces rozwojowy miast to niekontrolowany rozrost powodujący zawłaszczanie Natury, zużywanie energii przeznaczanej na proces tworzenia warunków podtrzymywania życia oraz wytwarzania coraz większej ilości szkodliwych odpadów.

Wartościowanie zależy od Wiary w to, czy Prawda istnieje, czy nie, oraz od tego, jak ją definiujemy i jak do niej dążymy. Należę do tych, którzy są głęboko przekonani, że przestrzeń powłoki Kuli Ziemskiej została wykreowana jako bardzo mały fragment procesu, który my ludzie nazywamy Stwarzaniem Boskim lub Ewolucją. Wykreowana po to, by tworzyła warunki powstawania i utrzymywania życia (może dlatego tak mały, że jest to pierwsza próba?). Ewolucja to ludzkie, racjonalne odczytywanie przeszłości dostępnych nam fragmentów procesu Stwarzania. Jestem przekonany, że podstawą tego procesu jest równoważenie walki dobra i zła, jak również wszelkich innych przeciwieństw, w taki sposób, by powstawała przewaga organizacji nad entropią, tworzenia nad niszczeniem. Dlatego wartościowanie przestrzeni miasta uważam za jedynie słuszne z pozycji, uwzględniającej odpowiedź na pytanie, czy przestrzeń miasta podtrzymuje Życie. Bo życie człowieka podtrzymuje zarówno Natura, jak i Kultura.

Nie znamy przyszłości Stwarzania, ale człowiek ma wolną wolę i rozum. Poznaliśmy trochę zasad Stwarzania, sformułowanych w myślach o Ewolucji, które utrwaliły się mimo różnych załamania ciągłości. Nie mamy żadnych podstaw do ich zaprzeczenia, a tylko do twórczej, ostrożnej (zrównoważonej) interpretacji w Kontynuacji, dlatego każde działanie i planowanie działania musi być oparte o nadrzędną świadomość, że współżycie człowieka z człowiekiem, jak i człowieka z Naturą, jest już obecnie czynnikiem determinującym utrzymanie życia na ziemi.

Miasto traktuję jako organizm ukształtowany przez świadome i spontaniczne działania człowieka. Organizm kształtowany przez poczucie więzi i bezpieczeństwa oraz większą możliwość wymiany dóbr. Miasto te wartości daje w zamian za ograniczenia, wynikające z potrzeby zabezpieczenia interesów każdego z mieszkańców. Miasto i jego sieciowy układ jest kolebką demokracji, opartą na świadomości podziału ról i jednocześnie świadomości równości obywatelskiej.

Historyczne miasto wytworzyło swój organiczny krajobraz oparty o zdefiniowane reguły procesów

w nim zachodzących. Stworzyło zdefiniowane formy przestrzeni — ulica, plac. Stworzyło wyrazistą jedność własności i przestrzeni — indywidualną i wspólną. Procesy, jakie w tym podziale własności się toczyły, zaciebrały jednocześnie tę wyrazistość. Status urzędów i budowli publicznych, prywatnych sklepów, usług, kościoła czy szpitala stwarzały złożoność wzajemnych reguł dostępności i jej ograniczania.

Globalizacja stosunków międzyludzkich, a głównie ekonomii i przepływu informacji, powoduje wiele destrukcyjnych zjawisk w mieście. Liczba ludności miast stale rośnie, a wraz z nią wzrasta liczba samochodów i skala pojedynczych inwestycji. Elektronika niweluje związki człowieka z przestrzenią. Z jednej strony miasto się intensyfikuje, z drugiej — rozlewa się w otaczający je krajobraz. Nowe inwestycje są w znaczącej części efektem woli ponad miejskich i ponad narodowych grup finansowych. Są efektem doraźnych, korzystnych finansowo, lokalizacji w danym miejscu. Buduje się zgodnie z doktryną by inwestować pieniądze w miejscu, gdzie najbardziej się to opłaca. I choć może być to korzystne z punktu widzenia makroekonomicznego, to z punktu widzenia jakości życia w mieście jest destrukcyjne. System ekonomiczny, nastawiony na stale wzrastającą konsumpcję jednostki, nie zauważa słabo lub w ogóle nieartykułowanych lokalnych potrzeb wspólnych. Duży dom handlowy czy siedziba filii korporacji z równym powodzeniem może być zlokalizowana w śródmieściu, co w lesie czy polu podmiejskim, ważne by klient lub pracownik mógł przekazać i otrzymać niezbędną informację lub inny produkt.

W pewnym okresie (w Europie w czasie Średniowiecza i Oświecenia, a w niektórych przypadkach w XIX wieku, w Chinach przez ostatnie 2000 lat) miasta osiągnęły taką organiczność, że żadna ich część nie zaprzeczała innej. Krajobraz miasta z tego okresu jest najpowszechniej akceptowany, odwiedzany. Potrafi się oprzeć zarówno ekspansji samochodu, jak i megaskali inwestycji. Relacja między skalą całości miasta, formą przestrzeni publicznej, skalą inwestycji prywatnej i jej formą, wynikającą ze zrównoważenia potrzeb wspólnych, możliwości inwestora i możliwości technicznych, oraz prostą relacją z człowiekiem i jego cechami psychofizycznymi stworzyła pewien wzorzec całościowy. Nie wiem czemu, mimo tak powszechnej aprobaty dla struktury krajobrazu miasta tradycyjnego, nie próbujemy procesów, które go ukształtowały, wykreować jako świadome, demokratyczne ograniczenia tworzące przestrzeń względnie powszechnie aprobowaną.

Oczywiście nie wszędzie, ale w określonych proporcjach i miejscach. Współczesny proces wielokoinwestycyjny, który jest faktycznie globalizmem, też musi mieć swoje ujście, dlatego dobrze byłoby wyznaczyć dla niego miejsce i określić zasady kształtowania odpowiadającej mu przestrzeni.

Jeszcze socrealizm używał zdefiniowanych form przestrzeni miejskiej. Modernizm i związany z nim socjalizm realny, a obecnie globalizm, te reguły i formy swobodnie omijają i doprowadzają do zaniku. Niszczą również tradycyjnie ukształtowany krajobraz miasta.

Miasto rozlewa się niszcząc otaczającą przyrodę, a jednocześnie nie tworzy wewnętrznego mikroklimatu sprzyjającego zdrowiu. Nie tylko nie tworzy przestrzeni społecznych, ale istniejące często niszczy. Zanika przestrzeń wspólna – przestrzeń publiczna. Droga czy plac wypelniony samochodami to jedynie urządzenie, użytkowane demokratycznie przez posiadaczy samochodów, które nie wytwarza więzi między nimi. Galerie handlowe „wysysają” handel z ulic. Miasto coraz częściej tworzy chaos ruchów dla ruchu i niekontrolowanych inwestycji, których podstawowym celem jest przepływ coraz większego kapitału.

Miejskość utrzymywana jest tylko przez bezwład stanu istniejących części historycznych. Miasto typu historycznego nie powstaje już „naturalnie”, ale nadal jest przestrzenią, organizmem, w którym żyjemy. W miastach takich jak Warszawa, w których miasto historyczne zostało zniszczone, tylko wyraźnie sformułowana **wola obywateli miasta** może przeciwdziałać temu rozpadowi przestrzeni publicznych.

System makroekonomiczny jest właściwie całkowicie uniezależniony od organizmu miejskiego. Nie jest też państwowy, jak to było w socjalizmie. Obecnie jest globalny. W mojej ocenie ekonomia socjalizmu realnego i globalizmu mają identyczny stosunek do przestrzeni i organizmu miasta – nie zauważają ich. Działają pomimo.

Lokalność współcześnie bierze się tylko z tradycji i związanych z nią zwyczajów, historycznie ukształtowanych kontaktów międzyludzkich twarzą w twarz i związanych z nimi form przestrzennych i nastrojów przestrzeni placu, ulicy z ruchem pieszym ze sklepami, restauracjami itp. Te formy przestrzeni jednak z dużą aprobatą społeczną zastępowane są przez super budowle biurowe, handlowe, hotelowe czy mieszkaniowe. Budowle, które stanowią samoistny organizm o wysokim standardzie użytkowym, ale zaprzeczają organizmowi miasta historycznego.

Współczesna cywilizacja dąży do doskonałej atomizacji osoby i jej przestrzeni. Obok osoby istnieje ludzkość, a w niej stratyfikacja poszczególnych osób wynikająca z wielkości obracanych przez nią pieniędzy. Inne wartości, a w szczególności te, które wytwarzają pojęcie „My”, czyli tworzące kapitał społeczny, są w sposób naturalny dezawuowane.

Pieniądże, informacja, produkcja, usługa przepływają niezależnie od związków ludzi z przestrzenią. Państwa i miasta jako samoistne organizmy zanikają. Demokracja i w historii, i w jej obecnej formie, wiąże się wprost z określonymi grupami ludzi i ze związaną z nimi przestrzenią i organizacją – gminą, państwem. Wszystko wskazuje na to, że globalizacja powoduje nie tylko zanik miasta i zanik państwa, ale w sposób naturalny powoduje zanik demokracji.

Pewną Nadzieję budzi fakt, że rodząca się społeczność wolnego przepływu informacji oparta jest na pojęciu sieci. Sieć ułożona na powierzchni kuli (Ziemi) podkreśla znane prawo geometrii przestrzennej, że każdy znajdujący się na niej punkt jest jej środkiem. Jest Centrum sieci. Gdy wchodzimy do sieci, miejsce Każdego z nas jest w jej Centrum. Dzięki sieci każde miejsce, w którym żyjemy może stać się naszym Centrum życia na Ziemi. Zależy to od naszej Woli i umiejętności zrozumienia tego faktu. Wymaga to aktywnego włączenia się w sieć i równie aktywnego włączenia się w działania w miejscu życia. To połączenie Naszego miejsca tu w Polsce w Warszawie ze Świadomością, Wolą i przestrzenią świata. To droga do demokratyzacji globalizacji. Wymaga ona równoległego tworzenia wokół nas dokonań organizacyjnych i materialnych, budzących nasze możliwości i rozwijających naszą Wolę i Świadomość. Tworzenia Naszych wartości w Naszej przestrzeni tak, by nasze życie tu i teraz odnosiło się do rzeczywistych spraw, w tym również, a może przede wszystkim, do współdziałania z Naturą i Kulturą w tworzeniu warunków podtrzymujących Życie, a nie tylko do wirtualnych reklam, mitów i celów wskazywanych przez innych z naturalnej potrzeby dominacji.

Drogę do demokratyzacji globalizacji trzeba rozpocząć od rzeczy najprostszych, od tworzenia wymiany lokalnych informacji, pomysłów czy poglądów, związanych z Miejscem, w którym żyjemy twarzą w twarz, w którym tworzy się kontakt z siecią, komputerem, programem i nauką korzystania z nich. Należy zacząć od tworzenia Naszej rzeczywistej przestrzeni w sposób Nam przydatny i swojsko piękny. Od tworzenia prze-

strzeni powiązanej wirtualnie z resztą świata. Te działania mogą stać się Wzorem.

Wiemy, że między nami a naszą przestrzenią istnieją bezpośrednie związki. Pan Bóg nas stworzył z tego samego budulca. Stworzył On niezwykle subtelną i złożoną organizację procesów tworzących i podtrzymujących życie. Wiemy również, że proces stwarzania trwa. Globalizacja jest jego częścią. Od naszej Woli, chęci i umiejętności działania zależy, czy będzie to Nasza Globalizacja, czy będziemy w niej twórczo, choć lokalnie, uczestniczyć. Od tego zależy czy będąc Polakami staniemy się również Ziemiakami, i czy staniemy się twórcami naszej rzeczywistości.

Jan Paweł II w Encyklice *Wiara i rozum* pisze:

Łaska nie niszczy natury ale ją doskonali.

Kiedy Kościół styka się po raz pierwszy z wielkimi kulturami nie może wyrzec się tego co zyskał dzięki inkulturacji w myśli grecko-lacińskiej. Odrzucając to dziedzictwo sprzeciwiłby się opatrnościowemu zamysłowi Boga, który wiedzie swój Kościół po drogach czasu i historii. To kryterium zresztą obowiązuje Kościół każdej epoki, także Kościół jutra, który będzie bogatszy o to, co zyska dzięki dzisiejszym kontaktom z kulturami wschodu i z tego dziedzictwa zaczerpnie nowe wskazania, aby nawiązać owocny dialog z Kulturami, jakie ludzkość zdołała wytworzyć i rozwinąć w swojej wędrówce ku przyszłości.

Opat Zenkei Shibayama w książce *Milczenie Kwiatu* podaje cztery maksy tradycyjnie przyjmowane jako wyjaśnienie właściwości i ideałów Zen. Wydały mi się one bardzo bliskie i ogromnie przydatne w pracy zawodowej, również jako pedagoga. To proste chińskie zwroty:

- *Przekaz poza pismami;*
- *Nie opieranie się na literaturze (doświadczenie);*
- *Bezpośrednie wskazywanie na umysł;*
- *Osiągnięcie Stanu Buddy przez wgląd we własną Naturę.*

Pisze tam również:

Prawdziwy człowiek zen jest mistrzem czasu i przestrzeni, żyjącym w absolutnym punkcie „tutaj-teraz”...

Przez fakt religijnego doświadczenia człowiek osiąga swoją Naturę Buddy — satori — (oświecenie)...

Zen naucza:

Nawet jeśli dam Ci odpowiedź, będzie to moja odpowiedź. Nie będzie mieć nic wspólnego z twoim zrozumieniem, które powinieneś uzyskać sam i którego sam powinieneś doświadczyć.

Aleksander Wallis w eseju *Przestrzeń jako wartość* stwierdza:

Jednostka dostrzegając różne wartości danej przestrzeni świadomie lub nieświadomie ocenia własną w niej sytuację.

Florian Znaniecki definiuje:

Przestrzeń społeczną danej zbiorowości stanowi użytkowany i kształtowany przez nią obszar, z którym wiąże ona system wiedzy, wyobrażeń, wartości i reguł zachowania, dzięki którym identyfikuje się najpełniej z tym właśnie obszarem.

Definicję tę uzupełnia stwierdzeniem:

co dla jednych jest święte, dla drugich może być obce lub wrogie.

Jeżeli naprawdę uważamy, że jesteśmy demokratami i widzimy bezpośrednie związki między demokracją, społecznością i jej przestrzenią, że nadal są prawdziwe i obecnie funkcjonują definicje Floriana Znanieckiego, oraz, że lubimy naturę przestrzeni miasta historycznie ukształtowanego to: **Twórzmy Miasta lokalne w globalnej sieci.**

Duże miasta dzielimy na małe, o określonych naturalnych preferencjach ekonomicznych i przestrzennych, o wielkościach od kilkunastu do kilkudziesięciu tysięcy obywateli. Niektóre z tych części należy odrestaurować, w innych odbudować układ urbanistyczny ulic i placów określonymi w skali inwestycjami, a w jeszcze innych dopuśćmy spontaniczne narastanie budowy. Dążmy do tego, by dzielone części miasta miały swój środek ciężenia życia lokalnego, a jednocześnie tworzyły związki z otaczającą Naturą, oraz włączały się w stworzony proces podtrzymywania życia z *a priori* założonym celem drobnoziarnistości i różnorodności akcji inwestycyjnych, by kształtowany obszar również kształtował społeczność i wiązał ze sobą jej „system wiedzy, wyobrażeń i wartości”.

Założenia społeczne

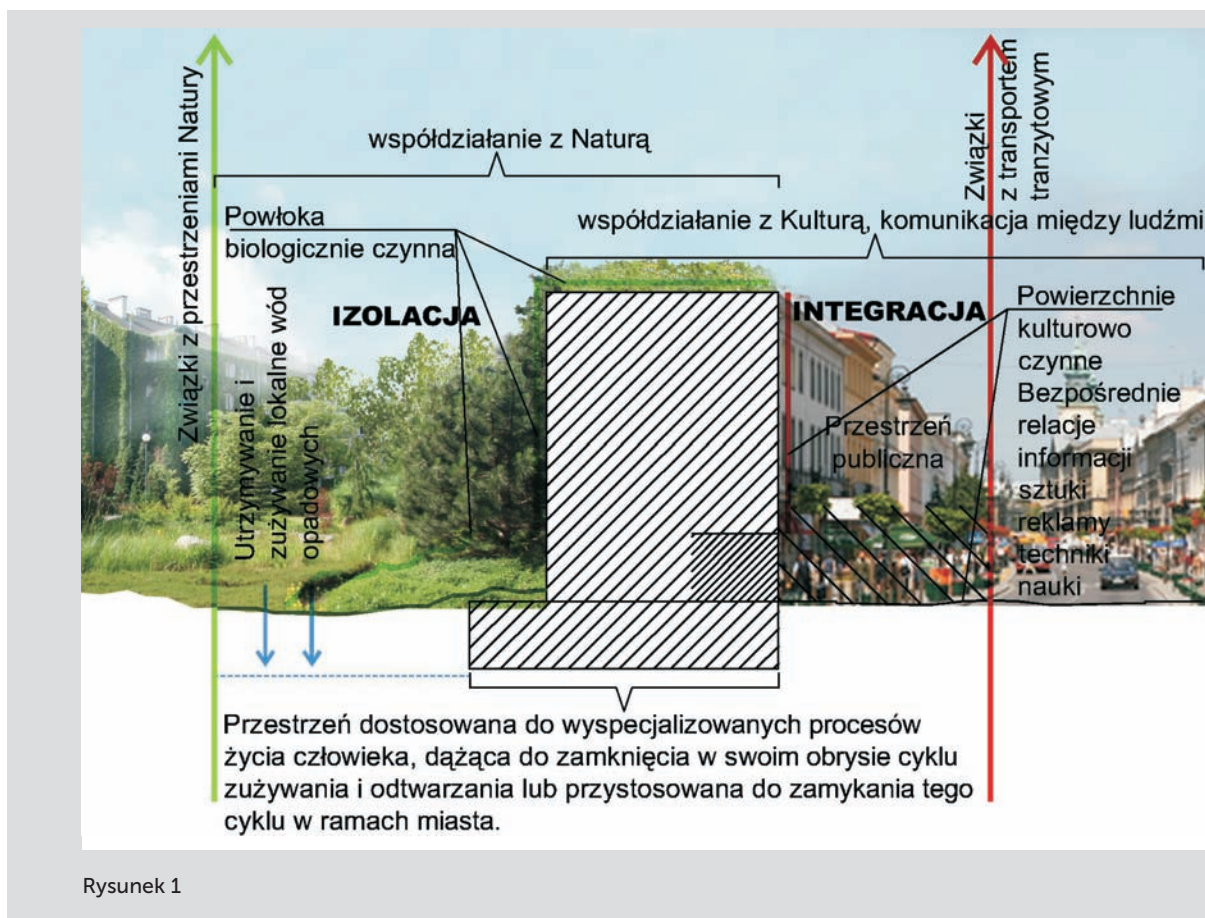
- umożliwienie satysfakcjonowania wielu — wielu determinuje kształt całości;
- umożliwienie narastania i scalania różnych indywidualnych działań;

- zapewnienie względnie szerokiej możliwości wyboru, współuczestnictwa, współdziałalności w tworzeniu własnego środowiska, jego przekształceniu i dostosowywaniu do nowych potrzeb;
- zrównoważenie jednoczesnego występowania przeciwieństw, np.: izolacja–integracja, wybór–determinacja, produkcja–konsumpcja, obsługujący–obsługiwany, specjalizacja–uniwersalizacja, zagrożenie–bezpieczeństwo;
- zmniejszenie specjalizacji stref aktywności społecznej (względne zaspokojenie w jednej przestrzeni strefy produkcji, usług, mieszkania i rekreacji);
- dostosowanie rozwiązań przestrzennych i organizacyjnych do stale wzrastającego zakresu i komplikacji potrzeb jednostki w ramach rodziny i grup społecznych, a także ich zmienności w czasie znacznie krótszym, niż techniczna trwałość przestrzeni uformowanej dla pierwotnego zakresu tych potrzeb (nowo zrealizowane będzie przekształcone).

Założenia przestrzenne

Celem polityki kształtowania miast lokalnych powinno być stosowanie środków nadających nowemu lub przekształcanemu zespołowi charakteru miasta tradycyjnego (europejskiego), w tym również strefy miasta globalnego (amerykańskiego) oraz wciągnięcie go do współpracy z Naturą przez:

- **Każda budowla** w mieście powinna być przestrzenią dostosowaną do wyspecjalizowanych procesów życia człowieka i dążąca do zamknięcia w swoim obrysie cyklu zużywania i odtwarzania lub być przystosowana do zamykania tego cyklu w ramach miasta. Jednocześnie z jednej strony przez zewnętrzną powłokę biologicznie czynną dąży do związków z Naturą, współdziałając z nią w procesie podtrzymywania życia biologicznego i osiągania izolacji. Z drugiej strony jej powłoka zewnętrzna i jej wewnętrzna funkcja w strefie przestrzeni publicznej współpracuje z Kulturą, tworząc warunki wspierające związki między ludźmi, współtwo-



Rysunek 1

rząc przestrzeń bezpośrednich relacji informacji, sztuki, reklamy, techniki – wspiera integrację.

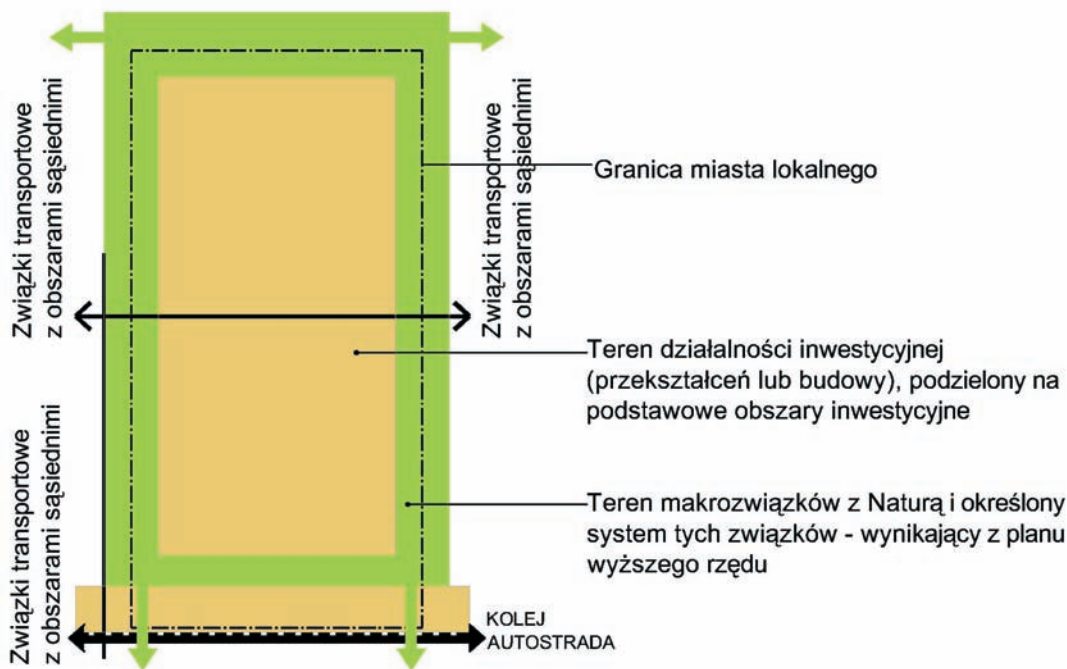
Trzeba oprzeć się na rozumieniu, że z jednej strony zależność budynku od budynku, funkcji zawartych w jednym od funkcji zawartych w drugim budynku, ich wzajemna relacja z ulicą jako przestrzenią publiczną i dojazdami samochodów lub innego środka transportu, a z drugiej strony z zielenią jako przestrzenią własną lub z zielenią jako przestrzenią publiczną, a obie te zależności w relacji do człowieka i jego cech psychofizycznych są kluczem do tworzenia przestrzeni współżycia Natury i Kultury. Z tych zależności wynikają dopiero ludzkie i przyrodnicze formy cechujące współczesność, które mogą otwierać drogę do przestrzeni miasta, tworzącej warunki podtrzymywania życia.

■ **Miasto lokalne** ma wyznaczone w planie regulacyjnym (przekształcenia lub budowy nowego):

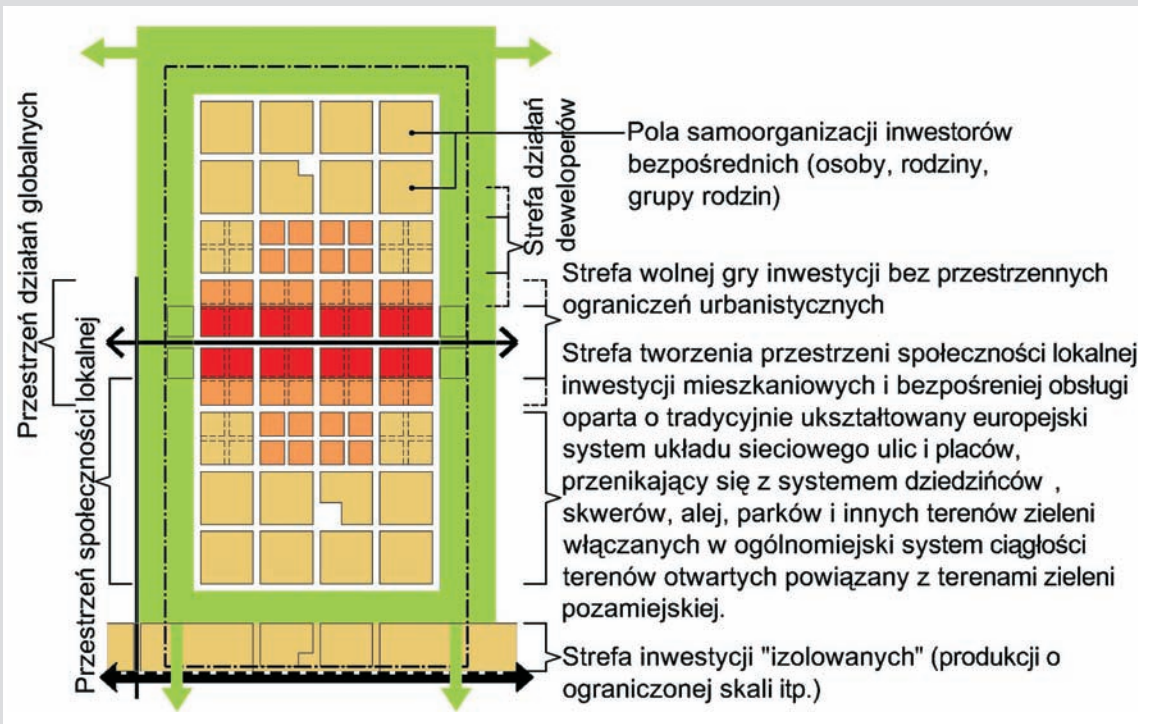
- obszar (granice). W tym obszarze są określone związki transportowe z obszarami sąsiednimi (tworzące układ względnie niezależny od terenu miasta lokalnego) oraz wyznaczony teren makrozwiązków

z Naturą i określony system tych związków, wynikający z planu wyższego rzędu i podziału obszaru urbanizacji na miasta lokalne.

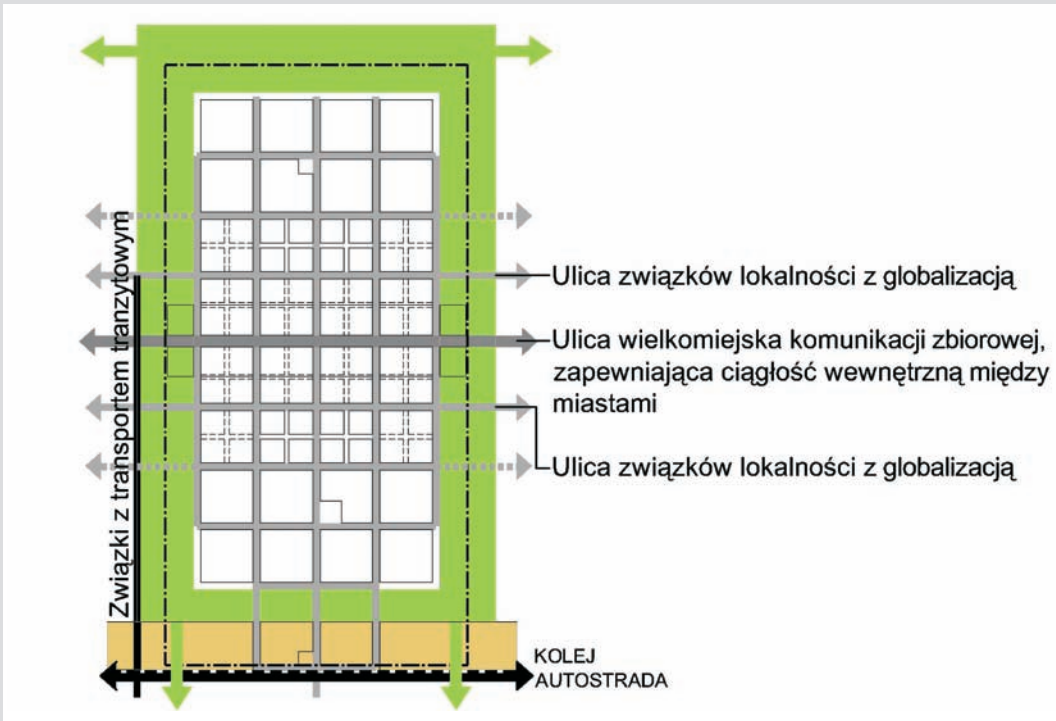
- obszar miasta lokalnego musi być podzielony na:
 - tereny działalności inwestycyjnej (przekształceń lub budowy nowego) – umożliwiające różnorodny i drobnoziarnisty system inwestowania i użytkowania,
 - tereny inwestycji miejskich określające reguły walki inwestora o teren i jego użytkowanie:
 - sieciowy układ pasów magistralnych miejskich (teren dróg i infrastruktury technicznej), otaczający tereny inwestycyjne oraz wyznaczający powiązania z obszarami sąsiednimi – a przede wszystkim włączanie się inwestorów w system Kultury podtrzymujący życie,
 - koncentryczno-dośrodkowy liniowy system wytwarzania przestrzeni publicznej (ulic i placów),
 - odśrodkowy system ciągłości układów przyrodniczych (roślinno-wodny).



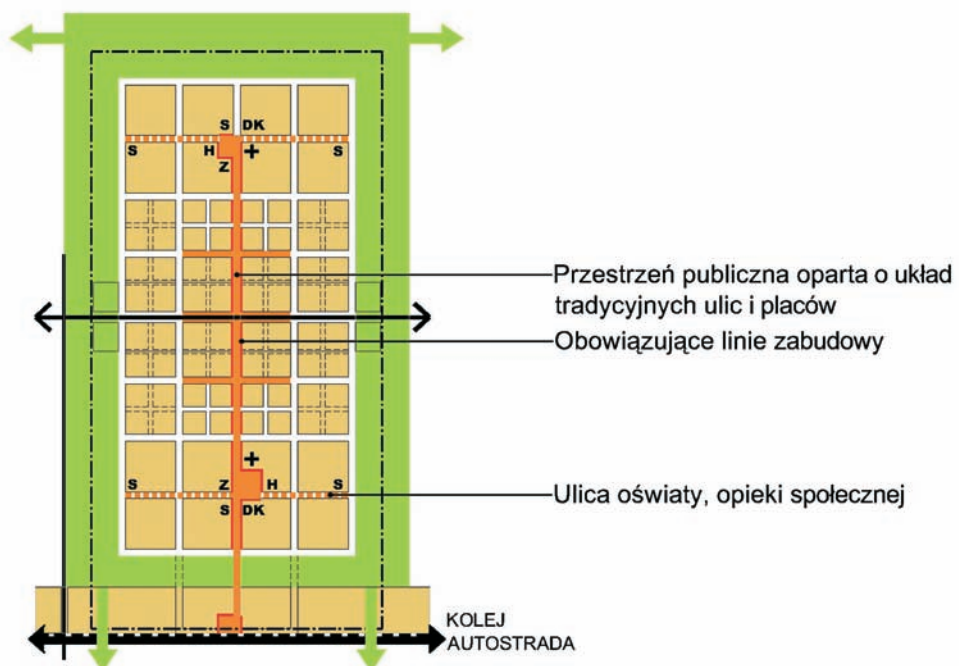
Rysunek 2



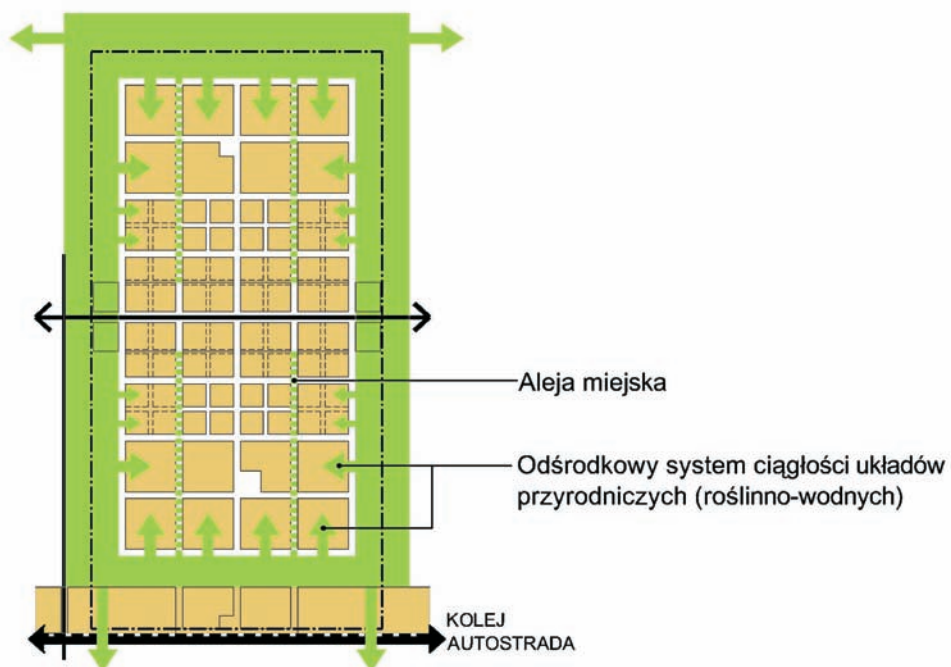
Rysunek 3



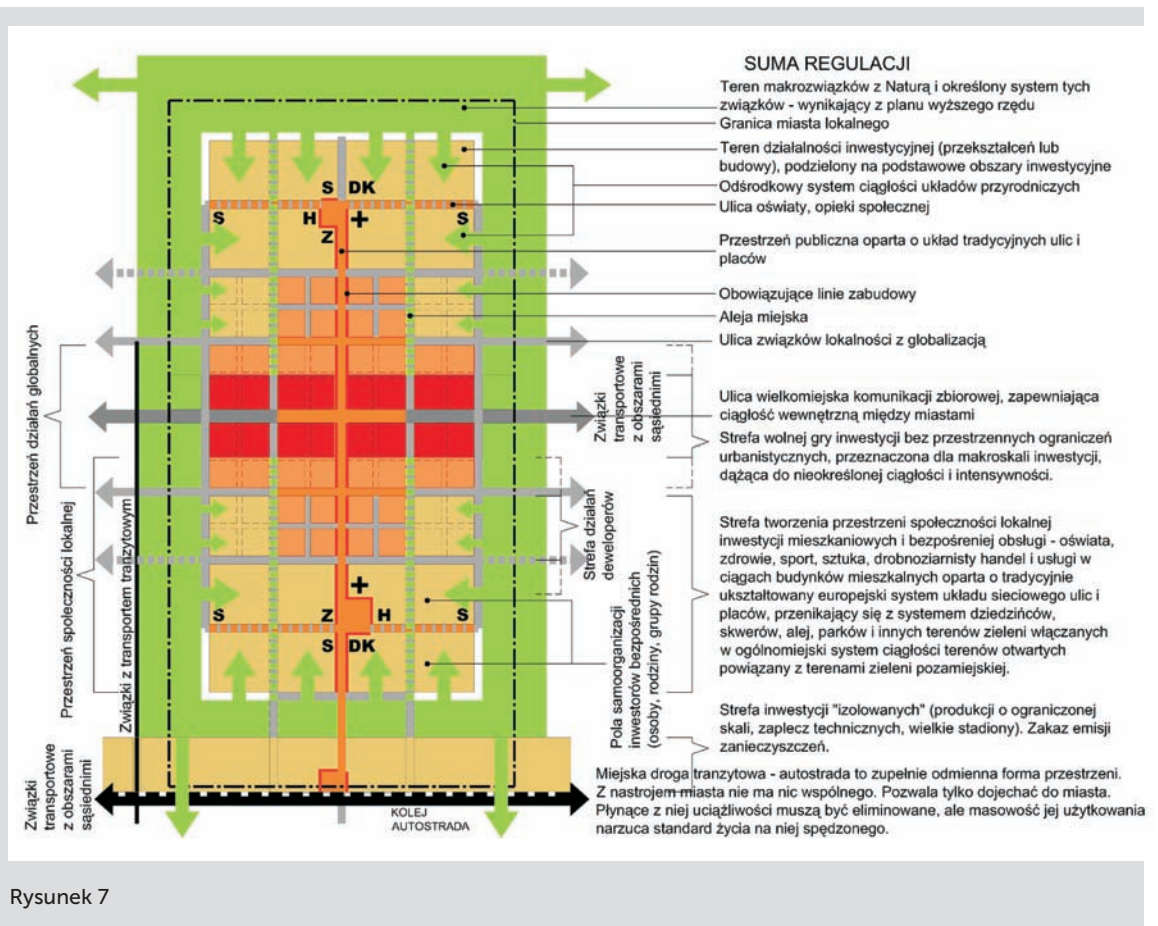
Rysunek 4



Rysunek 5



Rysunek 6



Rysunek 7

Miejska droga tranzytowa — autostrada to zupełnie odmienna forma przestrzeni. Z nastrojem miasta nie ma nic wspólnego. Pozwala tylko dojechać do miasta. Płynące z niej uciążliwości muszą być eliminowane, ale masowość jej użytkowania narzuca standard życia na niej spędzonego.

Założenia polityki i administracji

W poprawnie realizowanym mieście lokalnym (budowanym lub przekształcanym) występują, zgodnie z historycznym rozwojem miast, dwie komplementarne siły:

- Scentralizowana — demokratycznie wybierana władza miasta — kształtująca szkielet, determinująca „miejsca”, zabezpieczająca, regulująca i kontrolująca rozwój „całości”, dążąca do pełnego podporządkowania działalności w imię interesów „całości”.
- Zdecentralizowana — inwestorzy, użytkownicy — wielośrodkowa, wypełniająca szkielet, realizująca

określone inwestycje, dążąca do niekontrolowanego rozwoju i zabezpieczenia pełni potrzeb własnych z pominięciem interesu całości.

Zrównoważona, prowadzona według określonych reguł walka między tymi dwoma zespołami sił stanowi o prawdziwości rozwoju miasta.

- Cechą działań konkretnego inwestora musi być to, że budując swój dom (swoje domy) buduje miasto, współtworzy warunki podtrzymujące życie — system współdziałania Kultury i Natury — i to musi być od niego wyegzekwowane.
- Istotą działań odpowiedzialnego za całość (władza miejska) musi być to, że realizuje miasto lokalne, stwarzając warunki do budowy wielu niezależnych inwestycji współtworzących warunki życia — system współdziałania Kultury i Natury — i to musi być od niego wyegzekwowane.

Te dwie siły muszą mieć wyraźnie wyznaczone pola działań i reguty współdziałania w przestrzeni. Dzielić przeskalowane powiązania, łączyć, ukierunkowywać odizolowane i sprzeczne działania.

Uważam, że lokalność musi być w naszym życiu równie ważna jak i globalność, a myślenie holistyczne jednocześnie z redukcjonistycznym.

Osobność wynikająca z osobowości powinna być jednocześnie i równorzędna z więzią budującą synergię niezbędną do przeżycia, jak i wartościowania naszej osoby.

W swoim życiu zawodowym urbanisty i architekta uważam, że doświadczenie, a więc bezpośrednie uczestniczenie w procesie Stwarzania (co w myśl Dogmatyki Katolickiej wchodzi w zakres rządów Opatrzności Bożej), jest wiodące.

W *Refleksjach o etyce pracy* Ks. Józef Tischner napisał:

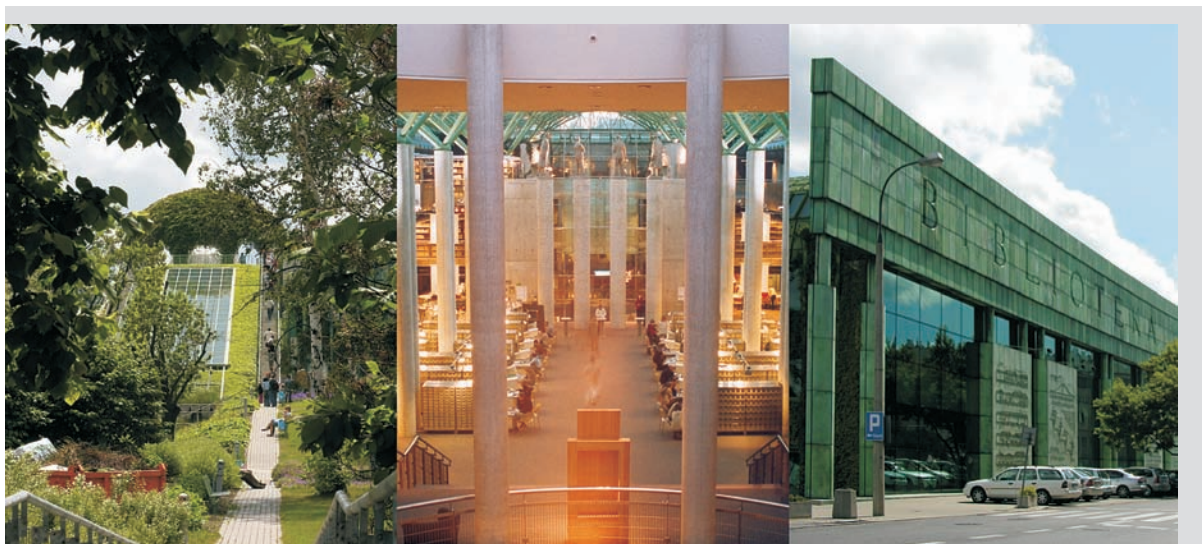
Prawda poznania jest wartością szczególną. Jeżeli chcesz realizować w świecie jakiegokolwiek wartości, budować domy, leczyć ludzi, sądzić zważnionych musisz znać prawdę o tym świecie, musisz znać prawdę o wartościach, które chcesz w świecie ucieleśnić. Prawdziwe poznanie świata jest warunkiem podstawowym moralnego działania w świecie.

Jestem przekonany, że pełne poznanie jest poza zasięgiem człowieka i zapewne dlatego tak nienajle-

piej jest z moralnością jego działania. Twórczość człowieka byta, jest i zawsze będzie w większej mierze intuicyjna (objawiona) niż płynąca z pełni poznania rozumowego, choć ta intuicja jest też oparta na poznaniu świadomym i nieświadomym. Jednak objawienie, imperatywu intuicji musi być sprawdzony rozumem i dopiero osiągnięte cząstkowe poznanie rozumowe i intuicyjne może się stać podstawą działań twórczych.

Teoretyzuję tylko w celu przekazania doświadczeń z walki o Życie. Ponieważ ta moja walka jako architekta już trwa prawie 50 lat zacytuję kilka myśli o przestrzeni niewątpliwie inspirowanych myślami Tischnera o etyce pracy oraz pokażę kilka dokonań (doświadczeń) zbudowanych w ramach przekształcania przestrzeni w budowlę lub fragment miasta w nadziei, że będą tworzyły warunki podtrzymywania Życia, budowały Nadzieję:

- Ludzkie kształtowanie przestrzeni zakłada i rozwija porozumienie.
- Proces kształtowania przestrzeni jest procesem wznoszenia tworzywa do poziomu mniej lub bardziej jednoznacznie sprecyzowanej myśli.
- Tak jak słowa mogą być prawdziwe lub fałszywe, tak prawdziwe lub fałszywe mogą być wytwory ukształtowania przestrzeni. Przestrzeń jest prawdziwa, gdy zgadza się z ideą, przyświecającą twórcy w trakcie jej powstawania.



Rysunek 8



Rysunek 9

- Kryterium „wewnętrznej prawdy rzeczy” stanowi jeden z podstawowych mierników ukształtowanej przez człowieka przestrzeni.
- Przestrzeń formowana przez człowieka jako produkt pracy jest integralnym elementem komunikacji społecznej, jest językiem i tak jak język pod-

lega kwalifikacji, której kryterium jest Prawda lub Fałsz.

Instynkty, wiara i poznanie, w tym nauka, buduje personalny światopogląd. Buduje personalną prawdę poznania. To daje podstawę do świadomego włączania się w proces stwórczy – do tworzenia syntezy Przeszłości i Teraźniejszości kształtującej Przyszłość w sensach pozytywnych. Mam nadzieję, że oczekiwany krajobraz trzeciego tysiąclecia to krajobraz jedności prawdy Objawionej i prawdy Rozumu, świadomego współistnienia przeciwieństw, współżycia natury i kultury, przeszłości i przyszłości, intymności i masowości, związków z tradycją i zerwania z tradycją, dostosowanie się do rzeczywistości i próba pozytywnego wyjścia poza nią.

Żeby powstawała architektura i urbanistyka, jak poezja lub literatura jej litery, słowa muszą znaczyć to samo dla wszystkich.

Proces stwarzania przestrzeni Życia trwa – architektura i urbanistyka jest jego częścią.

Profesor Marek Budzyński kierownik pracowni Projektowania Miejskiego Wydziału Architektury Politechniki Warszawskiej oraz prezes Pracowni Badowski, Budzyński, Kowalewski. Autor lub współautor wielu prac, np.: Koncentracja liniowa, 1969; Parcelacja grupowa, 1985; Zakłady Azotowe Włocławek, 1968 (główny projektant); Zespół Osiedli Ursynów, 1972–1980 (generalny projektant); Zespół Osiedli Młodych, 1978–1983 (generalny projektant); Kościół na Ursynowie, 1980–1985; Biblioteka Uniwersytetu Warszawskiego, 1994–1999; Siedziba Sądu Najwyższego, 1991–1999 (ze Zbigniewem Badowskim); Dom XXI wieku, 2001–2004; Europejskie Centrum Muzyki i Sztuki w Białymstoku, 2005 (w budowie). W czasie swojej wieloletniej pracy wygrywał liczne konkursy architektoniczne i został uhonorowany wieloma nagrodami i odznaczeniami.



Daleka podczerwień (THz) w półprzewodnikach — fizyka i aplikacje

Na podstawie odczytu wygłoszonego w dniu 23 lutego 2012 roku

Marian Grynberg

Instytut Fizyki Doświadczalnej Uniwersytetu Warszawskiego

183

W obszarze dalekiej podczerwieni energie kwantów promieniowania elektromagnetycznego są bardzo małe. Zatem, „w naturalny sposób” spektroskopia dalekiej podczerwieni jest narzędziem do badania pobudzeń elementarnych niskoenergetycznych. W fizyce półprzewodników pobudzeń takich jest bardzo wiele, np. energie wiązania płytkich donorów czy energie drgań plazmy w słabo domieszkowanych półprzewodnikach. Energie fotonów dalekiej podczerwieni to obszar 3–15 meV (długość fali 1 mm–100 μm).

Ponieważ w tym obszarze nie ma intensywnych źródeł promieniowania ciągłego, używa się monochromatycznych źródeł (lasery molekularne, karcinotrony, lasery półprzewodnikowe). Zamiast zmiany długości fali umieszcza się badany obiekt w zewnętrznym polu magnetycznym, które zmienia energię stanów kwantowych, „dostrajając je” do energii padającego fotonu. Energia jonizacji płytkiego, wodoropodobnego donora w półprzewodniku jest około 1000 razy mniejsza niż w atomie wodoru (mniejsza masa efektywna i większa stała dielektryczna) 4 meV–20 meV, zatem w obszarze dalekiej podczerwieni. Metodami absorpcji w polu magnetycznym lub fotoprzewodnictwa udało się zmierzyć stany domieszkowe w półprzewodniku. Jest to analogon widma absorpcyjnego atomu wodoru tylko w obszarze spektralnym około 1000 razy dłuższych fal. W uzyskanych widmach „zakodowanych” jest wiele informacji o stanach donorowych, a tym samym o półprzewodniku. W układach dwuwymiarowych (2D) – studniach kwantowych – problem donora jest znacznie bardziej złożony.

Położenie energetyczne płytkiego donora bardzo wyraźnie zależy od szerokości i głębokości studni. Dodatkowo dla studni o szerokości $\sim 100 \text{ \AA}$ energia wiązania płytkiego donora wynosi $4 Ry^*$, dla donora w środku studni (Ry^* energia wiązania w kryształach trójwymiarowych), a jeśli donor znajduje się na powierzchni studni – bariera energia ta wynosi $1 Ry^*$, zatem w zależności od położenia donora w studni energia ta zmienia się 16 razy. Dokonując umiejętnego domieszkowania studni kwantowej (metodą *Molecular Beam Epitaxy* – MBE) udało się zjawisko to zaobserwować w wybranych materiałach półprzewodnikowych.

Badania płytkich stanów domieszkowych (stanów wodoropodobnych) w studniach kwantowych pozwoliły zaobserwować w laboratorium stan ujemnie naładowanego płytkiego donora, D^- (analogon ujemnie naładowanego atomu wodoru H^-). Astrofizycy od wielu lat obserwowali w widmie korony słonecznej w bliskiej podczerwieni (około 1,5 μm) silną linię spektralną, której natury fizycznej nie udawało się wyjaśnić. W latach czterdziestych ubiegłego stulecia Chandrashekar wykazał, że atom wodoru potrafi wiązać drugi elektron tworząc stan H^- . Obydwa związane elektrony muszą mieć przeciwnie skierowane spiny, a drugi elektron posiada tylko jeden stan związany. Energia wiązania tego stanu wynosi 5,5% Rydberga (13,6 eV). Przejście między tym stanem a „poziomem jonizacji” dokładnie odpowiada obserwowanej linii z korony słonecznej.

Rozpatrując atom wodoropodobny w półprzewodniku (np. płytki donora) należałoby się spodziewać

powstawania ujemnie natadowanego donora D^- . W strukturach trójwymiarowych (3D), mimo usilnych prób, takiego stanu nie udało się zaobserwować. Energia wiązania drugiego elektronu, podobnie jak stanu H^- , powinna wynosić 5,5% energii wiązania pierwszego. Z tym tylko, że zamiast Ry w przypadku donora w półprzewodniku, energia wiązania wynosi Ry^* (około 1000 razy mniej niż dla atomu wodoru), a 5,5% z 5 meV (energia wiązania elektronu na donor w GaAs) wynosi około 300 μeV .

Domieszkując struktury tylko w środku bariery obserwuje się (w polu magnetycznym), że przejście z D^- zachodzi w małej energii. Domieszkując tylko w środku studni obserwuje się takie przejście w znacznie wyższej energii. W strukturach domieszkowanych zarówno w środku studni, jak i w środku bariery obserwuje się 3 linie absorpcyjne – dwie o energiach odpowiadających poprzednim przypadkom i trzecią (równie intensywną), która jak się okazało jest przejściem ze stanu D^- .

W temperaturze wzrostu struktury (300–4000 °C) wszystkie stany donorowe są zjonizowane. Przy ochłodzeniu do studni „wpadają” nie tylko elektrony z donorów ze studni, ale również część elektronów pochodzących z donorów w barierze. W ten sposób w studni znajduje się więcej elektronów niż centrów donorowych. Część z tych „dodatkowych” elektronów wiąże się na neutralnym donorze D^0 , tworząc stan D^- . W obecności silnego pola magnetycznego energia stanu związanego D^- silnie rośnie i dlatego może być łatwo zaobserwowana. Zatem w specyficznym domieszkowanych strukturach dwuwymiarowych powstają stany D^- , któ-

rych nie udało się uzyskać w kryształach trójwymiarowych (3D).

W ostatnim dwudziestoleciu rozpoczęły się intensywne badania mikronowych i submikronowych pojedynczych tranzystorów polowych, traktowanych jako źródła i detektory promieniowania THz. W tranzystorach tych występuje, w kanale tranzystora, dwuwymiarowa warstwa plazmy, w której istnieją drgania plazmonów. Drgania te skutkują emisją fal elektromagnetycznych (dla submikronowych tranzystorów) rzędu pojedynczych THz. Emisja ta zachodzi zarówno w temperaturach helowych, jak i, co ważne, również w temperaturze pokojowej. Podobny tranzystor (podobnie spolaryzowany) oświetlony promieniowaniem THz generuje między źródłem a drenem napięcie, które linowo zależy od intensywności padającego promieniowania. Jest zatem detektorem THz, pracującym nie tylko w niskich temperaturach (helowych), ale również w temperaturze pokojowej. Dodatkowo „odpowiedź” detektora na promieniowanie o częstotliwości THz jest napięciem stałym (DC), a zatem detektor ten jest silnie nieliniowym elementem elektronicznym, z czego wynika, że pojedyncze tranzystory polowe mogą być źródłami i detektorami promieniowania THz. Potencjalne zastosowanie tych elementów jest ogromne w różnych dziedzinach życia. Niektóre doczekały się już zastosowań. Mechanizmy fizyczne – zarówno emisji, jak i detekcji – wymagają jeszcze szczegółowego zrozumienia. Wiele z tych mechanizmów fizycy już wyjaśnili, pozostaje jednak jeszcze wiele pracy. Pełne zrozumienie tego, co dzieje się w dwuwymiarowej plazmie w kanale tranzystora, jest nadal tematem zajmującym fizyków.

Profesor Marian Grynberg pracownik naukowy Uniwersytetu Warszawskiego. Pełnił funkcje dyrektora i zastępcy dyrektora Instytutu Fizyki Doświadczalnej UW. Przez niemal 20 lat kierował Zakładem Fizyki Ciała Stałego (IFD UW), przez 12 był członkiem, a następnie sekretarzem Komisji Półprzewodników Międzynarodowej Unii Fizyki Czystej i Stosowanej (IUPAP). Zasiada w Komitecie Wydawniczym

tygodnika „Solid State Communications” (Elsevier). Przez 13 lat był wiceprezesem Fundacji na Rzecz Nauki Polskiej. Działalność naukowa profesora jest związana z badaniami półprzewodników, początkowo trójwymiarowymi, a następnie o obniżonej wymiarowości oraz fizyką doświadczalną, głównie spektroskopią i magnetospektroskopią w bardzo dalekiej podczerwieni.

Jasne i ciemne strony Wszechświata

Na podstawie odczytu wygłoszonego w dniu 6 października 2011 roku

Marek Demiański

Wydział Fizyki Uniwersytetu Warszawskiego

Wiek dwudziesty przeszedł do historii, jako wiek niezwykłych odkryć naukowych i oszałamiającego postępu technologicznego, które dramatycznie zmieniły nasze życie i nasze wyobrażenia o otaczającym nas świecie. Badanie atomów doprowadziło do powstania mechaniki kwantowej, wnikliwa analiza rozchodzenia się światła była dla Einsteina inspiracją do stworzenia szczególnej teorii względności, która nie tylko zburzyła newtonowską koncepcję absolutnego czasu i przestrzeni, ale spowodowała burzliwy rozwój badań cząstek subatomowych, co ostatnio zakończyło się wielkim sukcesem – odkryciem bozonu Higgsa i wyjaśnieniem pochodzenia masy cząstek elementarnych. Zbudowanie pierwszego tranzystora spowodowało niezwykle szybki, niespotykany dotychczas rozwój elektroniki i informatyki. Wystrzelenie sputnika Ziemi otworzyło drogę do eksplorowania kosmosu, co doprowadziło do spektakularnego lądowania człowieka na Księżycu. Rozwój technik satelitarnych umożliwił wykorzystanie kosmosu do badań astronomicznych i obserwacji nieba w zakresie promieniowania rentgenowskiego i gamma. Rozszyfrowanie kodu genetycznego przyczyniło się do bardzo szybkiego rozwoju biologii molekularnej i medycyny. Postęp technologiczny umożliwił budowanie nowych coraz lepszych i dokładniejszych instrumentów naukowych. Jednym z efektów tego postępu było niewyobrażalne rozszerzenie możliwości badań od najmniejszych obiektów – cząstek elementarnych do tych największych i całego Wszechświata.

Jeszcze w latach dwudziestych minionego wieku astronomowie wyobrażali sobie, że cały Wszechświat

to ogromny układ gwiazd tworzący Drogę Mleczną. Wprawdzie na niebie znajdowano różne obłokopodobne obiekty, niektóre z powodu kształtu i struktury nazywane mgławicami spiralnymi, ale uważano, że są one składnikami Galaktyki.

Przełomu dokonano dopiero po uruchomieniu na Mount Wilson w Stanach Zjednoczonych największego wówczas teleskopu optycznego o średnicy zwierciadła 254 cm. Korzystając z tego teleskopu w 1923 roku Edwin Hubble wypatrył gwiazdy w mgławicy spiralnej w Andromedzie, a następnie, korzystając z zauważonych tam gwiazd zmiennych – cefeid oszacował odległość do tej mgławicy i wykazał, że znajduje się ona daleko poza granicami Drogi Mlecznej. W ten sposób Hubble nie tylko znacznie rozszerzył granice obserwowalnego Wszechświata, ale pokazał, że podstawowymi składnikami Wszechświata nie są gwiazdy, lecz galaktyki. Typowa galaktyka spiralna, których we Wszechświecie jest najwięcej, składa się z około 150 miliardów gwiazd tworzących dyskopodobną strukturę o średnicy około stu tysięcy lat świetlnych (rok świetlny to odległość, jaką sygnał świetlny poruszający się z prędkością trzystu tysięcy kilometrów na sekundę przebywa w ciągu roku). W okolicach Słońca grubość dysku Drogi Mlecznej wynosi około tysiąca lat świetlnych. Obecnie szacuje się, że w obserwowalnej części Wszechświata znajduje się około 100 miliardów galaktyk. Po odkryciu świata galaktyk Hubble resztę swojego życia poświęcił na badanie galaktyk i ich klasyfikację. Kolejnym wielkim odkryciem Hubble'a było stwierdzenie, że galaktyki oddalają się od siebie a prędkość,

z jaką się rozbiegają jest wprost proporcjonalna do ich wzajemnej odległości. Okazało się, że największy obiekt fizyczny, jaki możemy obserwować – cały Wszechświat zmienia się, ewoluuje.

W 1917 roku, zaraz po sformułowaniu ogólnej teorii względności Albert Einstein zaproponował model wszechświata zgodny z relatywistyczną teorią grawitacji. Einstein zaakceptował ówczesny pogląd większości astronomów, że Wszechświat składa się z gwiazd i jest statyczny. Aby zapewnić statyczność Wszechświata Einstein musiał jednak zmodyfikować równania ogólnej teorii względności i wprowadzić dodatkowy człon nazywany obecnie stałą kosmologiczną, który równoważy przyciąganie grawitacyjne. Jeszcze zanim Hubble odkrył, że Wszechświat się rozszerza, rosyjski matematyk i meteorolog Alexander Friedman znalazł inne rozwiązanie równań Einsteina, które opisuje jednorodny i izotropowy Wszechświat wypełniony materią, ale bez stałej kosmologicznej. Wszechświat Friedmana zmienia się w czasie – ewoluuje. Model Friedmana, choć został potwierdzony przez odkrycia Hubble'a, budził kontrowersje, gdyż przewidywał osobliwy początek Wszechświata. Jeżeli materia we Wszechświecie spontanicznie nie powstaje i nie znika, to średnia gęstość Wszechświata powinna rosnąć, w miarę jak będziemy się cofali w czasie do coraz wcześniejszych okresów ewolucji Wszechświata i w końcu osiągniemy moment, kiedy gęstość staje się nieskończona. Ten moment identyfikujemy obecnie z momentem powstania Wszechświata, a kosmologowie ten osobliwy początek Wszechświata nazywają Wielkim Wybuchem.

Początkowo bardzo sceptycznie odnoszono się do modelu Wielkiego Wybuchu i nawet powątpiewano w realność rozszerzania się Wszechświata. Dopiero w drugiej połowie lat czterdziestych XX wieku konsekwencjami osobliwego początku Wszechświata zainteresował się George Gamow. Gamow zdał sobie sprawę z tego, że w początkowym okresie ewolucji Wszechświata materia była tak gęsta i tak gorąca, że nie mogły wówczas istnieć nie tylko atomy, ale również jądra atomowe, innymi słowy, materia była wówczas rozłożona na swoje najbardziej elementarne składniki. W owym czasie za te elementarne składniki uznawano proton, neutron, elektron, foton i neutrino. Ta początkowa gorąca i gęsta mieszanina tych elementarnych składników szybko rozszerzała się i ostygła. Gdy temperatura obniżyła się do około kilku milionów stopni, zaczęły zachodzić reakcje termojądrowe takie, jakie zachodzą obecnie w centrum Słońca. Powstały wów-

czas jądra kilku lekkich pierwiastków, głównie helu, trytu i deuteru. Gamow wraz ze swoimi doktorantami Alpherem i Hermanem obliczyli, że pramateria, z której mogły następnie powstać pierwsze gwiazdy składała się w około 75% z wodoru i w 25% z helu z małą domieszką innych lekkich pierwiastków. Gamow przewidział też, że po tym początkowym gorącym i gęstym etapie ewolucji Wszechświata powinien do dziś pozostać ślad w postaci mikrofalowego tła promieniowania, którego obecna temperatura powinna wynosić kilka stopni powyżej absolutnego zera.

Odkrycie promieniowania relikowego w 1964 roku przez A. Penziasa i R. Wilsona, oszacowanie jego temperatury na około trzy stopnie powyżej absolutnego zera oraz określenie składu chemicznego najstarszych gwiazd i obłoków materii międzygwiazdowej było ostatecznym triumfem modelu Wielkiego Wybuchu. Rozpoczął się złoty okres rozwoju kosmologii.

Coraz dokładniejsze dane obserwacyjne stymulowały pojawienie się nowych pytań. Dlaczego Wszechświat jest niemal płaski albo wręcz płaski? W jaki sposób powstały początkowe zaburzenia gęstości, z których następnie powstały galaktyki? Na początku lat osiemdziesiątych XX wieku, na te pytania znaleziono bardzo sprytną odpowiedź, która na dodatek była związana z ogromnym postępem, jaki się w tym czasie dokonał w teorii cząstek elementarnych i próbami znalezienia teorii unifikującej wszystkie oddziaływania fundamentalne. Alan Guth zauważył, że gdyby w bardzo wczesnych fazach ewolucji energia Wszechświata była zdominowana przez energię potencjalną wolno zmieniającego się w czasie samooddziałującego pola skalarnego, to wówczas Wszechświat rozszerzałby się wykładniczo. Takie wykładnicze rozszerzanie, gdyby trwało odpowiednio długo, spowodowałoby wyptaszczenie się Wszechświata i zanik wszelkich początkowych niejednorodności. Ten okres bardzo szybkiego wykładniczego rozszerzania się Wszechświata nazwano epoką inflacyjną. Przez kilka lat wydawało się, że problemy związane z wyjaśnieniem ewolucji Wszechświata zostały rozwiązane. Jedynym problemem, jaki pozostał do rozwiązania był tak zwany problem brakującej masy. Równania Friedmana, opisujące ewolucję Wszechświata, wiążą tempo rozszerzania się Wszechświata i jego krzywiznę ze średnią gęstością materii we Wszechświecie. Wynikająca z tego związku średnia gęstość materii była jednak wyraźnie większa od oszacowań obserwacyjnych. Astronomowie rozpoczęli poszukiwania ukrytej materii.

Częściowe rozwiązanie tego problemu przyniosły dynamiczne oszacowania masy galaktyk spiralnych. W pierwszym przybliżeniu masę galaktyki można oszacować, dokonując pomiaru jej jasności. Dzieliąc jasność galaktyki przez jasność Słońca otrzymujemy w przybliżeniu liczbę gwiazd, z których składa się galaktyka, a mnożąc tę liczbę przez masę Słońca, otrzymujemy masę galaktyki. Innymi słowy masa galaktyki wyrażona w masach Słońca jest równa jasności galaktyki wyrażonej w jednostkach jasności Słońca. Dominującym świecącym składnikiem galaktyki spiralnej jest dysk galaktyczny. Gwiazdy i obłoki gazowe tworzące dysk galaktyczny wirują wokół centrum dysku. Zwykle centralna część dysku jest najbardziej jasna skąd można wywnioskować, że zawiera też największą część masy dysku. Gdyby tak było, to mierząc prędkość liniową, z jaką gwiazdy poruszają się wokół centrum dysku można oszacować masę, która się tam znajduje. W takim przypadku prędkość gwiazd powinna maleć odwrotnie proporcjonalnie do pierwiastka z ich odległości od centrum. W połowie lat siedemdziesiątych XX wieku Vera Rubin wraz ze swoimi współpracownikami zauważyła, że gwiazdy w zewnętrznej części dysku w galaktyce Andromeda poruszają się względem środka ze statymi prędkościami liniowymi. Oszacowana w ten sposób masa Andromedy była kilka razy większa od masy oszacowanej na podstawie jej jasności. Obserwacje innych galaktyk potwierdziły to odkrycie. Okazało się, że galaktyki spiralne są otoczone niemal sferycznie symetrycznym halo, które zawiera znaczną część masy galaktyki, ale materia tworząca halo nie świeci.

Gdy w drugiej połowie lat osiemdziesiątych rozpoczęły się obserwacje nieba w zakresie promieniowania rentgenowskiego, w centralnych obszarach każdej gromady galaktyk stwierdzono występowanie bardzo gorącego gazu, którego temperatura sięgała kilkunastu milionów stopni. Aby taki bardzo gorący gaz mógł utrzymać się w centralnych obszarach gromad galaktyk przez miliardy lat grawitacyjna studnia potencjalna gromady powinna być odpowiednio głęboka. Ten warunek pozwala na oszacowanie całkowitej masy gromady. Masy gromad galaktyk, oszacowane w taki sposób, są kilkaset razy większe od sumy mas tworzących je galaktyk. Kiedy dokonano bilansu masy we Wszechświecie, otrzymano zaskakujący wynik — całkowita masa była sześć razy większa od masy zwykłej materii, z której zbudowane są gwiazdy i obłoki gazowe. To było szokujące odkrycie. Okazało się, że domi-

nującym materialnym składnikiem Wszechświata jest jakaś egzotyczna materia złożona najprawdopodobniej z masywnych elektrycznie obojętnych cząstek, które wprawdzie oddziałują grawitacyjnie, ale ze znanymi obecnie cząstkami oddziałują bardzo słabo. Ten nowy składnik Wszechświata nazwano ciemną materią. Trwające od ponad dwudziestu lat próby wykrycia cząstek ciemnej materii nadal kończą się niepowodzeniem. Duże nadzieje wiązano z możliwością znalezienia cząstek ciemnej materii wśród produktów zderzeń wysoko energetycznych protonów w potężnym akceleratorze LHC działającym w CERN-ie koło Genewy. Po przeanalizowaniu danych z pierwszych dwóch lat obserwacji, nie znaleziono jednak śladów cząstek ciemnej materii.

Odkrycie ciemnej materii złagodziło nieco rozbieżności między szacowaną teoretycznie średnią gęstością materii we Wszechświecie i danymi obserwacyjnymi oraz zniwelowało trudności, na jakie napotykały próby wyjaśnienia procesu powstawania wielkoskalowej struktury rozkładu materii we Wszechświecie. Początkowe zaburzenia gęstości ciemnej materii mogły zacząć narastać znacznie wcześniej, niż zaburzenia w zwykłej materii, które przez pierwszych trzysta tysięcy lat istnienia Wszechświata były tłumione dzięki oddziaływaniu z promieniowaniem. Dopiero po okresie rekombinacji, gdy powstały neutralne atomy zaburzenia w zwykłej materii, mogły zacząć narastać.

W drugiej połowie lat dziewięćdziesiątych XX wieku ciekawość kilku astronomów doprowadziła do kolejnego wielkiego odkrycia. S. Perlmutter, A. Riess i B. Schmidt postanowili sprawdzić, czy tempo rozszerzania się Wszechświata zmienia się z czasem. W modelu Friedmana stała Hubble'a, czyli współczynnik proporcjonalności między prędkością oddalania się galaktyki a jej odległością wcale nie jest stała, jak sugeruje nazwa, ale zmienia się i we wczesnych epokach ewolucji Wszechświata jego wartość była większa. Jest to związane z hamowaniem tempa rozszerzania się Wszechświata, spowodowanym wzajemnym przyciąganiem grawitacyjnym.

Aby sprawdzić, czy tempo rozszerzania się Wszechświata maleje z czasem, trzeba znaleźć sposób na wyznaczenie odległości do bardzo odległych obiektów. Początkowo do pomiaru odległości do niezbyt odległych galaktyk wykorzystywano cefeidy — gwiazdy zmienne, które w charakterystyczny sposób zmieniają swoją jasność. Na początku XX wieku Henrietta Leavitt, korzystając z danych obserwacyjnych o cefeidach

odkrytych w Obłokach Magellana, zauważyła, że maksymalna jasność cefeid jest proporcjonalna do okresu zmian ich jasności. Okres zmian jasności cefeid łatwo daje się zmierzyć, a korzystając z empirycznej zależności znalezionej przez Panią Leavitt, można wyznaczyć ich maksymalną jasność absolutną. Natężenia strumienia światła, które do nas dociera od dalekiej gwiazdy, jest wprost proporcjonalne do jej jasności absolutnej a odwrotnie proporcjonalne do kwadratu jej odległości. Korzystając z tej zależności oraz z tego, że cefeidy są jasnymi gwiazdami, astronomowie wykorzystywali je do wyznaczania odległości do niezbyt odległych galaktyk. W galaktykach, które znajdują się od nas bardzo daleko nie można jednak wypatrzyć pojedynczych gwiazd i wobec tego trzeba poszukiwać innych obiektów, które mogłyby spełniać rolę „standardowych świec”. Oczywiście, aby można było je zauważyć, powinny to być obiekty bardzo jasne. Najjaśniejszymi, choć tylko przez pewien czas, obiektami na niebie są wybuchy supernowych. Supernowa to spektakularny kres ewolucji masywnej gwiazdy. Cykl przemian termojądrowych zachodzący w centrum gwiazdy kończy się, gdy produktem reakcji jest żelazo. Spalanie – a raczej przekształcanie jąder żelaza w jądra cięższych pierwiastków – wymaga wkładu energii, więc taki proces nie może zachodzić spontanicznie we wnętrzu gwiazd. Żelazo, jako końcowy produkt naturalnych przemian termojądrowych, zaczyna się stopniowo gromadzić w centrum gwiazdy. Masa żelaznego jądra gwiazdy narasta i w końcu ciśnienie wytwarzane głównie przez zdegenerowany gaz elektronowy nie jest w stanie równoważyć sił grawitacyjnych i jądro gwiazdy zaczyna się bardzo szybko, w ciągu zaledwie kilku sekund, zapadać. W przybliżeniu można powiedzieć, że jądra żelaza zostają zmiażdżone i w centrum gwiazdy powstaje mieszanina protonów, neutronów i elektronów. Podczas tego bardzo szybkiego procesu kurczenia wyzwolane są ogromne ilości energii. Znaczna część tej energii jest unoszona przez bardzo intensywny strumień neutrin, a reszta zostaje przetworzona na energię promieniowania i energię kinetyczną odrzucanych zewnętrznych warstw gwiazdy. Gwiazda rozbłyскуje i przez kilka dni świeci bardzo jasno. Dzięki temu wybuchy supernowych można obserwować nawet, gdy zachodzą bardzo daleko. Jasność supernowej zależy od masy wybuchającej gwiazdy, a tego parametru niestety nie można obserwacyjnie wyznaczyć. Innymi słowy zwykłe supernowe, choć są widoczne z bardzo daleka, nie mogą być wykorzystywane jako standardowe świece.

Analiza widm i krzywych zmian blasku supernowych pozwoliła na rozdzielenie ich na dwa typy. W widmach supernowych typu Ia nie występują linie wodoru – najpopularniejszego pierwiastka we Wszechświecie. Od niedawna wiemy, że takie supernowe są generowane przez wybuchające białe karzele. Biały karzeł to zwykle niewielka gwiazda o promieniu porównywalnym z promieniem Ziemi, lecz o masie w przybliżeniu równej połowie masy Słońca. Średnia gęstość takiej gwiazdy jest około milion razy większa od gęstości wody. Przy takich ogromnych gęstościach do głośu zaczynają dochodzić efekty kwantowe i pojawia się nowe źródło ciśnienia – ciśnienie gazu elektronowego. Na początku lat trzydziestych ubiegłego wieku Subrahmanian Chandrasekhar przewidział, że głównym źródłem ciśnienia w białych kartach może być gaz elektronowy. Okazało się przy tym, że ciśnienie gazu elektronowego we wnętrzach białego karta jest w stanie równoważyć siły grawitacyjne tylko wówczas, gdy masa białego karta jest mniejsza od $1.4 M_{\odot}$. Typowy biały karzeł ma masę $\sim 0.6 M_{\odot}$. Jeżeli jednak biały karzeł wchodzi w skład układu podwójnego, którego drugim składnikiem jest normalna gwiazda, to biały karzeł może przyciągnąć materię z tej gwiazdy i w ten naturalny sposób jego masa może wzrastać. Gdy biały karzeł osiągnie masę $1.4 M_{\odot}$, gaz elektronowy w jego jądrze nie jest już w stanie przeciwdziałać siłom grawitacyjnym i jądro gwiazdy zaczyna się gwałtownie kurczyć, co powoduje szybki wzrost temperatury, następuje zapalenie tlenu i węgla, co prowadzi do termojądrowej detonacji i rozerwania całej gwiazdy. Ponieważ masa wybuchającej gwiazdy jest zawsze zbliżona do M_{\odot} , można przypuszczać, że podczas wybuchu takiej supernowej zawsze będą generowane podobne ilości energii, innymi słowy takie supernowe powinny świecić tak samo jasno. W rzeczywistości sytuacja nie jest tak idealna i jasności supernowych typu Ia nie są dokładnie takie same.

Dokładniejsza analiza krzywych zmian blasku supernowych typu Ia, które wybuchają w galaktykach o znanej odległości, doprowadziła do znalezienia zależności między tempem zaniku jasności supernowej a jej jasnością maksymalną. Okazało się, że im jaśniejsza jest supernowa, tym szybciej jej jasność maleje. Dzięki tej zależności można było odpowiednio przekalować obserwowaną jasność, co pozwoliło na wykorzystanie supernowych typu Ia, jako standardowych świec. Jednak supernowe typu Ia pojawiają się nie tylko bardzo rzadko, ale na dodatek w przypadkowych

galaktykach. Trzeba było niezwyklego kunsztu dyplomatycznego i siły perswazji, aby skłonić komitety rozdzielające czas na największych teleskopach optycznych, w tym na teleskopie satelitarnym Hubble'a, do przyznania czasu na przeprowadzenie głębokich przeglądów nieba w poszukiwaniu supernowych typu Ia. W 1998 roku po przeanalizowaniu danych z zaledwie czterech supernowych o przesunięciu ku czerwieni większym od 0.3 i mniejszym od 0.86 zauważono pierwsze sygnały świadczące o tym, że stała kosmologiczna jest różna od zera. Wykorzystanie satelitarnego obserwatorium Hubble'a pozwoliło w ciągu następujących dziesięciu lat na zaobserwowanie kilkudziesięciu supernowych o przesunięciu ku czerwieni > 1 ($z_{\max} = 1.415$). Te dane pozwoliły potwierdzić niezbicie, że obecnie Wszechświat rozszerza się coraz szybciej. Najprostszym wyjaśnieniem obserwowanego przyspieszonego tempa rozszerzania się wszechświata jest założenie, że wprowadzona przez Einsteina w 1917 roku stała kosmologiczna jest różna od zera. Z danych obserwacyjnych wynika, że obecnie stała kosmologiczna wnosi dominujący wkład do średniej gęstości energii we Wszechświecie. Trójka astronomów, którzy dokonali tego odkrycia — Saul Perlmutter, Brian Schmidt i Adam Riess w 2011 roku otrzymali Nagrodę Nobla z fizyki.

Wiek XXI zaczyna się od bardzo ważnych odkryć astronomicznych, które obnażają naszą bardzo ograniczoną wiedzę o podstawowych składnikach materii wypełniających cały Wszechświat. Materia, z której jesteśmy zbudowani — materia barionowa, jest dominującym składnikiem Układu Słonecznego, ale stanowi zaledwie 4% średniej gęstości materii i energii we Wszechświecie. Aż 96% tego, co istnieje we Wszechświecie to są składniki, które dają o sobie znać, ale ich natura nie jest znana. Te dominujące ciemne składniki Wszechświata to ciemna materia — najprawdopodobniej złożona z egzotycznych cząstek i ciemna energia, która jest kojarzona ze stałą kosmologiczną, ale nie można obecnie wykluczyć, że jest to energia jakiegoś nieznanego jeszcze pola skalarnego lub, na przykład, oddziaływanie pochodzące z innych wymiarów.

Na progu XXI wieku jasne strony Wszechświata odstąpiły przed nami otchłani ciemności.

Profesor Marek Demiański specjalista z dziedzin astrofizyki relatywistycznej i kosmologii, pracownik naukowy w Instytucie Fizyki Teoretycznej Wydziału Fizyki Uniwersytetu Warszawskiego od 1962 roku. Na początku swojej drogi naukowej zajmował się matematycznymi i fizycznymi aspektami ogólnej teorii względności, od lat siedemdziesiątych zeszłego stulecia zajmuje się głównie astrofizyką relatywistyczną i kosmologią, w szczególności procesem powstawania struktury we wszechświecie, własnościami promieniowania relikтового oraz naturą ciemnej energii i ciemnej materii. Współpracuje z wieloma zagranicznymi ośrodkami naukowymi, np. Instytutem Badań Przestrzeni Kosmicznej Rosyjskiej Akademii Nauk w Moskwie, Instytutem Nielsa Bohra w Kopenhadze i Uniwersytetem Frederico II w Neapolu. Był profesorem wizytującym w Kalifornijskim Instytucie Technologicznym oraz na Uniwersytecie Stanu Teksas w Austin. Od wielu lat współpracuje z Williams College w Massachusetts, USA.

