

DELTA

MATEMATYKA – FIZYKA – ASTRONOMIA – INFORMATYKA

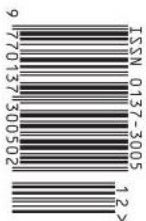
$h=6,63 \cdot 10^{-34} \text{ J}\cdot\text{s}$ $\pi=3,141592\dots$

NR 12 (595) 2023

CENA 8 ZŁ - VAT 8%




PL ISSN 0137-3005 | NR IND 35-550 X
MIESIĘCZNIK

www.deltami.edu.pl

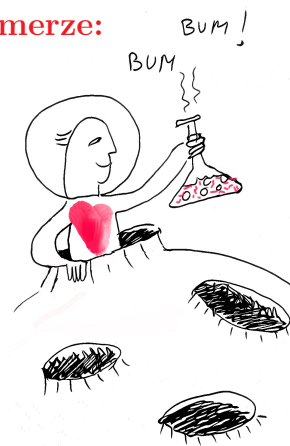


Nakład: 2800 egz.

SPIS TREŚCI NUMERU 12 (595)

W styczniu 1974 roku...	str. 1
Jeszcze o taliach <i>Marek Kordos</i>	str. 2
Prawo powszechnego ciężenia <i>Witold Sadowski</i>	str. 4
 Zadania	str. 7
Czego nie da się zobaczyć? <i>Wojciech Czerwiński</i>	str. 8
Prostota zrozumianego problemu <i>Stanisław Mrówczyński</i>	str. 10
Renesans słów Lyndona <i>Jakub Radoszewski</i>	str. 12
37 lat później <i>Krzysztof Ciesielski</i>	str. 16
Twierdzenie Białynickiego-Biruli o rozkładzie <i>Maria Donten-Bury, Joachim Jelisiejew, Jarosław A. Wiśniewski</i>	str. 19
Współbieżność jest nieintuicyjna <i>Tomasz Idziaszek</i>	str. 22
 Moja Delta <i>Magdalena Fikus</i>	str. 24
Harriot, Kepler, Hales i inni <i>Paweł Strzelecki</i>	str. 26
Tym razem się nie uda... <i>Tomasz Kazana</i>	str. 29
Jak na wyświetlaczu <i>Joanna Jaszewska</i>	str. 32
Berstery rentgenowskie dziesięć lat później <i>Agnieszka Majczyna</i>	str. 34
Liczba chromatyczna z komputera <i>Michał Adamaszek</i>	str. 36
Mało przekątnych – duży problem <i>Jerzy Bednarczuk</i>	str. 38
Dementi <i>Piotr Zalewski</i>	str. 40
Klub 44	str. 43
Prosto z nieba: Co trzęsie Księżycem?	str. 44
Niebo w grudniu	str. 44
Aktualności	
Pół wieku osławiania Wszechświata	str. 46
 Różnice kwadratów w równaniach diofantycznych <i>Bartłomiej Bzdega</i>	str. 48

W następnym numerze:
**Ekstremalne
kosmiczne
laboratoria**



Miesięcznik Delta – matematyka, fizyka, astronomia, informatyka założony został w 1974 roku przez Marka Kordosa. Wydawany jest przez Uniwersytet Warszawski przy współpracy towarzystw naukowych: Polskiego Towarzystwa Matematycznego, Polskiego Towarzystwa Fizycznego, Polskiego Towarzystwa Astronomicznego i Polskiego Towarzystwa Informatycznego.

Komitet Redakcyjny: dr Waldemar Berej, doc. dr Piotr Chrzastowski-Wachtel, dr Krzysztof Ciesielski, prof. UJ – przewodniczący, prof. dr hab. Bożena Czerny, dr Andrzej Dąbrowski, dr Tomasz Greczyło, prof. UW, dr Adam Gregosiewicz, dr Andrzej Grzesik, prof. dr hab. Agnieszka Janiuk, dr hab. Artur Jeż, prof. UW, prof. dr hab. Bartosz Klin, prof. dr hab. Andrzej Majhofer – wiceprzewodniczący, dr Adam Michalec, prof. dr hab. Damian Niwiński, prof. dr hab. Krzysztof Oleszkiewicz, dr hab. Krzysztof Pawłowski, prof. PAN, dr Milena Ratajczak, dr hab. Radosław Smolec, prof. PAN, prof. dr hab. Paweł Strzelecki, prof. dr hab. Andrzej Wysmołek.

Redaguje kolegium w składzie: Wiktor Bartol, Michał Bejger, Szymon Charzyński – red. nac., Agnieszka Chudek, Anna Durkalec, Katarzyna Małek, Michał Miśkiewicz, Wojciech Przybyszewski, Łukasz Rajkowski – z-ca red. nac., Anna Rudnik, Krzysztof Rudnik, Oskar Skibski, Marzanna Wawro – sekr. red.

Adres do korespondencji:
Redakcja *Delty*, ul. Banacha 2, pokój 4020, 02-097 Warszawa
e-mail: delta@mimuw.edu.pl tel. 22-55-44-402.

Okładki i ilustracje:
Anna Ludwicka Graphic Design & Serigrafia.
Skład systemem L^AT_EX wykonała Redakcja.
Druk: Poligrafia NOT poligrafianot.pl

Prenumerata:
Garmond Press: www.garmondpress.pl
Kolporter: www.kolporter.com.pl (tylko instytucje)
RUCH S.A.: www.prenumerata.ruch.com.pl

Numery archiwalne (od 1987 r.) można nabyć w Redakcji osobiście lub zamówić przez e-mail.

Cena 1 egzemplarza: z ostatnich 12 miesięcy 8 zł;
wcześniejsze egzemplarze 3 zł



Strona internetowa (w tym artykuły archiwalne, linki itd.):
deltami.edu.pl

Można nas też znaleźć na
facebook.com/Delta.czasopismo

Wydawca: Uniwersytet Warszawski



Delta, styczeń 1974

W styczniu 1974 roku pierwszy numer *Delty* trafił do rąk czytelników (prosto z Drukarni im. Rewolucji Październikowej). Zeszyt, po który właśnie sięgnąłeś, Szanowny Czytelniku, zamyka więc półwiecze istnienia naszego miesięcznika, a następny (1/2024) rozpocznie drugie półwiecze.

Bardzo prawdopodobne, że czytając te słowa, nie trzymasz jednak w rękach pachnącego farbą papierowego zeszytu, tylko tablet czy smartfon, albo czytasz tekst z monitora LCD swojego komputera, a plik z *Delta* ściągasz sobie ze strony deltami.edu.pl. Te urządzenia i kanały komunikacji nie istniały 50 lat temu, podobnie jak narzędzia, które nam, członkom redakcji, wydają się teraz nieodzowne w naszej pracy. Autorzy przysyłają nam propozycje tekstów e-mailem. Składamy je w LaTeXu. Pliki z artykułami trzymamy w redakcyjnej chmurze, dzięki czemu wszyscy redaktorzy jednocześnie mają do nich dostęp i widzą zmiany i komentarze wprowadzane przez innych. Praca redakcji kończy się co miesiąc wygenerowaniem pliku, który wysyłamy do drukarni i publikujemy w Internecie.

Ten skok technologiczny był możliwy dzięki odkryciom naukowym dokonywanym przez ludzi, którzy kiedyś w młodości zafascynowali się nauką. Żeby dotrzeć potem do frontu badań naukowych, musieli przyswoić masę zgromadzonej wcześniej wiedzy w wybranej dziedzinie. Wymagało to mozolnej pracy, studiowania podręczników, prac naukowych, uczestnictwa w wykładach, dyskusjach, rozwiązywania tysięcy zadań. Aby wybrać taką drogę i konsekwentnie się jej trzymać, niezbędna jest odpowiednia motywacja – ciekawość, fascynacja i dążenie do poszerzenia horyzontów naszej wiedzy naukowej.

Delta od 50 lat zaprasza na swoje łamy autorów zadań i artykułów, którzy sami kiedyś zafascynowali się nauką, żeby mogli zarażać swoją pasją kolejne pokolenia. Redakcji zależy na tym, by naukowcy piszący w *Delcie* zarówno o aktualnych badaniach i odkryciach, jak i o klasycznych wynikach, robili to w sposób jak najbardziej przystępny. Dla autorów bywa to nie lada wyzwaniem. Niejednokrotnie spotkałiśmy się z opinią, że napisanie artykułu do *Delty* może być trudniejsze niż napisanie pracy do branżowego czasopisma naukowego.

Przez pół wieku nie tylko warsztat pracy redakcji uległ zmianie. Kolejny aspekt to gigantyczna zmiana w dostępności źródeł informacji, która się dokonała. Dawniej młodzi pasjonaci mieli utrudniony dostęp do materiałów pozwalających rozwijać ich zainteresowania. Dziś natomiast jesteśmy bombardowani informacjami – najróżniejsze serwisy i aplikacje konkurują o naszą uwagę. Odbiorca stoi więc przed problemem wyboru z całego tego bogactwa tych treści, które mają szansę być dla niego wartościowe i interesujące. Mnogość źródeł informacji, niestety, nie zawsze idzie w parze z jakością. Każdy może wypowiadać się na dowolny temat, a Internet pozwala błyskawicznie docierać do ogromnej liczby odbiorców. Nie ulega więc wątpliwości, że czasopisma takie jak *Delta* odpowiadają na trochę inne zapotrzebowanie niż w czasach przedinternetowych. Działając w takich okolicznościach, redakcja *Delty* tym bardziej czuje ciężar odpowiedzialności za rzetelność materiałów, które publikuje. Jedną z zasad, którą przejeśliśmy

od założyciela naszego pisma, Marka Kordosa, i wciąż jesteśmy jej wierni, jest to, że „autorzy piszą o rzeczach, na których się znają”. Kolejna wypracowana przez prof. Kordosa zasada to, że każdy artykuł jest czytany i analizowany przez kilku różnych redaktorów. Z kolei nad pracą redakcji czuwa Komitet Redakcyjny, który zbiera się co pół roku, żeby podzielić się swoimi cennymi uwagami i obserwacjami z członkami redakcji.

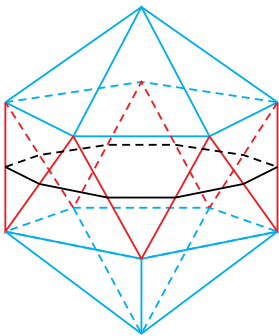
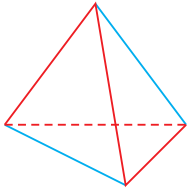
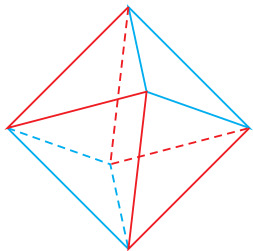
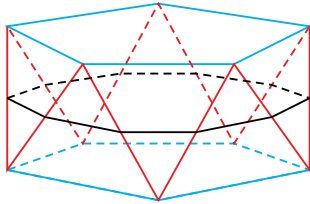
Przecięcie aktualnego zespołu redakcyjnego z zespołem, który redagował pierwszy numer *Delty*, jest zbiorem pustym. Mimo tego pewne elementy w pracy redakcji nie zmieniły się przez te pół wieku. Na przykład to, że nadal, tak samo jak pięćdziesiąt lat temu, spotykamy się raz w tygodniu w redakcji. Czytamy na głos artykuły, dyskutujemy o nich. Potem redagujemy, proponujemy autorom zmiany. Niektóre artykuły przychodzą w formie praktycznie gotowej do publikacji (wystarczy tylko dostawić przecinki), ale są też takie, nad którymi praca trwa miesiącami i wymaga wielu iteracji wymiany korespondencji z autorem. Te dyskusje redakcyjne i praca w zespole złożonym z ludzi napędzanych ciekawością i zafascynowanych nauką jest dla piszącego te słowa najprzyjemniejszym elementem tej pracy. Przez redakcję *Delty* przewinęło się już dobre kilkadziesiąt osób, których listę prezentujemy na ostatniej stronie. Każdy był redaktor, którego miałem przyjemność poznać, miło wspomina czas spędzony w redakcji. W jubileuszowym numerze publikujemy przygotowane specjalnie na tę okazję teksty Deltowych Weteranów: byłych członków kolegium redakcyjnego oraz autorów wieloletnich cykli. Żeby pomieścić całe to bogactwo, numer na 50-lecie ma 50 stron (licząc, jak w informatyce, od strony zero, czyli spisu treści). Przypominamy też 50 archiwalnych okładek – po jednej z każdego rocznika.

Z okazji 50-lecia na naszej stronie deltami.edu.pl znajdzie się pełne archiwum wszystkich numerów *Delty*. Efekty pracy ponad tysiąca autorów i dziesiątek redaktorów będą dostępne dla wszystkich szukających rzetelnych informacji w Internecie.

Szymon CHARZYŃSKI

*Twórca *Delty*, w latach 1974–2018 jej redaktor naczelny

W tym tekście występować będą wielościany mające wszystkie krawędzie równej długości, niech to będzie 1.



Delta, marzec 1975

Jeszcze o taliach

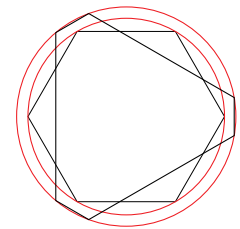
Marek KORDOS*

W znakomitym artykule *O wieszaniu bombek na choince*, Δ_{11}^{12} , Krzysztof Rudnik definiuje talie: **talia** jakiejś bryły to najmniejszy okrąg otaczający ją i niedający się z niej zsunąć. Kluczowymi bohaterami opowieści są foremne **antygraniastosłupy**.

Antygraniastosłup to bryła ograniczona przez dwa jednakowe wielokąty foremne (podstawy) leżące w płaszczyznach równoległych i nieco względem siebie obrócone, tak by można je było połączyć paskiem trójkątów równoramiennych (a nie, tak jak w graniastosłupie, prostokątami), które też antygraniastosłup ograniczają. Antygraniastosłup jest foremny, gdy trójkąty mają wszystkie boki tej samej długości co jego podstawy. Obok narysowany jest **5-antygraniastosłup foremny**. Jego talia to okrąg opisany na przecięciu tego antygraniastosłupa płaszczyzną równoległą do jego podstaw i w tej samej od nich odległości – przecięciem tym jest dziesięciokąt foremny o boku (przy naszej przedstawionej na marginesie umowie) $\frac{1}{2}$.

Co więcej – tak samo otrzymujemy talie dowolnego n -antygraniastosłupa foremnego: są to okręgi opisane na $2n$ -kątach foremnych o boku $\frac{1}{2}$. To śmiała teza i wypadaloby jej dowieść. Tym bardziej że wszystkie przekroje danego antygraniastosłupa płaszczyznami równoległymi do podstaw mają obwód tej samej długości. Czemu więc obręcz nałożona na „środkowy” przekrój nie miałaby się zsunąć?

Argumentem uzasadniającym ten fakt jest spostrzeżenie, że spośród otrzymanych w przecięciach wielokątów najmniejszy okrąg opisany ma ten, który ma boki jednakowej długości. Miłośnicy rachunków zapewne potrafią to obliczyć, ale my obejrzymy to dla 3-antygraniastosłupa, czyli (prawda?) **ośmiościanu foremnego**, gdzie talia jest sześciokątem (przy większej liczbie boków różnica między wielokątem foremnym a okręgiem jest mało widoczna).



Czarne obwody są równe

Zwróćmy jeszcze uwagę na fakt, że antygraniastosłup leży w całości po jednej stronie płaszczyzny każdego ograniczającego go wielokąta, więc (jako przecięcie półprzestrzeni) jest wypukły. Może zatem dziwić, że bryła wypukła ma talię, choć... (ale o tym potem).

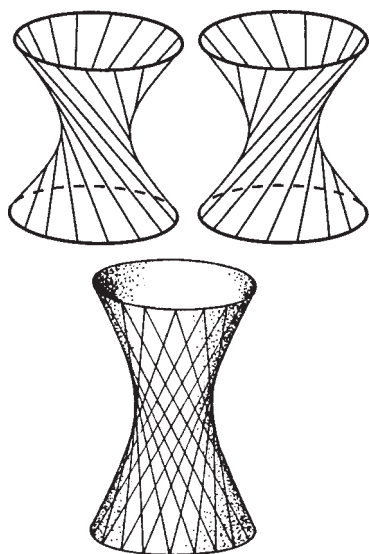
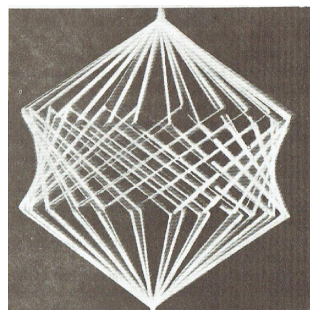
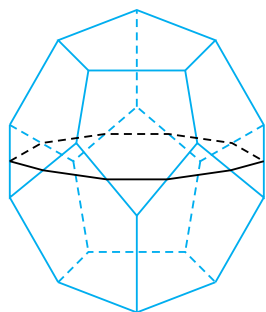
Autor wspomnianego na początku artykułu zaleca wklejanie paska trójkątów równobocznych w rozciętą wzdłuż równika kulistą bombkę, by zyskiwała tym sposobem talię, za którą można by ją wieszać na choince. Mnie jednak to się nie udawało – szklane bombki nie chciały współpracować, więc przyjrzałem się innym doskonałościom: wielościanom foremnym, zwanym też platońskimi.

Jeden z nich to wymieniony już ośmiościan foremny. Dzięki swoim symetriom ma on cztery talie, a nie jedną, jak większość antygraniastosłupów. Inne bryły foremne już antygraniastosłupami nie są. Ale śmiało uogólniając, możemy rozważyć 2-antygraniastosłup foremny: jego podstawy wyglądają jak odcinki, a różnią się od odcinków tym, że są „podwójne” – mają prawy bok i lewy bok. Jeśli zbudujemy taki 2-antygraniastosłup, to okaże się, że mamy **czworościan foremny**, i jest od razu (?) widoczne, że ma trzy talie opisane na kwadratach.

Za pomocą 5-antygraniastosłupa, narysowanego na górze strony, można jeszcze rozstrzygnąć problem talii **dwudziestościanu foremnego** – wystarczy na podstawach zafundować „czapeczki” – ostrosłupy z trójkątów równobocznych. Dwudziestościan będzie wtedy miał talię taką jak 5-antygraniastosłup (czyli opisaną na dziesięciokącie), ale nie jedną, tylko sześć.

Pozostałe dwa wielościany już tak prosto nie poddadzą się. W sprawie ich talii mamy jednak potężnego sojusznika – to Hugo Steinhaus. Rozwiązania można znaleźć w *Kalejdoskopie matematycznym*.

Kalejdoskop matematyczny Hugona Steinhausa został wydany w 1938 roku we Lwowie równocześnie z zamówioną przez Amerykanów angielską wersją (*Mathematical snapshots*). Ma wiele wydań w najrozmaitszych językach, w tym cztery polskie („najnowsze” wydanie, z 1989 roku, jest najsłabsze edytorsko – radzę poszukać wcześniejszych). Przytoczony tu eksperyment jest w *Kalejdoskopie* oznaczony numerami 243–247, a w *Delcie* znalazł się w Δ_{11}^{11} na stronie 0.



Studenci Autora omawianego artykułu wykonali całą serię prawdziwych, bo kartonowych, wielościanów wyposażonych w prawdziwe, bo miedziane, talie, a niemających związków z hiperboloidami.

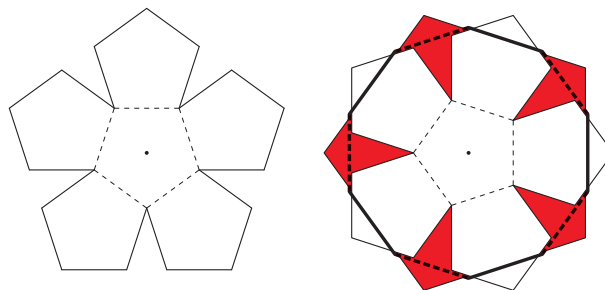
W *Kalejdoskopie* można znaleźć zachętę do takiego eksperymentu.

• Z tekturki wytnij dwie takie figury, jak poniżej (bok 3 cm) i powyginaj wzdłuż przerywanych linii.

• Połóż jedną na drugiej tak, by rogi dolnej jednakowo wystawały spod górnej.

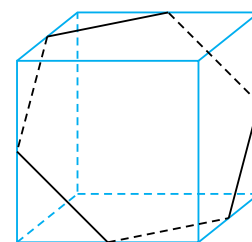
• Przyciskając środek palcem, nałóż recepturkę tak, by była nad rogami górnej i pod rogami dolnej.

• PUSĆ!



Wynik będzie mniej więcej taki, jak widać obok. Otrzymamy **dwunastościan foremny** (każdy zauważył, że tekturki składały się z pięciokątów foremnych). Można na to spojrzeć jak na modyfikację antygraniastosłupów: tym razem podstawy pięciokątne, jak trzeba, lekko obrócone, łączymy paskiem nie trójkątów równobocznych, lecz foremnych pięciokątów. Identyczna jak w przypadku antygraniastosłupów argumentacja wskazuje, że okrąg opisany na przecięciu płaszczyzną tak samo odległą od obu podstaw jest talią. Przecięcie to jest dziesięciokątem foremnym, o bokach będących połową przekątnej pięciokąta. Symetrie zaś dwunastościanu dowodzą, że takich talii jest sześć.

W *Kalejdoskopie* pod numerem (203) znajdujemy zdjęcie sześciianu obracającego się względem swojej dużej przekątnej. Widzimy tu dwie czapeczki (jak w dwudziestościanie) i fragment hiperboloidy obrotowej (jest ona wynikiem obracania prostej względem skośnej do niej osi). Tu słowo „talia” samo się narzuca – wręcz widoczny jest okrąg o promieniu będącym odległością obracanej prostej od osi obrotu (bo proste skośne mają jedną wspólną prostopadłą realizującą minimum odległości ich punktów).

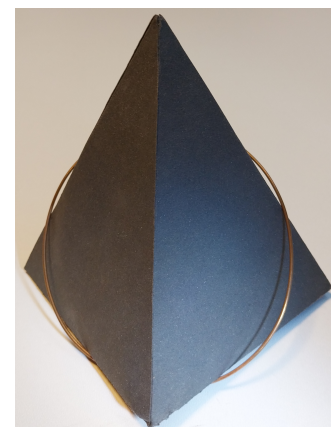


Powstanie tego zdjęcia udziela odpowiedzi na pytanie o talie sześcianu – da się go „wyjąć” z tego zdjęcia: talia sześcianu to okrąg opisany na jego przecięciu z płaszczyzną połowiącą prostopadłe jego przekątną. To przecięcie to sześciokąt foremny o boku $\frac{\sqrt{2}}{2}$. Tali jest tyle co przekątnych, czyli cztery.

Gdy spojrzymy na hiperboloidę w ten sposób, że powstaje ona z walca, którego podstawy są nieco względem siebie obrócone, to zobaczymy, że tworzące staną się skośne względem jego osi. A to demonstruje fakt, że konstrukcja hiperboloidy jest identyczna z konstrukcją antygraniastosłupa (tyle że jest „ciąglą”). Co więcej, łatwo wskazać osie, podczas obrotu względem których każdy z występujących w tym tekście wielościanów może zafundować sobie taką fotkę, jaką Steinhaus zafundował sześcianowi – z kawałkiem hiperboloidy w środku.

I tak się okazało, że wszystkie przykłady talii wielościanów przytoczone w tym tekście to po prostu ukryta obrotowa hiperboloida jednopowłokowa.

Oczywiście istnieją wielościany wypukłe mające talię, a niespokrewnione z hiperboloidą, jak choćby ten z miedzianą talią widoczny na zdjęciu obok.



Prawo powszechnego ciążenia

Witold SADOWSKI*

* University of Bristol

Redaktor *Delta* w latach 2000–2011.

Jedną z najbardziej zdumiewających cech matematyki to fakt, że obiekty czy też zjawiska z pozoru niczym ze sobą niepowiązane okazują się ni stąd ni zowąd bliskimi krewnymi. Wiadomo na przykład, że jedyne trajektorie ciał niebieskich w polu grawitacyjnym Słońca to dokładnie te krzywe, które otrzymuje się z przecięcia stożka płaszczyzną, to znaczy okrąg, elipsa, parabola lub hiperbola.

Jak jednak spojrzeć związek między cięciem stożka płaszczyzną a ruchem ciał niebieskich?! Pełna odpowiedź na to pytanie to jeden z największych triumfów w historii matematyki. Jego autorem jest oczywiście Izaak Newton (choć prelude napisał Johannes Kepler).

W kultowych „Wykładach z historii matematyki” Marka Kordosa (a także w Δ_{81}^8 , w artykule Piotra Mankiewicza) znaleźć można rekonstrukcję rozumowania Newtona oraz słynny dialog pomiędzy Newtonem a odwiedzającym go Halleyem:

„Halley: Jak wyglądałaby siła, która powoduje, że planety krążą po orbitach eliptycznych?

Newton: Odwrotność kwadratu.

Halley: Skąd pan to wie?

Newton: Po prostu obliczyłem.”

W tym artykule nie będziemy odtwarzać rozumowania angielskiego geniusza. Pokażemy tylko, jak w możliwie najprostszy sposób wydedukować z eliptycznego ruchu planet (Kepler) siłę, która na te planety działa (Newton).

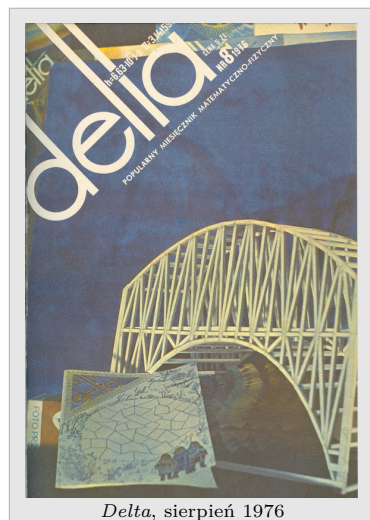
Elipsa na trzy sposoby

Żeby zorientować się, jakie równanie ma owalna krzywa powstająca z odpowiedniego przecięcia stożka płaszczyzną, wyobraźmy sobie dwie sfery, z których każda jest jednocześnie styczna do stożka i do płaszczyzny tnącej. W przekroju zawierającym oś symetrii stożka wyglądać to będzie jak na rysunku 1.

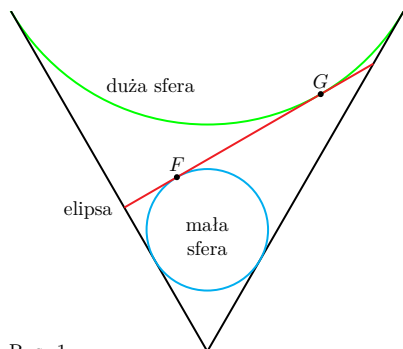
Oznaczmy przez F i G punkty styczności sfer z płaszczyzną (punkty te nazywać będziemy ogniskami elipsy). Zauważmy też, że punkty styczności sfer ze stożkiem tworzą dwa okręgi na powierzchni stożka. Niech odległość między tymi okręgami będzie równa $2a$. Niech też $F'G'$ będzie jakimkolwiek odcinkiem długości $2a$ łączącym oba okręgi styczności. Oznaczmy przez X punkt na odcinku $F'G'$, który należy do elipsy. Korzystając z tego, że dwa odcinki o tym samym początku styczne do tej samej sfery muszą być równej długości (rys. 2), mamy $|FX| = |F'X|$ oraz $|GX| = |GX|$ (rys. 3), czyli

$$|FX| + |GX| = |F'X| + |GX| = 2a.$$

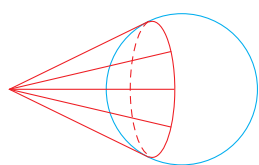
A zatem krzywa, którą otrzymaliśmy z przecięcia stożka, jest złożona z punktów, których suma odległości od punktów F i G jest stale równa $2a$ (rys. 4). Jest to jedna z możliwych definicji elipsy (punkty F i G nazywamy wówczas jej ogniskami).



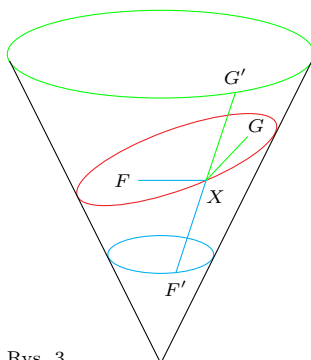
Delta, sierpień 1976



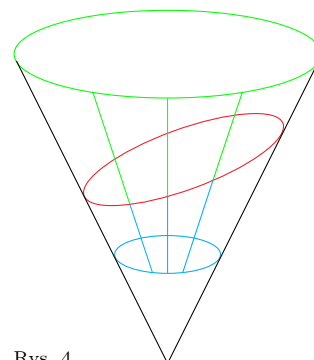
Rys. 1



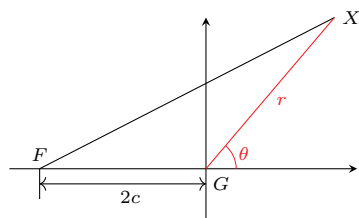
Rys. 2



Rys. 3



Rys. 4



Rys. 5

Znajdźmy teraz równanie elipsy w biegunowym układzie współrzędnych o środku w ognisku G . Niech X będzie punktem na elipsie w odległości r od G i niech kąt pomiędzy GX a osią xków będzie równy θ (rys. 5). Z wyprowadzonej własności elipsy wynika, że $|FX| = 2a - r$.

Oznaczając teraz odległość ognisk F i G przez $2c$ i korzystając z twierdzenia cosinusów w trójkącie FGX , mamy

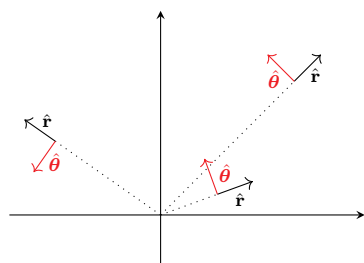
$$(2a - r)^2 = (2c)^2 + r^2 - 4cr \cos(180^\circ - \theta).$$

Po standardowych przekształceniach dostajemy:

$$r = \frac{\alpha}{1 + \varepsilon \cos \theta},$$

gdzie $\alpha = \frac{a^2 - c^2}{a}$, $\varepsilon = \frac{c}{a}$.

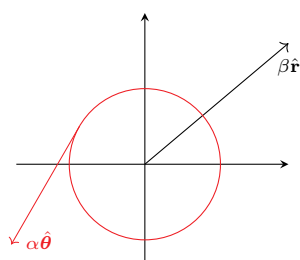
Jest interesujące, że wszystkie krzywe stożkowe są opisywane tym samym równaniem. Dla elipsy mamy $0 \leq \varepsilon < 1$ (zero dla okręgu), dla paraboli $\varepsilon = 1$, a dla hiperboli $\varepsilon > 1$. Uniwersalność tego wzoru podkreśla, że rzeczywiście mówimy o krzywych z tej samej rodziny.



Rys. 6

Prędkość i energia

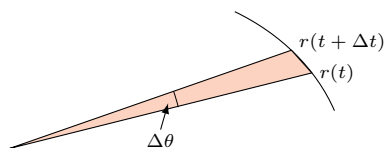
Ponieważ rozważać będziemy sytuację w biegunowym układzie współrzędnych, przypomnijmy, jak najwygodniej opisywać prędkość w tym układzie. W tym celu zdefiniujemy w każdym punkcie płaszczyzny wektory \hat{r} oraz $\hat{\theta}$. Wektor \hat{r} w punkcie (x, y) ma ten sam zwrot co wektor o współrzędnych (x, y) , ale długość zawsze równą jeden. Wektor $\hat{\theta}$ to wektor jednostkowy prostopadły do \hat{r} , co obrazuje rysunek 6. Rozważmy teraz dwa specjalne przypadki ruchu (rys. 7):



Rys. 7

1) Ruch wzdłuż półprostej wychodzącej ze środka układu współrzędnych. Ruch jest wciąż w kierunku $\pm \hat{r}$ i odbywa się z taką szybkością, z jaką zmienia się r – odległość od środka układu współrzędnych. Prędkość w takim ruchu jest zatem równa: $\mathbf{v} = \frac{dr}{dt} \hat{r}$.

2) Ruch po okręgu o promieniu r . Tym razem cały czas poruszamy się w kierunku $\pm \hat{\theta}$. Szybkość w tym ruchu jest tym większa, im większy jest promień i prędkość kątowa, a zatem $\mathbf{v} = r \frac{d\theta}{dt} \hat{\theta}$.



Rys. 8

Składając te dwa przypadki w jeden i nieco machając rękami, dostajemy ogólny wzór na prędkość:

$$\mathbf{v} = \frac{dr}{dt} \hat{r} + r \frac{d\theta}{dt} \hat{\theta}.$$

Wynika stąd, że energia kinetyczna dana jest poniższym wzorem (korzystamy tu z faktu, że jednostkowe wektory \hat{r} oraz $\hat{\theta}$ są prostopadłe):

$$E_K = \frac{1}{2} m |\mathbf{v}|^2 = \frac{1}{2} m \left[\left(\frac{dr}{dt} \right)^2 + r^2 \left(\frac{d\theta}{dt} \right)^2 \right].$$

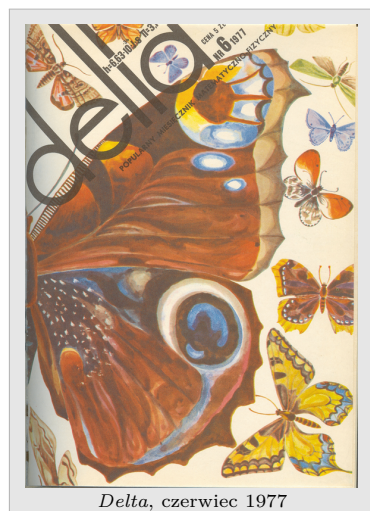
Wykorzystamy teraz odkryte przez Keplera prawa ruchu planet do wykazania, że energia kinetyczna planety zależy tylko od jej odległości od Słońca! Jest to zaskakujące, bo okazuje się, że z ogólnego wzoru można w przypadku planet wyrugować zupełnie pochodne czasowe r i θ .

Kepler i Newton

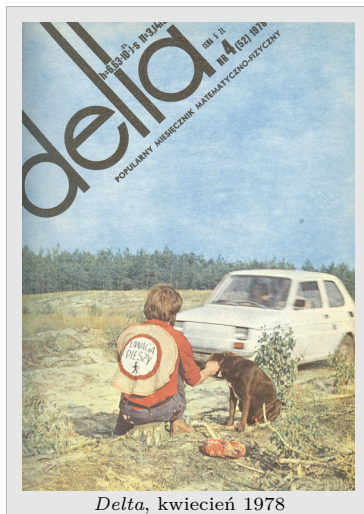
Z trzech praw Keplera, użyjemy dwóch pierwszych:

- 1) Planety poruszają się po orbitach eliptycznych.
- 2) Odcinek łączący planetę ze Słońcem wymiata to samo pole w jednakowych przedziałach czasu.

Na początek wykorzystajmy II prawo (o którym można więcej przeczytać w artykule Grzegorza Łukaszewicza i Mikołaja Sierżęgi, Δ_{20}^{10}). Rozpatrzmy moment, gdy planeta w chwili t jest w odległości r od środka układu współrzędnych. Po bardzo krótkim czasie Δt jej odległość od Słońca praktycznie się nie zmieni: $r(t) \approx r(t + \Delta t)$ (rys. 8), a jej współrzędna kątowa zmieni się



Delta, czerwiec 1977



o małej $\Delta\theta$. Pole wymiecione w czasie Δt jest więc w przybliżeniu równe polu trójkąta:

$$\frac{1}{2}r^2 \sin(\Delta\theta) \approx \frac{1}{2}r^2 \Delta\theta,$$

więc po podzieleniu przez Δt i odpowiednim przejściu do granicy otrzymujemy $\frac{1}{2}r^2 \frac{d\theta}{dt}$ jako prędkość „wymiatania pola”. Wynika stąd, że jeśli zdefiniujemy teraz moment pędu:

$$L = mr^2 \frac{d\theta}{dt},$$

to pozostanie on niezmienny w czasie ruchu planety. To dla nas dobra wiadomość, bo pozwala pozbyć się ze wzoru na energię kinetyczną pochodnej kąta θ . Wystarczy tylko podstawić $\frac{d\theta}{dt} = \frac{L}{mr^2}$.

Teraz potrzebujemy jeszcze pozbyć się pochodnej: $\frac{dr}{dt}$. W tym celu równanie elipsy zapiszmy w postaci:

$$(*) \quad \frac{\alpha}{r} - 1 = \varepsilon \cos \theta$$

i zróżniczkujmy obie strony po czasie, dostając prostą zależność między pochodnymi funkcji $r(t)$ i $\theta(t)$:

$$-\frac{\alpha}{r^2} \frac{dr}{dt} = -\varepsilon \sin \theta \frac{d\theta}{dt}.$$

Po szybkich przekształceniach mamy:

$$\frac{dr}{dt} = \frac{\varepsilon}{\alpha} r^2 \frac{d\theta}{dt} \sin \theta = \frac{L\varepsilon}{m\alpha} \sin \theta.$$

Wynika stąd, że:

$$E_K = \frac{1}{2}m \left[\left(\frac{dr}{dt} \right)^2 + r^2 \left(\frac{d\theta}{dt} \right)^2 \right] = \frac{L^2}{2m} \left(\frac{\varepsilon^2}{\alpha^2} \sin^2 \theta + \frac{1}{r^2} \right).$$

Korzystając z tego, że $\sin^2 \theta = 1 - \cos^2 \theta$ oraz z (*), dostajemy po zupełnie automatycznych rachunkach:

$$\begin{aligned} E_K &= \frac{L^2}{2m} \left(\frac{\varepsilon^2 - \varepsilon^2 \cos^2 \theta}{\alpha^2} + \frac{1}{r^2} \right) = \frac{L^2}{2m} \left(\frac{\varepsilon^2}{\alpha^2} - \frac{1}{\alpha^2} \left(\frac{\alpha}{r} - 1 \right)^2 + \frac{1}{r^2} \right) = \\ &= \frac{L^2}{2m} \left(\frac{\varepsilon^2 - 1}{\alpha^2} + \frac{2}{\alpha r} \right) = B + \frac{C}{r} \end{aligned}$$

dla pewnych stałych B i C .

Aby wywnioskować teraz, jaka siła działa na planetę, wystarczy skorzystać z prawa zachowania energii. Energia potencjalna musi być tak dobrana, by jej suma z energią kinetyczną pozostawała stała. Musi więc być równa (z dokładnością do stałej):

$$V(r) = -\frac{C}{r}.$$

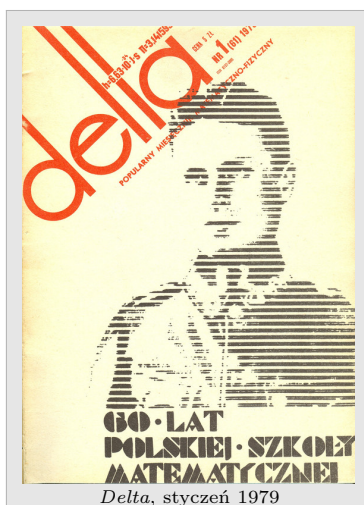
Siła musi popychać cząstkę/planetę w kierunku najszybszego spadku potencjału, bowiem siła przyspiesza początkowo nieruchome obiekty tak, że zyskują energię kinetyczną – energię potencjalną muszą więc wtedy tracić. Najszybszy spadek potencjału jest w kierunku $-\hat{\mathbf{r}}$. Tempo spadku potencjału w tym kierunku, a więc wielkość siły grawitacyjnej, dostajemy z różniczkowania $V(r)$: $\frac{dV}{dr} = \frac{C}{r^2}$. Stąd siła grawitacyjna działająca na planetę musi być równa:

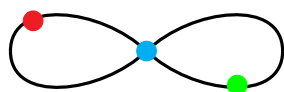
$$\mathbf{F} = -\frac{C}{r^2} \hat{\mathbf{r}}.$$

Otrzymaliśmy zatem słynny wynik Newtona („odwrotność kwadratu”!), przebywając całą drogę od cięcia stożka do prawa powszechnego ciążenia. Jest to stan wiedzy z roku 1666.

Wiele ciał

Elegancja i prostota teorii opisującej ruch planety wokół Słońca mogłyby zachęcić nas do tego, byśmy zaczęli szukać podobnych rezultatów opisujących ruch trzech lub więcej ciał niebieskich. Niestety tak uniwersalnych wyników po prostu nie ma. Zagadnienie trzech ciał zazwyczaj prowadzi do chaotycznych trajektorii. W niektórych wyjątkowych przypadkach daje się jednak znaleźć





Rys. 9

zaskakujące szczególne rozwiązania. Okazuje się na przykład, że przy odpowiednim wyborze warunków początkowych trzy ciała, oddziałujące na siebie tylko grawitacyjnie, mogą podążać za sobą, kreśląc trajektorię „ósemki” (rys. 9). Numerycznie zaobserwował to Chris Moore (1993 r.), a w 2000 roku udowodnili to formalnie Richard Montgomery i Alain Chenciner. Ta w gruncie rzeczy ciekawostka zainspirowała niektórych do poszukiwania rozmaitych „choreografii”, w których cząstki oddziałujące grawitacyjnie pływają po specyficznych trajektoriach.

Wiemy też dzisiaj, że w przypadku większej liczby obiektów mogą dziać się jeszcze bardziej dziwne rzeczy. W 1988 roku Jeff Xia pokazał, że można skonstruować taki układ pięciu ciał, w którym cztery ciała krążą po eliptycznych orbitach, a piąte oscyluje między nimi tak, że w skończonym czasie osiąga nieskończoną szybkość, przy czym wszystko to odbywa się bez żadnych zderzeń. Podobny wynik udało się niedawno osiągnąć już dla czterech ciał.

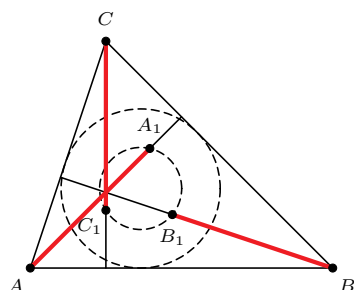
Te publikowane niedawno wyniki, o których tu wspominamy, są ciekawe i pobudzają wyobraźnię, ale są niepomierne trudniejsze do wprowadzenia i jednocześnie nie są już tak rewolucyjne, jak genialny wynik Newtona. Cóż, Amerykę odkrywa się tylko raz...



Zadania

Przygotował Dominik BUREK

Redaktor zadań matematycznych od 2021 roku.



M 1765. Na wysokościach AA_0 , BB_0 , CC_0 trójkąta ostrego nierównobocznego ABC zaznaczono, odpowiednio, punkty A_1 , B_1 , C_1 w taki sposób, że $AA_1 = BB_1 = CC_1 = R$, gdzie R jest promieniem okręgu opisanego na trójkącie ABC . Udowodnić, że środek okręgu opisanego na trójkącie $A_1B_1C_1$ pokrywa się ze środkiem okręgu wpisanego w trójkąt ABC .

Rozwiązanie na str. 10

M 1766. Łazik księżycowy porusza się po powierzchni planety, która ma kształt kuli o długości równikowej 400 km. Łazik obejmuje swoim radarem obszar w promieniu 50 km (licząc po powierzchni planety) od punktu, w którym się znajduje. Czy łazik księżycowy może w pełni zbadać planetę, pokonując nie więcej niż 600 km?

Rozwiązanie na str. 11

M 1767. Parę (m, n) różnych liczb naturalnych m i n nazywamy *dobrą*, jeśli mn i $(m+1)(n+1)$ są kwadratami liczb całkowitych. Udowodnić, że dla dowolnej liczby całkowitej dodatniej m istnieje liczba całkowita $n > m$ taka, że para (m, n) jest dobra.

Rozwiązanie na str. 21

Przygotował Andrzej MAJHOFER

Redaktor zadań fizycznych od 1984 roku.

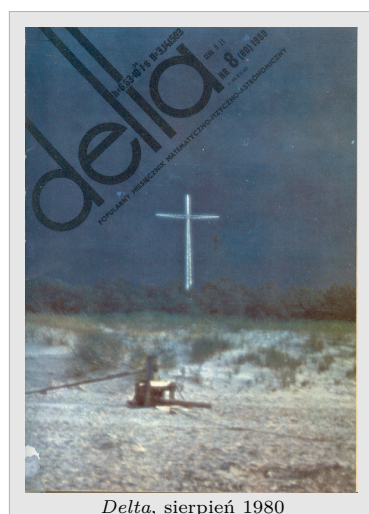
F 1085. Ile wynosi ciepło właściwe w stałej objętości na jednostkę masy azotu (N_2) ogrzanego do temperatury $T = 5000 \text{ eV}/k$ (k oznacza stałą Boltzmanna)? Liczba masowa azotu $A = 14,1 \text{ eV}/k \approx 1,16 \cdot 10^4 \text{ K}$.

Rozwiązanie na str. 20

F 1086. Oszacuj, jaka energia jest potrzebna do całkowitego zjonizowania atomów azotu ($Z = 7$, $A = 14$). Energia jonizacji atomu wodoru wynosi $E_1 = 13,6 \text{ eV}$.

Wskazówka: model Bohra z dobrym przybliżeniem odtwarza energie stanów elektronowych lekkich atomów.

Rozwiązanie na str. 42



Delta, sierpień 1980

* Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

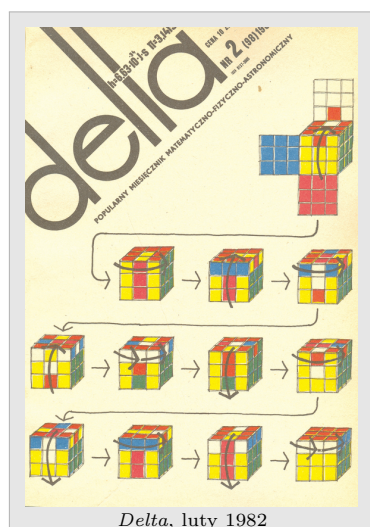
Redaktor *Delta* w latach 2014–2022.



Delta, marzec 1981

Starsi Czytelnicy mogli przeczytać o tym 48 lat temu w artykule Jana Gaja *Oko i barwa* $\Delta \frac{1}{16}$.

Jako ciekawostkę dodajmy, że jedną z przyczyn daltonizmu może być brak jednego z typów czopków. Z drugiej strony niektóre kobiety mają cztery rodzaje czopków i w związku z tym dostrzegają więcej barw. Taką cechę, czy też umiejętność, nazywa się tetrachromatyzmem.



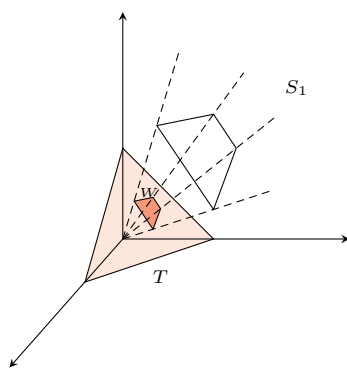
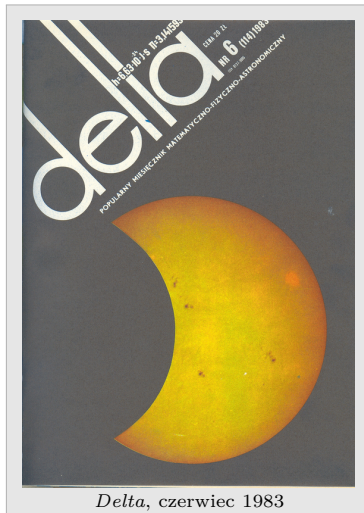
Delta, luty 1982

Czy wszystko da się wygooglować? Albo przynajmniej – czy będzie się dało za parę lat, gdy sztuczna inteligencja będzie zdecydowanie bardziej rozwinięta niż ostatni czat GPT-4? Niestety, okazuje się, że wobec niektórych pytań nawet wyszukiwarka Google okazuje się całkowicie bezradna. Doświadczyłem tego ostatnio, chcąc się dowiedzieć, jakiego koloru nie pokażą nam ekrany komputerów. Bo okazuje się, że ekrany nie wyświetlają wszystkich barw, które możemy zobaczyć ludzkim okiem, a tylko niektóre. I to nie dlatego, że są źle robione, ale z pewnych głębszych, bardzo interesujących przyczyn. Ale zacznijmy opowieść od początku.

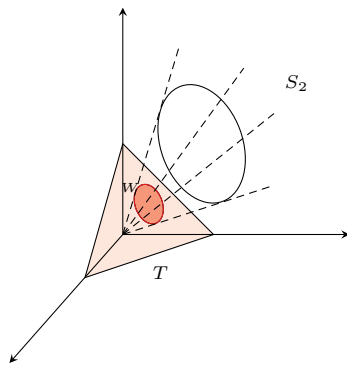
Jak zapewne większość Czytelników wie, barwy na ekranie komputerów, telewizorów, komórek i innych urządzeń wyświetlane są w systemie RGB (od ang. *Red, Green, Blue*, czyli kolorów czerwonego, zielonego i niebieskiego). Każdy piksel na ekranie ma dla każdej z trzech barw podaną intensywność w skali od 0 do 255. Przykładowo trójka intensywności RGB (255, 0, 0) to kolor czerwony, (0, 0, 255) to niebieski, (255, 255, 0) to żółty, (0, 0, 0) to czarny, a (0, 0, 100) to ciemnoniebieski. Przyzwyczailiśmy się do myślenia, że każdy kolor można reprezentować za pomocą trzech składowych: czerwonej, zielonej i niebieskiej, ale warto zadać sobie pytanie, skąd się to bierze. Przecież światło widzialne składa się z fal o różnych częstotliwościach i na przykład fala o długości 580 nm, którą ludzkie oko odbiera jako kolor żółty, nie jest w żadnym sensie sumą fal o częstotliwościach 650 nm (czerwony dla ludzkiego oka) i 550 nm (zielony dla ludzkiego oka). A jednak kolor żółty reprezentujemy jako trójkę (255, 255, 0), czyli jako sumę kolorów zielonego i czerwonego. Dlaczego?

Zapewne doświadczeni Czytelnicy wiedzą lub domyślają się, że rozwiązanie zagadki tkwi w tym, jak nasze oczy odbierają kolory. Mianowicie: w ludzkim oku, na siatkówce, znajdują się światłoczułe komórki zwane czopkami. Większość ludzi ma trzy rodzaje czopków, wrażliwych na nieco różne długości fali. Czopki D (od: długofalowe) są najbardziej wrażliwe na światło czerwone, czopki Ś (średniofalowe) są najbardziej wrażliwe na światło zielone, a czopki K (krótkofalowe) są najbardziej wrażliwe na światło niebieskie. Nasz odbiór barwy światła zależy od tego, jak mocno pobudzone są czopki różnych rodzajów, i tylko w ten sposób jesteśmy w stanie odróżnić jedne kolory od drugich. A więc rozważane przez nas dla przykładu światło barwy żółtej pobudza mocno czopki typów D i Ś (czerwone i zielone), a mniej czopki typu K (niebieskie). Wyświetlany przez monitor kolor (255, 255, 0) również pobudza mocno czopki D i Ś, a słabo czopki K. Dlatego nasze oko widzi ten kolor jako żółty, mimo że nie ma on nic wspólnego z monochromatycznym światłem żółtym.

Po tym wstępie wróćmy jednak do początkowego pytania: czy monitory oparte na systemie RGB wyświetlają wszystkie kolory widoczne dla ludzkiego oka? A raczej poprawniej byłoby zapytać: czy te monitory są w stanie zasymulować dowolną możliwą reakcję czopków naszego oka na padające światło? Odpowiedź, jak już powiedzieliśmy na wstępie, brzmi: nie. Co więcej, każdy monitor działający na podobnej zasadzie, to znaczy wyświetlający trzy lub dowolną inną liczbę podstawowych kolorów, będzie miał tę samą wadę. Żeby zrozumieć przyczynę tego zjawiska, przyjrzyjmy się dokładniej reakcjom naszego oka. Każde światło jakoś pobudza trzy typy naszych czopków – powiedzmy, że światło o widmie s powoduje sygnał od czopków czerwonych (D) w nasileniu R_s , od czopków zielonych (Ś) w nasileniu G_s , a od czopków niebieskich (K) w nasileniu B_s . A więc całość reakcji na światło s to trójwymiarowy wektor (R_s, G_s, B_s) . Warto zadać pytanie: jak wygląda zbiór wszystkich możliwych do uzyskania wektorów – reakcji na światło. Oznaczmy go przez S_r (stożek reakcji), a zatem $S_r = \{(R_s, G_s, B_s) : s \text{ jest falą elektromagnetyczną}\} \subseteq \mathbb{R}_{\geq 0}^3$. Przypomnijmy, że czopki D, Ś i K reagują nie tylko na światło czerwone, zielone i niebieskie, ale na wiele różnych częstotliwości światła, tylko w różnym natężeniu. Czyli na przykład światło niebieskie pobudza wszystkie czopki, tyle że najbardziej niebieskie, a czerwone i zielone mniej. Dlatego zbiór S_r okazuje się mieć ciekawe geometryczne własności i nie jest po prostu zbiorem wszystkich



Rys. 1



Rys. 2

Stożek S_r to zbiór wszystkich możliwych reakcji na światło, więc żaden fizyczny obiekt nie może mieć koloru, który wywołałby reakcję spoza S_r . W szczególności nie jest możliwe stworzenie monitora, który generowałby zbiór większy niż stożek S_r .

Więcej o stożkach barw można poczytać np. tu <https://photographylife.com/srgb-vs-adobe-rgb-vs-prophoto-rgb>.

możliwych wektorów o nieujemnych wartościach $\mathbb{R}_{\geq 0}^3$. Zauważmy najpierw, że zbiór S_r spełnia dwie własności: (1) jeśli $v \in S_r$, $\lambda \in \mathbb{R}_+$, to również $\lambda v \in S_r$, oraz (2) jeśli $v_1, v_2 \in S_r$, to również $v_1 + v_2 \in S_r$. Istotnie, jeśli wektor v jest reakcją na jakieś światło s , to wektor λv będzie reakcją na światło λs , czyli światło o tym samym spektrum, tylko λ razy mocniejsze. Natomiast jeśli v_1 jest reakcją na światło s_1 , a v_2 jest reakcją na światło s_2 , to $v_1 + v_2$ będzie reakcją na światło s_1 świecące razem ze światłem s_2 , czyli na światło $s_1 + s_2$. To oznacza, że zbiór S_r istotnie spełnia warunki (1) i (2) sformułowane powyżej, zbiory takie nazywamy stożkami. Żeby sobie wyobrazić, jak taki stożek może wyglądać, przyjrzyjmy się zbiorowi $T = \{(x, y, z) \in \mathbb{R}_{\geq 0}^3 : x + y + z = 1\}$, czyli trójkątowi rozpiętemu pomiędzy punktami $(1, 0, 0)$, $(0, 1, 0)$ oraz $(0, 0, 1)$. Każdy stożek S ma tę własność, że jego przecięcie z trójkątem T jest zbiorem wypukłym, bo jeśli $u, v \in S$, to również $\alpha u + (1 - \alpha)v \in S$ dla dowolnego $\alpha \in [0, 1]$. Co więcej, każdy zbiór wypukły W zawarty w trójkącie T jednoznacznie wyznacza stożek: ten stożek to po prostu wektory ze zbioru W pomnożone przez jakąś liczbę $\lambda \in \mathbb{R}_+$. A zatem żeby wyobrazić sobie stożki, wystarczy patrzeć na ich przecięcia z trójkątem T . Na rysunkach 1 i 2 przedstawione są dwa przykłady stożków: S_1 i S_2 .

Stożki S_1 i S_2 różnią się od siebie fundamentalnie. Przecięcie stożka S_1 i trójkąta T to czworokąt: jego wierzchołkami są wektory v_1, v_2, v_3, v_4 . Natomiast przecięcie stożka S_2 i trójkąta T to koło. Zauważmy, że każdy punkt stożka S_1 może być wygenerowany z wektorów v_1, v_2, v_3, v_4 poprzez aplikację reguł (1) oraz (2). Istotnie, możemy w ten sposób łatwo przedstawić każdy punkt we wnętrzu czworokąta $S_1 \cap T$ jako średnią ważoną wektorów v_1, \dots, v_4 , a potem ten wektor dowolnie wydłużyć lub skrócić. W takiej sytuacji mówimy, że stożek S_1 jest *skończenie generowany*, bo da się go wygenerować za pomocą reguł (1) i (2) ze skończonej liczby wektorów. Można łatwo zauważyć, że stożek S jest skończenie generowany wtedy i tylko wtedy, gdy jego przecięcie z trójkątem T jest po prostu wielokątem. Wtedy jest on generowany przez wierzchołki tego wielokąta. Z drugiej strony, jeśli skończona liczba wektorów v_1, \dots, v_n generuje stożek S , to powstaje on jako otoczka wypukła punktów v_1, \dots, v_n , która następnie zostaje przeskalowana (możemy skracać lub wydłużać wektory, używając reguły (1)). Czyli stożek S_2 nie jest skończenie generowany, bo jego przecięcie z trójkątem T to koło, a nie wielokąt. Intuicyjnie jest to jasne: żeby go wygenerować musielibyśmy użyć wszystkich wektorów na brzegu koła $S_2 \cap T$, a tych jest nieskończenie wiele.

Jaki ma to związek z wyświetlaniem barw przez monitor? Otóż okazuje się, że zbiór S_r wszystkich możliwych reakcji oka ludzkiego na światło jest nieco zaokrąglony, a zatem nie jest skończenie generowany. Jaki z tego wniosek? Ano taki, że nie możemy go wygenerować ze skończenie wielu możliwych reakcji przez dodawanie i mnożenie. A przecież to dokładnie próbuje zrobić monitor. Światło $r = (255, 0, 0)$ to kolor czerwony, wywołuje on jakiś wektor reakcji $v_r = (R_r, G_r, B_r)$ w naszych oczach. Światło zielone wywołuje jakiś inny wektor reakcji v_g , a światło niebieskie jeszcze inny wektor reakcji v_b . W związku z tym wszystkie możliwe reakcje, które monitor może wywołać w naszych oczach, to stożek generowany przez trzy elementy: wektory v_r, v_g oraz v_b (tak naprawdę jeszcze dodatkowo ograniczony, bo monitor ma tylko 256 różnych nasileni światła i nie może też świecić dowolnie mocno, ale to z naszej perspektywy nieistotne). W szczególności niektóre możliwe reakcje z S_r nie dadzą się wywołać przez monitor. I to nie tylko taki, który ma trzy rodzaje pikseli, ale również taki, który miałby 10 albo 1000 rodzajów pikseli. Wciąż znajdują się jakieś reakcje na brzegu stożka reakcji S_r , których nie da się wywołać przez sumowanie tych 1000 rodzajów pikseli!

Specjaliści od grafiki starają się radzić sobie z tym fenomenem, jak mogą. Tworzą w szczególności różne skończenie generowane możliwe stożki barw: sRGB, AdobeRGB, ProPhotoRGB i inne, które lepiej lub gorzej przybliżają S_r . Ale nigdy nie zrobią tego idealnie, bo jest to niemożliwe!

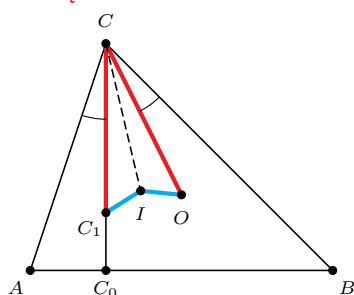
Na koniec zapytajmy: to jakich kolorów nie przedstawia dobrze monitor? Generalnie stożkowi sRGB brakuje wielu kolorów na pograniczu niebieskiego i zielonego, przykładowo kolorów w okolicy cyjanowego. Tę sprawę polecam jednak badać na żywo, bo nie wszystko da się dostrzec w świecie wirtualnym.

* Uniwersytet Jana Kochanowskiego w Kielcach i Narodowe Centrum Badań Jądrowych w Warszawie

Redaktor *Delty* w latach 1991–1994.



Rozwiązanie zadania M 1765.



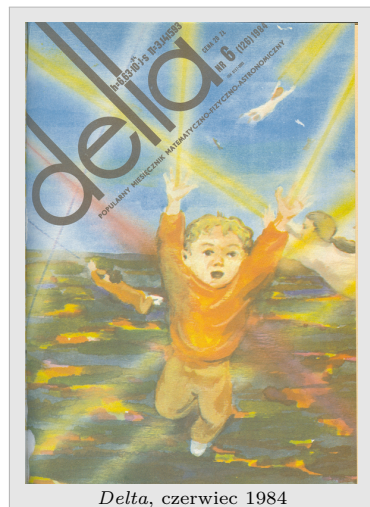
Jeśli O jest środkiem okręgu opisanego na trójkącie ABC , to

$$\sphericalangle BCO = 90^\circ - \sphericalangle BAC = \sphericalangle C_0CA.$$

Zatem punkty O i C_1 są symetryczne względem dwusiecznej kąta C , w związku z czym $IC_1 = IO$, gdzie I jest środkiem okręgu wpisanego w trójkąt ABC . Podobnie uzasadniamy, że

$$IA_1 = IO = IB_1,$$

czyli I jest środkiem okręgu opisanego na trójkącie $A_1B_1C_1$.



Delta, czerwiec 1984

W 1982 roku, wkrótce po ukończeniu studiów, zostałem wysłany do Zjednoczonego Instytutu Badań Jądrowych w Dubnej pod Moskwą, aby zajmować się zderzeniami jąder atomowych przy najwyższej wówczas dostępnej energii. Celem tego rozpoczynającego się właśnie programu badawczego było wytworzenie w laboratorium i poznanie własności egzotycznego stanu materii – plazmy kwarkowo-gluonowej, która występowała we wczesnym gorącym Wszechświecie, krótko po Wielkim Wybuchu.

Prawie 20 lat później cel został osiągnięty, a przynajmniej mocno się do niego zbliżyliśmy. W eksperymentach prowadzonych w Europejskim Centrum Badań Jądrowych w Genewie i Narodowym Laboratorium Brookhaven w stanie Nowy Jork zaobserwowano – prawda, że w dosyć pośredni sposób – krople plazmy kwarkowo-gluonowej, powstającej w zderzeniach najcięższych jąder atomowych rozpędzonych do energii wielokrotnie większej niż ta, która była dostępna w Dubnej.

Właśnie tam, w Rosji, na początku swej naukowej drogi przeczytałem niewielką popularnonaukową książeczkę „Szturm termojądrowej twierdzy” Gienadija Woronowa, opowiadającą o próbach – podejmowanych już wtedy od trzech dziesięcioleci – doprowadzenia do termojądrowej syntezy lekkich jąder atomowych, czyli odtworzenia w ziemskich warunkach procesu, dzięki któremu Słońce dostarcza nam życiodajne ciepło.

Gdy jądra atomowe lekkich pierwiastków, takich jak wodór, łączą się ze sobą, wydzielana jest energia. Aby do takiej syntezy doprowadzić, należy odpowiednio mocno podgrzać gaz tworzony przez atomy, powiedzmy, wodoru, tak aby elektrony oderwały się od jąder atomowych. Wówczas dochodzi do zderzeń jąder i staje się możliwa upragniona synteza. Materia w postaci jąder atomowych obdarzonych, jak pamiętamy, ładunkami dodatnimi i ujemnie naładowanych elektronów uwolnionych z atomów nazywana jest plazmą, dokładniej, plazmą elektromagnetyczną.

Ze wspomnianej książeczki dowiedziałem się, że plazma wykazuje zdumiewające własności, zupełnie odmienne od gazów, np. powietrza. Bardzo trudno ją kontrolować, uwięzić w pułapce. Plazma nie zachowuje się jak kulka umieszczona we wgłębieniu, która, jeśli ją potrącić, oscyluje wokół minimum, by po pewnym czasie wrócić do położenia na dnie wgłębienia. Plazma przypomina raczej kulę umieszczoną na szczycie wzniesienia. Wystarczy ją nieco wychylić z położenia chwiejnej równowagi, by potoczyła się w dół, szybko oddalając się od pierwotnego stanu. Fachowo mówimy, że plazma jest zwykle niestabilna – niewielkie zaburzenie jej stanu, fluktuacja nie zanikają z czasem, lecz gwałtownie narastają.

To jest właśnie główną przyczyną trudności w zbudowaniu termojądrowego reaktora. Historia kolejnych prób kontrolowania plazmy – licząca sobie teraz już nie trzy, a aż siedem dziesięcioleci – to historia odkrywania coraz to nowych rodzajów niestabilności – mechanizmów umożliwiających szybkie narastanie niewielkich zaburzeń.

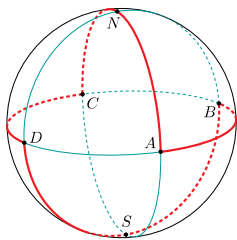
Lektura owej niewielkiej książki miała istotny wpływ na moją naukową działalność. Skoro plazma elektromagnetyczna wykazuje bogactwo niestabilności, myślałem, to podobnie jest pewnie z plazmą kwarkowo-gluonową. Wszak tworzą ją kwarki i gluony – najmniejsze, najprostsze z poznanych dotychczas składników materii – oddziałujące wzajemnie dzięki ładunkom podobnym do ładunków elektrycznych zwanych kolorami. Te, tak jak kolory podstawowe, są trzech rodzajów i tak, jak kolory podstawowe łączą się w kolor biały, tak kwarki o trzech kolorach wzajemnie znoszą swoje ładunki i całość jest neutralna. Teoria opisująca siły występujące między kwarkami i gluonami nazywa się chromodynamiką i wykazuje istotne podobieństwa z elektrodynamiką, rządzącą zachowaniem ładunków elektrycznych.



Rozwiązanie zadania M 1766.

Odpowiedź: Tak.

Zaznaczmy bieguny N (północny) i S (południowy) na planecie i niech punkty A, B, C i D podzielią odpowiedni równik na cztery równe łuki, jak na rysunku.



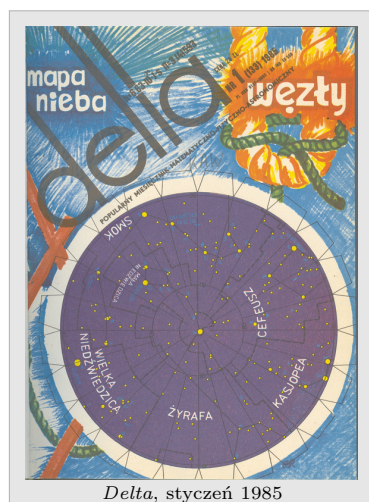
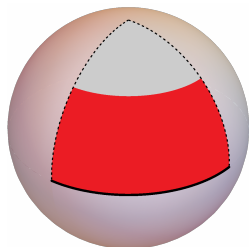
Rozważmy zamkniętą ścieżkę

$$A \rightarrow B \rightarrow S \rightarrow D \rightarrow C \rightarrow N \rightarrow A$$

wzdłuż powierzchni, składającą się z łuków wielkich kół planety. Ta ścieżka składa się z 6 identycznych łuków równych $\frac{1}{4}$ równika, więc długość ścieżki jest równa 600 km.

Pokażemy teraz, że dla każdego punktu planety istnieje na tej ścieżce punkt oddalony od niego o nie więcej niż 50 km. Podzielmy powierzchnię planety na 8 identycznych trójkątów sferycznych z wierzchołkami w zaznaczonych punktach. Łazik odwiedził wszystkie wierzchołki i wszystkie punkty co najmniej jednego boku każdego trójkąta.

Rozważmy jeden taki trójkąt, jak na rysunku niżej. Jego bok ma długość 100 km. Czerwona część tego trójkąta została zbadana przez łazik księżycowy, ponieważ podróżował on wzdłuż wyróżnionego boku, a szara część została zbadana, ponieważ odwiedził on przeciwległy wierzchołek. To samo rozumowanie dla dowolnego innego trójkąta podziału dowodzi, że podana ścieżka spełnia warunki zadania.



Obserwacja, że plazma kwarkowo-gluonowa może być niestabilna podobnie jak plazma elektromagnetyczna, nie była szczególnie oryginalna. Nie ja jeden o tym pomyślałem. Pytanie było raczej inne: czy niestabilności mogą odgrywać istotną rolę w zachowaniu plazmy kwarkowo-gluonowej wytwarzanej w zderzeniach rozpędzonych jąder atomowych? A należy pamiętać, że plazma kwarkowo-gluonowa występuje w zupełnie innych warunkach niż elektromagnetyczna.

Właśnie wtedy, gdy zajmowało mnie pytanie o stabilność plazmy kwarkowo-gluonowej, zostałem redaktorem *Delty*, co wspominam jako świetną, trwającą kilka lat przygodę. Brałem aktywny udział w dyskusjach o kolejnych numerach pisma, redagowałem teksty wielu autorów, sam dużo pisałem. W dwóch artykułach, „Drgania plazmowe” i „Niestabilności plazmowe”, które ukazały się, odpowiednio, w Δ_{93}^2 i Δ_{93}^9 , starałem się wyjaśnić w możliwie najprostszy sposób to, co mnie w owym czasie faktycznie pochłaniało.

W tym samym 1993 roku opublikowałem, jeśli nie najlepszy, to swój najczęściej cytowany artykuł naukowy, w którym dowodziłem, że pewna niestabilność, zwana filamentacyjną, może występować w plazmie kwarkowo-gluonowej powstałej w zderzeniu jąder. Ku memu rozczarowaniu rzecz nie wywołała żadnego zainteresowania, więc na jednej z konferencji złapałem znanego fizyka Larry'ego McLerrana i przymusiłem niejako do wysłuchania opowieści o tym, co udało mi się zrobić.

– Dlaczego ta niestabilność miałyby występować? – Larry zapytał, gdy skończyłem.

– Tak wychodzi z rachunków – mówię.

– To żadne tłumaczenie. Jeśli ten wynik ma sens, to powinienesz umieć go tak wyjaśnić, żeby zrozumiała moja żona.

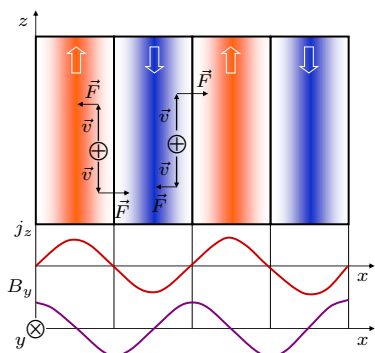
Dodam, że Alice McLerran nie była fizykiem, lecz antropologiem i autorką książek dla dzieci.

Krytyka McLerrana była dotkliwie trafna. Praca w *Delcie* dostarczała licznych przykładów, że skomplikowane problemy dają się prosto wyjaśnić, jeśli je dobrze zrozumieć. Wspomniane teksty dokumentują moje własne wysiłki uchwycenia mechanizmów zjawisk plazmowych w najprostszymi możliwych terminach. Jednak z przewidywaną niestabilnością filamentacyjną nie było łatwo. Po jakichś dwóch latach rozmyślań i studiowania literatury udało mi się w końcu rzecz zrozumieć, i opublikowałem w czasopiśmie naukowym elementarne objaśnienie. Z *Delty* już się wtedy rozstałem i nie było okazji, aby przedstawić je na łamach tego pisma. Rocznicowy numer stwarza świetną okazję, by wrócić do problemu.

Wyobraźmy sobie tedy gaz cząstek naładowanych dodatnio i ujemnie. Niech to będą elektrony o ładunku ujemnym i ich antycząstki – pozytony o ładunku dodatnim. Dodam, że taka właśnie plazma występuje w różnych układach astrofizycznych, więc przykład nie jest czysto akademicki. Dalej przyjmijmy, że średnie liczby elektronów i pozytonów w każdym punkcie układu są sobie równe, więc plazma jest lokalnie elektrycznie neutralna. Nie występują też w niej prądy elektryczne, gdyż wraz z elektronami o danej prędkości mamy też w układzie pozytony ze zbliżonymi prędkościami.

Kluczową okolicznością umożliwiającą wystąpienie niestabilności jest anizotropowy charakter rozkładu prędkości elektronów i pozytonów. Anizotropowość oznacza, że prawdopodobieństwo znalezienia elektronu lub pozytonu z daną prędkością zależy od jej kierunku. Załóżmy zatem, że średnia prędkość cząstek naszego układu w kierunku osi z jest większa niż w kierunkach x i y . Wyobraźmy sobie teraz, że na skutek fluktuacji kilka elektronów lub pozytonów poruszających się w kierunku osi z znalazło się obok siebie. Oznacza to, że w układzie pojawia się prąd elektryczny płynący w kierunku z .

Zwróćmy uwagę, że wystąpienie takiego fluktuacyjnego prądu jest bardziej prawdopodobne w kierunku z niż w kierunkach x i y . Dzieje się tak dlatego,



że wielkość prądu jest proporcjonalna do unieszonego ładunku i jego prędkości. Średnia zaś prędkość wzdłuż osi z jest z założenia większa niż wzdłuż osi x lub y .

Przyjęliśmy, że prądy w układzie, średnio rzecz biorąc, nie występują, co oznacza, że nadmiarowi elektronów w jednym miejscu będzie towarzyszył niedomiar w drugim. A zatem jeśli w danym miejscu pojawił się fluktuacyjny prąd w kierunku osi z , to gdzieś obok wystąpi prąd płynący w przeciwnym kierunku.

Tę pierwszą część opowieści możemy podsumować następująco. W neutralnym układzie elektronów i pozytonów, w którym cząstki mają prędkości w kierunku z średnio większe niż w kierunkach x i y , występują znaczące fluktuacje prądu w kierunku z . Innymi słowy, w układzie spontanicznie pojawiają się włókna z prądami płynącymi w przeciwnie strony kierunku z , pokazane na rysunku.

W drugiej, „dynamicznej”, części swojej opowieści postaram się wykazać, że fluktuacyjne prądy nie zanikają z upływem czasu, lecz przeciwnie – narastają.

Prądy elektryczne, jak wiemy, wytwarzają pole magnetyczne, a korzystając z prawa Ampère’a, nietrudno wykazać, że włókna prądowe pokazane na rysunku będą generowały pole magnetyczne w kierunku osi y o amplitudzie również tam uwidocznionej.

Jak wiemy, na cząstkę o ładunku q poruszającą się z prędkością \vec{v} w polu magnetycznym \vec{B} działa siła Lorentza $\vec{F} = q\vec{v} \times \vec{B}$. Siła wywierana na dodatnio naładowany pozyton lecący równoległe do osi z pokazana jest na rysunku. Widzimy, że siła ta kieruje pozyton poruszający się zgodnie z kierunkiem prądu w danym włóknie ku centrum włókna. Natomiast pozyton poruszający się przeciwnie do kierunku prądu włókna stara się przerzucić do włókna sąsiedniego. Następuje więc spontaniczne sortowanie cząstek tak, że prąd w każdym włóknie rośnie i wytwarza coraz silniejsze pole magnetyczne, które z kolei coraz silniej wpływa na cząstki tworzące prąd. W ten sposób rozwija się niestabilność filamentacyjna.

W 2016 roku Larry McLerran otrzymał doktorat honoris causa Uniwersytetu Jagiellońskiego. Odbyła się piękna uroczystość w Collegium Maius, a potem gratulacje i rozmowy przy białym winie z uniwersyteckich winnic. Przypomniałem i podziękowałem honorowemu doktorowi za lekcję, jakiej mi przed laty udzielił. Dzięki przedstawionemu powyżej rozumowaniu, elementarnemu, choć wymagającemu uwagi przy jego śledzeniu, możliwości występowania filamentacyjnej niestabilności w plazmie kwarkowo-gluonowej została zaakceptowana i stała się elementem – prawdą, że drobnym – fizyki zderzeń jądrowych przy najwyższych dostępnych energiach.



Renesans słów Lyndona

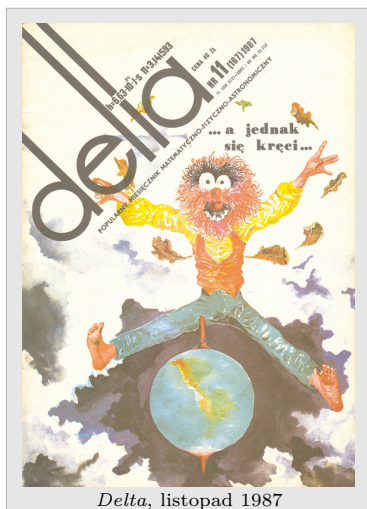
Jakub RADOSZEWSKI*

* Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Redaktor *Delta* w latach 2009–2014.

Mówimy, że słowo u jest mniejsze leksykograficznie niż słowo v , jeśli albo słowo u jest właściwym prefiksem (czyli początkowym fragmentem) słowa v , albo na pierwszej pozycji, na której słowa u i v różnią się, w słowie u występuje litera mniejsza niż w słowie v .

W artykule pt. *Słowa pierwsze* opublikowanym w Δ_{10}^{12} opisałem różne własności słów Lyndona. Żeby przypomnieć, czym są słowa Lyndona, trzeba podać kilka definicji dotyczących słów (skończonych). *Rotacja* słowa, zwana też obrotem cyklicznym, polega na przerzuceniu dowolnej (potencjalnie zerowej) liczby liter z początku słowa na koniec. Na słowach określa się *porządek leksykograficzny*, czyli słownikowy. I teraz *słowem Lyndona* nazywamy słowo, które jest ściśle najmniejsze w porządku leksykograficznym wśród swoich rotacji. Przykładowo słowo **aaab** jest słowem Lyndona, ponieważ wszystkie jego pozostałe rotacje, **aaba**, **abaa** i **baaa**, są od niego większe leksykograficznie. Innymi przykładami słów Lyndona są **aabab** i **a**. Natomiast słowo **abaaab** nie jest słowem Lyndona, gdyż jego rotacja **aaabab** jest mniejsza leksykograficznie od niego. Również słowo **aabaab** nie jest słowem Lyndona, jako że jego rotacja o 3 litery jest



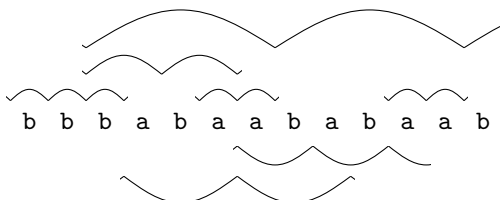
Delta, listopad 1987

mu równa. *Równoważna definicja* słów Lyndona jest taka, że słowa Lyndona są mniejsze od wszystkich swoich właściwych sufiksów (czyli końcowych fragmentów).

Słowa Lyndona nazywa się także słowami *pierwszymi*, przez analogię do liczb pierwszych; tak jak każda liczba naturalna przedstawia się jednoznacznie (z dokładnością do kolejności) jako iloczyn liczb pierwszych, tak każde słowo przedstawia się jednoznacznie jako sklejenie nierosnącego (leksykograficznie) ciągu słów Lyndona. Przykładowo słowo *abaababaabaaba* jest sklejeniem słów Lyndona *ab aabab aab aab a*. Dowód tej własności znalazł się we wspomnianym artykule wraz z innymi podstawowymi faktami dotyczącymi słów Lyndona oraz ich zadziwiającym związkiem z ciągami de Bruijna.

Po lekturze numeru Δ_{10}^{12} Czytelnik mógł przekonać się, że słowa Lyndona są perełką kombinatoryki słów. Natomiast pewnie mało kto mógł przewidzieć, że w ciągu kolejnych lat staną się one jednym z kluczowych narzędzi zarówno w tej dziedzinie, jak i w algorytmach tekstowych, a algorytmy bazujące na nich będą miały istotne zastosowania praktyczne. W tym artykule opowiem o dwóch takich zastosowaniach słów Lyndona.

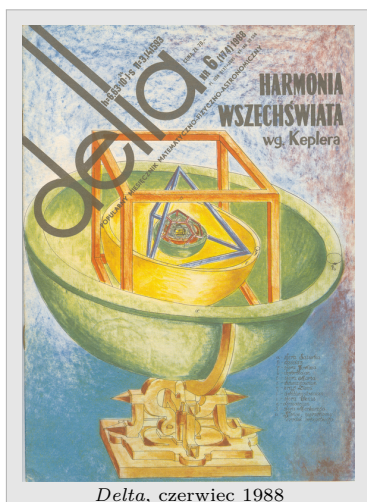
Pierwsze z nich dotyczy pojęcia *maksymalnych okresowości* (ang. *runs*) w słowie. Do ich zdefiniowania znów potrzebnych jest kilka pojęć. *Potęgą* słowa *w* o wykładniku *m*, oznaczaną w^m , nazywamy sklejenie kolejno *m* kopii słowa *w*. Powiemy, że słowo *s* ma *okres p*, jeśli *s* jest prefiksem potęgi w^m dla pewnego słowa *w* o długości *p*. Jako *p* wybieramy zazwyczaj najmniejszy okres słowa. Jeśli okres *p* jest co najmniej dwa razy mniejszy niż długość słowa *s*, powiemy, że słowo *s* jest *okresowe*. Przykładowo, słowo *abcabcabcab* jest okresowe z okresem 3, ponieważ jest prefiksem słowa $(abc)^4$. *Okresowością* w słowie *s* nazwiemy dowolny spójny fragment słowa *s*, który jest okresowy. Wreszcie okresowość nazwiemy *maksymalną*, jeśli nie da się jej rozszerzyć o jedną pozycję (ani w lewo, ani w prawo) z zachowaniem jej (najkrótszego) okresu. Wszystkie maksymalne okresowości w przykładowym słowie można znaleźć na poniższym rysunku:



Słowo *bbbabaababaab* ma 7 maksymalnych okresowości. Każdy łuk oznacza jedno powtórzenie okresu w okresowości; niektóre powtórzenia są ucięte. Przykładowo na trzeciej pozycji słowa zaczynają się dwie maksymalne okresowości, *baba* o okresie 2 i *babaababaab* o okresie 5.

Po co rozważać maksymalne okresowości w słowie? Przede wszystkim reprezentują one każdy możliwy fragment okresowy słowa. Faktycznie, każdy fragment okresowy albo sam już jest maksymalną okresowością, albo można go maksymalnie rozszerzyć w obie strony aż do uzyskania maksymalnej okresowości. W szczególności wszystkie kwadraty (potęgi o wykładniku 2) w słowie, o których pisałem w numerze Δ_{11}^3 , są reprezentowane przez maksymalne okresowości. Kluczową, a jednocześnie zupełnie nieoczywistą własnością maksymalnych okresowości jest to, że w słowie o długości *n* jest ich tylko $O(n)$. To z kolei przekłada się na liczne zastosowania maksymalnych okresowości w algorytmach tekstowych, w co Czytelnik musi mi już uwierzyć... na słowo.

Dowód faktu, że liczba maksymalnych okresowości w słowie jest co najwyżej liniowa względem jego długości, został podany przez Romana Kolpakova i Gregory'ego Kucherova jeszcze w 1999 roku. Nie potrafili oni jednak podać żadnego „sensownego” współczynnika w tej zależności liniowej, choć na podstawie eksperymentów komputerowych wyszło im, że ten współczynnik powinien być równy po prostu 1. Przez kolejne lata różni badacze podawali dowody konkretnych oszacowań górnych z coraz to lepszymi stałymi, nierzadko posiłkując się obliczeniami komputerowymi. W szczególności pierwszą konkretną stałą w oszacowaniu, równą 5, uzyskał Wojciech Rytter w 2006 roku. W 2011 roku Maxime Crochemore wraz ze współpracownikami, wykorzystując cały klaster komputerów, uzyskał oszacowanie górne ze współczynnikiem 1,029.



Delta, czerwiec 1988



Delta, luty 1989

W przypadku alfabetu o dowolnej liczbie liter definicja pozycji specjalnej nieco się komplikuje – decyzję o wyborze przypadku (a) lub (b) podejmujemy, porównując litery $s[j+1]$ oraz $s[j+1-p]$.

Pozycje specjalne maksymalnych okresowości oznaczono gwiazdkami.

Na przykład dla „górnej” maksymalnej okresowości $s[3..13]$ na rysunku rozważane fragmenty długości p to kolejno babaa, abaab, baaba, aabab i ababa.

W formalnym uzasadnieniu tego spostrzeżenia może pomóc udowodniony w 1965 roku przez Finego i Wilfa lemat o okresowości: jeśli słowo s o długości n ma okresy p i q oraz $p+q \leq n + \text{nwd}(p, q)$, to s ma też okres $\text{nwd}(p, q)$.

Na przykład dla maksymalnej okresowości $s[3..6]$ z rysunku i $k=4$ najdłuższym słowem Lyndona zaczynającym się na pozycji $k=4$ w słowie s jest ab.

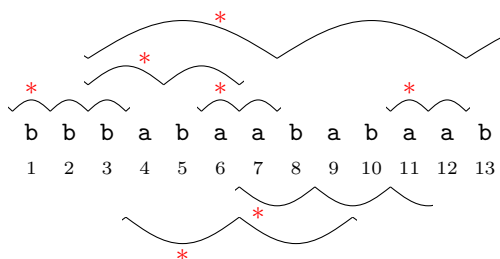
Wreszcie w 2017 roku zespół badaczy z Japonii, pod kierunkiem Hideo Bannai, udowodnił, że słowo długości n ma co najwyżej n maksymalnych okresowości. Udało im się to zrobić bez wykonywania jakichkolwiek obliczeń komputerowych, tylko właśnie za pomocą słów Lyndona.

Ostateczny dowód jest na tyle elementarny, że możemy go w tym miejscu naszkicować. Główny pomysł polega na tym, żeby każdej maksymalnej okresowości przypisać jedną pozycję w słowie, zwaną *pozycją specjalną*, tak aby różnym maksymalnym okresowości przypisać różne pozycje specjalne. W ten sposób automatycznie otrzymamy żądane ograniczenie górne. Aby nie męczyć się z pewnymi szczegółami technicznymi, w tym artykule udowodnimy jedynie słabszą wersję tezy, że każda pozycja specjalna może być przypisana co najwyżej dwóm maksymalnym okresowości; stąd wyniknie, że w słowie o długości n jest co najwyżej $2n$ maksymalnych okresowości. Przyjmiemy także, że rozpatrywane słowo s jest nad alfabetem binarnym $\{a, b\}$ (co trochę uprości argumentację).

Pozycje dowolnego słowa s będziemy numerowali od 1. Przez $s[i]$ oznaczmy literę znajdującą się na i -tej pozycji słowa s , a przez $s[i..j]$ oznaczmy fragment $s[i] \dots s[j]$. Oprócz słów Lyndona, w konstrukcji używa się słów *anty-Lyndona*, które są (ściśle) *największe* w porządku leksykograficznym wśród swoich rotacji. Niech $u = s[i..j]$ będzie maksymalną okresowością o okresie p w słowie s , a $x = s[j+1]$ będzie literą, która występuje w słowie s tuż za fragmentem u . Wtedy *pozycją specjalną* maksymalnej okresowości u będzie taka pozycja $k \in \{i, \dots, i+p-1\}$, że:

- (a) jeśli $x = a$ lub x nie istnieje (bo j jest ostatnią pozycją w słowie), to $s[k..(k+p-1)]$ jest słowem Lyndona,
- (b) jeśli $x = b$, to $s[k..(k+p-1)]$ jest słowem anty-Lyndona.

Przykład przypisania pozycji specjalnych można znaleźć na poniższym rysunku.



Potrzeba chwili, żeby tę definicję przetrawić. Zacznijmy od uzasadnienia, że każda maksymalna okresowość ma pozycję specjalną. Jeśli maksymalna okresowość $s[i..j]$ ma okres 1 (np. $s[1..3]$ na rysunku), to nie mamy wyboru, i jej pozycją specjalną musi być pozycja $k=i$. W ogólności zauważmy, że fragmenty długości p postaci $s[k..(k+p-1)]$ dla $k \in \{i, \dots, i+p-1\}$ są kolejno wszystkimi rotacjami słowa $s[i..(i+p-1)]$. Wynika to z faktu, że okres w maksymalnej okresowości musi się powtarzać co najmniej dwukrotnie. Żadne dwie takie rotacje, dla różnych $k \in \{i, \dots, i+p-1\}$, nie mogą być równe. Można zauważyć, że gdyby tak było, to słowo w o długości p , takie że maksymalna okresowość u jest prefiksem potęgi w^m , musiałoby być potęgą krótszego słowa. A zatem dokładnie jedna z tych rotacji będzie najmniejsza, a jedna największa w porządku leksykograficznym, i to one wyznaczą szukane słowo Lyndona i słowo anty-Lyndona.

Wykażemy teraz, że jeśli pozycja k jest pozycją specjalną maksymalnej okresowości $u = s[i..j]$ o okresie p wyznaczoną przez słowo Lyndona według przypadku (a), to $s[k..(k+p-1)]$ jest najdłuższym słowem Lyndona zaczynającym się na pozycji k . Załóżmy przez sprzeczność, że słowo $s[k..r]$ dla pewnego $r \geq k+p$ jest słowem Lyndona. Skorzystamy tu z równoważnej definicji słów Lyndona, według której muszą one być mniejsze leksykograficznie od wszystkich swoich właściwych sufiksów. Jeśli $r \leq j$, to słowo $s[k..r]$ ma okres p . Wtedy jednak sufiks $s[(k+p)..r]$ byłby prefiksem $s[k..r]$ (czyli w szczególności

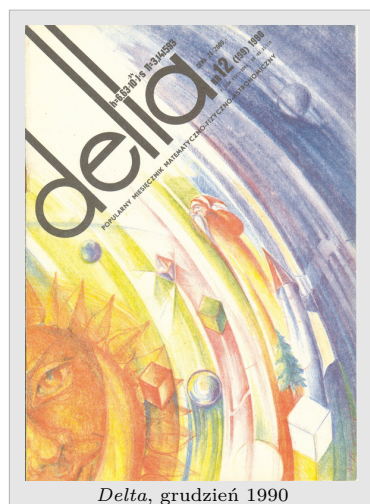
Dla rozważanego przykładu słowo $s[4..6] = aba$ nie jest słowem Lyndona, gdyż jego sufiks $s[6..6] = a$ jest od niego leksykograficznie mniejszy (bo jest jego prefiksem).

Przykładowo dla dowolnego $r > j = 6$ słowo $s[4..r] = abaa\dots$ jest większe od słowa $s[6..r] = aa\dots$

Jedna pozycja może być pozycją specjalną dwóch maksymalnych okresowości. Okazuje się, że wtedy jedna z tych maksymalnych okresowości ma okres 1. Przykładowo ma to miejsce dla pozycji 6 na rysunku powyżej. Jeśli pozycję specjalną wybierzemy ze zbioru $\{i + 1, \dots, i + p\}$ zamiast ze zbioru $\{i, \dots, i + p - 1\}$, można udowodnić, że wtedy już każda pozycja słowa będzie pozycją specjalną co najwyżej jednej maksymalnej okresowości.

Przykładowo, tablica sufiksowa słowa $s = abaababa$ to 8, 3, 6, 1, 4, 7, 2, 5:

sufiks	pozycja początkowa
a	8
aababa	3
aba	6
abaababa	1
ababa	4
ba	7
baababa	2
baba	5



Delta, grudzień 1990

byłby od niego mniejszy), co przeczy równoważnej definicji słów Lyndona. Jeśli zaś $r > j$, to sufiks $s[(k + p)..r]$ zaczyna się od $s[(k + p)..j]a$, a słowo $s[k..r]$ zaczyna się od $s[k..(j - p)]b$. Skąd ta litera **b**? Otóż gdyby tam była litera **a**, to maksymalną okresowość można by rozszerzyć o jedną pozycję w prawo, gdyż zachodziłoby $s[j + 1] = s[j + 1 - p]$. Mamy $s[k..(j - p)]b = s[(k + p)..j]b$, czyli znów sufiks jest mniejszy. Otrzymana sprzeczność dowodzi, że rzeczywiście $s[k..(k + p - 1)]$ jest najdłuższym słowem Lyndona zaczynającym się na pozycji k .

Podobnie można wykazać, że w przypadku (b) fragment $s[k..(k + p - 1)]$ jest najdłuższym słowem anty-Lyndona zaczynającym się na pozycji k . W tym celu wygodnie jest rozważyć porządek leksykograficzny, w którym **b** jest mniejsze od **a**; słowa anty-Lyndona są tak naprawdę słowami Lyndona w tym poniekąd odwróconym porządku leksykograficznym.

Oczywiście na każdej pozycji w słowie zaczyna się jedno najdłuższe słowo Lyndona i jedno najdłuższe słowo anty-Lyndona. Każde z tych słów możemy jednoznacznie rozszerzyć okresowo w lewo i w prawo, z okresem równym jego długości, otrzymując maksymalną okresowość, jeśli tylko okres powtórzy się w niej co najwyżej dwukrotnie. To ostatecznie dowodzi, że każda pozycja jest pozycją specjalną co najwyżej dwóch maksymalnych okresowości, i kończy nasz dowód.

Na koniec warto zaznaczyć, że Bannai i in. w swojej pracy podali nowy, prostszy algorytm wyznaczania maksymalnych okresowości, a kluczową rolę w tym algorytmie pełni twierdzenie o jednoznacznym rozkładzie słowa na niemające słowa Lyndona.

Opowiemy teraz pokrótce o drugim nowym zastosowaniu słów Lyndona. Niech słowo s ma długość n . Wyobraźmy sobie, że wszystkie sufiksy słowa s zostały uporządkowane rosnąco w kolejności leksykograficznej. *Tablica sufiksowa* słowa s zawiera pozycje początkowe sufiksów w tej właśnie kolejności. Jest ona zatem permutacją zbioru $\{1, \dots, n\}$.

Tablica sufiksowa, obok drzewa sufiksowego, jest jedną z podstawowych struktur danych w algorytmach tekstowych. Na przykład jeśli mamy tablicę sufiksową słowa s o długości n , to dla danego słowa-wzorca p o długości m możemy w czasie $O(m \log n)$ znaleźć wszystkie jego wystąpienia w s . W tym celu należy zauważyć, że zbiór sufiksów, na początku których występuje wzorec p , odpowiada spójnemu fragmentowi tablicy sufiksowej. Końce tego fragmentu możemy wyznaczyć za pomocą wyszukiwania binarnego.

Tablica sufiksowa została wprowadzona w 1990 roku przez Udiego Manbera i Gene'a Myersa. Korzystając z drzew sufiksowych, można ją skonstruować w czasie liniowym w przypadku słowa nad małym alfabetem. Jest też znany algorytm autorstwa Juhy Kärkkäinen i Petera Sandersa (2003 r.), który konstruuje ją w czasie liniowym bez użycia drzew sufiksowych. Tyle w teorii; natomiast praktyka pokazuje, że wszystkie te algorytmy są wolniejsze i bardziej pamięciochłonne niż algorytm o nazwie DivSufSort autorstwa Yuty Moriego działający w czasie $O(n \log n)$. Co ciekawe, Yuta Mori przedstawił swój algorytm w postaci gotowego programu, z niewielką ilością komentarzy, opublikowanego w publicznym repozytorium.

Ten rozdział między złożonością czasową a czasem działania na praktycznych danych był dość kłopotliwy dla badaczy zajmujących się algorytmami tekstowymi. Dopiero w 2021 roku Nico Bertram i in. w swojej pracy pt. *Lyndon Words Accelerate Suffix Sorting* przedstawili algorytm poparty poważną analizą teoretyczną i działający w praktyce szybciej niż DivSufSort. Jak tytuł pracy wskazuje, w ich algorytmie istotną rolę pełnią właśnie słowa Lyndona. Algorytm Bertrama i in. ma jednak wciąż ponadliniową złożoność czasową (choć korzysta z liniowej pamięci). Jego złożoność została poprawiona do liniowej, z jednoczesnym dodatkowym przyspieszeniem w praktyce, w pracy autorstwa Jannika Olbricha i in., opublikowanej bardzo niedawno, bo w 2022 roku.

37 lat później

Krzysztof CIESIELSKI*

*Instytut Matematyki, Uniwersytet Jagielloński

W latach 1991–1997 redaktor naczelny samodzielnej kolumny *Epsilon* w *Delcie*.

Od 1996 roku członek Komitetu Redakcyjnego *Delty*, w latach 2004–2019 jego wiceprzewodniczący, od roku 2019 przewodniczący.

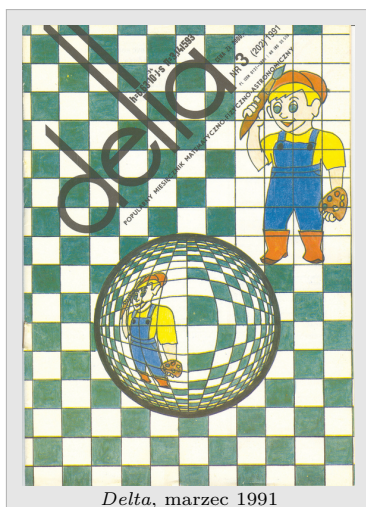
O związku między sześcianem a hiperboloidą obrotową pisze również Marek Kordos na stronie 3 – przyp. red.

Jednym z moich ulubionych numerów *Delty* jest ten z czerwca 1986 roku. Z okazji jubileuszu (był to numer 150) został praktycznie w całości wypełniony krótkimi ciekawostkami matematycznymi, fizycznymi i astronomicznymi, których, jak nietrudno zgadnąć, Redakcja umieściła 150.

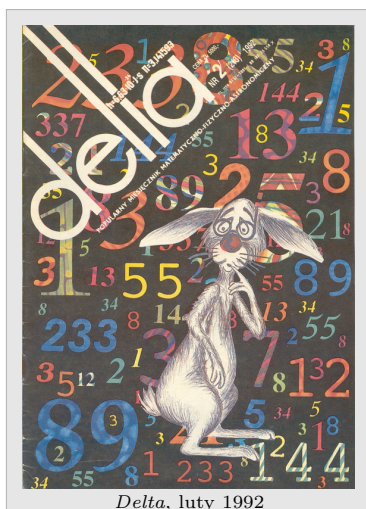
Czy coś opisanego w opublikowanych wówczas matematycznych ciekawostkach mogło po 37 latach ulec zmianie? Jedną z największych zalet matematyki jest dla mnie to, że twierdzenie raz udowodnione (poprawnie) pozostaje prawdziwym do końca świata. Na przykład w ciekawostce nr 122 podano, że jeśli obrócimy sześcian wokół jego przekątnej, to otrzymamy bryłę złożoną z dwóch stożków i hiperboloidy obrotowej. Tak było wówczas, gdy terminu „hiperboloida obrotowa” jeszcze nie wprowadzono, tak było w 1986 roku, tak jest teraz i tak będzie zawsze. Jednak w ciągu tych 37 lat w matematyce także co nieco (a nawet wiele) się zmieniło. Wtedy nie wiedzieliśmy jeszcze, że Wielkie Twierdzenie Fermata jest prawdziwe (o nim w numerze 6/1986 mowy nie było), a największą znaną liczbą pierwszą (tego tematu też nie poruszono) była wówczas $2^{216091} - 1$; we wrześniu 2023 tę rolę pełni $2^{82589933} - 1$. Postęp w tym zakresie zobrazowany jest na przykład przez to, że gdybyśmy największą znaną w 1986 roku liczbę pierwszą zapisali na pasku papieru taką czcionką, jaką wydrukowany jest ten tekst w *Delcie*, zajęłoby to około 114 metrów. W przypadku obecnej rekordzistki – ponad 43 kilometry. A liczby to wielkie nad wyraz – napisana na takim pasku szacunkowa liczba atomów we Wszechświecie nie wypełniłaby nawet jednej linijki kartki A4.

Także w przypadku kilku „deltowych ciekawostek” sprzed 37 lat można dziś powiedzieć coś więcej... O tym będzie mowa.

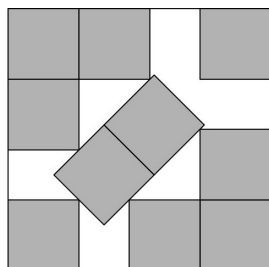
Oczywiste jest, że w kwadracie o wymiarach $n \times n$ możemy tak ułożyć n^2 kwadratów o boku 1, by ich wnętrza były parami rozłączne – nazwijmy takie umieszczenie małych kwadratów „dobrym”. Czy w sytuacji, gdy nieznacznie (o mniej niż 1) zwiększymy bok dużego kwadratu, można weń w dobry sposób włożyć więcej kwadratów jednostkowych? Nie da się tego zrobić, gdy będziemy je układać „standardowo”, czyli tak, by ich boki były równoległe do boków dużego kwadratu. Można je jednak umieścić inaczej... Na rysunkach poniżej pokazane są schematy takiego ułożenia. W kwadracie o boku $3 + \frac{\sqrt{2}}{2}$ (w przybliżeniu 3,71) zmieścimy $3^2 + 1$ kwadratów jednostkowych (rys. 1). W kwadrat o boku $5 + \frac{5\sqrt{2}}{2}$ (w przybliżeniu 8,54) można włożyć $8^2 + 1$ kwadratów jednostkowych (rys. 2). W tej sytuacji niepokryte pole w kwadracie wynosi $\frac{50\sqrt{2}-55}{2}$, czyli w przybliżeniu 7,86.



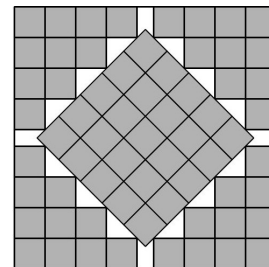
Delta, marzec 1991



Delta, luty 1992



Rys. 1



Rys. 2

Gdy n jest odpowiednio duże, można wkładać małe kwadraty jeszcze bardziej oryginalnymi metodami, a w konstrukcji odpowiednio wykorzystywać równoległoboki, trapezy... Powstaje naturalne pytanie: jak małe może być pole (oznaczmy je przez $W(x)$), które musi pozostać niepokryte przy umiejętnym ułożeniu kwadratów jednostkowych w kwadracie o boku x , przy czym x nie musi być liczbą naturalną. Formalnie:

$$W(x) = x^2 - \sup\{P : P \text{ jest liczbą kwadratów jednostkowych, które możemy umieścić w „dobry” sposób w kwadracie o boku } x\}.$$

Wymienieni w tekście matematycy mają w swym dorobku liczne znaczące osiągnięcia. Na marginesie podam jedynie ich daty życia.

Paul Erdős (1913–1996)
 Ronald Lewis Graham (1935–2020)
 Hugh Lowell Montgomery (ur. 1944)

Klaus Friedrich Roth (1925–2015)
 Robert Vaughan (ur. 1945)
 Fan Chung (ur. 1949)



Delta, lipiec 1993

A α	B γ	C β
C γ	A β	B α
B β	C α	A γ

Rys. 3

Leonhard Euler (1707–1783)
 Gaston Tarry (1843–1913)
 Raj Chandra Bose (1901–1987)
 Sharadchandra Shankar Shrikhande (1917–2020)
 Ernest Tilden Parker (1926–1991)

B	C	J	D	F	E	A	I	H	G
β	δ	α	ϑ	ε	ι	η	γ	λ	ζ
A	J	G	F	D	I	E	C	B	H
ϑ	λ	δ	α	β	ε	γ	η	ζ	ι
E	A	F	H	I	D	C	G	J	B
η	β	ζ	δ	α	λ	ε	ϑ	ι	γ
G	E	G	I	B	C	D	H	F	J
ε	ϑ	λ	ι	δ	α	ζ	β	γ	η
D	H	E	A	C	J	G	B	I	F
ι	ε	β	ζ	γ	δ	α	λ	η	ϑ
H	D	B	E	A	G	F	J	C	I
α	γ	ε	λ	ι	η	δ	ζ	ϑ	β
I	B	D	J	E	A	H	F	G	C
δ	α	η	ε	ζ	γ	ϑ	ι	β	λ
J	F	I	C	G	H	B	D	A	E
γ	η	ϑ	β	λ	ζ	ι	δ	α	ε
F	I	C	G	H	B	J	A	E	D
λ	ζ	ι	γ	η	ϑ	β	ε	δ	α
C	G	H	B	J	I	F	E	D	A
ζ	ι	γ	η	ϑ	β	λ	α	ε	δ

Rys. 4

Tę liczbę zaczęli badać Paul Erdős i Ronald Graham. W 1974 roku wykazali oni istnienie takiej stałej C , że dla odpowiednio dużych x zachodzi oszacowanie $W(x) \leq C \cdot x^{7/11}$ (jeśli użyjemy ciut bardziej zaawansowanej terminologii, zapisujemy to jako $W(x) = O(x^{7/11})$). W 1986 roku znano oszacowanie nieznacznie wzmocnione (o tym była mowa w ciekawostce nr 4). Liczba $x^{7/11}$ została zastąpiona przez $x^{(3-\sqrt{3})/2}$; dokonał tego Hugh Montgomery, który jednak swojego rezultatu nie opublikował.

Można rozważać oszacowanie niepokrytego pola od dołu. Klaus Roth (medalista Fieldsa z 1958 roku) i Robert Vaughan podali w 1978 roku odpowiednie ograniczenia, rozwiązując problem postawiony przez Erdősa i Grahama. Z ich rezultatu wynikało, że $W(n + \frac{1}{2})$ zawsze jest większe od \sqrt{n} .

Oszacowanie Montgomery’ego w kolejnych latach wzmocniano. Ronald Graham i jego żona Fan Chung udowodnili, że $W(x) = O(x^{(3+\sqrt{2})/7} \cdot \log x)$. A „jeszcze cieplej” wynik Chung i Grahama, opublikowany w 2020 roku, mówi, że $W(x) = O(x^{3/5})$.

Liczba $\frac{7}{11}$ to w przybliżeniu 0,6364, w oszacowaniu Montgomery’ego mamy w wykładniku 0,6334. Chung i Graham doszli do liczby 0,6. W swojej pracy stawiają kolejne problemy. W szczególności stwierdzili, że rezultat Rotha i Vaughana pozwala zapytać, czy przypadkiem nie zachodzi związek $W(x) = O(x^{\frac{1}{2}})$, niemniej oni w to nie wierzą... Za wykazanie (lub obalenie) tego, że $W(x) = O(x^{\frac{3}{5}-c})$ dla pewnej stałej c , zaoferowali 250 dolarów.

Przejdźmy teraz do ciekawostki nr 86. Zacytujmy: *Spróbujmy ustawić 36 wojskowych (po 6 z sześciu jednostek i po sześciu w tym samym stopniu, przy czym reprezentację jednostki stanowią wojskowi różnych stopni) na polach szachownicy 6 x 6 tak, by wszyscy stojący w szeregu czy w kolumnie byli różnych stopni i z różnych jednostek. Albo lepiej nie próbujmy – to na pewno się nie uda.* Podana została też informacja, że w przypadku pewnych kwadratów $n \times n$ da się to zrobić, ale w przypadku np. 10-osobowych reprezentacji dziesięciu jednostek problem jest wciąż otwarty.

Nie zostało wtedy w *Delcie* wspomniane, że autorem zadania o 36 wojskowych jest nie kto inny, jak Leonhard Euler. Pytanie to postawił w 1779 roku. W przypadku dwóch jednostek łatwo zauważyć, że odpowiednie ustawienie nie jest możliwe, w przypadku trzech zadanie jest bardzo proste (rys. 3). Euler potrafił stosownie rozmieścić żołnierzy dla $n = 4$ i $n = 5$, jednak po bezskutecznych próbach konstrukcji dla $n = 6$ wyraził w 1782 roku przypuszczenie, że w tym przypadku jest to niewykonalne. Na dowód trzeba było czekać ponad 100 lat... Opublikował go w 1901 roku Gaston Tarry. Obecnie za pomocą komputerów poradziłibyśmy sobie z problemem bez większych kłopotów.

Już Euler wiedział, że można odpowiednio rozlokować żołnierzy w przypadkach n nieparzystych oraz podzielnych przez 4. Sądził, że jest to niemożliwe dla liczb postaci $4k + 2$. Okazało się jednak, że jest inaczej.

W 1986 roku w Polsce dostęp do rozmaitych czasopism matematycznych był nad wyraz ograniczony – okazuje się, że Redakcja *Delty* nie dysponowała pełną informacją, gdyż problem w przypadku $n = 10$ był już wtedy (jak zresztą we wszystkich innych przypadkach) rozstrzygnięty. W 1959 roku Raj Bose, Sharadchandra Shrikhande i Ernest Parker udowodnili, że można odpowiedniego ustawienia dokonać dla wszystkich parzystych n niepodzielnych przez 4 i większych od 6. Warto wspomnieć, że artykuł (artykuł, nie krótka wzmianka) opowiadający o ich wyniku został opublikowany 26 kwietnia 1959 roku na łamach dziennika *New York Times*.

Odpowiednie ustawienie dla $n = 10$ pokazane jest na marginesie.

O kwantowej wersji problemu ustawiania wojskowych można przeczytać w artykule *36 splełtanych oficerów z Petersburga* w Δ_{23}^3 .

Lothar Collatz (1910–1990)
Stanisław (Stan) Ulam (1909–1984)
Shizuo Kakutani (1911–2004)
Bryan Thwaites (ur. 1923)
Helmut Hasse (1898–1979)
Harold Scott MacDonald Coxeter (1907–2003)
David Bařina (ur. 1984)



Delta, październik 1994

Szczegóły: Z Markiem Kordosem o *Delcie*, *Epsilon 11a*, [w]: D.Ciesielska, K.Ciesielski, Z.Pogoda, *Epsilon*, Wydawnictwo Szkolne Omega, 2002. Angielska wersja wywiadu: *European Mathematical Society Newsletter* No. 41 (dostępne na stronie www *European Mathematical Society*).

Kolejną, której poświęcimy tu uwagę, jest ciekawostka nr 52. Rozważamy liczbę naturalną n i tworzymy rekurencyjnie ciąg: jeśli wyraz k jest liczbą parzystą, to następnym jest $\frac{k}{2}$, jeśli nieparzystą – to $3k + 1$. W znanych przypadkach zawsze w pewnym momencie pojawia się jedynka (i potem konsekwentnie, okresowo, układ 4, 2, 1). Czy tak jest zawsze? W 1986 roku wiadomo było, że tak się dzieje, jeśli pierwszy wyraz jest mniejszy od 2^{40} .

Hipotezę, że zawsze dojdziemy do jedynki, postawił w 1937 roku Lothar Collatz. Nazywana jest hipotezą Collatza, ale łączona jest też w nazwach z Ulamem, Kakutanim, Thwaitesem i Hassem. . . Żadne z tych nazwisk nie pojawiło się wtedy w *Delcie*, ale ciekawostki musiały być krótkie. Stanisław Ulam oraz Shizuo Kakutani zainteresowali się problemem i go rozpowszechniali, a Bryan Thwaites postawił go niezależnie od Collatza w 1952 roku. Można też w języku angielskim spotkać nazwę *Syracuse Problem* (tę nazwę zaproponował Helmut Hasse, gdy przedstawił go w latach pięćdziesiątych podczas wizyty na uniwersytecie w Syracuse; według innej wersji, problem przyciągnął uwagę matematyków z Syracuse na Międzynarodowym Kongresie Matematyków w Cambridge w 1950 roku – wtedy Collatz mówił o nim różnym uczestnikom Kongresu). Inna angielska nazwa to *Hailstone Problem* (*hailstone* to ziarenko gradu, a wyrazy ciągu się zwiększają i zmniejszają, jak ziarenka gradu w chmurze). Pojawiają się również terminy *hailstone numbers* oraz *wondrous numbers* (*wondrous* to cudowny, niezwykły). Paul Erdős, który wielokrotnie fundował nagrody za rozwiązanie rozmaitych problemów, za rozstrzygnięcie hipotezy Collatza oferował 500 dolarów. Nie był jedyny; przed nim, w 1970 roku, H.S.M. Coxeter ogłosił nagrodę w wysokości 50 dolarów, a później, w 1982 roku, Thwaites – 1000 dolarów.

Jedyną rzeczą, która w tym przypadku się zmieniła, jest katalog pierwszych wyrazów ciągów, dla których hipoteza jest prawdziwa. We wrześniu 2023 wiadomo, że dojdziemy do jedynki, jeśli zaczniemy od jakiegokolwiek liczby mniejszej od 2^{70} (w przybliżeniu $1,18 \cdot 10^{21}$). Prace nad większymi liczbami trwają (oczywiście z wykorzystaniem komputerów); intensywnie prowadzi je David Bařina z Brna.

Ciąg badany w hipotezie Collatza związany jest z wieloma pasjonującymi problemami i wynikami, ale to oddzielna sprawa.

Ostatnią ciekawostką, do której nawiążę, jest ta o numerze 115. Zacytuję w dokładnym brzmieniu: *Niektórzy nazywają twierdzenie o trzech ciągach (jeśli $a_n \leq b_n \leq c_n$ oraz $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} c_n = g$, to ciąg (b_n) jest zbieżny i $\lim_{n \rightarrow \infty} b_n = g$) twierdzeniem o milicjantach i formułują je tak: Jeśli znajdziesz się między dwoma milicjantami idącymi do tego samego komisariatu, to też tam trafisz.*

Tu też mamy istotną zmianę, bo milicji od dawna w Polsce nie ma. . .

Lat temu 37 – oprócz tego, że mi się to bardzo spodobało – zastanawiałem się, jak cenzura dopuściła do opublikowania tych słów. Wskazane jest wyjaśnienie dla młodszych Czytelników: przed 1989 rokiem wszystko, co miało zostać wydane drukiem, musiało być zatwierdzone przez Główny Urząd Kontroli Prasy, Publikacji i Widowisk, popularnie nazywany cenzurą. A kontrolę przeprowadzano skrupulatnie i dotyczyły one także czasopism popularnonaukowych. Marek Kordos w wywiadzie udzielonym *Epsilonowi* opowiadał, że ingerencja cenzury pojawiła się już przy okazji tworzenia pierwszego numeru *Delty*. Umieszczono tam artykuł o Instytucie Badań Jądrowych w Krakowie i podano między innymi, ile Instytut zajmuje hektarów. Ta informacja okazała się wagi państwowej i należało ją skreślić. A na temat milicji – wracam do ciekawostki – można było wtedy pisać wyłącznie absolutnie pozytywnie, żarty nie wchodziły w grę. . .

Jak się później dowiedziałem, analogiczna nazwa istnieje (i nawet jest używana) w języku włoskim – *teorema dei due carabinieri*. Nie wiem, czy te nazwy powstały niezależnie od siebie i nie wiem, która jest wcześniejsza.

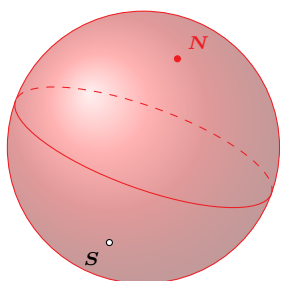
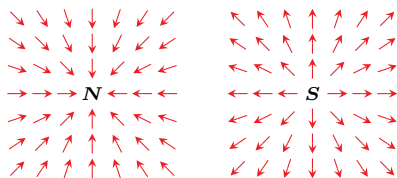
Może z okazji sześćdziesięciolecia *Delty* warto będzie sprawdzić, co się ewentualnie zmieniło w przypadku ciekawostek fizycznych i astronomicznych?

Twierdzenie Białynickiego-Biruli o rozkładzie

Maria DONTEN-BURY*, Joachim JELISIEJEW*, Jarosław A. WIŚNIEWSKI*

* Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Maria DONTEN-BURY była redaktorką *Delta* w latach 2009–2014.

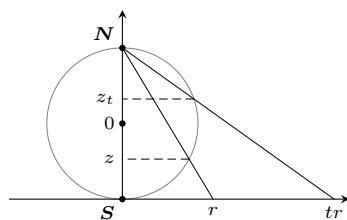


Wyprawa na biegun północny. Wyobraźmy sobie podróżników na powierzchni Ziemi wyposażonych w kompasy wskazujące północ (przyjmijmy, że północ magnetyczna i geograficzna to to samo i że nie ma żadnych lokalnych anomalii). Wskazania kompasów w okolicach biegunów północnego N i południowego S pokazują rysunki na marginesie. Każdy z podróżników chce dotrzeć na biegun północny. Jeśli podróżnik nie znajduje się akurat na biegunie, to wie, co ma robić: wędruje wzdłuż południka, na którym się znajduje, zgodnie ze wskazaniem kompasu. Ale na biegunie kompas jest bezużyteczny, bo nie pokazuje jednego kierunku – bieguny są punktami krytycznymi dla zadania podróżnika. Mimo to podróżnik będący na biegunie północnym umie poradzić sobie z zadaniem: po prostu stoi w miejscu.

W ten sposób zadanie dotarcia do bieguna i sam proces podróży dają nam podział powierzchni Ziemi (którą matematyk widzi jako sferę) na części, nazywane komórkami. Jedna to prawie cała sfera wraz z punktem N , druga jest jednopunktowa i składa się z bieguna S . Zauważmy, że komórki są w pewnym sensie prostsze niż wyjściowa sfera: ta większa komórka to właściwie płaszczyzna, tylko trochę wygięta.

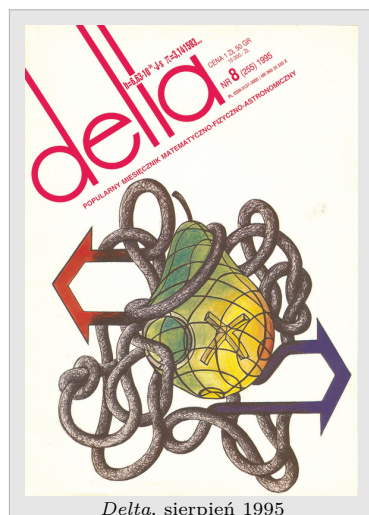
Twierdzenie o rozkładzie. Powyższy przykład wygląda na łatwy, ale Czytelnik może sobie wyobrazić, że jeśli zamiast sfery próbujemy podzielić na proste kawałki przestrzeni o mniej oczywistej strukturze, to problem robi się znacznie ciekawszy. Dla pewnej klasy przestrzeni rozwiązaniem jest *twierdzenie Andrzeja Szczepana Białynickiego-Biruli o rozkładzie*, które można wypowiedzieć jednym zdaniem:

Jeśli pogrupujemy punkty rozmaitości X względem granic przy działaniu grupy \mathbb{K}^ , to otrzymamy zaskakująco ładne i mniej skomplikowane podrozmaitości.*



To, że X jest *rozmaitością*, znaczy, że w bliskiej okolicy każdego swojego punktu przypomina zwykłą przestrzeń euklidesową: prostą, płaszczyznę, przestrzeń trójwymiarową... być może trochę zakrzywioną. Taką strukturę mają na przykład okrąg, sfera, powierzchnia torusa, ale też bardziej skomplikowane formy, jak przestrzeń rzutowa (o której za chwilę).

Oprócz rozmaitości mamy *grupę*, czyli zbiór elementów, na których możemy wykonywać pewne działanie oraz je odwracać. Na przykład grupę \mathbb{R}^* niezerowych liczb rzeczywistych, które da się mnożyć i dzielić. Taka grupa *działa* na pewnej rozmaitości, jeśli każdy jej element coś robi z punktami rozmaitości, jakoś je przestawia. Co więcej, te przestawienia mają być zgodne z mnożeniem w grupie, ale tu nie będziemy potrzebowali tego dokładnie wyjaśnić.



Spróbujmy w tym języku opowiedzieć o wyprawie na biegun północny. Jak zwykle, zamiast o powierzchni Ziemi myślimy o sferze S^2 zanurzonej w trójwymiarowej przestrzeni \mathbb{R}^3 . Opiszmy tę sferę równaniem $x^2 + y^2 + z^2 = 1$. Działa na niej grupa \mathbb{R}^* w następujący sposób. Trzecią współrzędną punktu na sferze możemy zapisać jako $z = \frac{r^2 - 4}{r^2 + 4}$ dla pewnego $r \in \mathbb{R}$. Wtedy liczbie $t \in \mathbb{R}^*$ przypiszemy przekształcenie S^2 , które przesuwają punkt wzdłuż południka do punktu o trzeciej współrzędnej $z_t = \frac{t^2 r^2 - 4}{t^2 r^2 + 4}$. Skąd te wzory? Pochodzą od mnożenia przez liczby z \mathbb{R}^* , jeśli najpierw rozprostujemy południk przekształceniem odwrotnym do rzutu stereograficznego. Geometryczny charakter przedstawionych zależności prezentuje rysunek na marginesie.

Oczywiście współrzędne x i y też się zmieniają – w taki sposób, żeby podróżnik został na tej samej długości geograficznej – ale nie potrzebujemy znać dokładnie ich nowych wartości. Zauważmy jeszcze, że bieguny moglibyśmy (niezbyt ściśle) opisać, w miejsce r wstawiając 0 oraz ∞ .



Rozwiązanie zadania F 1085.

Zacznijmy od przypomnienia skal energii.

Energia wiązań chemicznych to co najwyżej około 10 eV – wiązanie cząsteczki N₂ o energii 9,79 eV należy do najsilniejszych. Energia wiązania elektronów wewnętrznych powłok atomowych lekkich pierwiastków sięga pojedynczych kiloelektronowoltów (1362 eV dla neonu). Oznacza to, że azot ogrzany do temperatury $T = 5000 \text{ eV}/k \approx 5,8 \cdot 10^7 \text{ K}$ jest mieszaniną jonów azotu N⁷⁺ i elektronów e⁻ w liczbie 7 razy większej niż liczba jonów azotu, a średnia energia kinetyczna każdego z jonów i elektronów wynosi 2000 eV. Z każdego mola azotu cząsteczkowego o masie $m = 2 \cdot 14 \text{ g} = 28 \text{ g}$ powstaje 16 moli mieszaniny gazów jednocząstkowych. Wartość ciepła właściwego w stałej objętości na jednostkę masy tej mieszaniny wynosi:

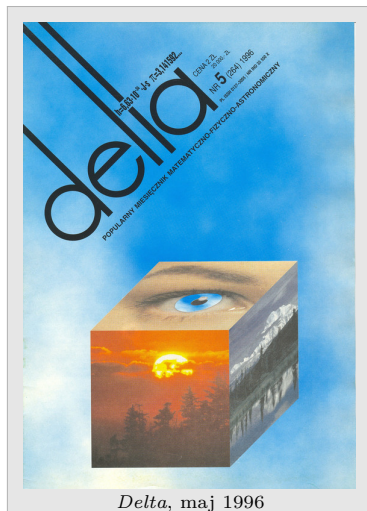
$$c_v = \frac{3}{2} 16R/m.$$

W powyższym wzorze

$R \approx 8,314 \text{ J}/(\text{mol}\cdot\text{K})$ oznacza stałą gazową. Liczbowo: $c_v \approx 7,13 \text{ J}/(\text{g}\cdot\text{K})$ i jest około 1,7 raza większe od ciepła właściwego ciekłej wody. Podana temperatura 5000 eV/k to około połowy temperatury w pobliżu centrum wybuchu jądrowego.

Definicję na klasach punktów można ująć w zgrabnej formie:

$$\mathbb{RP}^2 := (\mathbb{R}^3 \setminus \{(0, 0, 0)\})/\mathbb{R}^*.$$



Delta, maj 1996

Twierdzenie o rozkładzie mówi o grupowaniu punktów S^2 względem *granic* przy działaniu \mathbb{R}^* . Możemy zapisać $z_t = \frac{r^2 - 4/t^2}{r^2 + 4/t^2}$, skąd widać, że dla każdego $r \neq 0$ (czyli $z \neq -1$) jeśli t dąży do ∞ , to z_t dąży do 1. A jedyny punkt na sferze, w którym $z = 1$, to biegun N . Wobec tego grupowanie punktów S^2 względem *granic* przy badanym działaniu wyodrębnia, jak przewidzieliśmy, dużą komórkę złożoną z całej sfery oprócz bieguna S , dla której granicą jest biegun N . Druga komórka jest jednopunktowa, złożona z bieguna S , który jest swoją własną granicą przy działaniu \mathbb{R}^* , czyli tak zwanym *punktem stałym*.

Granice, których... brakuje. Jak widać, granice przy działaniu grupy są kluczowym elementem twierdzenia o rozkładzie. W powyższym przykładzie mieliśmy szczęście, i udało nam się je dość łatwo obliczyć. Spójrzmy teraz na sytuację, która pomoże nam lepiej zdefiniować granice przy działaniu grupy w ogólnym przypadku.

Niech $\mathbb{K} = \mathbb{R}$ i określmy działanie \mathbb{R}^* na płaszczyźnie \mathbb{R}^2 wzorem

$$(1) \quad t \circ (x, y) := (t^{-1}x, ty),$$

gdzie $t \in \mathbb{R}^*$. Jaka jest granica punktu (x, y) przy tak określonym działaniu? Jeśli $y = 0$, to sprawa jest prosta:

$$\lim_{t \rightarrow \infty} t \circ (x, y) = \lim_{t \rightarrow \infty} (t^{-1}x, ty) = \lim_{t \rightarrow \infty} (t^{-1}x, 0) = (0, 0).$$

Ale jeśli $y \neq 0$, to ta granica nie istnieje na płaszczyźnie... Fakt, że czasem tak się dzieje, jest irytujący. Mamy kłopot, ponieważ użyliśmy płaszczyzny \mathbb{R}^2 , której brakuje „punktów w nieskończoności”. Problem ten odczuwało już wielu matematyków i istnieje standardowe lekarstwo: zamiast płaszczyzny \mathbb{R}^2 powinniśmy rozważyć *płaszczyznę rzutową*.

Płaszczyzna rzutowa to wersja płaszczyzny z „wbudowaną nieskończonością”. W *Delcie* o niej nie raz już była mowa – po krótki opis można sięgnąć do Δ_{19}^4 , ale lepiej poznać ją z różnych punktów widzenia w Δ_{13}^5 . Tutaj zdefiniujemy ją jako zbiór punktów $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$ (tak!), gdzie utożsamiamy punkty (x_0, y_0, z_0) oraz (x_1, y_1, z_1) , jeśli są proporcjonalne, czyli istnieje taka liczba $\lambda \in \mathbb{R}^*$, że $x_1 = \lambda x_0$, $y_1 = \lambda y_0$, $z_1 = \lambda z_0$. Klasę tak utożsamionych punktów oznaczamy przez $(x_0 : y_0 : z_0)$, a zbiór wszystkich tych klas to właśnie płaszczyzna rzutowa, oznaczana jako \mathbb{RP}^2 . Dodajmy, że z przedstawionej definicji wynika, iż $(x_0 : y_0 : z_0) = (\lambda x_0 : \lambda y_0 : \lambda z_0)$ dla każdej niezerowej liczby λ .

Z tej definicji nie widać jednak zupełnie, dlaczego \mathbb{RP}^2 nazywa się *płaszczyzną rzutową*. Otóż w każdej klasie $(x_0 : y_0 : z_0) \in \mathbb{RP}^2$, gdzie $z_0 \neq 0$, jest dokładnie jeden punkt z ostatnią współrzędną równą 1 – to $(x_0/z_0, y_0/z_0, 1)$. Możemy więc utożsamić zbiór

$$U := \{(x_0 : y_0 : z_0) \in \mathbb{RP}^2 \mid z_0 \neq 0\}$$

ze zbiorem \mathbb{R}^2 , gdzie punktowi $(x_0 : y_0 : z_0) \in \mathbb{RP}^2$ odpowiada $(x_0/z_0, y_0/z_0) \in \mathbb{R}^2$. Pozostałe punkty \mathbb{RP}^2 , te postaci $(x_0 : y_0 : 0)$, uważamy za dodane do płaszczyzny U *punkty w nieskończoności*.

Wracając do naszego przykładu, możemy zdefiniować działanie \mathbb{R}^* na \mathbb{RP}^2 poprzez $t \circ (x : y : z) = (t^{-1}x : ty : z)$. Jeśli popatrzymy na U , to $t \circ (x : y : 1) = (t^{-1}x : ty : 1)$, co dokładnie zgadza się ze wzorem (1), którym powyżej definiowaliśmy działanie na \mathbb{R}^2 . Teraz wreszcie możemy znaleźć granicę dowolnego punktu przy tak ustalonym działaniu. Mianowicie, jeśli weźmiemy $(x : y : z) \in \mathbb{RP}^2$, gdzie $y \neq 0$, to

$$\begin{aligned} \lim_{t \rightarrow \infty} t \circ (x : y : z) &= \lim_{t \rightarrow \infty} (t^{-1}x : ty : z) = \lim_{t \rightarrow \infty} (t^{-1} \cdot t^{-1} \cdot x : t^{-1} \cdot ty : t^{-1}z) = \\ &= \lim_{t \rightarrow \infty} (t^{-2}x : y : t^{-1}z) = (0 : y : 0) = (0 : 1 : 0). \end{aligned}$$

Prościej można obliczyć granicę punktu $(x : 0 : z) \in \mathbb{RP}^2$, gdzie $z \neq 0$:

$$\lim_{t \rightarrow \infty} t \circ (x : 0 : z) = \lim_{t \rightarrow \infty} (t^{-1}x : 0 : z) = (0 : 0 : z) = (0 : 0 : 1).$$

Ostatni nierozważony jeszcze punkt \mathbb{RP}^2 to $(1 : 0 : 0)$. Ale w tym przypadku $t \circ (1 : 0 : 0) = (t^{-1} : 0 : 0) = (1 : 0 : 0)$, czyli mamy punkt stały działania.



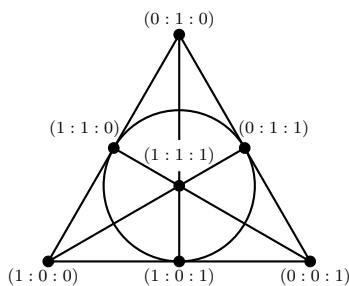
Rozwiązanie zadania M 1767.
Zauważmy, że

$$(m+1)(m(4m+3)^2+1) = \\ = ((m+1)(4m+1))^2,$$

wobec tego para $(m, m(4m+3)^2)$ spełnia warunki zadania.

Co by było, gdyby te liczby nie były różne?

Istnieją też dodatnie komórki zdefiniowane podobnie, ale gdy granice bierze się przy $t \rightarrow 0$.



Ilustracja płaszczyzny rzutowej $\mathbb{Z}_2\mathbb{P}^2$

Jak wobec tego wygląda pogrupowanie punktów względem granic przy działaniu \mathbb{R}^* ? Zbiór punktów takich, że $y \neq 0$, których granicą jest $(0:1:0)$, można utożsamić z płaszczyzną \mathbb{R}^2 (podobnie jak powyżej punkty, dla których $z \neq 0$). Zbiór punktów takich, że $y = 0$, ale $z \neq 0$, dla których granicą jest $(0:0:1)$, utożsamia się z prostą \mathbb{R} . Ostatni zbiór składa się z punktu $(1:0:0)$. Otrzymaliśmy więc podział płaszczyzny rzutowej na bardzo ładne części – ładne w tym przypadku oznacza, że można je utożsamić z \mathbb{R}^n dla pewnego n , zależnego od granicy.

Tytułowe twierdzenie, które wreszcie możemy dokładnie sformułować, mówi, że to samo będzie się działo dla każdego (algebraicznego) działania na każdej gładkiej podrozmaitości w każdej przestrzeni rzutowej (nie tylko na płaszczyźnie). Teraz zamiast \mathbb{R} będziemy dla większej ogólności pisać \mathbb{K} , bo twierdzenie działa nie tylko dla ciała liczb rzeczywistych. Ustalmy też różne liczby całkowite a_0, \dots, a_n i zdefiniujmy działanie \mathbb{K}^* na $\mathbb{K}\mathbb{P}^n$ wzorem

$$t \circ (x_0 : x_1 : \dots : x_n) = (t^{a_0}x_0 : t^{a_1}x_1 : \dots : t^{a_n}x_n).$$

Zbiór punktów o ustalonej granicy p nazywamy (*ujemną*) *komórką Białynickiego-Biruli* stowarzyszoną z p .

Twierdzenie (ASBB, 1973). *Niech $X \subseteq \mathbb{K}\mathbb{P}^n$ będzie gładką podrozmaitością taką, że $\mathbb{K}^* \circ X \subseteq X$, czyli działanie nie wyprowadza punktów X poza X . Wtedy każda komórka Białynickiego-Biruli w X jest izomorficzna z przestrzenią liniową \mathbb{K}^{n_i} dla pewnego $n_i \in \mathbb{N}$ zależnego od komórki.*

Inne ciała. Jak wspomnieliśmy powyżej, nie korzystamy z żadnych szczególnych własności ciała liczb rzeczywistych. Analiza przykładu przebiegałaby właściwie tak samo, gdyby za \mathbb{R} wszędzie podstawić \mathbb{Q} . Ba, ten argument działa nawet, jeśli zamiast \mathbb{R} wybierzemy ciało liczb zespolonych \mathbb{C} lub ciało \mathbb{Z}_p , które ma p elementów, dla liczby pierwszej p . Musimy jedynie zachować konwencję, że $\lim_{t \rightarrow \infty} t^{-n}x$ jest równe zero dla każdego naturalnego $n > 0$ i każdego $x \in \mathbb{C}$ czy \mathbb{Z}_p . To wiedzie do paradoksalnego, ale słusznego wniosku, że używane powyżej pojęcie granicy nie wymaga żadnej „ciągłości”. Przykładowo, Czytelnik może obliczyć, że $\mathbb{Z}_2\mathbb{P}^2$ składa się jedynie z siedmiu punktów:

$$(0:0:1), (0:1:1), (1:0:1), (1:1:1), (1:0:0), (0:1:0), (1:1:0),$$

a mimo to równość $\lim_{t \rightarrow \infty} t \circ (1:1:1) = (0:1:0)$ dla działania z naszego przykładu nadal zachodzi.

Gładkość. Komórki Białynickiego-Biruli można też zdefiniować w ogólniejszym przypadku, bez zakładania gładkości. Ustalmy tylko $X \subseteq \mathbb{K}\mathbb{P}^n$ spełniający $\mathbb{K}^* \circ X \subseteq X$. Wtedy definiujemy X^+ jako zbiór wszystkich \mathbb{K}^* -niezmienniczych odwzorowań z \mathbb{K} do X . Nie jest jasne, dlaczego na tak określonym zbiorze da się zbudować strukturę przestrzeni. Dopiero W. Drinfeld w roku 2013 dowiódł, że to możliwe. Przestrzeń ta okazała się bardzo użyteczna w badaniu przypadku odległego od gładkiego – w tzw. przestrzeniach moduli. Ale to już temat na inny artykuł.

Zakończenie. Chcielibyśmy zakończyć niniejszy artykuł fragmentem odpowiedzi Andrzeja Szczepana Białynickiego-Biruli na pytanie o pożytki z powszechnego nauczania matematyki, będące częścią przeprowadzonej przez Deltę ankiety (Δ_{95}^{10}):

(...) roli matematyki nie można sprowadzać do znajomości wzorów i faktów. Sądzę, że ważna jest szkoła myślenia, którą daje matematyka każdemu, pewna dyscyplina formułowania i wypowiedzania poglądów i argumentów. Brak takiej dyscypliny może bardzo w życiu przeszkadzać.

Na trzech stronach tego artykułu przebyliśmy dość daleką drogę – od wprowadzającego przykładu o podróżnikach na sferze do analizy bardzo abstrakcyjnych struktur. Mamy nadzieję, że lektura była dla Czytelników dobrym treningiem wspomnianego wyżej matematycznego myślenia.

Andrzej Szczepan Białynicki-Birula zmarł w 2021 r. Od 2022 r. działa Fundacja Jego imienia, wspierająca studentów matematyki i informatyki UW z Białorusi i Ukrainy:
www.mimuw.edu.pl/~fundacja_asbb

Współbieżność jest nieintuicyjna

Tomasz IDZIASZEK*

*Redaktor *Delta* w latach 2010–2016, współautor rubryki „Informatyczny Kącik Olimpijski”

O tym, jak różne rozwiązania techniczne stosowane w komputerach mają wpływ na wydajność uruchamianych na nich programów, pisaliśmy na łamach *Delta* niejednokrotnie.

Pierwsze komputery miały pojedynczy procesor. Można było zakładać, że każda z jego podstawowych operacji (tj. wczytanie z pamięci, wykonanie jednostki obliczeń i zapis do pamięci) trwa podobną ilość czasu.

O przetwarzaniu potokowym pisaliśmy w artykule *O sierotce, co chciała się mózgiem elektronowym wyręczyć*, Δ_{12}^8 .

W kolejnych dekadach inżynierowie konstruowali coraz bardziej skomplikowane procesory: tranzystory były mniejsze, gęściej upakowane, co pozwalało na implementację wydajniejszych mechanizmów umożliwiających wykonywanie wielu instrukcji jednocześnie (np. przetwarzanie potokowe).

O szczegółach komputerowej pamięci można przeczytać w artykułach: *Pamięć cache w praktyce*, Δ_{09}^{10} , *Kolejność ma znaczenie*, Δ_{11}^9 , *Pamięć w komputerze*, Δ_{16}^5 .

Potrzebowano również coraz większych pamięci przechowujących dane w komputerze. Problem w tym, że duże pamięci są albo wolne, albo drogie. Poradzono sobie z nim, tworząc tzw. hierarchię pamięci: procesor ma dostęp do pamięci podręcznej (kilka poziomów *cache*), pamięci głównej (RAM), a w końcu pamięci masowej (dyski twarde).

Proste modele są pożądane, bo łatwo analizować, co się w nich dzieje. Model procesora typu „wykonuję po kolei pojedyncze instrukcje, kończąc wykonanie instrukcji, zanim zacznę następną” jest intuicyjny dla programisty i wygodny, jeśli chcemy formalnie udowodnić poprawność działania naszego programu. Jednak chęć wprowadzenia do modelu dodatkowych mechanizmów mających na celu zwiększenie wydajności systemu powoduje, że musimy dobrze przemyśleć, jak te mechanizmy będą współdziałały z modelem i z pozostałymi dodatkami. Za przykład niech posłużą tutaj dwa mechanizmy: zmiana kolejności instrukcji (*reordering*) oraz efektywne korzystanie z pamięci podręcznej.

Zmiana kolejności może być wykonana przez kompilator, który tłumaczy program z języka wysokiego poziomu na kod maszynowy, a następnie optymalizuje ten kod. Przykładowo w kodzie $a=c+1$; $b=1$ kompilator może zamienić miejscami te dwie instrukcje, gdyż operują one na niezależnych miejscach w pamięci.

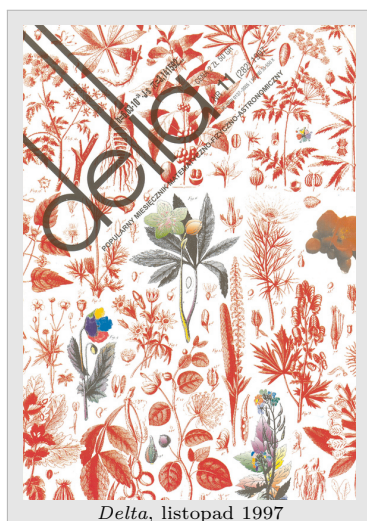
Ale ta zmiana może być również wykonana przez procesor, w ramach wykonywania instrukcji poza kolejnością (*out-of-order execution*). Procesor może, czekając na dostęp do pamięci dla zmiennej c , najpierw wykonać operację $b=1$, dla której ma wszystkie dane.

Podstawową zasadą jest tu, że zmiany te muszą być niewidoczne dla programisty. A konkretnie: wynik uruchomienia *jednowątkowego* programu przy zmianach kolejności musi być taki sam jak oryginalnego. Kluczowe jest tutaj założenie o jednowątkowości, tzn. program jest uruchamiany na jednym procesorze.

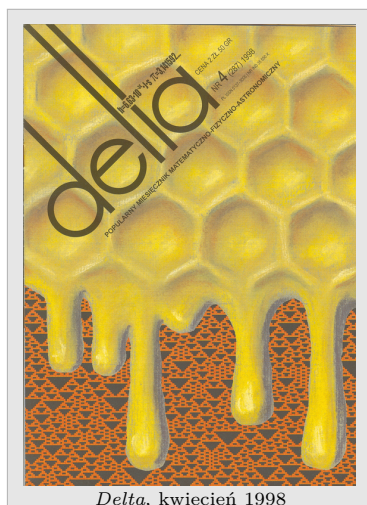
Idea hierarchii pamięci jest taka, że jeśli potrzebujemy wczytać dane, to najpierw szukamy ich w mniejszej i szybszej **pamięci podręcznej**. Dopiero, jeśli ich tam nie znajdziemy, to czytamy je z pamięci głównej i zapisujemy również do pamięci podręcznej (aby kolejny odczyt był szybszy; przy okazji zapisujemy też trochę sąsiednich komórek pamięci, zakładając lokalność odwołań).

Działa to całkiem efektywnie w przypadku *odczytu* z pamięci. A co w przypadku *zapisu*? Możemy założyć, że zapisu dokonujemy jedynie w pamięci podręcznej, a jeśli będziemy musieli jakieś dane z niej usunąć (bo brakuje nam miejsca), to wtedy dopiero zostaną one zapisane również w pamięci głównej (o ile były modyfikowane). Ten sposób „leniwego zapisu” ma swoją konsekwencję taką, że pamięć główna niekoniecznie jest zsynchronizowana z pamięcią podręczną, przy czym znowu: dla programu jednowątkowego jest to niewidoczne.

Jeszcze jednym sposobem przyspieszenia pracy procesora jest stworzenie w nim dodatkowego bufora, w którym kolejgowane są operacje zapisu (*store buffer*).



Delta, listopad 1997



Delta, kwiecień 1998

Tak więc nawet pamięć podręczna nie musi być w pełni zsynchronizowana (każdy odczyt z pamięci najpierw sprawdza bufor zapisu, a dopiero potem pamięć podręczną).

O problemach, jakie niesie ze sobą wieloprocesorowość, wspominaliśmy w artykule *Programowanie na platformie CUDA*, Δ_{11}^9 .

Upakowywanie coraz większej liczby tranzystorów ma swoje fizyczne granice. W związku z tym aktualną metodą na przyspieszenie komputerów jest instalowanie w nich **wielu procesorów** (lub procesorów o wielu rdzeniach). Idea jest taka, że można na nich niezależnie uruchamiać różne programy (czym automatycznie zajmuje się system operacyjny). Kłopot zaczyna się wtedy, kiedy chcemy wiele procesorów wykorzystać do przyspieszenia działania pojedynczego programu. Nie dostaniemy tego za darmo – nasz program musi być odpowiednio napisany, aby mógł być uruchomiony **współbieżnie**.

Problematyczna jest sytuacja, gdy wiele wątków (części naszego programu uruchamianych na różnych procesorach) próbuje uzyskać dostęp do tej samej komórki pamięci. Jeśli wszystkie z niej czytają – jest w porządku. Ale gdy choć jeden procesor do niej zapisuje, to możemy mieć do czynienia z wyścigiem (*race condition*). Zadaniem programisty jest zagwarantować, że taka sytuacja nigdy nie wystąpi.

Inną możliwością jest skorzystanie z wysokopoziomowych konstrukcji synchronizujących, tzw. muteksów (*mutex*).

Jednym ze sposobów jest użycie tzw. instrukcji atomowych (*atomic instructions*), które gwarantują na poziomie sprzętu, że na czas działania tej jednej instrukcji, procesor uzyskuje wyłączność na dostęp do komórki pamięci. Z wykorzystaniem takich instrukcji możemy zsynchronizować dostęp do pamięci, czego przykładem jest tzw. *message passing*:

```
T1: a=c+1    T2: while (ok != 1) { }
    ok=1      assert(a==c+1)
```

Mamy tu do czynienia z dwoma wątkami, T_1 i T_2 , które są uruchomione na dwóch procesorach. Pierwszy wątek zapisuje dane do pamięci dzielonej (zmienna a), a następnie ustawia flagę, że już skończył, korzystając z instrukcji zapisu do atomowej zmiennej ($ok=1$). Drugi wątek czeka, dopóki flaga nie jest ustawiona, a następnie czyta dane.

W programie mamy dwa wątki i dwie zmienne dzielone, początkowo równe ($x=y=0$):

```
T1: y=1    T2: x=1
    a=x      b=y
```

Gdyby model „przeplotu” był poprawny (tzn. prawdą byłoby, że wykonanie równoległe dwóch wątków odpowiada pewnemu przetasowaniu ich instrukcji), to instrukcja $b=y$ wykonałaby się po instrukcji $y=1$ albo instrukcja $a=x$ po instrukcji $x=1$. To pokazuje, że na końcu działania programu co najmniej jedna ze zmiennych, a lub b , będzie równa 1.

W praktyce okazuje się jednak, że nawet jeśli kolejność instrukcji nie zostanie zmieniona ani przez kompilator, ani procesor, to nadal wynik $a=0$, $b=0$ jest możliwy.

Jest to spowodowane buforem zapisu. Instrukcja $y=1$ kolejkuje się w buforze; zatem zmienna y dla wątku T_1 ma wartość 1, a dla wątku T_2 wartość 0 (bo ani pamięć główna, ani pamięć podręczna wspólna dla procesorów nie została jeszcze zsynchronizowana). Z symetrii zmienna x dla wątku T_2 ma wartość 1, a dla wątku T_1 wartość 0. Zatem dwa wątki nie zgadzają się co do kolejności, w której następują przypisania do zmiennych x i y .

Powyższy program nie musi być niestety poprawny w przypadku agresywnych operacji optymalizacji kompilatora lub procesora. Co będzie bowiem, gdy kompilator zamieni kolejność operacji w T_1 ? Z perspektywy tego wątku nic się nie zmieni, ale program współbieżny zadziała nieprawidłowo: wątek T_2 uzyska dostęp do nie w pełni zainicjowanych danych. Również procesor może doprowadzić do błędu. Ponieważ zmienne ok , a i c są niezależne, może on w wątku T_2 wczytać zawartość zmiennej a , zanim jeszcze zakończy się pętla.

Zabezpieczenie się przed takimi błędami należy do programisty. Jednym z rozwiązań jest umieszczenie w kodzie programu **barier** (*memory barrier* lub *fence*), który informuje kompilator i procesor, że nie można zmieniać kolejności instrukcji, pomiędzy którymi znajduje się bariera:

```
T1: a=c+1    T2: while (ok != 1) { }
    mb()      mb()
    ok=1      assert(a==c+1)
```

Barier należy jednak używać rozważnie, bo ich nadpodaż ogranicza możliwe optymalizacje i powoduje spadek wydajności programu. To na programiście spoczywa trud zapewnienia minimalnej liczby barier potrzebnych do poprawnego działania programu, bez zbytniego nadwężania jego wydajności. Jest to trudne zadanie, tym bardziej, że współbieżność wymaga od programisty wyrobienia sobie zupełnie innego rodzaju intuicji na temat programowania.

Przykładowo, typowa intuicja, że wynik działania programu wielowątkowego można uzyskać, rozważając pewien „przeplot” instrukcji uruchamianych wątków, nie jest w ogólności prawidłowa (patrz przykład na marginesie).

Drukiem wytłuszczonym wyróżniono nazwiska laureatów Nagrody Nobla.

Moja *Delta*

Jestem wdzięczna Profesorowi Markowi Kordosowi, który w 2010 roku zaproponował mi pisanie felietonów „Życie na Żywo” do *Delty*.

Z popularyzatorskiej aktywności znalazłam Profesora, i wiedziałam, że propozycja jest poważna. Przyjęłam ją po namyśle, ponieważ w *Delcie* rozumiałam jedynie słowa z niektórych tytułów; treść artykułów umykała mojemu biologicznemu pojmowaniu świata zwykle po pierwszym zdaniu. Jeśli tak – pomyślałam – to jak ja mam pisać, żeby zrozumieli to astronomowie, fizycy, informatycy czy matematycy?

Pierwszy numer *Delty* ukazał się 50 lat temu. Mniej więcej w tym samym czasie rozpoczął się lawinowy rozwój genetyki molekularnej, w której zamyka się całe moje życie naukowe. Nawet w telegraficznym skrócie nie mogę tego okresu w „mojej” nauce opisać. Proszę uruchomić wyobraźnię!

W 1953 roku zdałam maturę w LO im. Narcyzy Żmichowskiej w Warszawie. Pisałam wtedy obowiązujący egzamin z biologii – o ewolucji! Wiedziałam już, że pójdę na studia biologiczne. W tym samym roku w prestiżowym tygodniku *Nature* ukazało się kilka artykułów **Jamesa Watsona**, **Francisa Cricka**, **Maurice’a Wilkinsa** i Rosalind Franklin opisujących strukturę kwasu deoksyrybonukleinowego, DNA. Watson i Crick napisali w swoich powściągliwych tekstach zdanie, iż nie mają wątpliwości, że jest to molekularny zapis cech genetycznych, jakoś transkrybowany do aparatu syntezy białek.

Zaczynała się nowa epoka BIOLOGII MOLEKULARNEJ.

Gdy drukowano 1. numer *Delty* (1974), istniała już ogólna wiedza w tym zakresie. W DNA zawarta jest pełna wiedza o wszystkich cechach danego organizmu. Ten zapis jest realizowany za pośrednictwem innych cząsteczek: RNA, drugiego powszechnie występującego w komórkach kwasu nukleinowego i białek. RNA jest bardzo podobny do DNA, choć nie identyczny. Syntezę RNA (przepisanie informacji) nazwano transkrypcją. Reakcje syntezy, metabolizmu i rozpadu kwasów nukleinowych przebiegają w komórkach pod „dykcją” enzymów. Liczba i rodzaj enzymów w większości procesów były wówczas nieznanymi.

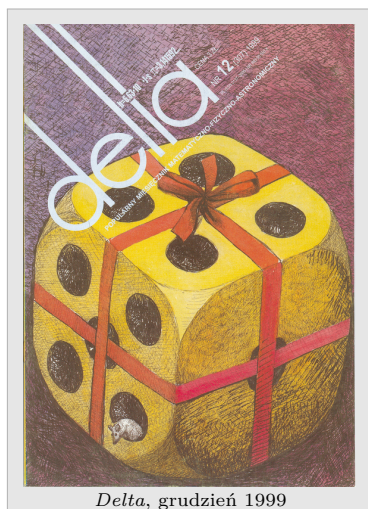
W 1967 roku zgromadzeni na Kongresie Biochemicznym w Warszawie uczeni dowiedzieli się o podstawowych procesach, w których realizowana jest informacja genetyczna zapisana w DNA kończąca się syntezą białek enzymatycznych (ekspresja genetyczna) – opisywanych wspólną nazwą kodu genetycznego. Białka to „wykonawcy” instrukcji z DNA.

To w Warszawie **Ghobind Khorana** ogłosił, że kod genetyczny jest trójkowy, tzn. że określona trójka nukleotydów w DNA (RNA) powoduje włączenie do nowopowstającego białka określonego składnika – aminokwasu. Każdy białkowy aminokwas ma swoją trójkę.

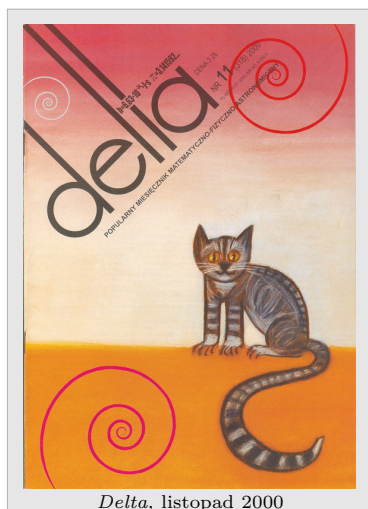
Szczegółów procesów ekspresji genów nie znano. Na szczęście okazało się dość szybko, że ogólnie są takie same w całym świecie ożywionym. Myślano: niezależnie, czy jest to wirus, czy człowiek, kierunek przepływu bioinformacji jest taki sam: DNA – RNA – białko.

Gdy wychodziły pierwsze numery *Delty*, już było jasne, że warunkiem zrozumienia tych procesów jest szczegółowe poznanie kolejności ułożenia nukleotydów w DNA. Jest to niciowa cząsteczka, w której liniowo ułożone są podjednostki (nukleotydy) 4 rodzajów, a ich kolejność (sekwencja) to właśnie owa informacja genetyczna. Cząsteczki bywają ogromne (kilkaset, miliony, a nawet miliardy nukleotydów). Nie istniały wówczas metody oznaczania tej kolejności (sekwencjonowanie), a bez tej wiedzy dalszy postęp był niemożliwy.

W drugiej połowie lat 70. opracowano dwie różne metody sekwencjonowania (**F. Singer** i wsp., 1977 i **W. Gilbert** i wsp., 1977). W dalszych latach metoda Singera, po licznych modyfikacjach, została zautomatyzowana, a automaty (sekwencjatory) są do dziś ulepszane, głównie w kierunku zwiększania czułości



Delta, grudzień 1999



Delta, listopad 2000

oznaczeń, zmniejszania wielkości koniecznej próbki. O sekwencjonowaniu DNA wie już każdy czytelnik kryminałów, wydaje się to prostym zabiegiem możliwym do wykonania w dowolnej policyjnej placówce, także w Polsce. Ogrom pracy, która doprowadziła do takiego postępu, jest nie do wyobrażenia. Pierwsze pełne sekwencje DNA bakterii (*Haemophilus influenzae*, 1,83 mln nukleotydów, i *Mycoplasma genitalium*, 580 tys. nukleotydów) oznaczono w 1995 roku. W roku 1997 zsekwenjonowano 5-milionowy genom najczęściej stosowanej w badaniach laboratoryjnych bakterii *Escherichia coli*. I wreszcie najpilniej i najgoręcej oczekiwana sekwencja ludzkiego DNA (3,2 mld nukleotydów) ogłoszona została kilkakrotnie – z pewnym przybliżeniem w lutym 2001 roku przez konsorcjum HUGO w *Nature*, a przez firmę Celera Genomics w *Science*. Od tego czasu często pojawiają się prace, w których sekwencje ludzkiego DNA oznaczane są coraz pełniej, z użyciem coraz nowocześniejszych i bardziej dokładnych technik.

Postępy w sekwencjonowaniu indywidualnych DNA doprowadziły do zadziwiającego i niezwykle stymulującego odkrycia, że budowa DNA prokariotów (brak jądra w komórce, bakterie) różni się diametralnie od DNA eukariotów (w ich komórkach DNA upakowany jest w jądrze). W dodatku w odcinkach kodujących białka eukariotyczne (potocznie identyfikowanych z genami) odcinki kodujące (eksony) przedzielane są odcinkami niekodującymi (introny), często dłuższymi niż te poprzednie (**R. Roberts, P. Sharp**). Gen eukariotyczny jest podzielony, co ogromnie utrudnia znajdowanie genów kodujących białka w poznanej sekwencji DNA. Jest jeszcze trudniej: w komórkach eukariotycznych odcinki DNA kodujące białka są w sumie krótsze niż cały pozostały DNA. Ponieważ w tamtych latach jedyną rozpoznawalną funkcją DNA było kodowanie białek, to temu pozostałemu DNA w końcu XX wieku nie umiano przypisać funkcji. Nawet pojawiła się nazwa „junk DNA” (DNA ze śmietnika). W DNA ludzkim np. na eksony+introny przypada około 27% DNA, w tym na eksony tylko 1,5%. A po co jest reszta? – słyhać okrzyki do dziś.

Równoległe do sekwencjonowania DNA nastąpił szybki postęp poszukiwań, identyfikacji i preparatyki enzymów metabolizmu kwasów nukleinowych. Zapoczątkowały go odkrycia licznych enzymów syntezy DNA (**A. Kornberg** i wsp., od 1950) ale – co pobudziło aktywność wielu światowych ośrodków – odkrycie w bakteriach tzw. enzymów restrykcyjnych (**W. Arber, H. Smith, D. Nathans**, koniec lat 70.) przecinających nici DNA w obrębie lub sąsiedztwie określonych, krótkich sekwencji (najczęściej 4 lub 6 nukleotydów).

Setki odkrywanych następnie enzymów restrykcyjnych stały się narzędziami do mapowania długich sekwencji DNA, wycinania fragmentów do sekwencjonowania i dalszej obróbki. Chyba stały się też one najliczniejszą grupą komercyjnie dostępnych odczynników biologicznych używanych na całym świecie.

W 1972 roku naukowcy z Zakładu Biochemii Uniwersytetu Stanforda w USA (**P. Berg** i wsp.) opracowali pierwsze doświadczenia, później ochrzczone nazwą inżynierii genetycznej, polegające na wycinaniu określonych sekwencji DNA i wprowadzaniu ich do innych niż dawcy komórek. Inżynieria genetyczna,

fachowo zwana rekombinacją DNA *in vitro*, uwieczniona została nie tylko jako podstawowa technika analizy DNA, ale również tematyka wzbudzająca liczne kontrowersje społeczne, ekologiczne, kończące się nawet demonstracjami ulicznymi.

Z poznawania kolejnych sekwencji DNA, całych genomów tysięcy organizmów wyrosły takie odkrycia, jak: klonowanie zwierząt, modyfikacje embrionów zwierząt (**M. Capecchi, M. Evans, O. Smithes**, N.N. 2007), uzyskiwanie do klonowania określonych sekwencji genomów (metoda PCR, **K. Mullis**, 1983), wiedza o komórkach macierzystych i ich reprogramowanie (**J. Gurdon, S. Yamanaka** N.N. 2012), liczne techniki terapeutyczne w zakresie immunoterapii nowotworów, wiedza o cyklach życiowych wirusów (np. HIV, 1983, **F. Barre-Sinoussi, L. Montagnier**), odkrycie i poznanie prionów (**S. Prusiner**, N.N. 1997). Hitem ostatnich lat jest technika CRISPR/Cas9, w której, korzystając z naturalnych procesów obrony bakterii przed wirusami, **J. Doudna** i **E. Charpentier** stworzyły w 2012 roku system redagowania genów o najwyższej precyzji. Umożliwia on precyzyjne wprowadzanie zmian w sekwencjach DNA, także *in vivo*. System ten, stale ulepszany, daje nadzieje na molekularne terapie w medycynie. Wreszcie Nagroda Nobla w 2023 roku dla **K. Kariko** i **D. Weissmana** przyznana za badania nad wirusowym RNA oraz wynalazek szczepionki RNA.

Od tego czasu często pojawiają się prace, w których sekwencje ludzkiego DNA oznaczane są coraz pełniej, z użyciem coraz nowocześniejszych i bardziej dokładnych technik. Sensacyjne są również oznaczenia sekwencji ludzkiego archaicznego DNA (DNA człowieka neandertalskiego) (**S. Paabo**, N.N. 2022).

Delta wkracza w kolejne 50 lat istnienia; wiedzę o molekularnych podstawach genetyki zaczyna wzbogacać sztuczna inteligencja, porównując w trybie automatycznym podobieństwa i różnice w sekwencjach DNA. Szybciej uzyskuje się ważne, zapewne nieoczekiwane, wyniki niż płynące z dotychczasowych „ręcznych” badań. Zazdroszczę ludziom, którzy tego poznania dostąpią.

Magdalena FIKUS (magda.fikus@gmail.com)

Od roku 2010 redaktor stałego działu „Życie na żywo”.

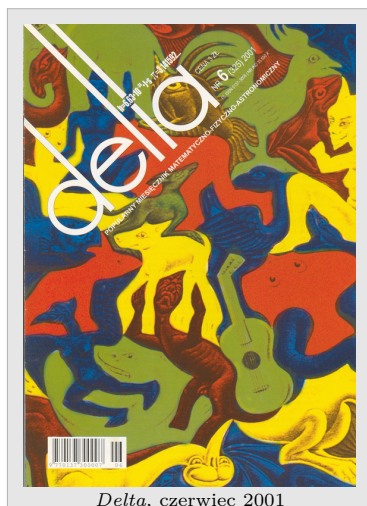
Harriot, Kepler, Hales i inni

Paweł STRZELECKI*

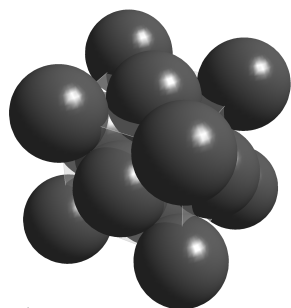
*Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Redaktor *Delta* w latach 1994–1999, od roku 2008 członek Komitetu Redakcyjnego.

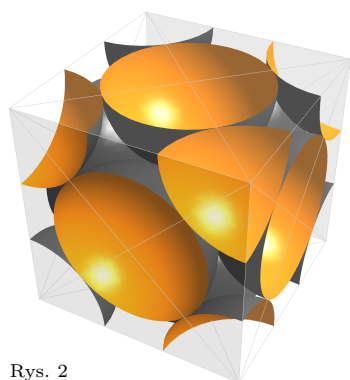
Niniejszy artykuł jest uwspółcześnioną i skróconą wersją rozdziału 6. z mojej książki *Matematyka współczesna dla myślących laików* [2].



Dostępne jest polskie wydanie *Noworocznego podarku* z 2006 roku (Wydawnictwa UW), z obszernym, bardzo ciekawym wstępem Zdzisława Pogody.



Rys. 1



Rys. 2

9 kwietnia 1585 roku z portu w Plymouth wypłynęła flotylla, której trzon tworzyły: *Tygrys*, *Czerwony Lew*, *Rogacz*, *Elżbieta* i *Dorota*. Okręty pozęgłowały w stronę dzisiejszej Wirginii. Organizatorem wyprawy był Sir Walter Raleigh, któremu rok wcześniej królowa Elżbieta I nadała przywilej kolonizowania i rządzenia „... wszelkimi odległymi, pogańskimi i barbarzyńskimi terytoriami, które nie są w posiadaniu chrześcijan ani nie są przez nich zamieszkane”, w zamian za jedną piątą złota i srebra, które można będzie tam wydobywać.

Nie potrafię stwierdzić, na którym z okrętów płynął Thomas Harriot, astronom, etnograf, nawigator i matematyk – dwudziestopięcioletni asystent i zarazem protegowany Raleigha, absolwent St. Mary Hall w Oxfordzie z 1580 roku. Jednak kontakty Harriota z Raleighem, a później jego korespondencja z Johannesem Keplerem, leżą u źródeł sławnego problemu matematycznego o długiej, nieoczekiwanej historii.

Otóż Raleigh zlecił Harriotowi opracowanie optymalnego sposobu składowania kul armatnich na pokładzie. Harriot przygotował tabelę możliwych ułożeń, podając liczbę kul w stosie jako funkcję jego wysokości. Zaczął też badać wzory na sumę kwadratów kolejnych liczb naturalnych, a także rozważać, jaka jest struktura materii. Wiadomo o tym dzięki zachowanej korespondencji Harriota z Keplerem, z lat 1606–1607, w której obaj dyskutowali m.in., dlaczego promień świetlny, padając na powierzchnię przezroczystego ośrodka, zostaje częściowo odbity, a częściowo załamany. Teza Harriota (wyrażona współczesnym językiem) brzmiała: materia ma strukturę atomistyczną i dlatego gładkie na pozór powierzchnie nie są wcale jednorodne. Kepler z czasem przyjął ten pogląd.

W 1611 roku w książeczce *Noworoczny podarek albo o sześciokątnych płatkach śniegu* Kepler stwierdził, że symetrię śnieżynek powoduje ziarnista struktura materii, pisał o sposobach gęstego, ścisłego układania kul, a także wyraził przypuszczenie, nazwane później *hipotezą Keplera*, głoszące, że lepiej układać identycznych kul, niż Harriot proponował Raleighowi, nie można. Na dowód przyszło matematyce czekać blisko 400 lat...

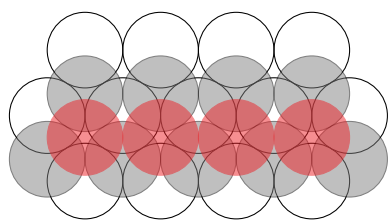
Spójrzmy na ułożenie identycznych kul w piramidę, nazywane *upakowaniem sześciennym, ściennie centrowanym* (krótka nazwa to *upakowanie fcc*, od angielskiego *face centered cubic lattice*). Środki kul jednej warstwy tworzą siatkę kwadratów; każda kula dotyka czterech innych. Kule z kolejnej warstwy leżą w zagłębieniach między kulami poprzedniej. Gdy weźmiemy $5 + 4 + 5 = 14$ kul z trzech warstw – dwie piątki od góry i na spodzie, każda ułożona w krzyż, oraz środkową czwórkę ułożoną w kwadrat, to środki ośmiu z tych kul będą w wierzchołkach pewnego sześcianu, a środki pozostałych sześciu – w środkach jego ścian (rys. 1). Prowadząc cięcia wzdłuż ścian sześcianu, otrzymalibyśmy klocek, a w nim sześć połówek kul i osiem fragmentów kul w wierzchołkach (rys. 2). Takimi klockami można wypełniać całą przestrzeń. Hipoteza Keplera wyrażona potocznie brzmi: gęściej układać kul nie można; próba użycia nieregularnych ułożeń nic nie da.

Obliczmy gęstość upakowania *fcc*. Przekątna sześcianu jest czterokrotnie dłuższa od promienia kuli równego 1. Krawędź takiego sześcianu to $4/\sqrt{2} = 2\sqrt{2}$, a jego objętość to $(2\sqrt{2})^3 = 16\sqrt{2}$. Objętość kuli to $\frac{4}{3}\pi$, więc łączna objętość fragmentów kul zawartych w sześcianie wynosi

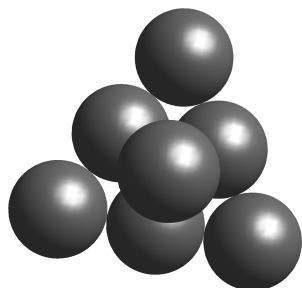
$$\left(8 \cdot \frac{1}{8} + 6 \cdot \frac{1}{2}\right) \frac{4}{3}\pi = \frac{16\pi}{3}.$$

Stosunek obu liczb to $\pi/(3\sqrt{2}) = \pi/\sqrt{18}$ i taka też jest szukana gęstość.

W ogólności *gęstość upakowania to granica górna* stosunku objętości tych (części) kul danego upakowania, które są zawarte w kuli $B(0, R)$, do objętości tej kuli, tzn. do $4\pi R^3/3$, przy $R \rightarrow +\infty$. Taka definicja działa również dla upakowań niesymetrycznych, gdzie w różnych obszarach przestrzeni kule są upakowane na przemian rzadziej i gęściej.



Rys. 3. Białe kule należą do warstwy typu a , szare do warstwy b , zaś kolorowe do c



Rys. 4

Upakowanie fcc można uzyskać, układając kule (na pozór) inaczej. Weźmy płaską warstwę gęsto ściśniętych kul, których środki są wierzchołkami siatki trójkątów równobocznych. Z góry zobaczymy koła wpisane w identyczne sześciokąty foremne wypełniające płaszczyznę. Zauważmy, że każdą taką warstwę kul ułożonych nad jednakowymi sześciokątami foremnymi można nałożyć na poprzednią na *dwa różne sposoby*. Nie da się wypełnić wszystkich zagłębień naraz; można wypełnić tylko co drugie. Patrząc od góry, możemy np. widzieć kolejne warstwy, powtarzające się w rytmie a, b, c, a, b, c, \dots lub w rytmie a, b, a, b, \dots (rys. 3). Można sprawdzić, że rytm walczyka, a, b, c, a, b, c, \dots daje opisane wyżej upakowanie fcc. Wykonajmy w tym celu następujący eksperyment myślowy: wyobraźmy sobie szóstkę tych kul, z jednej warstwy, ułożonych w trójkąt. Na wierzchu tego trójkąta – symetrycznie, w środkowym zagłębieniu – połóżmy siódmą kulę (rys. 4). Wyobraźmy sobie teraz, że cała siódemka została skleciona w sztywną konstrukcję. Biorąc dwie kopie takiej konstrukcji i stykając je podstawami (wierzchołki trójkątów powinny być skierowane w przeciwnie strony, a kule jednej warstwy trzeba wpasować we wgłębienia drugiej), uzyskamy 14 kul o środkach w wierzchołkach i środkach ścian sześcianu. Obie siódemki kul uważny Czytelnik może wskazać na rysunku na poprzedniej stronie.

Przy układzie warstw a, b, a, b, \dots środki kul tworzą tzw. *kratę sześciokątną gęstą*, a ułożenie nazywa się *upakowaniem sześciokątnym gęstym*. Gęstość tego upakowania jest równa gęstości upakowania fcc. Co więcej, *każdy* nieskończony („w obie strony”) ciąg liter a, b, c , w którym każde dwie sąsiednie litery są różne, odpowiada innemu układowi kolejnych warstw kul. Różnych ułożeń o tej samej gęstości jest więc continuum, a wśród nich – cała masa nieokresowych, bez powtarzalnej symetrii!

Przy dodatkowym wymaganiu, by środki kul tworzyły regularną kratę, hipoteza Keplera staje się łatwiejsza. W takiej wersji udowodnił ją Carl Gauss w 1831 roku. Jego dowód opiera się na tym, że założenie symetrii upakowania pozwala przechodzić od *lokalnych* wniosków o ułożeniu kul do wniosków globalnych: w optymalnym upakowaniu jakieś dwie kule muszą się dotykać (inaczej moglibyśmy nieco zwiększyć wszystkie kule, tzn. gęstość upakowania); później można wnioskować, że istnieją całe rzędy i warstwy dotykających się wzajemnie kul.

W uproszczonej, płaskiej wersji problemu odpowiednik hipotezy Keplera udowodnił w 1890 roku norweski matematyk Axel Thue. Jego dowód jest elementarny; patrz np. przeglądowy artykuł Halesa [1], a także rozdział 6 w [2]. Dziesięć lat po dowodzie Thuego, w roku 1900, David Hilbert umieścił hipotezę Keplera na swej słynnej liście problemów (jako część problemu XVIII).

Zanim wspomnę o samym dowodzie hipotezy Keplera, chcę napisać, co stanowi o jej trudności. Pierwszy powód jest banalny: geometria przestrzenna jest trudniejsza od płaskiej. Powód drugi to brak symetrii; wydaje się, że ten brak może tylko zmniejszyć gęstość upakowania, ale to jeszcze nie dowód, prawda? Każdy, kto dopychał walizkę kolanem, wie, że sukces wiąże się z zaburzeniem symetrii starannie ułożonych ubrań.

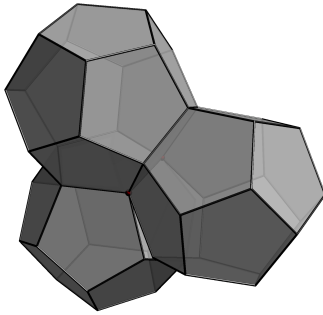
Najważniejszy jest trzeci powód. Otóż w płaskiej wersji hipotezy Keplera, rozwiązanej przez Thuego, jest tak, że to, co optymalne lokalnie, można po prostu zrealizować globalnie. W przestrzeni już tak nie jest; żeby to wyjaśnić, wytłumaczmy najpierw, co to są *komórki Woronoja*. Otóż komórka Woronoja danej kuli w upakowaniu to zbiór tych punktów przestrzeni, które są bliżej środka tej kuli niż środków innych kul z upakowania. Zbiór punktów równoodległych od ustalonych punktów A, B w przestrzeni to płaszczyzna; dlatego komórki Woronoja każdego upakowania kul są wielościanami wypukłymi.

Jeśli upakowanie ma mieć dużą gęstość, to jego komórki Woronoja powinny być małe. Jak jest na płaszczyźnie dla upakowań kół o promieniu 1? Tam komórki Woronoja to wielokąty wypukłe. Okazuje się, że komórką Woronoja



Georgij Fieodosjewicz Woronoi żył w latach 1868–1908; od 1894 roku wykładał na Uniwersytecie Warszawskim. Był promotorem Wacława Sierpińskiego i wg. danych *Mathematics Genealogy Project* ma ponad 4 tys. „potomków” (osób, które łączy z nim skończona liczba więzów doktorant–promotor). Jedną czwartą z nich to potomkowie Antoniego Zygmunta, który po wojnie pracował na Uniwersytecie w Chicago.

o najmniejszym możliwym polu jest sześciokąt foremny opisany na kole. Taką komórką można parkietować płaszczyznę i na tym opiera się dowód Thuego; gęstość optymalnego upakowania kół jest *taka sama* jak najlepszy stopień wypełnienia pojedynczej komórki.



Rys. 5

W przestrzeni jest inaczej. Okazuje się mianowicie, że kula wpisana w dwunastościan foremny zajmuje nieco ponad 0,754 jego objętości, a to *więcej* niż $\pi/\sqrt{18} = 0,74048\dots$ (Dwunastościan foremny to komórka Voronoja, którą jedna kula wypełnia *najbardziej jak się da*, ale to osobna opowieść). Kłopot w tym, że dwunastościanami foremnymi *nie da się* wypełniać przestrzeni (rys. 5). Niemniej zasadne jest pytanie: a może jeśli w jakimś upakowaniu wciśniemy, kosztem lekkiego luzu tu i ówdzie, dużo miejsc, gdzie komórki Voronoja kul będą dwunastościanami foremnymi, to okaże się, że globalnie zyskamy przewagę nad upakowaniem fcc? Cała trudność polega na tym, żeby wykluczyć wszystkie takie sytuacje, godząc się jednak z tym, że upakowanie może być pozbawione symetrii.

W 1953 roku węgierski matematyk Fejes Tóth wskazał w istocie pewną drogę do dowodu hipotezy Keplera – udowodnił mianowicie, że wynika ona z pewnego układu ważonych nierówności, jakie mają spełniać objętości komórek Voronoja danego upakowania. Przez kolejnych 40 lat z hipotezą Keplera nie stało się nic; w 1976 roku John Milnor pisał, że to skandal, że wciąż jest nieudowodniona.

W latach 1993–1998 Thomas Hales, wspomagany w części prac przez doktoranta Samuela Fergusona, przeprowadził dowód hipotezy Keplera, doskonaląc sugestię Tótha. Esencja pomysłu jest następująca: zamiast minimalizować objętości komórek Voronoja, trzeba minimalizować *nieco inną funkcję*, dodając do objętości każdej komórki poprawkę – taką aby suma poprawek dla wszystkich kul zawartych w wielkiej kuli $B(0, R)$ była znikoma w porównaniu z R^3 . Sedno dowodu to konstrukcja takich poprawek. Margines swobody konstrukcji jest spory, ale liczba niezbędnych obliczeń ogromna.

Cały dowód Halesa zajmuje kilka prac o łącznej długości mniej więcej 250 stron. Do tego dochodzą obliczenia wykonywane za pomocą komputera; ich dokumentacja zajmuje kilka gigabajtów danych. Metody użyte w dowodzie obejmują, prócz geometrii i topologii, teorię optymalizacji (liniowej i nieliniowej), teorię grafów i kombinatorykę.

W 2005 roku 120-stronicowa praca Halesa omawiająca bardzo szczegółowo ramy dowodu i jego strategię została opublikowana w *Annals of Mathematics*, które wielu matematyków uznaje za najbardziej prestiżowe na całym świecie czasopismo w naszej branży. Sprawdzał ją przez kilka lat dwunastoosobowy zespół recenzentów, który nie znalazł żadnych błędów, ale oczywiście nie mógł potwierdzić poprawności wszystkich użytych programów komputerowych. Redakcja *Annals* umieściła na okładce tomu odpowiednie zastrzeżenie. Uważam, że fakt ostatecznej publikacji pracy Halesa akurat w *Annals*, nie w czasopiśmie mniejszej rangi, należy uznać za znak czasu: dowody wspierane komputerowo są w matematyce akceptowane i akceptowalne. (Jest więcej takich znaków, nie tylko historie hipotezy Keplera i zagadnienia czterech barw).

W 2003 roku Hales rozpoczął realizację długofalowego projektu **FPK**; akronim pochodzi od angielskiego *Formal Proof of Kepler*. Celem było napisanie takiego

dowodu hipotezy Keplera, który mógłby zostać czysto formalnie sprawdzony z pomocą któregoś z komputerowych systemów wspomagania dowodzenia. (Przykładem takiego systemu jest polski Mizar; inne to Coq, Isabelle czy HOL Light; w systemie Coq podano w 2004 roku sformalizowany dowód zagadnienia czterech barw. Takich narzędzi używa się skądinąd nie tylko do formalnej analizy rozumowań, ale i do badania poprawności programów komputerowych stosowanych np. w przemyśle, energetyce, lotnictwie, organizacji ruchu kolejowego). Sukces **FPK** anonsowano latem 2014 roku; praca Halesa i dwudziestu kilku współautorów (jest wśród nich polski informatyk Cezary Kaliszyk) ukazała się w 2017 roku w *Forum Mathematicum*.

Czy to koniec historii? Nie. Po pierwsze jest otwarte i bardzo ważne dziś pytanie, czy i jak możliwości dowodów wspieranych komputerowo zmieniają – może za sprawą osiągnięć sztucznej inteligencji? – całą matematykę. Nie pokuszę się o odpowiedź. Po drugie bardzo naturalne dla matematyka jest inne pytanie: skoro wiadomo, jakie są optymalne, najgęstsze upakowania kul w wymiarach 2 i 3, to co się dzieje dla $n > 3$? Dla wymiarów $n \neq 1, 2, 3, 8, 24$ optymalnych upakowań nie znamy. Natomiast w \mathbb{R}^8 gęstość optymalnego upakowania kul wynosi $\pi^4/384$ (a więc około 0,25), a w \mathbb{R}^{24} jest równa $\pi^{12}/12!$. Dowody są klasyczne, bez wsparcia komputerowego. Za autorstwo pierwszego z tych wyników i udział w drugim Maryna Wiazowska 5 lipca 2022 roku w Helsinkach odebrała medal Fieldsa. Laudację o wynikach Wiazowskiej wygłosił jej współautor, Henry Cohn z MIT i Microsoft Research; druga afiliacja wiąże się z tym, że upakowania kul w wysokich wymiarach mają związek z konstrukcją tzw. kodów korekcji błędów, ale to znów temat na inną opowieść.

Bibliografia:

- [1] Thomas Hales. Cannonballs and honeycombs. *Notices Amer. Math. Soc.* **47**, no. 4 (2000).
- [2] Paweł Strzelecki. *Matematyka współczesna dla myślących laików*. Wydawnictwa Uniwersytetu Warszawskiego, Warszawa 2011.

Tym razem się nie uda, czyli epilog sagi kryptologicznej w odcinkach. W tej części: o tajnych głosowaniach elektronicznych.

Tomasz KAZANA*

Tajne wybory to sól demokracji.

Najwyższy rangą dokument obowiązujący w Polsce – Konstytucja RP – wspomina o nich kilkakrotnie (art. 96, 97, 127, 169), podając dość konkretne wymagania, takie jak powszechność, równość, bezpośredniość czy właśnie tajność. W tym artykule spróbujemy zastanowić się, czy w świecie cyfrowym możliwe jest uzyskanie tego ostatniego, a więc, czy realne jest przeprowadzenie tajnych wyborów elektronicznych.

Model analogowy

Zanim jednak wejdziemy w świat elektroniki, na chwilę jeszcze zostaniemy na ziemi i spróbujemy rzucić okiem na klasyczne wybory li tylko od strony ich bezpieczeństwa proceduralnego. Już w tym przypadku zapewnienie tajności wymaga precyzyjnej analizy. Oczywiście jest jasne, że główna idea polega na tym, że po prostu wchodzimy do kabiny pojedynczo i tam samodzielnie, bezpiecznie głosujemy. Ale czy na pewno nie ma tam ukrytych kamer? A czy długopis nie zostawia jakiegoś identyfikującego nas śladu?

Dalsze pytania można tylko mnożyć: Czy na pewno powinniśmy pozwalać wchodzić do kabin w dwie osoby, nawet jeśli są małżeństwem? A może jedna z nich jest pod presją drugiej? Czy nie jest problemem, gdy ktoś zrobi zdjęcie swojego głosu? A może chce ten głos sprzedać i udowodnić kupującemu, jak zagłosował?



Delta, grudzień 2003

Ktoś może powiedzieć, że to brzmi paranoicznie, gdyż oczywiście niby wiadomo, że istnieją takie zagrożenia, ale skomplikowanie ich wykonania oraz niewielki wpływ pojedynczego incydentu na ostateczny wynik pozwala założyć, że wszystko na koniec dnia i tak jest w porządku. Z opiniami tego typu, dotyczącymi wyborów tradycyjnych, akurat w tym miejscu polemizować nie będziemy. Zwróćmy jednak uwagę na ważną kwestię: nawet jeśli wierzymy, że w świecie niecyfrowym pojedyncze incydenty nas nie ruszają, to musimy sobie uświadomić, że w świecie elektroniki tak to nie działa. Tutaj wszelkie podatności i niedopatrzania da się zwykle łatwo wykorzystać – i to najczęściej stosunkowo tanio – na skalę masową. Przecież gdy raz stworzymy wirusa, który potrafi odtajnić czyjś głos, to zadziała on nie na jednym, ale na tysiącach komputerów stosujących podobne oprogramowanie, którego jakąś lukę wykorzystujemy.

Model cyfrowy

Nie ukrywamy, że drobiazgowość i plastyczność sekcji wyżej miały na celu zatruć głowę Czytelnika jedną konkretną myślą: otóż w przypadku potencjalnych głosowań elektronicznych skrupulatność granicząca z paranoją w podejściu do bezpieczeństwa nie jest zwyczajową manierą matematyków, ale potrzebą totalnie bezwzględną i musi być postawiona absolutnie na pierwszym miejscu.

W szczególności już na starcie powinniśmy odrzucać jakiegokolwiek pomysły związane z obecnością tzw. zaufanego serwera. To znaczy każdy model, w którym

istnieje jakiś jeden wyróżniony Ważny Komputer, do którego np. dostarczane są cyfrowo podpisane głosy wszystkich wyborców, następnie ten komputer zlicza wynik, po czym skutecznie usuwa informacje o głosujących (*utajnia* wybory), jest modelem kryptologicznie **nie do przyjęcia**. Zauważmy bowiem, że przy takim podejściu całe zaufanie spoczywa tak naprawdę na administratorze tego komputera. Musimy przecież uwierzyć, że ów administrator niczego sobie nie zapamięta z rzeczy, które ma usunąć, niczego nie usunie z rzeczy, które mają być policzone itp. Nie o taki system nam chodzi.

Gdy realnie myślimy o bezpiecznym tajnym głosowaniu elektronicznym, zastanawiamy się, czy da się je tak zaprojektować, aby całe bezpieczeństwo było oparte o prawa matematyki, które działają *pod spodem*, a nie o zaufanie do nawet najbardziej nieskalanego administratora, ważnego sądu czy innego podobnego organu. W kryptologii bowiem zawsze paranoicznie (a może zdroworozsądkowo) zakładamy, że w systemie są obecni (a najlepiej: przejęli pełną kontrolę nad całą infrastrukturą techniczną) złoczyńcy, którzy nieustannie sposobią się do sfalszowania wyborów bądź ich odtajnienia.

Czytelnicy *Delta* (a szczególnie Czytelnicy serii „A jednak się da” ukazującej się w latach 2018–2019) orientują się jednak, że nie takie cuda oferuje współczesna kryptologia. Jakie są więc najlepsze protokoły kryptologiczne do tajnych głosowań elektronicznych?

W 2020 roku poważni i poważani kryptolodzy wystosowali do władz USA *Joint Internet Voting Avoidance Appeal*, apel o niewprowadzanie wyborów elektronicznych.

Warta odnotowania jest ciekawa dualność: w wyborach tradycyjnych łatwiej zapewnić tajność, bardziej podejrzane jest liczenie głosów. W wyborach elektronicznych jest na odwrót: uzyskanie poprawnego wyniku jest zadaniem znacznie łatwiejszym niż zapewnienie tajności.

W rozumowaniu o Jasiu czynimy założenie, że system wie, kto aktualnie głosuje. Zakłada się bowiem, że nie istnieje żadna realna forma bezwzględnie bezpiecznej anonimowej komunikacji z systemem do głosowania. Czytelnik Nie Dający Za Wygraną może spróbować zgłębić istniejące próby realizacji anonimowej komunikacji np. w Δ_{17}^9 i zastanowić się nad ich wadami w kontekście głosowania elektronicznego.

Uzyskanie „minimalnego zaufania” dla matematyka jest trudne i nadzwyczaj efektowne. Ale niestety: specjaliści i prawnicy oceniają takie założenia jako wciąż niewystarczające.

Aby nieco obniżyć pułap lotu, warto zapoznać się z niektórymi artykułami z serii „A jednak się da”: w stosownych miejscach podajemy odnośniki.

Warto sobie uświadomić, że wybór parametrów (N, K) jest decyzją w zasadzie polityczną. Również można się zastanawiać, czy protokół wykonywać raz globalnie, czy np. niezależnie w każdym okręgu.

„Opublikować” oznacza tu pewne dodatkowe założenie: musimy założyć, że istnieje zaufane publiczne cyfrowe miejsce, do którego wszyscy mają łatwy dostęp do odczytu oraz gdzie można zapisywać dane, które nie mogą być usunięte w sposób niezauważony. To założenie nie uchodzi za bardzo podejrzane, ale też jego bezpieczna realizacja stanowi pewne wyzwanie.

Już tytuł tego artykułu przebił balon napięcia, więc nie będziemy udawać: świat kryptologów nie wierzy w wybory elektroniczne. I to nawet nie na zasadzie, że istnieją jakieś trudności, których do tej pory nie udało się pokonać, ale wciąż jest otwarte pytanie: czy się uda w przyszłości? Nie. Umieemy udowodnić, że nie da się lepiej, niż tak jak umiemy. A tak jak umiemy – choć i tak ma prawo zachwycać – większości fachowców nie zadowala. Dalsza część tego artykułu będzie więc próbą wyeksponowania najważniejszych cyfrowych ograniczeń tej dziedziny oraz pokazania, co jednak się da.

Królestwo za tajność

Okazuje się, że największym wyzwaniem jest właśnie tajność (a nie np. poprawność obliczonego wyniku) wyborów. Da się bowiem – i to dość zaskakująco prostym rozumowaniem – pokazać, że jeśli całkowicie nie ufamy systemowi – nieważne jak zaawansowany matematycznie by on nie był – to po prostu nie da się jej w pełni zapewnić. Dlaczego?

Rozważmy Jasia, który o wszystkim zapomniał i teraz szuka laptopa, aby zdążyć zagłosować tuż przed końcem. Jak to z Jasiem, możliwości będą tylko dwie: albo (a) Jaś nie zdąży, (b) Jaś będzie ostatnią osobą głosującą w wyborach. Zauważmy jednak, że poprawny system będzie w stanie podać wyniki wyborów w obu przypadkach. W systemie muszą więc być zawarte wszystkie informacje, które na to pozwalają. A to oznacza, że ktoś mający ciągle dostęp do wszystkich danych systemu będzie w stanie obliczyć obie te wartości: przed i po wizycie Jasia. Ale przecież z tych dwu danych da się odtworzyć głos Jasia!

Czy więc wszystko stracone? Czy cokolwiek możemy zrobić?

Możemy, o ile nieco poluzujemy założenia. Przy czym „nieco” nie jest tu bynajmniej eufemizmem. Otóż da się zapewnić tajność, gdy zaufanie do systemu jest z naszej strony naprawdę minimalne, a w zasadzie dowolnie małe, byle niezerowe. Bardziej konkretnie: założymy, że system działa w modelu rozproszonym i tworzy go nie jeden, ale np. tysiąc czy nawet milion komputerów. Wówczas, jeśli tylko ufamy w uczciwość choć jednego z nich (np. obsługiwanego przez nas samych albo przez przedstawicieli naszego obozu politycznego), to już tajność naszego głosu może zostać zapewniona!

Dokładnie na takim założeniu opiera się system głosowania elektronicznego, który zaprojektowali Ronald Cramer, Matt Franklin, Berry Schoenmakers i Moti Yung. Spróbujemy naszkicować ten dość skomplikowany protokół z lotu ptaka (i to niestety raczej takiego o wysokim pułapie lotu).

Protokół Cramer–Franklin–Schoenmakers–Yung, 1996

Zanim system zostanie uruchomiony, należy wybrać pewne parametry: skupimy się na dwóch najważniejszych: N i K , które spełniają nierówności $1 \leq K \leq N$.

System będzie składać się z N komunikujących się ze sobą komputerów (oznaczymy je jako s_1, \dots, s_N), co do których zakładamy, że uczciwych jest co najmniej K spośród nich. Przypadek $K = 1$ wydaje się więc najatrakcyjniejszy, ale jest tu też druga strona medalu – aby system *zawiesić* (nie dopuścić do wiarygodnego opublikowania wyników wyborów), wystarczy zmowa bądź awaria K komputerów. Tak więc zbyt małe K ma też swoje istotne wady.

Dalej wszystko będzie dość skomplikowane, ale jednak się uda:

- Jaś, aby oddać swój głos X , musi:
 - opublikować zobowiązanie $Commit(X)$ do swojego głosu (zob. Δ_{18}^{11}); zauważmy, że ten pozornie podejrzany krok bynajmniej nie zdradza wartości głosu (opublikowane zobowiązanie nigdy nie zostanie otwarte) – jest natomiast potrzebny później przy weryfikacji wyników;
 - opublikować dowód poprawności $Proof(X)$ swojego głosu (to znaczy udowodnić protokołem zero-knowledge (zob. Δ_{19}^1), że jego głos jest oddany w poprawnym formacie, np. ma wartość „NIE” lub „TAK”);
 - dokonać podziału swojego głosu protokołem Shamira (zob. Δ_{11}^2) na kawałki X_1, \dots, X_N ;
 - wysłać kawałek X_i do komputera s_i .

Dodajmy tu jednak łyżkę dziegciu. Wszystkie dowody, o których tu mowa, są *niepodrabialne* przy jeszcze jednym (powszechnie uważanym za rozsądne) dodatkowym założeniu – mianowicie: wierzymy, że logarytm dyskretny jest problemem trudnym obliczeniowo. Co ciekawe i ważne, da się pokazać, że nie da się skonstruować żadnego bezpiecznego protokołu głosowania elektronicznego bez tego typu obliczeniowych założeń. Nie jest to też coś zaskakującego – podobnie sprawa ma się w przypadku chociażby podpisu elektronicznego (zob. np. Δ_{18}^{10}).

Dodatkowo część protokołów (w tym prezentowany w tym artykule) jest skonstruowana tak, że aby handlować głosami, nawet nie trzeba robić nic specjalnego (typu instalowanie kamerek) w trakcie głosowania. Nawet już po głosowaniu da się stworzyć cyfrowy dowód (w naszym przypadku – wystarczy otwarcie zobowiązania $Commit(X)$), który można po prostu pokazać, aby udowodnić kupującemu (albo zastraszającemu), że głosowało się zgodnie z jego poleceniem. Na szczęście istnieją protokoły pozbawione akurat tej wady, np. protokół Canetti-Gennaro z 1997 roku.

Niektórzy postulują, aby nie używać do głosowań własnych komputerów, tylko dedykowanych superbezpiecznych urządzeń. Oczywiście naturalne jest pytanie, kto te urządzenia by wytwarzał i jak dowodził, że faktycznie działają poprawnie.



Delta, maj 2004

- gdy wszystkie głosy zostaną oddane, wówczas wszystkie komputery tworzące system do głosowania:
 - pracując razem (według protokołu MPC podobnego do opisanego w Δ_{19}^8), są w stanie obliczyć wynik wyborów, nie ucząc się przy tym (o ile nie więcej niż $(N - K)$ spośród nich jest nieuczciwych) niczego na temat wartości jakiegokolwiek oddanego głosu.
 - publikują nie tylko ostateczny wynik wyborów, ale i dowód (protokół jest tak zaprojektowany, że jest to możliwe), że jest on poprawny. Każdy, kto ma dostęp do tego dowodu oraz kompletu nieotwartych zobowiązań (takich jak $Commit(X)$ Jasia), jest w stanie upewnić się, czy ostateczny wynik jest poprawny. Jest to niezwykle silna cecha omawianego protokołu. Oznacza ona bowiem, że nawet gdy liczba oszustów przekracza pułap $(N - K)$, to wyniku wyborów i tak **nie da się sfałszować (!)**, a tylko wywołać zamieszanie i konsternację – wszyscy przecież i tak się zorientują, że ostateczny wynik nie jest zgodny z oddanymi głosami.

Epilog epilogu

W tym artykule chcieliśmy pokazać smak wyzwania i skalę problemów przy projektowaniu protokołów do głosowania. Jak na razie, tajność wyborów została wyeksponowana jako główny i w zasadzie dowodliwie nierozwiązywalny problem. Piszemy „w zasadzie”, bo wszystko zależy od subiektywnej opinii co do tego, na jakie założenia się godzimy. Choćby w rozważanym protokole – można dyskutować, które parametry (N, K) dają wiarygodną tajność oraz pozwalają wierzyć, że protokół wyborczy kiedykolwiek zakończy się bez zawieszenia. Wybór $K = 1$ oraz $N \approx 30\,000\,000$ (wówczas każdy uprawniony głosujący stałby się jednym z komputerów s_i , co niewątpliwie pozwala założyć, że każdy ufa przynajmniej jednej części systemu – mianowicie, samemu sobie) może i teoretycznie daje pełną tajność, ale przecież trudno uwierzyć, że taki protokół udałoby się przeprowadzić bez zakłóceń ze strony chociaż jednego komputera.

Co więcej, okazuje się, że – oprócz powyższego – istnieją jeszcze inne, trochę pozamatematyczne, ale niestety równie fundamentalne problemy, o których tutaj ledwie wspomniemy:

Po pierwsze – gigantycznym problemem jest potencjalny handel głosami (albo zastraszanie i wymuszanie – matematycznie to podobne zagadnienia). W świecie cyfrowym wydaje się on niestety znacznie łatwiejszy niż w świecie tradycyjnym. Głosujący może przecież zainstalować sobie kamerkę w domu i pozwolić komuś z zewnątrz na bieżąco upewniać się, że głosuje on zgodnie z jego wolą. Jest to na pewno wygodniejsze i bezpieczniejsze niż wszelkie podobne działania w rzeczywistym lokalu wyborczym.

Po drugie – pamiętajmy, że w świecie cyfrowym człowiek nie wykonuje obliczeń w głowie, ale korzysta z osobistego komputera, który pamięta za niego tajne dane, dokonuje wrażliwych obliczeń itp. Ten element jest zawsze podkreślany jako potencjalne ważne źródło problemów – nikt przecież tak do końca nie wie, jakie wirusy czy programy szpiegujące ma w swoim komputerze zainstalowane lub też nawet fabrycznie wbudowane. Oznacza to w szczególności, że pułap $(N - K)$ dotyczy nie tylko świadomych oszustów, ale też nieświadomych właścicieli podejrzanej sprzętu.

Melancholijne podsumowanie

Saga „A jednak się da” miała na celu przekonanie Czytelnika, że wiele z tego, co wydaje się cyfrowo niemożliwe do zrealizowania, nauka jednak zrealizowała i stworzyła protokoły, które dla zwykłego śmiertelnika są w zasadzie nieodróżnialne od magii. Ale niestety – pychą byłoby wierzyć (może trochę jak Hilbert?), że udawać się tak będzie zawsze. . .

W przypadku głosowań tajnych, mimo że te najlepsze protokoły aż jeżą się od zastosowanych perełek kryptologii, po prostu w pełni nie udało się. Musimy więc albo zrezygnować z głosowania cyfrowego, albo pogodzić się z którymś ze zgniłych kompromisów dotyczących definicji bezpieczeństwa takiego głosowania.

Wir müssen nicht alles wissen.

*Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

Redaktorka autorskiego działu *Deltoid*, który ukazywał się w *Delcie* w latach 2009–2019.

Jak na wyświetlaczu Joanna JASZUŃSKA*

W artykule *Kto da mniej?* z Δ_{17}^{10} przedstawiłam następującą zagadkę:

Mamy 10 worków z monetami. W dokładnie jednym z nich wszystkie monety są fałszywe, w pozostałych workach wszystkie są prawdziwe. Prawdziwa moneta waży 10 gramów, a fałszywa 11. Ile ważeń na wadze elektronicznej trzeba wykonać, aby wykryć worek z fałszywymi monetami?

Wystarczy, oczywiście, dziesięć ważeń: po jednej monecie z każdego worka. A nawet o jedno mniej, bo jeśli pierwszych dziewięć monet jest prawdziwych, dziesiąta musi być fałszywa. Można też sprytniej: wziąć na przykład po jednej monecie z pięciu worków i zobaczyć, czy razem ważą one 50 gramów, czy 51. Takie pomysły pozwalają zawęzić grupę podejrzanych worków i ograniczyć się do czterech ważeń (zachęcam do sprawdzenia).

Ciągle jednak nie jest to rozwiązanie optymalne, wystarczy bowiem jedno ważenie! Niech na wadze leżą:

$$\begin{aligned} & 1 \text{ moneta z worka nr } 1 + \\ & + 2 \text{ monety z worka nr } 2 + \\ & + 3 \text{ monety z worka nr } 3 + \\ & \quad + \dots + \\ & + 10 \text{ monet z worka nr } 10. \end{aligned}$$

Tych 55 monet, gdyby były prawdziwe, ważyłoby $55 \cdot 10 = 550$ gramów. Waga pokazuje wynik o n gramów większy wtedy i tylko wtedy, gdy leży na niej n fałszyfikatów. Ostatnia cyfra wyświetlacza wskazuje więc numer szukanego worka. □

Rozważmy teraz ogólniejszą wersję tej samej zagadki:

Mamy 10 worków z monetami. W niektórych workach wszystkie monety są fałszywe, w pozostałych wszystkie są prawdziwe. Prawdziwa moneta waży 10 gramów, a fałszywa 11. Ile ważeń na wadze elektronicznej trzeba wykonać, aby wykryć worki z fałszywymi monetami?

Niestety poprzednia metoda tym razem nie zadziała. Jeśli naszych 55 monet teraz waży np. o 7 gramów za dużo, to fałszyfikaty mogą równie dobrze być w workach o numerach 2 i 5, jak i w workach 1, 2 i 4 lub w jeszcze innych.

Nadal jednak wystarczy jedno ważenie. Niech tym razem na wadze leżą:

$$\begin{aligned} & 1 \text{ moneta z worka nr } 1 + \\ & + 10 \text{ monet z worka nr } 2 + \\ & + 100 \text{ monet z worka nr } 3 + \\ & \quad + \dots + \\ & + 10^9 \text{ monet z worka nr } 10 + \\ & + 1 \text{ odważniczek 1-gramowy.} \end{aligned}$$

Tych $1\ 111\ 111\ 111$ monet, gdyby były prawdziwe, ważyłoby $11\ 111\ 111\ 110$ gramów, a z odważniczką waga wyświetlałaby wynik $11\ 111\ 111\ 111$ gramów.

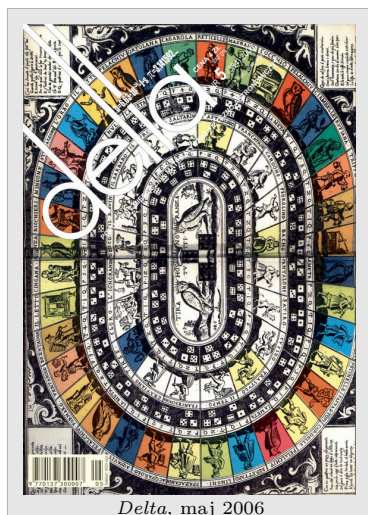
Jeśli przykładowo w worku numer 3 są fałszywe monety, to wynik jest o 100 g za duży: $11\ 111\ 111\ 211$. Jeśli ponadto np. w worku 7 też są fałszywe monety, mamy jeszcze o 10^6 g więcej: $11\ 112\ 111\ 211$. Z kolei fałszywa moneta z worka numer 1 daje dodatkowo o 1 g więcej: $11\ 112\ 111\ 212$.

Ogólniej, n -ta 1 od końca zmienia się na **2** wtedy i tylko wtedy, gdy w n -tym worku są fałszywe monety. Waga znów na swoim wyświetlaczu wskazuje odpowiedź, mianowicie: miejsca zajmowane przez **2**, liczone od prawej strony, to numery fałszywych worków. □

Czytelnik *Realista* może czuć się słusznie zaniepokojony propozycją ważenia naraz ponad miliarda monet. Szczęśliwie liczbę tę łatwo zredukować, przeprowadzając analogiczne rozumowanie z użyciem systemu dwójkowego zamiast dziesiętnego. Ważymy wówczas już tylko $1 + 2 + 2^2 + \dots + 2^9 = 1023$ monety, a przeanalizowanie związanych z tym dalszych niewielkich modyfikacji pozostawiam dociekliwym.



Delta, listopad 2005



Delta, maj 2006

Kolejna zagadka, na pierwszy rzut oka zupełnie inna, ma jednak z poprzednią zaskakująco wiele wspólnego.

Wielomian $w(x)$ to wyrażenie postaci $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$, gdzie n jest ustaloną liczbą całkowitą nieujemną, współczynniki $a_0, a_1, a_2, \dots, a_n$ są ustalonymi liczbami oraz $a_n \neq 0$. Za x można podstawiać dowolne liczby rzeczywiste i wyznaczać w ten sposób wartości wielomianu. Na przykład wielomian $w(x) = 4x^3 + 17x + 2$ dla $x = 1$ przyjmuje wartość $w(1) = 4 + 17 + 2 = 23$.

Jaś i Małgosia grają w następującą grę. Małgosia wymyśla sobie pewien niezerowy wielomian o współczynnikach całkowitych nieujemnych. Jaś może pytać o wartości tego wielomianu dla dowolnie wybranych liczb całkowitych. Wygra, gdy poda cały wzór wielomianu. Czy może tego dokonać? Jeśli tak, ile pytań musi zadać?

Pokażemy, że Jaś jest w stanie poznać cały wielomian już po dwóch pytaniach. Na początek Jaś pyta o $w(1)$ i Małgosia podaje mu wartość k , obliczoną ze wzoru

$$w(1) = a_n + a_{n-1} + \dots + a_2 + a_1 + a_0 = k.$$

Oznacza to, że każdy ze współczynników a_i (dla $i = 0, 1, \dots, n$) jest równy co najwyżej k (bo wszystkie a_i są nieujemne). Niech m będzie liczbą cyfr liczby k . Wówczas każdy ze współczynników a_i ma najwyżej m cyfr.

Następnie Jaś pyta o $w(10^m)$ i poznaje wartość

$$\begin{aligned} w(10^m) &= a_n \cdot (10^m)^n + a_{n-1} \cdot (10^m)^{n-1} + \dots + a_2 \cdot (10^m)^2 + a_1 \cdot 10^m + a_0 = \\ &= a_n \cdot 10^{mn} + a_{n-1} \cdot 10^{m(n-1)} + \dots + a_2 \cdot 10^{m \cdot 2} + a_1 \cdot 10^m + a_0. \end{aligned}$$

Okazuje się, że ta liczba pokazuje Jasiowi po kolei wszystkie współczynniki wielomianu $w(x)$, jak na wyświetlaczu. Zobaczmy na przykładzie, dlaczego tak jest.

Powiedzmy, że Małgosia wymyśliła sobie wielomian $w(x) = 4x^3 + 17x + 2$. Wówczas $w(1) = 4 + 17 + 2 = 23$ i ten wynik poznaje Jaś po pierwszym pytaniu. Wobec tego współczynniki wielomianu są najwyżej dwucyfrowe (bo ich suma to 23). Jaś w drugim pytaniu pyta zatem o $w(10^2)$. Małgosia oblicza:

$$w(100) = 4 \cdot 100^3 + 0 \cdot 100^2 + 17 \cdot 100 + 2, \quad \text{czyli}$$

4	00	00	00
0	00	00	
	17	00	
			2
			+
			4 00 17 02

i podaje Jasiowi ten wynik. Zauważmy, że powyższe dodawanie zawsze odbywa się bez przenoszenia, bo każdy ze współczynników a_i ma najwyżej m cyfr. Dlatego właśnie Jaś może po kolei odczytać wszystkie współczynniki wielomianu:

$$\begin{array}{cccc} \underline{4} & \underline{00} & \underline{17} & \underline{02} \\ \uparrow & \uparrow & \uparrow & \uparrow \\ a_3 & a_2 & a_1 & a_0 \end{array}$$

Pozostaje zauważyć, że Jaś nie może zagwarantować sobie zwycięstwa po jednym pytaniu. Dla każdej liczby całkowitej x_1 , o którą mógłby spytać, istnieją bowiem dwa różne wielomiany, które mogła wybrać Małgosia, dające tę samą wartość $w(x_1)$, a więc nierozróżnialne dla Jasia po tym jednym pytaniu. Są to mianowicie: wielomian stały $w(x) = |x_1| + 1$ oraz wielomian $w(x) = x + (|x_1| + 1 - x_1)$, oba niezerowe i o współczynnikach całkowitych dodatnich. Podsumowując, Jaś nie może zapewnić sobie zwycięstwa jednym pytaniem, ale dwoma już tak. □

Na zakończenie proponuję jeszcze jedną zagadkę, w której też można zastosować pomysł wyświetlacza prezentującego poszukiwaną informację.

Złośliwy czarodziej rzucił urok na jedną z 1000 beczek z winem – każdy, kto wypije choćby kroplę, zzielenieje w ciągu doby. Codziennie rano dysponujemy dokładnie 10 dzielnymi (i niezielonymi) rycerzami gotowymi ponieść ryzyko. Ile dni potrzeba, aby wykryć zaczarowaną beczkę?

Warto dodać, że odpowiedź „trzy dni” nie jest poprawna, a rozwiązanie zainteresowani odnajdą w tekście wspomnianym na początku niniejszego artykułu.



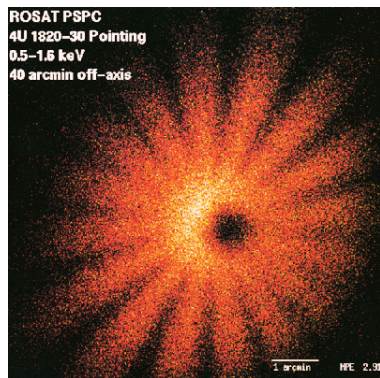
Czytelników Zaawansowanych zachęcam do zastanowienia się nad wariantem tej samej zagadki, w którym Jaś może pytać o wartości wielomianu dla dowolnych liczb, niekoniecznie tylko całkowitych.

Berstery rentgenowskie dziesięć lat później

* Narodowe Centrum Badań Jądrowych

Redaktorka *Delfy* w latach 2009–2012.

Berster rentgenowski (ang. *X-ray burster*) to specjalny układ podwójny gwiazd. Jeden składnik układu stanowi np. gwiazda podobna do naszego Słońca, zwana towarzyszem, a drugim jest gwiazda neutronowa (bardzo mała, ale niezwykle gęsta gwiazda będąca pozostałością po wybuchu supernowej). W takich układach materia przepływa z towarzysza na gwiazdę neutronową, ulegając na powierzchni ściśnięciu i ogrzaniu. Po osiągnięciu przez temperaturę i ciśnienie wartości krytycznych następuje zapoczątkowanie reakcji termojądrowych obserwowanych jako wybuch (stąd nazwa – *berster* – od angielskiego określenia *burst* – wybuch). Zjawisko to można obserwować w promieniowaniu rentgenowskim, skąd też drugi człon nazwy.



Ultra ciasny układ rentgenowski 4U 1820-30 znajdujący się w gromadzie kulistej gwiazd NGC 6624. Źródło: Max-Planck-Institut für extraterrestrische Physik (MPE)

Karłami nazywamy gwiazdy o stosunkowo małej masie i jasności (niektórych może zaskoczyć informacja, że Słońce jest karłem). Z czasem definicja rozciągnięta została na obiekty, które gwiazdami nie są. Na przykład białe karły to obiekty, które powstają po ustaniu reakcji termojądrowych i odrzuceniu otoczki wodorowej przez gwiazdy o małych i średnich masach.

Agnieszka MAJCZYNA*

Od czasu odkrycia *bersterów rentgenowskich* minęło niemal 47 lat. W tym czasie dokonano ogromnego postępu w rozumieniu natury tych obiektów. Dowiedzieliśmy się, jak bardzo są ciekawe i zróżnicowane. Na przykład w ciągu ostatnich lat odkryliśmy berstery o ekstremalnie krótkich okresach orbitalnych. Ich liczba była tak duża, że powstała oddzielna grupa: ultra ciasnych układów rentgenowskich. Ugruntowaliśmy naszą wiedzę o składzie chemicznym atmosfer gwiazd neutronowych należących do układu bersterów i rozważaliśmy możliwość używania bersterów rentgenowskich jako świec standardowych. Ponadto niedawna detekcja fal grawitacyjnych ożywiła zainteresowanie właściwościami supergęstej materii znajdującej się we wnętrzu gwiazd neutronowych będących częścią bersterów.

Czym są berstery?

Berstery rentgenowskie są układami podwójnymi złożonymi z gwiazdy neutronowej i towarzysza o masie mniejszej niż masa Słońca. Należą zatem do klasy obiektów zwanych mało masywnymi układami rentgenowskimi (LMXB od ang. Low Mass X-ray Binary). Obecnie znamy około 200 takich układów (w 2009 było ich 125), z których 80 to berstery rentgenowskie (wobec 40, o których pisałam w Δ_{09}^{12}). W układach tych materia przepływająca z towarzysza na gwiazdę neutronową nie spływa bezpośrednio, lecz opada po spirali, tworząc dysk akrecyjny. Następnie gromadzi się na powierzchni gwiazdy neutronowej, ulegając ściśnięciu i ogrzaniu. W momencie, gdy temperatura i ciśnienie gazu osiągną wartość krytyczną, rozpoczynają się reakcje syntezy termojądrowej powodujące znaczne i gwałtowane pojaśnienie nazywane wybuchem. Ważną i użyteczną cechą takich wybuchów (zwanych wybuchami typu I) jest fakt, że w trakcie gaśnięcia takiego wybuchu gwiazda neutronowa emituje promieniowanie w sposób charakterystyczny dla stygnącego obiektu zwartego. Obecność tego rodzaju wybuchów bezspornie wskazuje, że obiektem zwartym, na który opada materia, jest gwiazda neutronowa, a nie czarna dziura. Pozwala nam to na jednoznaczną ocenę, z jakim obiektem mamy do czynienia.

Berstery rentgenowskie jako układy podwójne mają okresy orbitalne krótsze niż 16 godzin. Rekordzistą w tej dziedzinie jest 4U 1820-30 z okresem orbitalnym równym tylko 11 minut! Z czasem, gdy odkrywano coraz więcej układów o bardzo krótkich okresach orbitalnych, stworzono oddzielną grupę obiektów nazwanych ultra ciasnymi układami rentgenowskimi (UCXB od ang. Ultra Compact X-ray Binaries). Zaliczają się do nich układy o okresach orbitalnych krótszych niż 80 min. Obecnie znamy 19 takich układów (mamy też około 30 niepotwierdzonych kandydatów na tego rodzaju obiekty). Tak krótki okres orbitalny oznacza, że składniki układu krążą bardzo blisko siebie. W konsekwencji towarzyszem gwiazdy neutronowej może być jedynie mała gwiazda typu karzeł lub biały karzeł, a dysk akrecyjny jest również bardzo mały. Obiekty te mają bardzo niską jasność zarówno w zakresie rentgenowskim, jak i w pozostałych zakresach widma promieniowania elektromagnetycznego. Większość UCXB może być obserwowana tylko przez instrumenty satelitów rentgenowskich. Dodatkowo pewna grupa UCXB należy do klasy tzw. źródeł przejściowych, czyli takich, gdzie epizody akrecji widoczne w promieniowaniu rentgenowskim rozdzielone są okresami, gdy ustaje akrecja, a jasność obiektu spada poniżej progu detekcji. Mimo to jesteśmy w stanie powiedzieć, jaki jest skład chemiczny akreowanego gazu. Ze względu na to, że w obiektach UCXB towarzyszem jest karzeł lub biały karzeł, gwiazda neutronowa akreuje materię ubogą w wodór. Widmo w zakresie od optycznego poprzez ultrafiolet po zakres rentgenowski zdominowane jest przez linie emisyjne pochodzące od helu, węgla, tlenu, neonu, brak jest natomiast charakterystycznych dla gwiazd ciągu głównego linii absorpcyjnych.

Więcej informacji na temat projektu Araucaria można znaleźć na stronie araucaria.camk.edu.pl/

Świece standardowe to obiekty astronomiczne o znanej absolutnej jasności gwiazdowej. Ponieważ ich jasność jest niezmienna, możliwe jest oszacowanie odległości do takiego obiektu. A ogólniej – do galaktyki, w której się znajduje.

Jasność Eddingtona jest to maksymalna jasność, jaką może mieć gwiazda, której atmosfera nie ekspanduje ani się nie zapada.

Berstery jako świece standardowe

Wyznaczanie odległości do obiektów astronomicznych jest skomplikowanym zadaniem, któremu poświęcono całe długoletnie projekty badawcze (np. w ramach projektu Araucaria). W przypadku bersterów rentgenowskich znajomość odległości jest krytycznie ważna, gdy chcemy weryfikować modele opisujące właściwości materii, z której zbudowana jest gwiazda neutronowa. Jeśli taki układ znajduje się w gromadzie kulistej, wówczas wyznaczamy odległość do gromady na podstawie dobrze znanych i dających precyzyjne wyniki metod. Jednak bersterów w gromadach kulistych jest niewiele (16 obiektów w 13 gromadach kulistych), dlatego poszukuje się takich cech fizycznych bersterów rentgenowskich, które pozwolą na opracowanie bezpośredniej, niezależnej od otoczenia układu, metody wyznaczania do nich odległości. Mówiąc dokładnie, chcieliśmy ustalić, czy berstery rentgenowskie mogą być świecami standardowymi. Pomocne okazały się wybuchy, od których przecież berstery wzięły swoją nazwę. Okazuje się, że jasność tych wybuchów jest bardzo bliska bądź równa jasności Eddingtona. Zatem jeśli jesteśmy w stanie zmierzyć strumień promieniowania w maksimum wybuchu f_{ob} i z modeli teoretycznych znamy jasność Eddingtona L_{Edd} dla gwiazdy neutronowej, to wówczas odległość możemy wyliczyć z wyrażenia: $d = \sqrt{L_{Edd}/(4\pi f_{ob})}$. Wyrażenie to jest w swej matematycznej postaci bardzo proste, jednak sama metoda taką nie jest. Wartość jasności Eddingtona zależy od nieprzezroczystości gazu, czyli od tego, w jakim stopniu pochłaniane są w nim fotony. Im nieprzezroczystość jest większa (fotony są bardziej pochłaniane), tym jasność Eddingtona jest mniejsza. Nieprzezroczystość zaś zależy między innymi od składu chemicznego gazu i jego temperatury. Obie te wielkości są słabo poznane w przypadku bersterów rentgenowskich. Obecnie wiemy, że jasność obserwowana bersterów rentgenowskich w maksimum wybuchów z ekspansją fotosfery (na granicy jasności Eddingtona) jest taka sama z dokładnością do kilku procent. Zatem można uznać, że obiekty w trakcie takich wybuchów mogą być używane jako świece standardowe.

Gwiazdy neutronowe w bersterach

Berstery rentgenowskie są w centrum uwagi astronomów nie tylko ze względu na fascynujące procesy zachodzące w układach podwójnych gwiazd, ale także ze względu na same gwiazdy neutronowe, które są ich częścią. Pomimo że od odkrycia gwiazd neutronowych minęło już ponad 50 lat, to zagadką pozostają właściwości materii, która je tworzy (stąd też tak istotne jest znalezienie metody wyznaczenia odległości do takiego układu). Dla gwiazdy neutronowej o promieniu rzędu 10 km i masie 1,4 masy Słońca gęstość materii w jądrze przewyższa kilkakrotnie gęstość jądra atomowego ($\rho_0 \approx 2 \times 10^{14} \text{ g/cm}^3$). Materii o takich własnościach nie jesteśmy w stanie otrzymać w żadnym z ziemskich laboratoriów. Jedynym zatem sposobem na poznanie jej właściwości jest stworzenie modeli teoretycznych i porównanie ich z obserwacjami astronomicznymi. Model ten może mieć np. postać zależności masy gwiazdy od jej promienia, dlatego ważnym zagadnieniem jest opracowanie metod wyznaczania zarówno odległości od układu, jak i mas obiektów i ich promieni. Dokładne wyniki oszacowań tych wielkości na podstawie obserwacji pozwolą na potwierdzenie bądź odrzucenie modeli teoretycznych.

Wybuchy na powierzchni gwiazd neutronowych będących częścią bersterów powodują, że gwiazdy te są gorące, a więc łatwo obserwowalne w zakresie rentgenowskim. Ponadto wybuchy z ekspansją fotosfery gwiazdy dają wyjątkową możliwość wyznaczenia promienia gwiazdy neutronowej. Podczas takiego wybuchu atmosfera gwiazdy (wzbogacona o gaz pochodzący z akrecji) pod wpływem ciśnienia promieniowania zwiększa swój promień do wartości maksymalnej, po czym atmosfera opada na powierzchnię gwiazdy. Właśnie ten moment wybierany jest aktualnie do wyznaczenia promienia gwiazdy neutronowej. Przez długi czas uważano, że w trakcie szczególnie silnych wybuchów wpływ akrecji na atmosferę można zaniedbać, gdyż przepływ materii jest bardzo mały. Dopiero w 2012 roku okazało się, że, istotnie, w trakcie



Delta, kwiecień 2008



Delta, kwiecień 2009



wybuchów przepływ materii jest bardzo mały, ale nie można go zaniedbać, gdyż prowadzi to do niewłaściwych oszacowań parametrów gwiazdy. W szczególności w bersterach rentgenowskich tempo przepływu materii z towarzysza na gwiazdę neutronową nie jest stałe, lecz zmienia się w zamkniętym cyklu, którego długość jest inna dla każdego obiektu. Zatem do wyznaczania masy i promienia gwiazdy neutronowej należy wybierać obserwacje, gdy tempo akrecji jest najmniejsze, by zminimalizować potencjalny wpływ tego zjawiska na pomiary.

Kolejne niespodzianki

Berstery rentgenowskie i rezydujące w nich gwiazdy neutronowe pomimo ogromnego postępu w rozumieniu ich natury wciąż skrywają tajemnice i są interesujące zarówno dla astronomów jak i fizyków. Detekcja fal grawitacyjnych ożywiła zainteresowanie tymi niezwykłymi obiektami i potrzebę lepszego poznania własności materii, z jakiej zbudowane jest wnętrze gwiazdy neutronowej. Ukonstytuowanie nowej podklasy obiektów, czyli ultra ciasnych układów rentgenowskich, pokazało, że na astronomów zajmujących się układami rentgenowskimi wciąż czekają niespodzianki. Liczne prace (w tym autorki tego artykułu) wskazują także na potrzebę stworzenia i wysłania w przestrzeń kosmiczną satelitów, których instrumenty dostarczą obserwacji obciążonych mniejszymi niepewnościami pomiarowymi. Wierzmy, że przyniosą one zarówno odpowiedzi na już postawione pytania, jak i nowe, niespodziewane odkrycia.

Liczba chromatyczna z komputera

Michał ADAMASZEK*

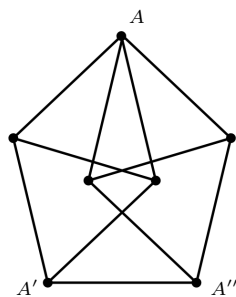
* Kopenhaga

Redaktor *Delty* w latach 2006–2007.

W 1976 roku, na drugie urodziny *Delty*, Kenneth Appel i Wolfgang Haken ogłosili dowód twierdzenia o czterech barwach. W nieco nieformalnej wersji głosi ono, że każdą mapę polityczną, na której wszystkie kraje mają spójne terytorium, można pokolorować czterema lub mniej kolorami tak, aby osiągnąć znany nam z kartografii efekt, w którym każde dwa graniczące ze sobą kraje mają różne kolory. Wynik ten był przełomowy z dwóch powodów. Po pierwsze, rozwiązywał znaną w teorii grafów hipotezę o bardzo długiej i (nota bene) barwnej historii. Po drugie, był to pierwszy, a przez to kontrowersyjny, poważny wynik, w którego dowodzie użyto komputera (po sprowadzeniu problemu do pewnej skończonej, lecz dużej liczby przypadków do sprawdzenia). O tej historii można przeczytać w Δ_{04}^6 .

Wynik Appela i Hakkena można przeformułować na stwierdzenie, że dla dowolnego planarnego grafu G mamy $\chi(G) \leq 4$. Tutaj $\chi(G)$ jest liczbą chromatyczną grafu G , czyli najmniejszą liczbą kolorów potrzebną do pokolorowania wierzchołków G tak, żeby żadne dwa wierzchołki połączone krawędzią grafu nie miały tego samego koloru. Problemy związane z kolorowaniem grafów i obliczaniem liczby chromatycznej są algorytmicznie trudne i, poza walorami estetycznymi, znajdują zastosowanie na przykład w problemach planowania i przydziału zadań. O tych zagadnieniach była już w *Delcie* nie raz mowa, patrz na przykład Δ_{14}^{11} .

Pojęcie liczby chromatycznej grafu można też odnieść do innych zagadnień kolorowania. W Δ_{08}^7 pytałem, jak wykazać, że jeśli każdy punkt płaszczyzny \mathbb{R}^2 pokolorujemy jednym z trzech kolorów, to znajdą się dwa punkty tego samego koloru odległe o dokładnie 1. Dowód znajduje się na rysunku, gdzie widzimy konfigurację 7 punktów z zaznaczonymi odcinkami długości 1, zwaną wrzeczkiem Mosera. Liczba chromatyczna tego grafu wynosi 4 (faktycznie: próba kolorowania 3 kolorami prowadzi do wniosku, że A i A' muszą mieć ten sam kolor; analogicznie A i A'' ; a więc punkty A' i A'' odległe o 1 mają ten sam kolor; sprzeczność). Jeżeli więc oznaczymy przez $\chi(\mathbb{R}^2)$ najmniejszą liczbę kolorów potrzebną do uniknięcia pary punktów tego samego koloru



Wrzeczono Mosera: 7 wierzchołków, wszystkie krawędzie długości 1



w odległości 1 przy kolorowaniu płaszczyzny, to mamy lewą z nierówności

$$4 \leq \chi(\mathbb{R}^2) \leq 7.$$

Prawa jest również łatwa – zachęcam do znalezienia odpowiedniego kolorowania płaszczyzny 7 kolorami.

W 2008 roku pisałem, że nic bardziej konkretnego nie wiadomo (pytanie o dokładną wartość $\chi(\mathbb{R}^2)$ jest znane jako problem Hadwiger–Nelsona) i zanosilo się na to, że tak już zostanie. Ale w 2018 roku Aubrey de Grey zaskoczył świat, publikując dowód nierówności

$$5 \leq \chi(\mathbb{R}^2).$$

Zamiast 7-wierzchołkowego gadżetu Mosera, który nie daje się pokolorować 3 kolorami, de Grey musiał skonstruować analogiczny gadżet, złożony z punktów płaszczyzny i krawędzi jednostkowych między niektórymi z nich, którego nie da się pokolorować 4 kolorami. Okazało się, że ma on 1581 wierzchołków (od tego czasu znaleziono przykłady o około 500 wierzchołkach) i z oczywistych przyczyn nie możemy go wydrukować. Można go znaleźć, wraz z systematyczną konstrukcją, w pracy [1]. Podobnie jak w przypadku gadżetu Mosera, towarzyszy mu dowód oraz wyjaśnienie, jak do tego przykładu dojść (kilka początkowych wskazówek można znaleźć w *Deltoïdzie* z Δ_{18}^{12}). Niektóre sprawdzenia wymagały jednak obliczeń komputerowych, a więc jest to już dowód wspomagany komputerowo.

O analogicznej liczbie chromatycznej $\chi(\mathbb{R}^n)$ wyżej wymiarowych przestrzeni wiemy jeszcze mniej (na przykład w trzech wymiarach wiemy, że $6 \leq \chi(\mathbb{R}^3) \leq 15$), a znane rezultaty wciąż opierają się na dość elementarnych pomysłach. Popatrzmy na przykład na następujący prosty dowód z pracy [2] najlepszego znanego dziś dolnego oszacowania dla \mathbb{R}^{10} , mianowicie: $\chi(\mathbb{R}^{10}) \geq 26$. Musimy znaleźć odpowiedni gadżet, tzn. zbiór punktów w \mathbb{R}^{10} i odcinków długości 1 pomiędzy nimi, którego nie da się pokolorować 25 kolorami. Jako punkty weźmiemy wierzchołki 10-wymiarowej kostki o krawędzi $\frac{1}{2}$, to znaczy 1024-elementowy zbiór złożony z wszystkich ciągów:

$$(x_1, x_2, \dots, x_{10}), \quad x_i \in \left\{0, \frac{1}{2}\right\}.$$

Dwa takie punkty są odległe o 1 wtedy i tylko wtedy, gdy różnią się na dokładnie 4 z 10 pozycji (bo te pozycje wnoszą do wzoru na odległość euklidesową $(0 - \frac{1}{2})^2 = \frac{1}{4}$, podczas gdy pozostałe pozycje, gdzie oba ciągi są równe, wnoszą 0). Możemy więc skonstruować graf odcinków jednostkowych pomiędzy tymi punktami czysto kombinatorycznie. Ma on dwie spójne składowe (dlaczego?), każda o 512 wierzchołkach, bierzemy więc jedną z nich. Dokładne obliczenie jej liczby chromatycznej przekracza nasze możliwości, ale da się obliczyć inny parametr, mianowicie licznosc największego zbioru niezależnego, tzn. największego zbioru wierzchołków, między którymi nie ma żadnej krawędzi. Oto kompletna komenda w popularnym systemie algebry komputerowej Sage konstruująca nasz graf i obliczająca poszukiwaną wartość:

```
Graph([[x,y] for x in [0..2^10-1] for y in [x+1..2^10-1]
      if sum((x^~y).bits())==4]
      .connected_components_subgraphs()[0].complement().clique_number()
```

Po kilkunastu minutach w wyniku otrzymujemy 20, co oznacza, że jeden kolor może być użyty dla co najwyżej 20 wierzchołków, a więc używając 25 kolorów, pomalowalibyśmy ich co najwyżej $20 \cdot 25 = 500$, a to za mało na cały 512-wierzchołkowy graf. To kończy nasz prosty, wspomagany komputerowo dowód.

Na zakończenie kluczowa notka biograficzna. Aubrey de Grey, autor pracy [1], nie jest zawodowym matematykiem. Jest bioinformatykiem i biogerontologiem, badaczem procesów starzenia, a kolorowaniem zajmował się amatorsko po godzinach. Z kolei publikacja [2] jest efektem pracy licencjackiej. W starciu z ciekawymi problemami otwartymi ma więc szansę każdy pomysłowy eksperymentator, a takich, jak sądzę, wśród Czytelników *Delty* nie brakuje.

Jak wiadomo, każdy program można zmieścić w jednym wierszu, jeśli tylko będzie on odpowiednio długi.

Literatura

- [1] de Grey, Aubrey D. N. J., „The chromatic number of the plane is at least 5”. *Geombinatorics* 28 (2018), także: arxiv.org/abs/1804.02385
- [2] Matthew Kahle, Birra Taha, „New lower bounds on $\chi(\mathbb{R}^d)$ for $d = 8, \dots, 12$ ”, *Geombinatorics* 24 (2015), także: arxiv.org/abs/1409.1278

Mało przekątnych – duży problem

Jerzy BEDNARCZUK*

*Redaktor *Delty* w roku 1974, pełniący wówczas funkcję zastępcy redaktora naczelnego.

Ille przekątnych może mieć wielościan wypukły?

Może ich nie mieć wcale. Na przykład ostrosłup nie ma przekątnych. Jeśli natomiast do jednej ściany bocznej ostrosłupa $(n + 2)$ -kątnego tak dokleimy czworościan, aby otrzymany wielościan był wypukły i miał $n + 5$ ścian, to ten wielościan będzie miał n przekątnych. Tak możemy w szczególności otrzymać wielościan z jedną przekątną oraz wielościan z dwiema przekątnymi.

Czy istnieją inne wielościany wypukłe z jedną przekątną? A z dwiema przekątnymi?

Wiemy, że istnieją co najmniej 3 wielościany wypukłe z jedną przekątną oraz co najmniej 8 wielościanów z dwiema przekątnymi. Poniżej podajemy ich opis. Wszystkie prezentowane wielościany otrzymujemy z graniastosłupa trójkątnego po odcięciu od niego jednego lub dwóch czworościanów.

Prawdopodobnie są to już wszystkie takie wielościany. Nie znamy dowodu, że nie ma ich więcej.

Podanie opisu wszystkich wielościanów z jedną i wszystkich wielościanów z dwiema przekątnymi wraz z dowodem, że to są rzeczywiście wszystkie takie wielościany, mogłoby być treścią ciekawej pracy, godnej zgłoszenia na Konkurs Uczniowskich Prac z Matematyki, którego 46. edycja zapowiadana jest na tylnej okładce niniejszego wydania *Delty*.

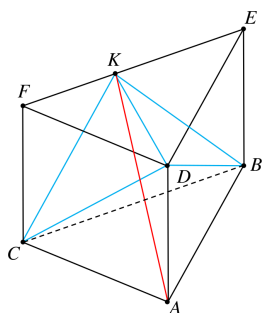


Wielościany wypukłe z jedną przekątną

Przykład 1

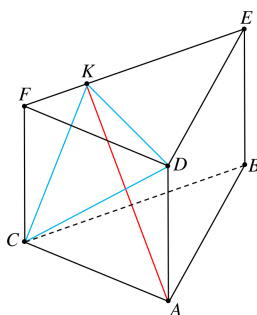
Od graniastosłupa trójkątnego odcinamy czworościany $CDKF$ i $BDKE$. Przekątna to AK .

Uwaga: Na wielościan ten można także patrzeć jak na dwa czworościany, $BCDA$ i $BCDK$, o wspólnej ścianie BCD .



Przykład 2

Od graniastosłupa trójkątnego odcinamy czworościan $CDKF$. Przekątna to AK .



Przyjmujemy oznaczenia:

- k : liczba krawędzi wielościanu,
- w : liczba wierzchołków,
- w_n : liczba wierzchołków, w których schodzi się n krawędzi,
- s : liczba ścian,
- s_n : liczba ścian n -kątnych.

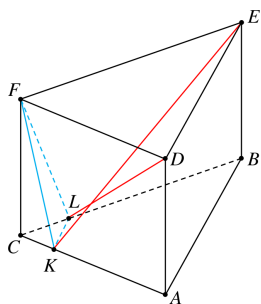
Zestawienie parametrów trzech wielościanów wypukłych z jedną przekątną.

przykład	w	w_3	w_4	k	s	s_3	s_4	s_5
1	5	2	3	9	6	6		
2	6	4	2	10	6	4	2	
3	7	6	1	11	6	3	2	1
?								

Wielościany wypukłe z dwiema przekątnymi

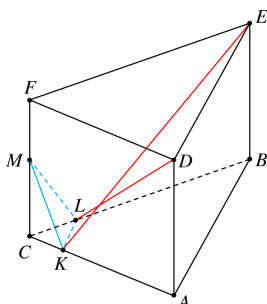
Przykład 4

Od graniastosłupa trójkątnego odcinamy czworoscian $CKLF$. Przekątne to KE i LD .



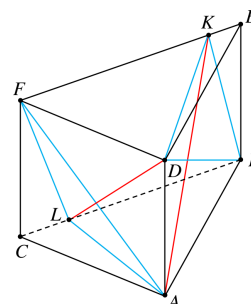
Przykład 5

Od graniastosłupa trójkątnego odcinamy czworoscian $CKLM$. Przekątne to KE i LD .



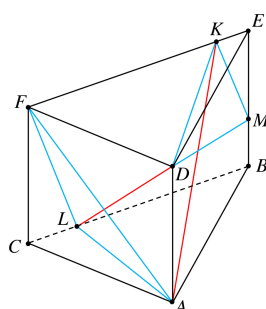
Przykład 6

Od graniastosłupa trójkątnego odcinamy czworosciany $CALF$ i $DBEK$. Przekątne to KA i LD .



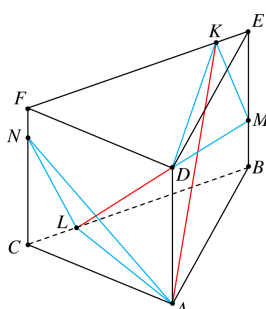
Przykład 7

Od graniastosłupa trójkątnego odcinamy czworosciany $CALF$ i $DMEK$. Przekątne to KA i LD .



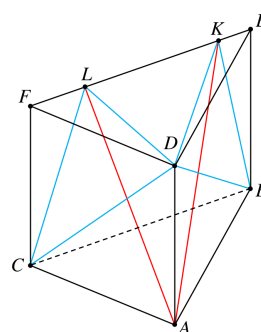
Przykład 8

Od graniastosłupa trójkątnego odcinamy czworosciany $CALN$ i $DMEK$. Przekątne to KA i LD .



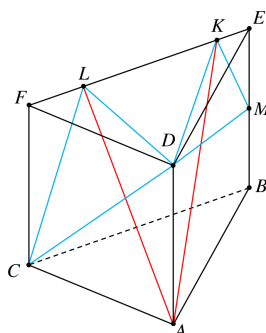
Przykład 9

Od graniastosłupa trójkątnego odcinamy czworosciany $CDLF$ i $DBEK$. Przekątne to KA i LA .



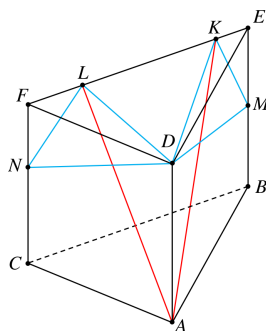
Przykład 10

Od graniastosłupa trójkątnego odcinamy czworosciany $CDLF$ i $DMEK$. Przekątne to KA i LA .



Przykład 11

Od graniastosłupa trójkątnego odcinamy czworosciany $NDLF$ i $DMEK$. Przekątne to KA i LA .



Poniżej zestawienie parametrów ośmiu wielościanów wypukłych z dwiema przekątnymi.

Typ X to taki wielościan, że dwie jego przekątne nie mają wspólnego wierzchołka, typ V to taki wielościan, że dwie jego przekątne mają wspólny wierzchołek.

przykład	w	w_3	w_4	w_5	k	s	s_3	s_4	s_5	s_6	typ
4	7	6	1		11	6	2	4			X
5	8	8			12	6	2	2	2		X
6	6	2	4		11	7	6	1			X
7	7	4	3		12	7	5	1	1		X
8	8	6	2		13	7	4	2		1	X
9	6	3	2	1	11	7	6	1			V
10	7	5	1	1	12	7	5	1	1		V
11	8	7		1	13	7	4	2		1	V
?											

* Narodowe Centrum Badań Jądrowych w Warszawie

Redaktor *Delty* w latach 1997–2018, z-ca redaktora naczelnego w latach 2000–2010.

Pogląd, jakoby utworzenie miesięcznika *Delta* (1973) było reakcją na odkrycie prądów neutralnych (1973), nie znalazł potwierdzenia. Pozornie nie ma w tym nic zaskakującego. Niewielu to wydarzenie naukowe w ogóle doceniło, a jeszcze mniej uznało za przełom. Dlaczego więc napisałem „pozornie” i co to są prądy neutralne? Wyjaśnijmy to sprawnie, ale po kolei.

Przewidywanie istnienia prądów neutralnych jest konsekwencją potraktowania na serio jednocześnie względności i kwantów. Jako pierwszy zrobił to Paul Dirac prawie pół wieku wcześniej (1928). Zrobił, ale nie od razu uwierzył w to, co zrobił.

Na kartce papieru odkrył spin połówkowy (bez którego nie byłoby ani chemii, ani nas) i antymaterię (wtedy, chwilowo, nie wiadomo, do czego potrzebną)... i nie potraktował tego całkiem serio. Dopiero odkrycie pozytonu (antycząstki elektronu, 1932) uwolniło odpowiednią uwagę.

Kolejnym (w telegraficznym skrócie) doświadczalnym bodźcem było zmierzenie nieprzewidywanej przez (samo) równanie Diraca różnicy poziomów energetycznych atomu wodoru, nazwanej *przesunięciem Lamba* (1947). Wyjaśnienie wymagało uznania dynamicznej natury próżni ujawniającej się przy skali rzędu masy elektronu, czyli 10^{-11} cm. To przy okazji rozwiązywało problem potencjalnie nieskończonej energii pola elektrycznego cząstek naładowanych. W miejsce nieskończoności kreowały się pary cząstka-antycząstka (po to jest antymateria). Otworzyło to drogę do elektrodynamiki kwantowej, czyli pierwszej kompletnej teorii o nieustalonej liczbie cząstek, w której nośnikiem oddziaływań jest foton, będący „lokalną kwantową reakcją czasoprzestrzeni” na obecność ładunków elektrycznych.

Worek z cząstkami (uznawanymi wtedy za elementarne) rozwiązał się w latach pięćdziesiątych. Do pewnego stopnia niezależnie obserwowano przemiany jednych cząstek w inne „spokrewnione”, ze zmianą ładunku o jednostkę (tzw. prądy naładowane). „Niechęć” do takich przemian zaczęła być wyjaśniana przez postulowaną bardzo dużą masę (dwóch przeciwnie naładowanych) nośników takiego „słabego oddziaływania”, odpowiadającą za jego „krótkozasięgowość” (rzędu 10^{-17} cm).

W 1967 roku Steven Weinberg zaproponował wspólny opis oddziaływania elektromagnetycznego i słabego. Wymagało to zapostulowania dodatkowego neutralnego nośnika oddziaływania słabego, który miał mieszać się z proto-fotonem, dając bezmasowy foton i masywny (dzięki mechanizmowi Brouta-Englerta-Higgsa) bozon Z^0 . Propozycja była nietrywialna. Wkrótce okazało się jednak (zgodnie z przypuszczeniem Weinberga), że taka teoria pozwala na precyzyjne rachunki (jest renormalizowalna) i można w nią włączyć oddziaływania silne (czyli opisać całą odkrywaną menażerię cząstek oddziałujących silnie, jak np. nukleony).

Jej pierwszym przewidywaniem było właśnie istnienie prądów neutralnych (oddziaływanie słabe bez zmiany ładunku) przenoszonych przez Z^0 . I to właśnie odkryto, obserwując oddziaływanie neutrin i anty-neutrin mionowych w komorze pęcherzykowej Gargamelle w CERN w 1973 roku. Tak narodził się model standardowy, świadkiem konstytuowania i potwierdzania którego była *Delta*.

Jest to najbardziej precyzyjny opis zjawisk, jaki udało się ludzkości opracować (zgodność z pomiarami nawet do 12. dziesiątnej miejsca znaczącego). Jest to czysto matematyczna konsekwencja skończonej prędkości przekazywania informacji oraz niezerowej wartości stałej Plancka. Model standardowy, opierający się na obserwowanych symetriach oraz mierzonych ładunkach i masach, przewiduje całą resztę.

Weinbergowi przypisywane jest powiedzenie, że do zajmowania się fizyką potrzebne są tylko trzy rzeczy: matematyka, matematyka i jeszcze raz



Delta, luty 2014



Delta, marzec 2015

Symetria CP oznacza złożenie sprzężenia ładunkowego C (zamiana cząstek na antycząstki) i odbicia przestrzennego P. Naruszenie symetrii oznacza, że w pewnych doświadczeniach układ poddany działaniu symetrii zachowuje się inaczej od wyjściowego układu.



Delta, październik 2016

O odkryciu przyspieszenia tempa ekspansji Wszechświata pisał w Δ_{12}^{10} Mateusz Iskrzyński.



Delta, luty 2017

matematyka... Pogląd, którego redakcja *Delty* może się i wstydzi, ale raczej nie wypiera.

W miesięczniku mamy jeszcze astronomię i informatykę, ale o tym za chwilę.

Pierwsze dziesięć lat *Delty* przypadło na ugruntowanie się modelu standardowego. Potrzebny był czwarty kwark: został odkryty. Potrzebna była trzecia rodzina fermionów (do wyjaśnienia naruszenia CP): odkryto kwark piękny oraz lepton tau. Przydałoby się bezpośrednio zaobserwować masywne bozony pośredniczące: dokonano tego, gdy tylko udało się zмайstrować odpowiednio potężne urządzenie (CERN, 1983).

Następna dekada (doświadczalnej fizyki cząstek) została naznaczona przygotowaniem i realizacją dokładnego badania Z^0 (głównie) w akceleratorze LEP (ang.: *Large Electron Positron collider*). Na początek (lata dziewięćdziesiąte) wykazano, że rodziny cząstek materii z prawie bezmasowymi neutrinami są tylko trzy. Natomiast precyzyjne pomiary pozwoliły na dokładne oszacowanie masy ostatniej brakującej w tych trzech rodzinach cząstki (kwarku top) na około 180 mas protonu albo liczbę masową tantalu. Dokładnie taki kwark top dostrzeżono w Tevatronie w 1994 roku (początkowo tylko garść przypadków, które bez wsparcia dokładnych pomiarów prawdopodobnie w ogóle nie zostałyby zareportowane już w 1994 roku).

W ten sposób w modelu standardowym pozostał już do odkrycia tylko bozon Higgsa związany z naruszeniem symetrii elektrosłabej generującym masy cząstek. Odkrycie topu ograniczyło spodziewaną masę bozonu Higgsa do wartości niewiele większej od masy samego topu.

I wreszcie zaczęło się coś jeszcze ciekawszego. Najpierw przywołajmy astronomię w postaci astrofizyki, czy też kosmologii, bo właśnie tam miało się wydarzyć coś nieoczekiwanego. W 1998 roku, w pomysłowy sposób badając populację supernowych IA, będących świecami standardowymi w odległościach kosmologicznych, odkryto przyspieszanie tempa ekspansji Wszechświata.

W połączeniu z dokładnymi pomiarami relikтового promieniowania tła (np. wyniki WMAP opublikowane przy okazji trzydziestolecia *Delty*) pozwoliły wyznaczyć gęstość energii fluktuacji próżni na około 10^{-3} eV/mm^3 .

Co w tym ciekawego? Potraktowanie na serio czasoprzestrzeni daje oszacowanie wartości tej gęstości energii tym większe, im mniejszy rozpatrujemy obszar, a ponieważ potencjalnie najmniejszym fizycznie dostępnym obszarem jest komórka o rozmiarach długości Plancka (próba dokładniejszego sprawdzania skończyłaby się utworzeniem czarnej dziury), to gęstość energii próżni powinna być rzędu masy Plancka, czyli 120 rzędów wielkości większa niż obserwowana. Dopóki jej nie zmierzono, to liczone na jakiś nieznaną mechanizm dokładnie niwelujący gęstość energii próżni. Jednak szansa na uzyskanie wyniku, który wyzeruje 120 cyfr znaczących, ale zostawi 121, wydaje się bardzo mała.

Jeden taki problem już się rozwiązał poprzez odkrycie antymaterii przy 10^{-11} cm . Natomiast ten problem sugeruje coś nowego przy odległościach rzędu milimetra – tylko nic takiego nie widać!

Ale to nie koniec. Do naruszania symetrii elektrosłabej trzeba użyć czegoś pozbawionego spinu. A coś takiego, jeżeli nie jest dodatkowo chronione, również zasysa poprawki rzędu masy Plancka.

Czyli powinniśmy oczekiwać czegoś nowego, co ustabilizuje masę bozonu Higgsa przy skali elektrosłabej (10^{-17} cm). Jaka to energia? Taka, jaka już była dostępna w LEP – tylko tam nic przecież nie znaleźliśmy.

Ten „problem hierarchii”, a dokładniej oczekiwanie jego rozwiązania poprzez odkrycie czegoś nowego przy skali naruszenia symetrii elektrosłabej, był jednym z głównych argumentów przy planowaniu i zatwierdzaniu LHC, czyli wielkiego zderzacza hadronów, który zbudowano w tunelu LEP w CERN.



Rozwiązanie zadania F 1086.

Zgodnie z modelem Bohra energie stanów elektronu poruszającego się wokół jądra atomowego o ładunku Z opisuje wzór:

$$E_n = -\frac{Z^2}{n^2} E_1,$$

przy czym $n = 1, 2, 3, \dots$ oznacza numer tzw. powłoki elektronowej. Poza liczbą n stany elektronu charakteryzują jeszcze liczby: spinowa $s = \pm 1/2$, orbitalnego momentu pędu $l = 0, 1, \dots, n - 1$ oraz liczba kwantowa m przyjmująca wartości całkowite od $-l$ do l . Stan opisany jednym zestawem wartości liczb n, l, m, s może być obsadzony tylko przez jeden elektron (zakaz Pauliego). W najprostszej wersji modelu energia stanu zależy tylko od wartości n zgodnie z podanym wzorem. Kolejnym energiom odpowiadają następujące liczby stanów: $n = 1$ to 2 stany, $n = 2 - 8$ stanów, $n = 3 - 18$ stanów itd. Dotychczas opisywaliśmy stany pojedynczego elektronu. Atom azotu ma jednak 7 elektronów. W obliczeniach energii należy więc uwzględnić także energie wzajemnego oddziaływania elektronów. Uczynimy to w sposób bardzo uproszczony. Będziemy budowali atom, dodając krok po kroku elektrony wypełniające kolejne powłoki n i po każdym kroku obliczając energię obsadzanego stanu dla liczby Z zmniejszonej o 1 – przybliżenie: kolejny elektron porusza się w polu ładunku punktowego równego sumie ładunku jądra i już związanych elektronów. Dla azotu mamy 2 elektrony w powłoce $n = 1$ i 5 w powłoce $n = 2$. Energia elektronów wynosi więc:

$$E = -E_1 \left(49 + 36 + \frac{25 + 16 + 9 + 4 + 1}{4} \right).$$

Liczbowo: $E \approx -1343$ eV. Całkowita jonizacja atomu azotu wymaga dostarczenia energii równej $-E = 1343$ eV. Przybliżenie jest bardzo dobre – zmierzona wartość wynosi 1486 eV. Zaproponowane tu przybliżenie dla atomów lekkich daje wartości równe około 90% wartości dokładnych (od 85% dla helu do 91% dla argonu). Niedoszacowanie energii wynika z faktu, że dla kolejnych elektronów ładunek już związany nie jest ładunkiem punktowym.



Delta, październik 2018

Podniesienie energii, efektywnie, o rząd wielkości oraz zmiana na zderzacz hadronowy (przeszukujący cały zakres energii zamiast jednej dobranej) powodowały, że bozon Higgsa nie miał już gdzie się chować. Ale obiecywano (nie tylko sobie) odkrycie tzw. nowej fizyki, która nie pokazała się w akceleratorze LEP.

Najciekawszą (jednak daleko niejedyną) możliwością, na którą liczono, było wykorzystanie przez Naturę ostatniej oryginalnej rzeczy, którą mogłaby ona zrobić, a czego jeszcze nie zaobserwowaliśmy. Chodzi o supersymetrię, która każdej cząstce kooptuje wzbudzenie o komplementarnym spinie (poprzez kolejny, antysymetryczny tym razem, nowy dyskretny wymiar). Jeżeli taka symetria nie jest naruszona, to wkłady do masy bozonu Higgsa się kasują. W dodatku najlżejsze takie wzbudzenie byłoby stabilne, więc byłoby świetnym kandydatem na cząstkę ciemnej materii, której domaga się kosmologia. A zupełnie niezależnie, aby w trakcie ewolucji Wszechświata wytworzyło się odpowiednio dużo ciemnej materii, odpowiadająca za nią cząstka powinna mieć masę też w okolicach skali 10^{-17} cm.

Akcelerator LHC miał być uruchomiony tuż po trzydziestych urodzinach *Delty*, ale budowa (LHC wraz z detektorami) trwała pięć lat dłużej.

Trzy scenariusze były możliwe.

„Medialny”: LHC odkrywa bozon Higgsa i „nową fizykę”, która uzasadnia jego masę.

„Kryzysowy”: LHC niczego nie odkrywa – trzeba zacząć wymyślać fizykę cząstek od zera (wyjaśniając przy okazji, dlaczego model standardowy daje tak dobre przewidywania pomimo braku bozonu Higgsa).

„Poważny”: LHC odkrywa „tylko” bozon Higgsa, ale niczego wyjaśniającego jego niską, w porównaniu do skali Plancka, masę.

Przy okazji czterdziestolecia *Delty* przyznano Nagrodę Nobla za wymyślenie bozonu Higgsa, dzięki temu, że tak wyglądającą cząstkę odkryliśmy w LHC.

Ostatecznie, musieliśmy samym sobie przyznać, że scenariusz „medialny”, jeżeli w ogóle był możliwy, to najpóźniej w LEP-ie. Odkrycie nowej fizyki w LHC i tak wymagałoby precyzyjnego dostrojenia w celu wyjaśnienia masy bozonu Higgsa. Inaczej mówiąc, nieprofesjonalne było argumentowanie, że masa Higgsa nie powinna podlegać dostrojeniu, skoro gęstość energii próżni i tak musi być dostrojona rzędy wielkości dokładniej.

Dlaczego musi? Bo gdyby była choć rząd wielkości większa, to nasz Wszechświat byłby pusty, a więc nie powstałibyśmy, żeby się mu dziwować. Zresztą z masą Higgsa jest podobnie. Jakby trochę ją zwiększyć, to z pierwiastków „mielibyśmy” tylko wodór (choć tu można by jeszcze innymi parametrami pokręcić, żeby pierwiastki próbować uratować).

LHC ma działać jeszcze do (mniej więcej) sześćdziesięciolecia *Delty*. Zwiększana będzie świetność, czyli liczba elementarnych zderzeń na jednostkę czasu. Czy coś nowego uda nam się wyłowić? Nie wiadomo, ale z każdą kolejną próbką przeanalizowanych danych jest to coraz mniej prawdopodobne.

Jednocześnie ciekawe jest np. to, że najmniej podstrojona wersja supersymetrii (tzw. Minimal Split SUSY) nie jest wykluczona, a przewiduje cząstki o masach tuż za granicą dostępności w LHC (oraz taką masę Higgsa jak obserwowana). Ma dobrą kandydatkę na cząstkę ciemnej materii, i to taką, która wyjaśnia fiasko prób jej tzw. bezpośredniej detekcji (komplementarny do LHC sposób poszukiwań rozwijający się na dużą skalę). Nie wiadomo, czy ludzkość da się przekonać do zafundowania następcy LHC, a jak będzie zwlekać, to już może nie być kim budować.

Podsumowując, półwiecze *Delty* odpowiada półwieczu traktowania na serio połączenia względności i kwantów w spójny matematyczny model zwany modelem standardowym. On już z nami zostanie na zawsze. Pójście dalej



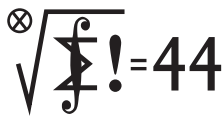
wymaga potraktowania matematyki jeszcze bardziej serio. Można powiedzieć, że trzeba choć trochę cofnąć się do „proto-matematyki” (czyli informatyki?).

Drogowskazami pozostają takie proste stwierdzenia jak $2 \neq 3$ (liczba spinowych stopni swobody bozonu bezmasowego i masowego). Jeżeli ograniczymy się (z konieczności lub wyboru) do kartki papieru, to trzeba będzie zrezygnować z czasoprzestrzeni (bo nie da się stawiać pytań czasoprzestrzeni przy skali Plancka). Czasoprzestrzeń powinna pojawić się (na tej kartce) jako niskoenerytyczny przejaw prostych przekształceń, np. permutacji. Jeszcze tylko nie mamy pewności, co trzeba permutować (oprócz pewności, że trudno to będzie ogarnąć nawet wyćwiczonym umysłem). Matematyka, która może posłużyć do jeszcze głębszego opisu świata, jest przydatna dopiero do odkrycia lub uświadomienia.

Tematów wystarczy na tysiąclecia *Delty*. Oby tylko wystarczyło Czytelników i Autorów.

Liga zadaniowa Wydziału Matematyki, Informatyki i Mechaniki, Wydziału Fizyki Uniwersytetu Warszawskiego i Redakcji *Delty*

Klub 44 M



Czołówka ligi zadaniowej **Klub 44 M** po uwzględnieniu ocen rozwiązań zadań 861 ($WT = 1,93$) i 862 ($WT = 2,64$) z numeru 5/2023

Michał Adamaszek	Kopenhaga	48,47
Radosław Kujawa	Wrocław	43,57
Paweł Najman	Kraków	43,16
Adam Woryna	Ruda Śl.	40,91
Janusz Fiett	Warszawa	38,18
Marek Spychała	Warszawa	36,43
Paweł Kubit	Kraków	36,11
Szymon Tur		35,35
Piotr Kumor	Olsztyn	35,26
Jerzy Cisło	Wrocław	32,97

Pan Michał Adamaszek pokonał był próg 44 p. (po raz siódmy) dokładnie rok wcześniej i od tego czasu nie stracił w żadnej serii ani ułamka punktu! Teraz mija 44 po raz ósmy.

Termin nadsyłania rozwiązań: 29 II 2024

Czołówka ligi zadaniowej **Klub 44 F** po uwzględnieniu ocen rozwiązań zadań 758 ($WT = 1,75$), 759 ($WT = 2,95$) z numeru 5/2023

Tomasz Rudny	Poznań	43,41
Marian Łupieżowicz	Gliwice	2-40,56
Jacek Konieczny	Poznań	38,28
Tomasz Wietecha	Tarnów	16-33,11
Konrad Kapcia	Poznań	2-31,17
Ryszard Baniewicz	Wrocław	1-24,97
Andrzej Nowogrodzki	Chocianów	3-20,05
Paweł Perkowski	Ożarów	5-16,52

Klub 44 F



Zadania z matematyki nr 871, 872

Redaguje Marcin E. KUCZMA, od 1980 roku

871. Dana jest liczba całkowita parzysta $n > 0$.

(a) Dowieść, że w przedziale $[n+1, 2n+1]$ zawiera się n -elementowy zbiór liczb całkowitych M taki, że żaden jego element nie jest dzielnikiem sumy wszystkich liczb zbioru M .

(b) Wyjaśnić, czy zawsze istnieją (w tym przedziale) co najmniej dwa różne zbiory n -elementowe o powyższych własnościach.

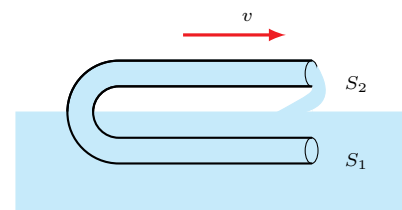
872. Wielomiany W_1, \dots, W_m (jednej zmiennej), stopni dodatnich, mają dodatnie współczynniki całkowite. Wykazać, że dla nieskończenie wielu liczb całkowitych $n > 0$ wartości $W_1(n), \dots, W_m(n)$ są jednocześnie liczbami złożonymi.

Zadanie 872 zostało opracowane na podstawie propozycji, którą przysłał pan Witold Bednarek z Łodzi.

Zadania z fizyki nr 768, 769

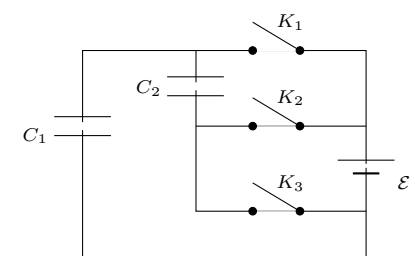
Redaguje Elżbieta ZAWISTOWSKA, od 2012 roku

768. Otwarta z dwóch stron rurka w kształcie litery U porusza się z prędkością v równoległe do powierzchni cieczy (rys. 1). Przekrój zanurzonej w cieczy dolnej części rurki wynosi S_1 , a górnej, znajdującej się nad cieczą, jest równy S_2 . Jaka siła zewnętrzna działa na rurkę w kierunku poziomym? Tarcie i powstawanie fal należy zaniedbać. Gęstość cieczy wynosi ρ .



Rys. 1

769. W układzie przedstawionym na rysunku 2 na początku wszystkie klucze są otwarte, a kondensatory o pojemnościach C_1 i C_2 nienaładowane. Klucze K_1 i K_3 zostały zamknięte i po ustaleniu równowagi otwarte, po czym zamknięty został klucz K_2 . Znaleźć napięcie końcowe na kondensatorze C_1 . Siła elektromotoryczna baterii wynosi \mathcal{E} .



Rys. 2



Doba księżycowa jest równa miesięcowi synodycznemu, czyli okresowi pomiędzy kolejnymi pełniami Księżyca widzianymi z Ziemi, i wynosi około 29,5 ziemskiego dnia.

F. Civilini i in., „Thermal Moonquake Characterization and Cataloging Using Frequency-Based Algorithms and Stochastic Gradient Descent”, *Journal of Geophysical Research: Planets* (2023).

Prosto z nieba: Co trzęsie Księżycem?

Księżyc ma masę około 1/80 masy Ziemi i jest od niej oddalony o średnio 384 tys. km, czyli 60 razy tyle co promień Ziemi, wynoszący średnio 6370 km. Przy zachowaniu proporcji układ Ziemia-Księżyc wygląda mniej więcej tak:

W odróżnieniu od Ziemi Księżyc nie wykazuje aktywności wulkanicznej ani ruchomych płyt tektonicznych. Wiemy jednak, że struktura wewnętrzna Księżyca jest niejednorodna; składa się on z geochemicznie odrębnej skorupy, płaszczka i jądra. Jądro jest niewielkie (promień około 350 km, przy promieniu Księżyca 1750 km), co oznacza, że jest dużo mniejsze niż na przykład ziemskie, które zaczyna się około 2900 km pod powierzchnią (promień Ziemi to około 6350 km). Struktura ta powstała w wyniku frakcyjnej krystalizacji oceanu magmy wkrótce po uformowaniu się układu Ziemia-Księżyc, około 4,5 miliarda lat temu.

W latach 70. w ramach projektu Lunar Seismic Profiling Experiment (LSPE) astronauta misji Apollo 17 umieścili na Księżycu zestaw trzech sejsmometrów (geofonów). Zbierały one dane przez okres ośmiu miesięcy, od października 1976 roku do maja 1977 roku. Okazuje się, że mimo braku wyrafinowanej struktury wewnętrznej Księżyc wykazuje się wewnętrzną aktywnością. Bez atmosfery, która by go izolowała, Księżyc doświadcza drastycznych wahań temperatury, od 120°C w księżycowe „południe” do –120°C w nocy. Te ekstremalne wahania powodują, że powierzchnia Księżyca rozszerza się, gdy jest gorąco, i kurczy, gdy jest zimno, a ruch ten prowadzi do niewielkich wstrząsów i pęknięć zewnętrznych warstw, które prowadzą do termicznych trzęsień Księżyca.

Dane LSPE zostały niedawno przeanalizowane przez geofizyków z Kalifornijskiego Instytutu Technicznego. Wyniki wskazują na to, że oprócz oczekiwanych termicznych trzęsień globu w danych znajdują się jeszcze dodatkowe, niespodziewane sygnały, które występują podczas księżycowych poranków. Niezwykle regularne – występujące co 5–6 minut przez okres paru godzin po wschodzie Słońca – minitrzęsienia są w istocie wynikiem rozgrzewania się w promieniach Słońca pozostałej na powierzchni części lądowiska księżycowego, od której oddzielił się wracający na Ziemię statek z astronautami.

Zrozumienie aktywności księżycowej jest ważne, ponieważ w przeciągu kilku lat na powierzchni Księżyca wylądują astronauta w ramach misji Artemis, by ostatecznie założyć bazę księżycową. Termiczne trzęsienia księżycowe, choć zbyt małe, aby mogły być odczuwalne, są istotne, ponieważ dostarczają wskazówek na temat procesów zachodzących na powierzchni Księżyca, w tym budowy kraterów i księżycowej „gleby”. To z kolei ma znaczenie dla planów umieszczenia na Księżycu różnego typu aparatury badawczej, np. dużych teleskopów.

Michał BEJGER

Centrum Astronomiczne im. Mikołaja Kopernika PAN,
Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Ferrara, Włochy
Redaktor Deltę od roku 2011.

Niebo w grudniu

Słońce zajmuje najniższe położenie na niebie w ciągu roku, górując w środkowej Polsce na wysokości około 15°. Jak co roku, 13 grudnia następuje najwcześniejszy zachód Słońca i w kolejnych dniach zmierzch następuje coraz później. Rano jednak Słońce wschodzi nadal coraz później i odstęp pomiędzy kolejnymi jego wschodami jest większy niż między zachodami, stąd dzień się skraca aż do przesilenia zimowego, które w tym roku wypada 22 grudnia o godzinie 4:27 naszego czasu. Najpóźniejszy wschód Słońca natomiast ma miejsce 30 grudnia.

Na początku miesiąca nachylenie ekliptyki wieczorem jest jeszcze słabe, co przełoży się na praktyczny brak możliwości obserwacji Merkurego po zmierzchu, mimo tego że 4 grudnia planeta osiągnie swoją maksymalną

elongację wschodnią, przekraczającą 21°. Wędruje jednak pod ekliptyką i dlatego zachodzi mniej więcej godzinę po Słońcu, zanim jeszcze niebo odpowiednio się ściemni.

Grudzień również zacznie się dobrą widocznością Księżyca na niebie porannym. Pierwszego dnia miesiąca jego tarcza w fazie 85% zajmie pozycję 2° od Polluksa w Bliźniętach i przetnie południk lokalny około godziny 3, na wysokości 65°. 4 grudnia Księżyc w fazie 60% przejdzie 3° nad Regulusem, a dobę później osiągnie ostatnią kwadrę, świecąc dalej na tle Lwa. W kolejnych dobach jego sierp stanie się coraz cieńszy, zdążając ku nowiu 13 grudnia tuż po północy naszego czasu. Po drodze Srebrny Glob cztery kolejne poranki spędzi na tle gwiazdozbioru Panny, gdzie w nocy

z 6 na 7 grudnia w fazie 33% zakryje jaśniejszą od 4^m gwiazdę Zaniah (η Vir). Tarcza Księżycza pokaże się na niebie niewiele przed północą, a wspomniana gwiazda zniknie za jej jasnym brzegiem około godziny 1:50, by godzinę później pojawić się ponownie po drugiej stronie.

Poranek 8 grudnia zostanie Srebrny Glob w sierpie zwięzonym do 23% mniej więcej 6° nad Spiką, najjaśniejszą gwiazdą Panny, 24 godziny później jego sierp zaprezentuje fazę 15%, docierając na pozycję 5° na prawo od Wenus. Druga planeta od Słońca w grudniu pokona na niebie 35°, zaczynając miesiąc 4,5° od Spiki i kończąc 2° od gwiazdy Graffias, czyli β Skorpiona, która jest szerokim układem podwójnym, widocznym już w lornetkach. Tutaj dwie gwiazdy o jasnościach +2,6^m i +4,9^m dzieli dystans 14'', czyli tak samo jak u Mizara w Wielkim Wozie, tylko z troszkę słabszymi składnikami. Przez cały miesiąc tarcza Wenus zmniejszy jasność od -4,2^m do -4,1^m i średnicę z 17'' do 14'', zwiększy za to fazę z 68% do 78%. 18 grudnia planeta przejdzie 2° na północ od Zuben Elgenubi, czyli drugiej co do jasności, ale oznaczanej na mapach nieba grecką literą α , gwiazdy Wagi.

10 grudnia o stopień dalej, lecz na południe od Zuben Elgenubi znajdzie się bardzo cienki już sierp Księżycza w fazie 8%. Tarczę Srebrnego Globu można jeszcze próbować dostrzec o świcie 11 grudnia, gdy w fazie 4% wszędzie on około godziny 6, a 45 minut później wzniesie się na wysokość 4°. Wspomniana wyżej gwiazda Graffias znajdzie się 7° na lewo od tarczy Księżycza, lecz do jej dostrzeżenia niezbędne mogą okazać się lornetka lub teleskop.

Grudniowy nów Księżycza bardzo dobrze zbiega się z maksimum aktywności corocznego roju meteorów Geminidów. Ich radiant znajduje się jakieś 2° na północny zachód od Kastora, drugiej co do jasności gwiazdy Bliźniąt. Geminidy promieniują od 4 do 17 grudnia, z maksimum 14 dnia miesiąca, gdy można się spodziewać nawet 150 zjawisk na godzinę. Geminidy są widoczne przez całą noc, a ich radiant góruje około godziny 2 na wysokości prawie 70°. Są to dość wolne meteory, ich prędkość zderzenia z naszą atmosferą wynosi 35 km/s.

Na niebie wieczornym dobrze można obserwować cztery planety Układu Słonecznego. Niżej i bliżej zachodniej części nieba znajduje się Saturn z Neptunem, które dzieli dystans około 22°. Wyżej i bardziej po stronie wschodniej znajduje się Jowisz z Uranem, odległe od siebie o mniej więcej 14°. Do końca miesiąca wszystkie planety odwiedzi Srebrny Glob dążący do pełni. Księżyc zacznie pojawiać się o zmierzchu od 15 grudnia, gdy o godzinie 16:30 zajmie pozycję na wysokości 6° nad południowo-zachodnią częścią firmamentu, prezentując tarczę w fazie 9%. Dobę później o tej samej porze jego faza zwiększy się do 16%, wzrośnie również wysokość nad widnokregiem do 17°.

17 i 18 grudnia Księżyc spotka się z planetą Saturn. Najpierw pokaże się 6° pod nią, a następnie 10° na wschód od niej, w fazie, odpowiednio, 26% i 37%. Planeta w grudniu pokona 2° na tle gwiazdozbioru

Wodnika, a 12 dnia miesiąca minie w odległości 20' gwiazdę 5. wielkości 42 Aqr. Blask samej planety spadnie w grudniu do +0,9^m, przy średnicy tarczy 16''. Saturn góruje jakieś 45 minut po zachodzie Słońca i znika z nieboskłonu około godziny 21.

19 grudnia przypada I kwadra Księżycza, który spotka się wtedy z Neptunem, mijając tę planetę w odległości 2°. Neptun na początku miesiąca zmieni kierunek swojego ruchu na prosty, zawracając na granicy między Rybami i Wodnikiem, niecałe 1,5° na zachód od świecącej z jasnością obserwowaną +5,5^m gwiazdy 20 Psc. W grudniu planeta świeci blaskiem +7,9^m i do jej dostrzeżenia potrzebna jest lornetka. Około godziny 18 Neptun góruje 10° wyżej od Saturna.

Jowisz z Uranem są miesiąc po opozycji i są widoczne prawie przez całą noc na tle gwiazdozbioru Barana. Jowisz przecina południk lokalny po godzinie 20, a Uran mniej niż godzinę później. Obie planety przekraczają wtedy wysokość 50°. Ostatniego dnia grudnia jaśniejsza z planet zmieni kierunek ruchu na prosty. Przez cały miesiąc jej jasność spadnie z -2,8^m do -2,6^m, tarcza zaś zmniejszy średnicę z 48'' do 44''. Uran świeci z jasnością +5,7^m i do jego dostrzeżenia przyda się lornetka.

W dniach 22 i 23 grudnia tę parę planet odwiedzi Księżyc. Pierwszego z wymienionych dni naturalny satelita Ziemi w fazie 81% zajmie pozycję 4° na północny wschód od Jowisza, dobę później natomiast przesunie się na pozycję także 4° na północny wschód, ale od Urana. Zwiększy przy tym fazę do 88%. Jednocześnie zakryje gwiazdę 4. wielkości Botein (δ Ari). Gwiazda zniknie za ciemnym brzegiem księżycowej tarczy około godziny 16:30, a zatem na jasnym jeszcze niebie. Pokaże się ponownie po jej drugiej stronie jakieś 55 minut później.

27 grudnia Księżyc przejdzie przez pełnię, świecąc między Bliźniętami, Bykiem i Woźnicą. Dwa dni wcześniej Srebrny Glob utworzy trójkąt równoramienny z Plejadami i Aldebaranem w Byku, kolejnej zaś nocy zbliży się na 1,5° do El Nath, gwiazdy wspólnej dla Byka i Woźnicy. Mieszkańcy pogranicza brazylijsko-urugwajsko-argentyńskiego dostaną możliwość obserwacji zakrycia tej gwiazdy przez naszego satelitę. Do końca miesiąca warto wspomnieć o wschodzie Księżycza 4° od Polluksa wieczorem 28 grudnia i ponownym zbliżeniu na 5° do Regulusa rankiem 31 dnia miesiąca.

21 grudnia przez opozycję przejdzie planetoida (4)Westa, czyli najjaśniejsza planetoida na naszym niebie. Tegoroczna opozycja należy niestety do tych małych, i Westa osiągnie jasność tylko +6,4^m. Planetoida kreśli swoją trajektorię na pograniczu Oriona, Bliźniąt i Byka. 15 grudnia Westa zbliży się na 9' do jaśniejszej od 5^m gwiazdy χ 2 Ori, 8 dni później natomiast przejdzie 22' od podobnie świecącej gwiazdy χ 1 Ori.

Ariel MAJCHER

Autor cyklu „Niebo w...” od roku 2017.

Pół wieku osvajania Wszechświata

Okragłe jubileusze są zawsze dobrą okazją do różnych podsumowań i refleksji. Zwłaszcza gdy, jak dokumentujemy to w tym numerze, dostojna jubilatka ma się świetnie i pełna młodzieńczej werwy nie daje po sobie poznać, że jest nestorką rynku czasopism popularyzujących naukę w Polsce. Pozwólmy sobie zatem na mały skok w przeszłość i zastanówmy się, jak w ciągu ostatnich pięciu dekad zmieniło się nasze wyobrażenie o tym, skąd wziął się Wszechświat, taki jakim go obecnie widzimy.

Mikrofalowe promieniowanie tła pozwala nam zajrzeć głęboko w przeszłość Wszechświata. Jest to możliwe dlatego, że gdy Wszechświat był dostatecznie gęsty i gorący, przemierzające go fotony miały dostatecznie duże energie, by jonizować atomy wodoru. Występowała wówczas równowaga między procesami łączenia się elektronów i protonów w neutralne atomy wodoru – czemu towarzyszyło wydzielanie energii w postaci fotonów – i jonizacją tych atomów, co wymagało dostarczenia energii właśnie w postaci fotonów. Jednak w pewnym momencie, gdy temperatura Wszechświata spadła poniżej pewnej krytycznej wartości i tym samym energia fotonów się obniżyła, jonizacja

w praktyce ustała. Nastąpiła rekombinacja protonów i elektronów. Wszechświat stał się przezroczysty dla promieniowania elektromagnetycznego, które, nie niepokozone oddziaływaniami, mogło już bez przeszkód przemierzać przestrzeń. Miliardy lat rozszerzania się Wszechświata przesunęły zakres tego promieniowania z dalekiego ultrafioletu do podczerwieni, ale poza tym zachowało ono pełną informację o tym, co działo się we Wszechświecie w chwili, gdy zostało wyemitowane. Badając je, sporządzamy w istocie fotografię plasterka Wszechświata w kształcie ogromnej sfery, odpowiadającej sferze niebieskiej.

Wszyscy wiedzą, że mikrofalowe promieniowanie tła zostało odkryte niejako przypadkiem i opisane w 1964 roku przez Arno Penziasa i Roberta Wilsona podczas prac badawczo-rozwojowych, które – jak się okazało po wielu latach – ostatecznie doprowadziły do powstania telefonii komórkowej. Mniej znany jest fakt, że już trzy lata później Ronald Bracewell i E.K. Conklin zmierzili udział poszczególnych częstości fal w tym promieniowaniu – czyli, mówiąc bardziej uczenie, jego widmo – i stwierdzili, że odpowiada ono widmu ciała doskonale czarnego o temperaturze około 2,73 K.

Temperatura ta nie jest jednak zupełnie stała na całej powierzchni sfery niebieskiej. Już George Gamow w latach czterdziestych XX wieku zauważył, że fluktuacje kwantowe powinny prowadzić do małych niejednorodności tej temperatury i – przez równania Einsteina – do małych niejednorodności gęstości materii we Wszechświecie. Podobnie jak dźwięk w powietrzu – będący przecież w istocie również niejednorodnościami gęstości materii i ciśnienia poruszającymi się w określonym kierunku – możemy te pierwotne niejednorodności rozłożyć na składowe o określonych częstościach i rozważać ewolucję każdego z tych „tonów” z osobna.

Jeden taki ton ma pewną szczególną własność. W czasie, jaki upłynął od najwcześniejszych chwil Wszechświata do momentu emisji mikrofalowego promieniowania tła, tworząca ten ton materia zdążyła wykonać pół oscylacji, tzn. miejsca gęstsze stały się maksymalnie gęste, a miejsca rzadsze – maksymalnie rozrzedzone. Gęstsze znaczy cieplejsze, a rzadsze – chłodniejsze. Oznacza to, że w mikrofalowym promieniowaniu tła powinniśmy obserwować przestrzenne korelacje miejsc gęstszych i rzadszych na określonych skalach kątowych, około jednego stopnia.

Zmierzenie tego nie było jednak takie łatwe. Dopiero w 1989 roku w przestrzeń kosmiczną został wysłany satelita Cosmic Background Explorer (COBE), który w latach 1992–1996 zebrał dane potwierdzające względne niejedności temperatury mikrofalowego promieniowania tła na poziomie 1 do 100 000. Za to odkrycie przyznano Nagrodę Nobla z fizyki – otrzymali ją George Smoot i John Mather.

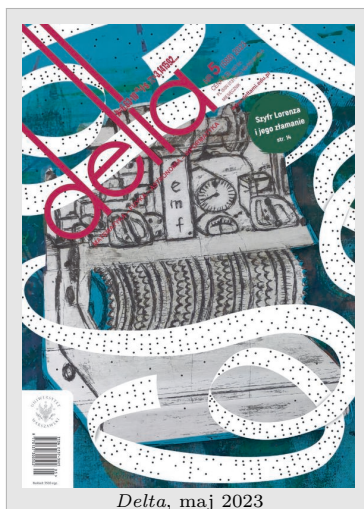
Dlaczego jednak skupiać się tylko na tym jednym tonie. Przecież wyższe składowe harmoniczne, tzn. zaburzenia gęstości, które wykonały jedną, półtorej, dwie itd. oscylacji, także zostawiają charakterystyczne korelacje w przestrzennym rozkładzie temperatury na sferze niebieskiej. Te korelacje występują na mniejszych skalach kątowych, więc trudniej je zmierzyć. Dlatego trzeba było czekać do 2002 roku, kiedy to swoją misję rozpoczął amerykański satelita Wilkinson Microwave Anisotropy Probe, który od 2003 roku ujawniał corocznie coraz dokładniejsze pomiary rozkładów temperatury mikrofalowego



Delta, kwiecień 2021



Delta, maj 2022



Delta, maj 2023

promieniowania tła. Od 2013 roku palmę pierwszeństwa w dokładności przejął europejski satelita Planck, który wyznaczył korelacje na skalach do kilku minut kątowych.

Od ćwierć wieku wiemy zatem, że we wczesnym Wszechświecie występowały małe niejednorodności gęstości, co potwierdza licząca sobie 75 lat przewidywania. Teoria grawitacji przewiduje jednak, że powinna wystąpić w tym przypadku specyficzna wersja efektu świętego Mateusza – miejsca, gdzie jest więcej materii, wykazują silniejsze przyciąganie grawitacyjne, i w ten sposób ściągają na siebie jeszcze więcej materii, co powoduje, że jeszcze silniej przyciągają itd. Czy w ten sposób mogły powstać obserwowane w dzisiejszym Wszechświecie skupiska materii, takie jak galaktyki i gromady galaktyk?

Taką tezę postawili w 1970 roku P.J.E. Peebles i J.T. Yu, którzy stwierdzili, że dostępna wówczas wiedza jest zgodna z hipotezą, że gromady galaktyk i duże galaktyki powstały właśnie na skalach kątowych zgodnych ze skalami niejednorodności temperatury mikrofalowego promieniowania tła – ponad 20 lat przed ich obserwacyjnym wykryciem! Innymi słowy, zasugerowali, że ten sam mechanizm spowodował różnice energii fotonów docierających do nas z rekombinacji i małe niejednorodności gęstości, z których wyłoniły się obserwowane dziś we Wszechświecie struktury materii.

W fizyce teza teoretyczna jest jednak zazwyczaj niewiele warta, jeśli nie da się jej zweryfikować empirycznie? Czy można jednak stwierdzić, co działo się we Wszechświecie miliardy lat temu, skoro tych procesów nie można dziś powtórzyć w kontrolowanych warunkach? Każdy miłośnik seriali kryminalnych wie, że wykrycie sprawców przestępstwa nie zawsze wymaga schwytania zлочyńców na gorącym uczynku. Dobry detektyw gromadzi i analizuje dane, dążąc do uzyskania niezależnych potwierdzeń swoich hipotez, i w ten sposób jest w stanie odtworzyć przebieg wydarzeń. Dobry fizyk też.

Pięćdziesiąt lat temu rozkład materii we Wszechświecie można było odtworzyć na podstawie katalogów obserwacji astronomicznych. Miały one postać opasłych tomów, w które wpisane były dane dotyczące położenia i własności świecących na niebie obiektów. Nie lada wysiłkiem było przetworzenie ich do postaci zrozumiałej przez ówczesne komputery, które ulegały znaczącej miniaturyzacji – zmniejszały swoje rozmiary z całych pokoiów do pojedynczych szaf. Groth i Peebles skorzystali z jednego takiego katalogu, zestawionego przez Shane'a, przeanalizowali statystycznie komputerowo zapisany w nim rozkład gwiazd i galaktyk na niebie, przedstawili swoje wyniki w formie schematu, na który naniesione były pozycje tych obiektów na niebie, i w 1977 roku byli gotowi do podtrzymania stwierdzenia o wspólnym pochodzeniu niejednorodności w mikrofalowym promieniowaniu tła i kosmicznych struktur materii. Kolejne dekady przyniosły nowe obserwacje – niektóre z nich zostały wspomniane powyżej – i nowe obliczenia. Gwałtowny rozwój komputerów pozwolił na coraz większe wyrafinowanie rachunków i śledzenie możliwych losów miliardów galaktyk dla dowolnie generowanych przez badaczy parametrów opisujących Wszechświat. Wszystkie te badania z coraz większą dokładnością potwierdzały zasadniczą tezę o prostej fizyce u podłoża złożonych struktur. Nic zatem zaskakującego, że James Peebles otrzymał w 2019 roku Nagrodę Nobla z fizyki za badania teoretyczne pozwalające na zrozumienie procesów kształtujących Wszechświat.

W ciągu ostatniego półwiecza w fizyce teoretycznej, zwłaszcza tej opisującej wczesny Wszechświat, zmieniło się wiele. Niemalże skończył się czas „samotnych wilków” nauki; badania prowadzone są zazwyczaj w większych grupach. Śmiało korzysta się z coraz potężniejszych komputerów. I tylko rola dobrych pomysłów czy też, mówiąc bardziej precyzyjnie, dobrych hipotez badawczych się nie zmieniła. Rewolucyjne idee są najczęściej bardzo proste i mocno zakorzenione w doświadczeniach i obserwacjach.

Krzysztof TURZYŃSKI

Wydział Fizyki Uniwersytetu Warszawskiego

Redaktor *Delty* w latach 2009–2016; w latach 2010–2016 pełnił funkcję zastępcy redaktora naczelnego.

Korzystałem m.in. z wykładu noblowskiego Jamesa Peeblesa.

Różnice kwadratów w równaniach diofantycznych



Kącik Początkującego Olimpijczyka ukazuje się już przez 10% czasu istnienia *Delty*. Z tej okazji Redakcja życzy jego Autorowi przekroczenia 50%!

Bartłomiej BZDEGA

Uniwersytet im. A. Mickiewicza w Poznaniu

Do klasyki równań diofantycznych należy następujące

Zadanie. Wyznaczyć wszystkie trójki (a, b, c) liczb całkowitych dodatnich spełniające równość $3^a + 4^b = 5^c$.

Tu od razu widać, że $4^b = (2^b)^2$ jest kwadratem liczby naturalnej. Chcielibyśmy uzasadnić, że 5^c również jest kwadratem. Wtedy po przeniesieniu 4^b na prawą stronę otrzymalibyśmy różnicę kwadratów równą 3^a , czyli iloczyn sumy i różnicy, co jest możliwe tylko wtedy, gdy oba czynniki są potęgami trójki.

Pamiętamy, że równość możemy zastąpić dowolną kongruencją (ale nie na odwrót!). Narzuca się kongruencja modulo 4 lub modulo 3 (bo wtedy znika 4^b lub 3^a). Pierwsza z nich pozwala wykazać, że a jest parzyste:

$$(-1)^a \equiv 3^a \equiv 3^a + 4^b = 5^c \equiv 1^c = 1 \pmod{4}.$$

To być może przyda się potem. Druga na początku sprawdza się zdecydowanie lepiej:

$$1 = 1^b \equiv 4^b \equiv 3^a + 4^b = 5^c \equiv (-1)^c \pmod{3}.$$

Wynika z niej, że $c = 2d$ dla pewnej liczby całkowitej dodatniej d , a zatem

$$3^a = 5^c - 4^b = (5^d)^2 - (2^b)^2 = (5^d + 2^b)(5^d - 2^b).$$

Z tego wnioskujemy, że

$$5^d + 2^b = 3^{a_1}, \quad 5^d - 2^b = 3^{a_2},$$

przy czym liczby a_1 i a_2 są całkowite nieujemne oraz $a_1 + a_2 = a$ i $a_1 > a_2$. Teraz dobrym pomysłem jest dodanie i odjęcie stronami tych dwu równości. Dostaniemy:

$$2 \cdot 5^d = 3^{a_1} + 3^{a_2} = 3^{a_2}(3^{a_1 - a_2} + 1),$$

$$2^{b+1} = 2 \cdot 2^b = 3^{a_1} - 3^{a_2} = 3^{a_2}(3^{a_1 - a_2} - 1).$$

Każda nich pozwala wywnioskować, że $a_2 = 0$ (bo lewa strona nie dzieli się przez 3), więc $a_1 = a$. Dalej popracujemy z drugą z nich, która właśnie przybrała postać $2^{b+1} = 3^a - 1$. Tu warto sobie przypomnieć, że a jest parzyste. Niech $a = 2f$. Otrzymujemy:

$$2^{b+1} = 3^{2f} - 1 = (3^f)^2 - 1^2 = (3^f + 1)(3^f - 1).$$

Liczby $3^f + 1$ i $3^f - 1$ są zatem potęgami dwójki o wykładnikach całkowitych nieujemnych. Różnica tych liczb wynosi 2, a jest tylko jedna para potęg dwójki o tej własności: 4 i 2.

Teraz wystarczy zebrać żniwo tych rozważań. Mamy oczywiście $f = 1$, więc $a = 2$ i $b = 2$. Po bezpośrednim podstawieniu do równania otrzymamy $c = 2$. Nic dziwnego, bo przecież 3, 4, 5 to nasza ulubiona trójka pitagorejska. Poza rozwiązaniem $(2, 2, 2)$ nie ma żadnych innych.

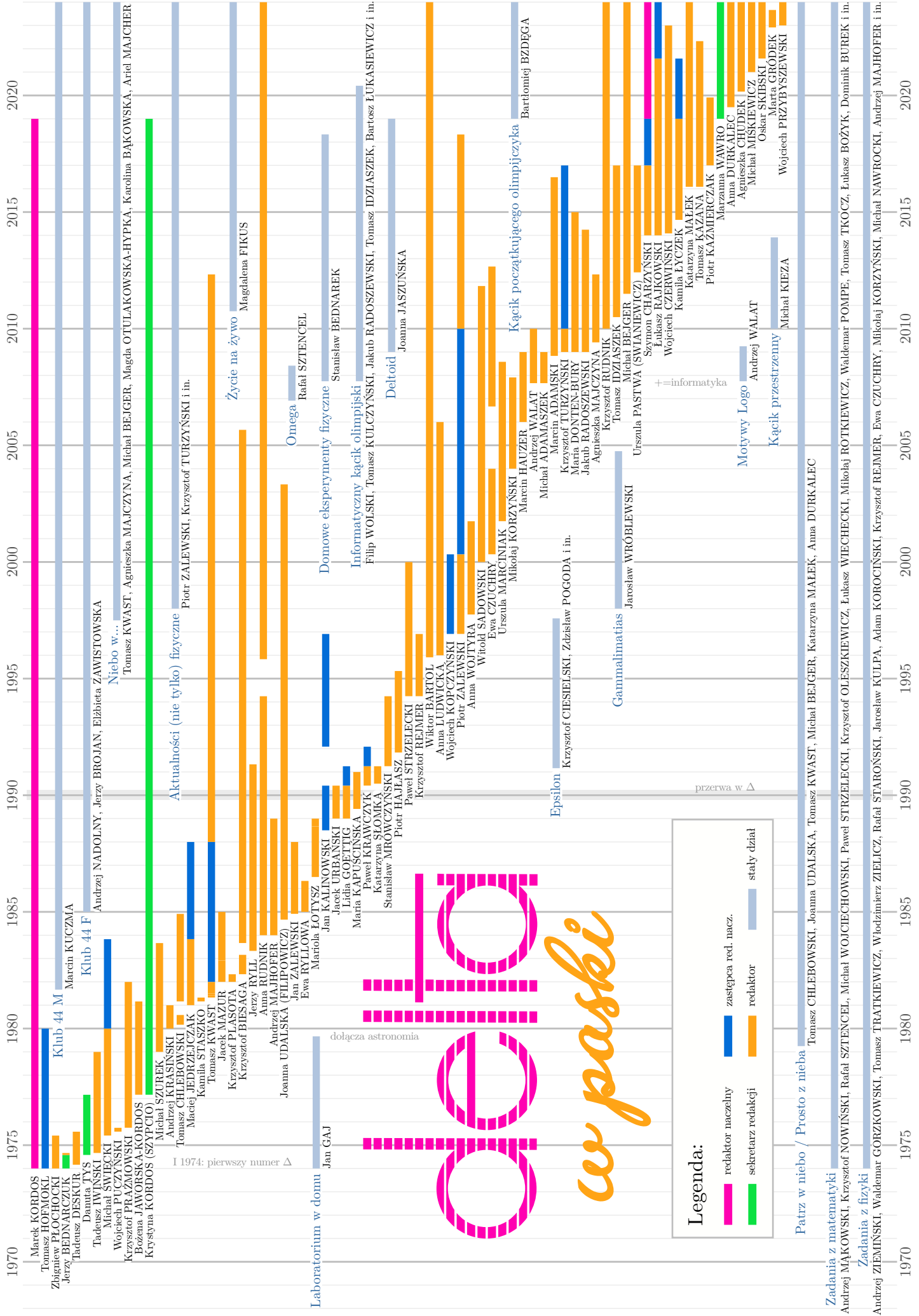
Zakończę dwiema uwagami. Po pierwsze – skąd wiadomo, jakie kongruencje będą działać? Odpowiedź brzmi: zazwyczaj nie wiadomo i trzeba stosować metodę prób i błędów. Są jednak pewne poszlaki. Gdy mamy w równaniu potęgę z konkretną podstawą i niewiadomą w wykładniku, to można spróbować modulo podstawa, ale także podstawa ± 1 . Na pewno warto próbować różnych kongruencji i notować, co można dzięki nim wykazać.

I po drugie – w roku 1844 Eugène Charles Catalan postawił hipotezę, że 8 i 9 to jedyna para kolejnych liczb naturalnych, które są pełnymi potęgami (liczbami postaci a^n , w której $a, n > 1$ są naturalne). Przypuszczenie to udowodnił Preda Mihăilescu w roku 2002. Dowód jest dość żmudny i wymaga zaawansowanych narzędzi matematycznych. Zachęcamy Czytelnika, by na czas rozwiązywania zadań z niniejszego kącika zapomniał o istnieniu tego twierdzenia.

Zadania

1. Wyznaczyć wszystkie liczby naturalne n , dla których $2^n + 1$ jest kwadratem liczby naturalnej.
2. Niech $a > 1$ będzie liczbą naturalną nieparzystą, a n liczbą całkowitą dodatnią. Dowieść, że jeśli $a^n + 1$ jest kwadratem liczby naturalnej, to $n = 1$.
3. Znaleźć wszystkie takie pary (x, y) liczb całkowitych dodatnich, że liczba $2^x + 5^y$ jest kwadratem liczby całkowitej. (63 OM, I etap)
4. Wyznaczyć wszystkie pary liczb całkowitych dodatnich (m, n) , dla których $2^n + 3^m$ jest kwadratem liczby naturalnej.
5. Wyznaczyć najmniejszą wartość wyrażenia $|20^n - 9^m|$ dla całkowitych dodatnich liczb n i m . (64 OM, I etap)
6. Dla jakich liczb naturalnych n oba ułamki: $\frac{1}{n}$ i $\frac{1}{n+1}$, mają skończone rozwinięcia dziesiętne?

Wskazówki do zadań
1. Niech $a = 2 = 2^n + 1$. Wtedy $2^n = (a - 1)(a + 1)$, więc liczbą $a - 1$ i $a + 1$ są potęgami dwójki.
2. Równanie $a^n + 1 = b^2$ jest równoważne równaniu $a^n = (b - 1)(b + 1)$. Liczba b jest parzysta, więc zgodnie z algorytmem Euklidesa $\text{NWD}(b - 1, b + 1) = 2$.
Z tego wynika, że $b - 1 = 2^c$ i $b + 1 = 2^d$ dla pewnych liczb całkowitych dodatnich c i d . Liczby c i d są względnie pierwsze oraz $2^c + 2^d = 2$. Z tego wynika, że $c = d = 1$ i $b = 2$.
3. Rozważmy kwadratowy modulo 5 są $0, 1, 4$ i 2 – pozwała to stwierdzić, że x jest parzyste. Mamy wtedy równanie postaci $5^n = (z + 2^k)(z - 2^k)$ i postępujemy podobnie jak w przytoczonym przykładzie.
4. Kwadrat liczby naturalnej daje resztę 0 lub 1 z dzielenia przez 3, więc nie może być $n = 1$ – dalej zakładamy $n \geq 2$.
Resztami kwadratowymi modulo 4 są 0 i 1, co pozwala wykażać, że m jest parzyste – wtedy mamy równanie postaci $2^n = (x + 3^k)(x - 3^k)$ i dalej postępujemy standardowo.
5. Niech $n = 9^m$ i $20^n = 9^m + b^{2k}$. Niech $r = 3^k$ i 11 lub $r = 1$. Wynika z tego, że $r = 1$ lub $r = 11$.
Równość $20^n - 9^m = 1$, równoważnie $20^n - 1 = 9^m$, jest niemożliwa, bo lewa strona dzieli się przez 19. Jeśli $20^n - 9^m = 11$, to po analizie modulo 7 wykazujemy, że n jest parzyste, co prowadzi do niemożliwej równości $(3^m + 20^k)(3^m - 20^k) = 11$.
Ostatecznie $r = 11$.
6. Ułamek $\frac{1}{n}$ ma skończone rozwinięcie dziesiętne wtedy i tylko wtedy, gdy $n = 2^a 5^b$ dla pewnych liczb całkowitych nieujemnych a i b . W zadaniu dodatkowym mamy $n + 1 = 2^c 5^d$. Mamy zatem równanie $2^c 5^d + 1 = 2^e 5^f$. Przynajmniej jedna z liczb c i d musi być równa zero, podobnie jest z liczbami e i f .
Otrzymujemy stąd $2^c + 1 = 5^f$ i $5^d + 1 = 2^e$. W pierwszym przypadku $2^c \leq 4$ i $5^f \leq 5$, co daje $c = 1$ i $f = 1$. W drugim przypadku $2^c \leq 2$ i $5^f \leq 25$, co daje $c = 2$ i $f = 2$.
Odpowiedź: $n = 1$ lub $n = 4$.



Legenda:

- █ redaktor naczelny
- █ zastępca red. nacz.
- █ sekretarz redakcji
- █ redaktor
- █ stały dział

Przegląd
we paski

Zadania z matematyki
Andrzej MAKOWSKI, Krzysztof NOWIŃSKI, Rafał SZTENCCEL, Michał WOJCIECHOWSKI, Paweł STRZELECKI, Krzysztof OLESZKIEWICZ, Łukasz WIECHECKI, Mikołaj ROTKIEWICZ, Waldemar POMPE, Tomasz TKOCZ, Łukasz BOŻYK, Dominik BUREK i in.

Zadania z fizyki
Andrzej ZIEMIŃSKI, Waldemar GORZKOWSKI, Tomasz TRATKIEWICZ, Włodzimierz ZIELIŃCZ, Rafał STAROŃSKI, Jarosław KULPA, Adam KOROCIŃSKI, Krzysztof REJMER, Ewa CZUCHRY, Mikołaj KORZYŃSKI, Michał NAWROCKI, Andrzej MAJHOFER i in.

Patrz w niebo / Prosto z nieba
Tomasz CHLEBOWSKI, Joanna UDALSKA, Tomasz KWAST, Michał BEJGER, Katarzyna MALEK, Anna DURKALEC

Zadania z fizyki
Krzysztof RUDNIK, Tomasz IDZIASZEK, Michał BEJGER, Urszula PASTWA (SWIANIEWICZ), Szymon CHARZYŃSKI, Lukasz RAJKOWSKI, Wojciech CZERWIŃSKI, Kamila LYCZEK, Katarzyna MALEK, Tomasz KAZANA, Piotr KAZMIERCZAK, Marzanna WAWRO, Anna DURKALEC, Agnieszka CHUDEK, Michał MISKIEWICZ, Michał SKIBSKI, Oskar GRODEK, Wojciech PRZYBYSZEWski

Laboratorium w domu
Jan GAJ

Laboratorium w domu
Jan KALINOWSKI, Janek URBANSKI, Lidia GOETTIG, Maria KAPUSCIŃSKA, Paweł KRAWCZYK, Katarzyna SŁOMKA, Stanisław MROWCZYŃSKI, Piotr HAJLĄSZ, Paweł STRZELECKI, Krzysztof REJMER, Wiktor BARTOL, Anna LUDWICKA, Wojciech KOPCZYŃSKI, Piotr ZALEWSKI, Anna WOJTYRA, Witold SADOWSKI, Ewa CZUCHRY, Urszula MARCINIĄK, Mikołaj KORZYŃSKI, Marcin HAUZER, Andrzej WALAT, Michał ADAMASZEK

Aktualności (nie tylko) fizyczne
Piotr ZALEWSKI, Krzysztof TURZYŃSKI i in.

Życie na żywo
Magdalena FIKUS

Omega
Rafał SZTENCCEL

Domowe eksperymenty fizyczne
Stanisław BEDNAREK

Informacyjny kącik olimpijski
Filip WOLSKI, Tomasz KULCZYŃSKI, Jakub RADOSZEWSKI, Tomasz IDZIASZEK, Bartosz LUKASIEWICZ i in.

Niebo w...
Tomasz KWAST, Agnieszka MAJCZYNA, Michał BEJGER, Magda OTULAKOWSKA-HYPKA, Karolina BĄKOWSKA, Ariel MAJCHER

Andrzej NADOLNY, Jerzy BROJAN, Elżbieta ZAWISTOWSKA

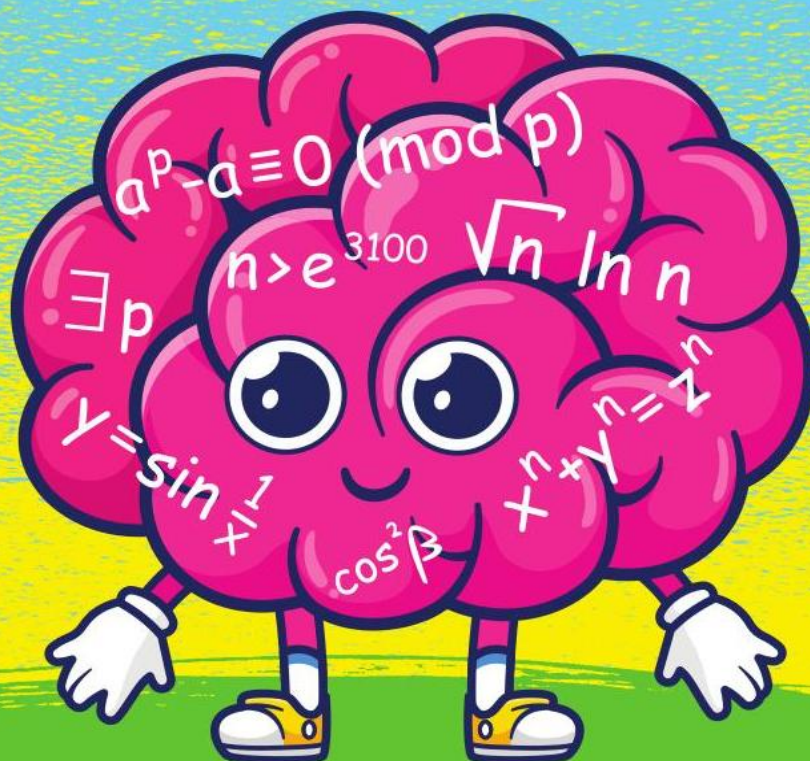
Kłuby
Klub 44 M, Klub 44 F

Redaktor naczelny
Marek KORDOS, Tomasz HOFMANN, Jerzy BŁOCHOCKI, Jerzy BEDNARCZAK, Radosław DESKUL, Damián JYS, Michał IWŃSKI, Tomasz PUZYŃSKI, Wojciech PRĄŻMOWSKI, Bożena JAWORSKA-KORDOS, Krystyna KORDOS (SZYPCIO), Michał SZUREK, Andrzej KRASIŃSKI, Tomasz CHLEBOWSKI, Andrzej JEDRZEJCZAK, Maciej Mactej, Tomasz KWAST, Kamila STASZKO, Jacek MAZUR, Krzysztof PŁASOTA, Krzysztof BIESAGA, Jerzy RYLL, Anna RUDNIK, Anna MAJHOFER (FILIPOWICZ), Jan ZALEWSKI, Ewa RYLLÓWA, Mariola ŁOTYSZ, Joanna UDALSKA

Przerwa w Δ

I 1974: pierwszy numer Δ

dolęca astronomia

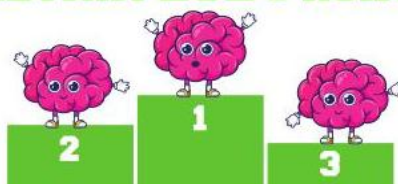


46. KONKURS UCZNIOWSKICH PRAC Z MATEMATYKI IM. PAWŁA DOMAŃSKIEGO

W konkursie biorą udział matematyczne, twórcze prace napisane przez uczniów szkół ponadpodstawowych oraz podstawowych w klasach 7-8. Mówiąc krótko: należy udowodnić coś, czego jeszcze nikt inny nie udowodnił!



**TERMIN ZGŁASZANIA PRAC:
30 KWIECZNIA 2024 ROKU**



Regulamin konkursu, lista dotychczasowych laureatów, niektóre prace oraz wiele innych przydatnych informacji:

www.deltami.edu.pl/delta/redakcja/konkurs_prac_uczniowskich



deltami@mimuw.edu.pl
f Delta.czasopismo