

Ki volt
igazából



Neumann ?
János !

MATEMATIKUS ?
SZÁMÍTÁSTECHNIKUS ?
FIZIKUS ?
KÖZGAZDASZÁRTANOS ?



Nemzeti Tankönyvkiadó

DÁVID GYULA-FORGÓ FERENC-GÖTZ GUSZTÁV-
KOVÁCS GYŐZŐ-LEGENDI TAMÁS-OBÁDOVICS J. GYULA-
SZELEZSÁN JÁNOS-ZALAI ERNŐ

Ki volt igazából Neumann János?

Alkotószerkesztő: Kovács Győző

Nemzeti Tankönyvkiadó, Budapest

Lektorálták
DR. CZELNAI RUDOLF
DÖMÖLKI BÁLINT
DR. GERGELY JÓZSEF
DR. GÖMÖRI ANDRÁS
DR. LÓCS GYULA
DR. MATOLCSI TAMÁS

Felelős szerkesztő
DR. KORECZNÉ KAZINCZI ILONA

Az ábrákat és fotókat készítették
DR. DÁVID GYULA
DR. FORGÓ FERENC
KOVÁCS GYÖZŐ
DR. LEGENDI TAMÁS
DR. OBÁDOVICS J. GYULA

Anyanyelvi lektor
DR. FALUSSY BÉLÁNÉ

ISBN 963 19 3931 6

Előszó

2003-ban ünnepeljük Neumann János születésének 100. évfordulóját. A Neumann János Számítógép-tudományi Társaság, számos iskola és civil szervezet, sőt a kormány is meghirdette a Neumann János-centenáriumú évét. E könyv megjelenése is az ünnepi év egyik jeles eseménye lesz.

Az év eseményeinek fő célja, hogy világszerte mindenki emlékezzék arra, valamikor közöttünk élt egy – Budapesten született – nagy matematikus, aki megszenvedte az elmúlt évszázad tragikus eseményeit, végül emigrált, így tudományos munkájának eredményeit a világ másik részén érte el.

Tragikusan rövid életét – 1957. február 8-án fejezte be.

1903. december 28-án született, a kor számos más kiválóságát is nevelő Fasori Evangélikus Gimnáziumban tanult, ahol Rácz László, matematika tanára azonnal felfigyelt a fiú tehetségére, és engedte, sőt támogatta, hogy ez a tehetség kibontakozzék.

Számos legenda terjed a fasori gimnázium és Neumann János kapcsolatáról, például az egyik, hogy Rácz tanár úr gyakorlatilag felmentette Neumannt a középiskolai matematikaórák látogatása alól. Az órákon való részvétel helyett inkább azt szorgalmazta, hogy a matematikáról eszmét cseréljen olyan kiváló matematikusokkal, mint Kürschák József, a Műegyetem kiváló matematikatanára, később rektora, vagy pedig Fekete Mihály matematikus, később a jeruzsálemi Hebrew University rektora, akivel – 18 éves korában – közösen írta meg élete első matematikai dolgozatát.

Érdekes módon az életéről szóló írásokban az egyetemi tanulmányairól is különféle történeteket lehet olvasni. A biztos pont, hogy édesapja kívánságára beiratkozott a zürichi Eidgenössische Technische Hochschuléra, ahol vegyész-mérnöki tanulmányokat folytatott (1921–1925). Az apa kívánságával nem volt ellentétben, hogy 1921-ben beiratkozott a budapesti tudományegyetemre is, ahol matematikát, fizikát és kémiát tanult, miközben az 1921–1923-as éveket jórészt Berlinben töltötte, ahol részt vett Einstein fizikai szemináriumain. 1925-ben vette át vegyész-mérnöki diplomáját Zürichben, majd 1926-ban visszatért Budapestre, ahol a tudományegyetemen „summa cum laude” doktorált matematikából.

Itt érdemes egy kicsit megszakítani Neumann János életrajzának ismertetését. Neumann életrajzíróinak a legtöbbje megemlíti, hogy ellentétben más nagy matematikusokkal, Neumann János után szinte semmi megoldatlan probléma nem maradt. Amihez hozzáférést, lett légyen az bármiféle, akár nem tisztán matematikai feladat is, megoldotta. Egy amerikai matematikussal beszélgetve, rájöttem a titokra: Neumann János nemcsak matematikus volt, hanem vegyész-mérnök is, ezért, mint mérnök az egyetemen hozzászókot a mérnöki gondolkodásmódhoz. A mérnöknek ugyanis a kapott feladatot meg kell oldania, mert egy félig kiszámolt híd leszakad, egy be nem fejezett elektronikus áramkör nem működik, egy előre ki nem számolt kémiai kísérlet felrobban. Neumann egyik kedvenc mondása – ha egy feladattal elméletileg végzett – a következő volt: *nézzük meg numerikusan is!* – ezzel munkatársait is arra ösztökélte, hogy számolják ki mindazt, amit elméletileg már megoldottak.

Ezért szoktam az előadásaimban Neumann Jánost mérnöki gondolkodású matematikusnak nevezni.

A fiatal tudós Göttingenbe igyekszik, ahol David Hilbert mellett gyűjt tapasztalatokat. Hermann Goldstine – Neumann János barátja – idézi könyvében Lothar Nordheimet,

akivel Neumann Göttingenben együtt dolgozott. Nordheim így jellemzi a két tudóst: ... *Hilbert lassú felfogású volt, Neumann viszont a leggyorsabban gondolkodott mindazok közül, akiket valaha ismertem.*

A húszas évek németországi eseményei a legtöbb tudóst, így Neumann Jánost is Németország elhagyására készítették.

1930–1933 között vendégprofesszor a princetoni egyetemen, 1933-tól 1957-ig (haláláig) a Princetoni Felsőfokú Tanulmányok Intézetének (Institute for Advanced Study) matematikaprofesszora.

Neumann nagyon sok probléma iránt érdeklődött, ezért tevékenysége is szerteágazó volt. Így történhetett meg, hogy több tudományág is *az egyik legnagyobb* alkotójának tekintik. A matematikusok Neumann Jánost tartják a XX. század egyik, ha nem *a legnagyobb* matematikusának. A meteorológia Neumann Jánost tekinti a *numerikus meteorológia*, azaz a matematikai módszerekkel való számítógépes időjárás-előrejelzés megteremtőjének. A közgazdászok többsége azt mondja, hogy az Oscar Morgensternnel közösen kidolgozott *játékelmélete* új irányba fordította a közgazdaság tudományát. A fizika is sokat köszönhet Neumann Jánosnak azzal, hogy megalkotta a *kvantummechanika matematikai alapjait*, a haditudományban pedig kidolgozta az első atombomba felrobbantásánál alkalmazott *berobbanási technikát*. Még az orvostudomány is profitált Neumann János: *A számítógép és az agy* című, utolsó tanulmányából, noha ezt néhányan kétségbe vonják.

Utoljára hagytam Neumann Jánosnak a számítógép kifejlesztésében végzett tevékenységét, aminek a jelentőségét az újkori informatikatörténészek általában vagy túl-, vagy pedig alulértékelik. Sok szerzővel ellentétben – véleményem szerint – Neumann Jánosnak három maradandó alkotása őrzi meg a nevét a számítástechnika legnagyobbjai között:

- Az egyik, a tárolt program elvének kitalálása és alkalmazása, még akkor is, ha az elvet Németországban körülbelül ugyanabban az időben, Konrad Zuse is alkalmazta, de a háború miatt erről a világnak nem vehetett tudomást.
- A másik, a *First Draft ...*, a számítógépnek az első részletes leírása. A dolgozatot egyedül Neumann írta meg és közölte. Ez a tanulmány volt hosszú ideig a számítógépek fejlesztésének az alapidokumentuma. Ma is minden megállapítása helytálló.
- Végül a harmadik, az első modern (mai) architektúrájú, egy című, párhuzamos működésű IAS (Neumann-) számítógép, minden mai számítógépnek a közvetlen *nagyapája*.

Ha a számítástechnikában Neumann János más eredményt nem ért volna el, csak ezt a hármat, már ezzel is az *informatika Parnasszusán lenne a helye*, erről szól ez a könyv.

Kovács Győző

Obádovics J. Gyula

Az első számítógép alkalmazásával megjelenő numerikus problémák

Bevezetés

A numerikus módszerek problémái a számítógép alkalmazása során

A probléma szemléltetése:

1. példa Számítsuk ki a

$$\frac{\left(4\frac{5}{9} - 4\frac{3}{7}\right) \cdot 9\frac{1}{6}}{4\frac{5}{9} + 4\frac{3}{7}} \cdot 10\,000$$

törtet először közös nevezőre hozva az utolsó osztásig, majd az osztással kapott eredményt 4 tizedesjegyre kerekítve szorozzuk 10 000-rel. Másodszor a számítást minden részeredmény két tizedesjegyre kerekítésével végezzük, és az utolsó osztás eredményét 4 tizedesjegyre kerekítve szorozzuk 10 000-rel. A két eredményt hasonlítsuk össze.

$$\begin{aligned} \frac{\left(4\frac{5}{9} - 4\frac{3}{7}\right) \cdot 9\frac{1}{6}}{4\frac{5}{9} + 4\frac{3}{7}} \cdot 10\,000 &= \frac{\left(\frac{41}{9} - \frac{31}{7}\right) \cdot \frac{55}{6}}{\frac{41}{9} + \frac{31}{7}} \cdot 10\,000 = \\ &= \frac{440}{3396} = 0,12956419 \cdot 10\,000 \approx 1296, \end{aligned}$$

$$\begin{aligned} \frac{(4,56 - 4,43) \cdot 9,17}{4,56 + 4,43} \cdot 10\,000 &= \frac{1,19}{8,99} \cdot 10\,000 = \\ &= 0,132\,369\,299\,2 \cdot 10\,000 \approx 1324. \end{aligned}$$

A kétféle számítási eljárás különbsége: $1324 - 1296 = 28$.

Az első eredmény minden jegye értékes jegy a kerekítési szabályok figyelembevételével (pontos megoldás), a második eredményt is a kerekítési szabályok figyelembevételével ké-

peztük, mégis 28-cal nagyobb értéket kaptunk. Az egyes lépések kerekítési hibái halmozódása okozta a második számítással kapott eredmény eltérését a valódi értéktől. A kerekítési hibák elsősorban a szorzás és osztás műveletei révén keletkeznek. Ha olyan feladat megoldásáról van szó, amelyben például az utóbbi eredmény igen sokszor szerepel szorzás és osztás műveletében, elképzelhető, hogy a kerekítési hibák halmozódása az eredményt annyira megfertőzi, hogy egyetlen jegy sem fogadható el értékes jegyként.

Felvetődik a kérdés, hogyan becsülhetjük az eredmény hibakorlátját a szorzás és osztás műveletszáma alapján?

A műszaki-gazdasági feladatok néhány esetben olyan lineáris egyenletrendszerre vezetnek, amelyben esetleg egyetlen együtthatónak mindössze a kerekítési hibával való megváltoztatásával a közelítő gyökök ugrásszerűen megváltoznak. Az ilyen tulajdonságú egyenletrendszereket **gyengén meghatározott** (*kevésbé stabil, rosszul kondicionált*) rendszereknek nevezzük (l. a 6. pontot).

A jelenség, a középiskolai ismeretekre alapozva, kétismeretlenes egyenletrendszer megoldásával, illetve grafikus megoldásával is jól szemléltethető.

2. példa

Oldjuk meg a

$$\left. \begin{array}{l} 2x + 2,000\,01y = 0 \\ 2x + 2y = 1 \end{array} \right\} \quad (1)$$

kétismeretlenes lineáris egyenletrendszert, és ábrázoljuk a két egyenlet által meghatározott egyenest. Vizsgáljuk meg, hogy a kiszámított gyökök mennyire változnak, ha figyelembe vesszük, hogy az első egyenlet y együtthatóját mérési sorozat eredményeként $\pm 2 \cdot 10^{-5}$ hibával kaptuk.

Megoldás

Az első egyenletből vonjuk ki a második egyenletet, és számítsuk ki y értékét:

$$0,000\,01y = -1, \quad \text{és ebből} \quad y = \frac{-1}{0,000\,01} = -99\,999,50.$$

Az y értékét helyettesítsük be a második egyenletbe, és számítsuk ki az x értékét:

$$2x + 2 \cdot (-99\,999,50) = 1, \quad \text{és} \quad x = \frac{1 - (-99\,999,50)}{2} = 100\,000,00.$$

Tehát az egyenletrendszer két gyöke:

$$x = 100\,000,00, \quad \text{és} \quad y = -99\,999,50.$$

A gyökök pontosságáról az egyenletrendszerbe való behelyettesítéssel meggyőződhetünk, mindkét egyenletet azonosan kielégítik.

Most oldjuk meg az egyenletrendszert, ha az első egyenlet y együtthatóját a hibakorlátal megváltoztatjuk: $2,000\,01 - 0,000\,02 = 1,999\,99$ és $2,000\,01 + 0,000\,02 = 2,000\,03$. A

$$\left. \begin{array}{l} 2x + 1,999\,99y = 0 \\ 2x + 2y = 1 \end{array} \right\} \quad (2)$$

egyenletrendszer megoldása: $x = -99\,999,50$, és $y = 100\,000,00$.

Megdöbbenéssel tapasztaljuk, hogy az előjelek ellentétesre változtak, a két gyök felcserélődött. Ne a számítógépben keressük a hibát, és mi sem követtünk el hibát.

A hibakorláttal megnövelt együtthatóval felírt

$$\left. \begin{aligned} 2x + 2,000\,03y &= 0 \\ 2x + 2y &= 1 \end{aligned} \right\} \quad (3)$$

egyenletrendszer gyökeit kiszámítva $x = 33\,333,333\,33$, $y = -33\,332,833\,34$ értékeket kapjuk. A három egyenletrendszer gyökeinek „ugrásszerű” megváltozása egyetlen együtthatójának 2 százezreddel való megváltozása miatt állt elő.

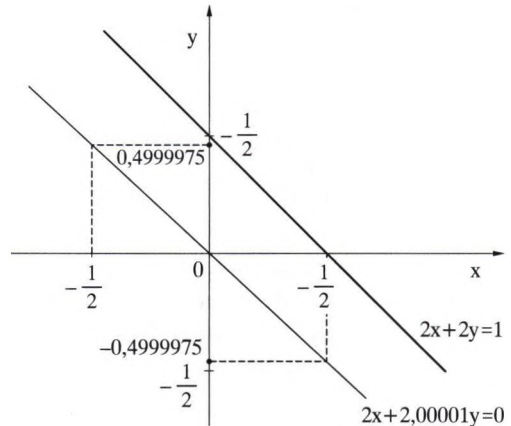
A jelenség az 1. ábra alapján egyszerűen magyarázható. A kétismeretlenes lineáris egyenletrendszer grafikus közelítő megoldását az egyenleteknek megfelelő egyenesek metszéspontja adja. Mivel az egyenletrendszer egyenleteihez tartozó egyenesek majdnem párhuzamosak, így az egyenesek kicsiny iránytangens-változtatás hatására is a metszéspontok nagy eltolódásához vezethetnek, sőt, mint a számítások mutatják, ellentétes irányban is metszhetik egymást.

A példa azt is jelzi, hogy közel merőleges egyenesek esetén az iránytangensek kis megváltoztatása nem vezet a metszéspont, illetve a gyökpár ugrásszerű megváltozásához. Viszont az egyenletrendszer jobb oldalának kicsi megváltozása is előidézheti a gyökök ugrásszerű megváltozását, melyre a következő fejezetekben mutatunk példát.

A két példával szemléltetett számítási problémák hatványozott mértékben jelentkeztek a számítógépek első alkalmazásakor, mivel a különböző tudományterületeken egy időben megjelent feladatok több száz ismeretlen tartalmazó nagyméretű lineáris rendszerek megoldását kívánták meg, és a számítógép műveletsebessége lehetővé is tette e feladatok megoldását.

A számítógépek megjelenése előtt mechanikus számológépek használatával oldották meg a feladatokat. Nagyon nagy műveletszámmal nem számolhattak, hiszen több ezer ismeretlenes egyenletrendszer megoldásába bele sem kezdhettek, pedig a gyakorlat megkövetelte már ilyen feladat megoldását is, mivel az elmélet a megoldási módszereket előállította. A valóságos folyamatokat leíró bonyolult egyenletek megoldását gyakran visszavezették jól közelítő lineáris egyenletrendszerre, mert annak megoldására már a négy alaplétre támaszkodó számos kidolgozott eljárás felhasználható volt.

Az 1940-es években folyó műszaki, fizikai, hadtudományi kutatások olyan problémák megoldását kívánták meg, amelyekhez több ezerszer annyi műveletet kellett elvégezni, mint amennyit korábban a valóságot jól közelítő modelleket leíró egyenletek megoldásához.



1. ábra

A matematika alkalmazása választ várt a következő kérdésekre:

Hány értékes jeggyel kell a kezdőértékeket, együttthatókat megadni, ha a számítási folyamattal előállított végeredmény pontosságát előre megszabtuk?

Milyen feltételeknek kell eleget tennie egy egyenletrendszer együttthatóiból alkotott mátrixnak ahhoz, hogy az egyenletrendszer együttthatóinak kicsi megváltoztatása a megoldásokban is csak kicsi megváltozást okozzon?

Melyek azok a számítási eljárások, amelyek a kerekítési hibák halmozódására kevésbé érzékenyek?

Neumann János zsenialitása nemcsak a nagy műveletsebességű számítógépek megszületését, hanem annak felhasználása által keletkezett matematikai problémák megoldását is elősegítette. Új szemlélettel fogott hozzá a numerikus módszerek alkalmazásának vizsgálatához, figyelemmel arra, hogy melyek azok az eljárások, amelyek több millió művelet elvégzése után is lehetővé teszik a végeredmény hibakorlátjának becslését. Ez azt jelentette, hogy a számítási eljárásokat pontosság és stabilitás szempontjából újra kellett értékelnie. A nagyméretű lineáris rendszerek vizsgálatát a számítógépek fejlesztésével egy időben felvázolta, eredményei és ötletei felhasználásával számos matematikus a numerikus módszerek, a numerikus analízis újabb eredményeit érte el.

Neumann és Goldstine ([1], [2], [3], [4]) munkáiban közzétett eredményeinek ismertetéséhez a mátrixok, determinánsok, mátrixsajátértékek, egyenletrendszerek elméletének és az ortogonalizálási eljárások alapjainak ismeretére van szükség. Ezt a következő fejezetekben a középiskolai matematikai ismeretekre alapozva röviden összefoglaljuk.

1. Mátrixok

A mátrix értelmezése

Elemeknek egy téglalap alakú táblázatban, azon belül sorokban és oszlopokban elrendezett rendszerét **mátrixnak** nevezzük [7].

A mátrix általános alakja:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Az a_{11} , a_{12} , ..., a_{mn} számok (esetleg függvények) a mátrix **elemei**. A kettős indexelés az elemek helyét mutatja. Az a_{ik} elem az i -edik sor k -adik eleme. A fenti mátrixnak m sora és n oszlopa van, ezért szokás azt mondani, hogy $m \times n$ típusú.

A mátrix jelölése általában **A**, **B**, **C** stb. módon (tehát félkövér nagybetűkkel) történik. De más jelölések is használatosak.

A fenti mátrix jelölése lehet pl.:

\mathbf{A} vagy $\mathbf{A}_{(m,n)}$ vagy $\mathbf{A}_{(m,n)}$ vagy $[a_{ik}]$ vagy (a_{ik}) .

Két mátrix egyenlő, ha mindkettő ugyanolyan típusú (méretű), és a megfelelő helyeken álló elemeik egyenlők, azaz

$$\mathbf{A}_{(m,n)} = \mathbf{B}_{(m,n)} \quad \text{ha} \quad a_{ik} = b_{ik} \quad (i = 1, 2, \dots, m; k = 1, 2, \dots, n).$$

Ha az \mathbf{A} mátrix sorait felcseréljük az oszlopaival, a mátrix **transzponáltját** kapjuk. Jelölése: \mathbf{A}^* vagy \mathbf{A}^T vagy \mathbf{A}^t vagy \mathbf{A}' . Például:

$$\mathbf{A} = \begin{bmatrix} 3 & -5 & 2 \\ 6 & 7 & 4 \\ 1 & 5 & 8 \end{bmatrix}, \quad \mathbf{A}^t = \begin{bmatrix} 3 & 6 & 1 \\ -5 & 7 & 5 \\ 2 & 4 & 8 \end{bmatrix}.$$

Ha a mátrixnak ugyanannyi sora van, mint ahány oszlopa, akkor azt **négyzetes (kvadratus) mátrixnak** nevezük. Az n sorból és n oszlopból álló kvadratus mátrixról azt mondjuk, hogy *n-edrendű*.

Ha az \mathbf{A} kvadratus mátrix megegyezik a transzponáltjával, azaz ha

$$\mathbf{A} = \mathbf{A}^t,$$

akkor az \mathbf{A} **szimmetrikus**. Szimmetrikus mátrix elemei a főátlóra nézve szimmetrikus elrendezésűek, azaz: $a_{ik} = a_{ki}$. A főátlót az $a_{11}, a_{22}, \dots, a_{nn}$ elemek alkotják.

Ha a \mathbf{D} kvadratus mátrix főátlóján kívül valamennyi eleme nulla, akkor \mathbf{D} **átlós (diagonális) mátrix**. Például:

$$\mathbf{D} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Ha egy diagonális mátrix főátlójában álló valamennyi elem 1, akkor **egységmátrixnak** nevezük. Jele: \mathbf{E} . Az n -edrendű egységmátrixot \mathbf{E}_n -nel is jelöljük.

$$\text{Például egy harmadrendű egységmátrix: } \mathbf{E}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

A csupa nulla elemből álló mátrixot **zérusmátrixnak (nullamátrixnak)** nevezük, és $\mathbf{0}$ -val jelöljük.

A mátrixok körében kitüntetett szerepük van az egyetlen oszlopból vagy egyetlen sorból álló mátrixoknak. Ezeket **oszlopmátrixoknak**, illetve **sormátrixoknak** nevezük, és legtöbbször félkövér kisbetűvel jelöljük. Szokás oszlopvektornak, illetve sorvektornak is nevezni az ilyen mátrixot. Egy ilyen oszlopvektor pl.:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Ennek transzponáltja egy sorvektor:

$$\mathbf{a}' = [a_1, a_2, \dots, a_n].$$

Vektorok **skaláris szorzatának** megfelelően értelmezzük egy sormátrixnak (sorvektor-nak) egy oszlop mátrixszal (oszlopvektorral) való szorzatát:

$$[a_1, a_2, \dots, a_n] \cdot \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n.$$

Az eredmény egy 1×1 típusú mátrix, amit **számnak** tekintünk.

Példa Számítsuk ki az $\mathbf{a} = [5, -1, 0, 9, 7]$ és $\mathbf{b} = [3, 2, 10, -4, 6]^t$ vektorok skaláris szorzatát!

$$\mathbf{a} \cdot \mathbf{b} = [5, -1, 0, 9, 7] \begin{bmatrix} 3 \\ 2 \\ 10 \\ -4 \\ 6 \end{bmatrix} = 5 \cdot 3 - 1 \cdot 2 + 0 \cdot 10 + 9 \cdot (-4) + 7 \cdot 6 = 19.$$

Mivel a mátrix sorokból és oszlopokból áll, ezért az $\mathbf{A}_{(m,n)} = [a_{ik}]$ mátrix felírható

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n] = \begin{bmatrix} \mathbf{a}^1 \\ \mathbf{a}^2 \\ \vdots \\ \mathbf{a}^m \end{bmatrix}$$

alakban is, ahol $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ oszlopvektorok, $\mathbf{a}^1, \mathbf{a}^2, \dots, \mathbf{a}^m$ pedig sorvektorok, és

$$\mathbf{a}_k = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix}, \quad \mathbf{a}^i = [a_{i1}, a_{i2}, \dots, a_{in}].$$

A gyakorlatban előforduló mátrixok nagy méretűek is lehetnek. Ilyenkor célszerű blokkokra bontani (particionálni).

Példa Az alábbi 4×5 típusú mátrix egyik lehetséges particionálása:

$$\mathbf{A} = \begin{bmatrix} 3 & -6 & 2 & 0 & 7 \\ 0 & 0 & 5 & 4 & 1 \\ 2 & 10 & -4 & 5 & 3 \\ 7 & 1 & 1 & 9 & -2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

$$\text{ahol} \quad \mathbf{A}_{11} = \begin{bmatrix} 3 & -6 & 2 \\ 0 & 0 & 5 \end{bmatrix}, \quad \mathbf{A}_{12} = \begin{bmatrix} 0 & 7 \\ 4 & 1 \end{bmatrix} \text{ stb.}$$

Műveletek mátrixokkal

a) Összeadás

Két mátrix csak akkor adható össze, ha azonos típusúak (méretűek). Mátrixok **összeget** úgy képezzük, hogy az azonos helyen álló (azonos indexű) elemeiket összeadjuk.

Legyen $\mathbf{A} = [a_{ik}]$, $\mathbf{B} = [b_{ik}]$ egy-egy $m \times n$ típusú mátrix. Összegük $(\mathbf{A} + \mathbf{B})$ olyan $\mathbf{C} = [c_{ik}]$, szintén $m \times n$ típusú mátrix, amelyre

$$c_{ik} = a_{ik} + b_{ik}, \quad i = 1, 2, \dots, m; \quad k = 1, 2, \dots, n.$$

Az értelmezés alapján látható, hogy az összeadás kommutatív és asszociatív, azaz $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$,

$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}.$$

Példa

$$\begin{bmatrix} 2 & -3 & 0 \\ 5 & 11 & 7 \end{bmatrix} + \begin{bmatrix} -1 & 2 & 4 \\ 3 & -9 & 6 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 4 \\ 8 & 2 & 13 \end{bmatrix}.$$

Az \mathbf{A} és \mathbf{B} mátrix **különbségét** $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-1)\mathbf{B}$ módon értelmezzük, ami azt jelenti, hogy kivonásnál a mátrixok azonos helyen álló elemeinek különbségét képezzük.

Példa

$$\begin{bmatrix} 2 & -4 & 5 \\ 1 & 3 & 7 \\ 2 & 4 & 6 \end{bmatrix} - \begin{bmatrix} -5 & 6 & 9 \\ 7 & 2 & 3 \\ 4 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 7 & -10 & -4 \\ -6 & 1 & 4 \\ -2 & -1 & 5 \end{bmatrix}.$$

Belátható továbbá, hogy

$$(\mathbf{A} + \mathbf{B})^t = \mathbf{A}^t + \mathbf{B}^t.$$

b) Mátrix szorzása számmal

Mátrixot egy számmal úgy szorzunk, hogy a mátrix mindegyik elemét szorozzuk a számmal.

$$\text{Legyen } \mathbf{A} = [a_{ik}]. \text{ Ekkor } \lambda \mathbf{A} = [\lambda a_{ik}]. \text{ Például } 5 \cdot \begin{bmatrix} 2 & 4 & 7 \\ 1 & 3 & 5 \\ 6 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 10 & 20 & 35 \\ 5 & 15 & 25 \\ 30 & 40 & 45 \end{bmatrix}.$$

c) Mátrixok szorzása

Az \mathbf{A} mátrixnak a \mathbf{B} mátrixszal való \mathbf{AB} szorzata csak akkor értelmezhető, ha \mathbf{A} -nak (a bal oldali tényezőnek) ugyanannyi oszlopa van, mint ahány sora van \mathbf{B} -nek (a jobb oldali tényezőnek). Ebből következik, hogy az $\mathbf{A}_{(m,p)}$ és $\mathbf{B}_{(p,n)}$ mátrixok \mathbf{AB} szorzata értelmezhető. Az $\mathbf{A} = [a_{ik}]$ és $\mathbf{B} = [b_{ik}]$ mátrixok ilyen sorrendben vett szorzata az a $\mathbf{C} = [c_{ik}]$ mátrix, amelyre

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \dots + a_{ip}b_{pk}, \quad i = 1, 2, \dots, m; \quad k = 1, 2, \dots, n,$$

azaz c_{ik} elem az \mathbf{A} mátrix i -edik sorvektorának és a \mathbf{B} mátrix k -edik oszlopvektorának skaláris szorzata: $\mathbf{a}^i \cdot \mathbf{b}_k$.

A C szorzatmátrix $m \times n$ típusú.

Példa

$$\begin{bmatrix} 4 & 2 & -1 \\ 3 & -7 & 1 \\ 2 & 4 & -3 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ -3 & 0 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 4 \cdot 2 + 2 \cdot (-3) + (-1) \cdot 1 & 4 \cdot 3 + 2 \cdot 0 + (-1) \cdot 5 \\ 3 \cdot 2 + (-7) \cdot (-3) + 1 \cdot 1 & 3 \cdot 3 + (-7) \cdot 0 + 1 \cdot 5 \\ 2 \cdot 2 + 4 \cdot (-3) + (-3) \cdot 1 & 2 \cdot 3 + 4 \cdot 0 + (-3) \cdot 5 \end{bmatrix} = \begin{bmatrix} 1 & 7 \\ 28 & 14 \\ -11 & -9 \end{bmatrix}.$$

Az értelmezés alapján nyilvánvaló, hogy a szorzás *nem kommutatív*, azaz

$$\mathbf{AB} \neq \mathbf{BA},$$

viszont *asszociatív*, és érvényes a *disztributív törvény*:

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \quad \text{és} \quad \mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}.$$

A szorzás asszociativitása folytán értelmezhető a kvadratikus mátrix pozitív egész kitevőjű *hatványa*:

$$\mathbf{A}^n = \mathbf{AA} \dots \mathbf{A},$$

ahol a jobb oldalon n darab tényező áll. A zéruskitevőjű hatványt *egységmátrixként* értelmezzük:

$$\mathbf{A}^0 = \mathbf{E}.$$

$$\text{Igazolható, hogy } (\mathbf{AB})^t = \mathbf{B}^t \mathbf{A}^t.$$

Ennek ismeretében belátható, hogy a négyzetes \mathbf{A} mátrix \mathbf{AA}^t szorzata *szimmetrikus mátrix*:

$$(\mathbf{AA}^t)^t = (\mathbf{A}^t)^t \mathbf{A}^t = \mathbf{AA}^t.$$

Itt felhasználtuk azt, hogy az \mathbf{A} transzponáltjának transzponáltja \mathbf{A} -val egyenlő, azaz

$$(\mathbf{A}^t)^t = \mathbf{A}.$$

Ha \mathbf{A} négyzetes mátrix, és \mathbf{E} , ill. \mathbf{O} vele azonos rendű egységmátrix, illetve zérusmátrix, akkor

$$\mathbf{AE} = \mathbf{EA} = \mathbf{A};$$

$$\mathbf{AO} = \mathbf{OA} = \mathbf{O}.$$

2. Determinánsok

A determinánsok értelmezése, kiszámításának műveletszáma

Az a_{11} , a_{12} , a_{21} , a_{22} elemekből képzett **másodrendű determinánson** az $a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$ különbséget értjük, és ezt így jelöljük:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12} a_{21}.$$

Az $a_{11}a_{22} - a_{12}a_{21}$ különbséget a **determináns értékének** is mondjuk.

A **harmadrendű determináns** jelölése és értelmezése:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Az $a_{11}, a_{12}, \dots, a_{33}$ elemek kettős indexelése olyan, hogy a_{ik} a determináns i -edik sorának k -edik eleme (ugyanakkor a k -edik oszlop i -edik eleme). Az a_{11}, a_{22}, a_{33} elemek a determináns **főátlóját**, az a_{13}, a_{22}, a_{31} elemek pedig a **mellékátlóját** alkotják.

Látható, hogy a harmadrendű determinánst másodrendű determinánsok segítségével értelmeztük. A jobb oldali kifejezés értéke kiszámítható, és ez lesz a determináns értéke.

A jobb oldali kifejezés a determináns (egyik) **kifejtése**, az ott szereplő másodrendű determinánsok rendre az a_{11}, a_{12}, a_{13} elemekhez tartozó **aldeterminánsok**. Az a_{ik} elemhez tartozó aldeterminánst úgy kapjuk, hogy a determinánsból töröljük az i -edik sort és a k -edik oszlopot (az a_{ik} elem sorát és oszlopát). Így keletkezik egy másodrendű (al-)determináns.

A jobb oldali kifejezésről szokás azt mondani, hogy az a harmadrendű determinánsnak az **első sora szerinti kifejtése**. Az eddigiek alapján igazolható, hogy akármelyik sor vagy oszlop szerint fejtjük ki a determinánst, annak értéke mindig ugyanaz lesz. Ezt röviden úgy mondjuk, hogy a determináns akármelyik sora vagy oszlopa szerint kifejtethető. A kifejtés előjelszabálya: az a_{ik} elemhez tartozó D_{ik} aldeterminánst meg kell szorozni $(-1)^{i+k}$ -val. Az aldeterminánsokhoz tartozó $+1, -1$ tényezők váltakozását szokás az „előjelek sakktáblaszabályának” nevezni:

$$\begin{bmatrix} + & - & + & \dots \\ - & + & - & \dots \\ + & - & + & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}.$$

Így például az előbbi D determináns második oszlop szerinti kifejtése:

$$D = -a_{12}D_{12} + a_{22}D_{22} - a_{32}D_{32}.$$

A sakktáblaszabálynak megfelelő előjellel ellátott aldeterminánst **előjeles aldeterminánsnak** nevezzük. Így az a_{ik} elemhez tartozó, A_{ik} -val jelölt előjeles aldetermináns:

$$A_{ik} = (-1)^{i+k} D_{ik}.$$

Példa

$$A_{32} = (-1)^{3+2} \cdot \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} = - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}.$$

Ennek megfelelően az előbbi kifejtés így is felírható (2. oszlop szerinti kifejtés):

$$D = a_{12}A_{12} + a_{22}A_{22} + a_{32}A_{32}.$$

A harmadrendű determinánshoz hasonlóan értelmezzük az n -edrendű determinánst.

Ennek n sora és n oszlopa van:

$$D = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Ez a determináns szintén kifejtéssel, $(n-1)$ -edrendű aldeterminánsok segítségével írható fel. Például a D determináns első sor szerinti kifejtése:

$$D = a_{11}D_{11} - a_{12}D_{12} + a_{13}D_{13} - a_{14}D_{14} + \dots + (-1)^{1+n}a_{1n}D_{1n},$$

ahol D_{ik} az a_{ik} elemhez tartozó aldetermináns, amit úgy kapunk, hogy a D determinánsból kitöröljük az i -edik sort és a k -adik oszlopot. Ugyanez a kifejtés:

$$D = a_{11}A_{11} + a_{12}A_{12} + \dots + a_{1n}A_{1n},$$

ahol A_{ik} az a_{ik} elemhez tartozó előjeles aldetermináns. A determináns számértékének ilyen módon való meghatározását a determináns **Laplace-féle kifejtésének** nevezzük.

Innen látható, hogy a determináns értékének kiszámítása igen sok szorzási és összeadási művelettel jár, hiszen az $(n-1)$ -edrendű aldeterminánsok kiszámítása újabb, most már $(n-2)$ -edrendű aldeterminánsok kiszámítását teszi szükségessé.

Példa

Az alábbi determinánst az első sor szerint fejtjük ki:

$$D = \begin{vmatrix} 2 & -3 & 1 \\ -3 & 4 & -2 \\ 5 & 0 & 4 \end{vmatrix} = 2 \cdot \begin{vmatrix} 4 & -2 \\ 0 & 4 \end{vmatrix} + 3 \begin{vmatrix} -3 & -2 \\ 5 & 4 \end{vmatrix} + \begin{vmatrix} -3 & 4 \\ 5 & 0 \end{vmatrix} = 6.$$

A determinánst szokás úgy is értelmezni, hogy az nem más, mint egy négyzetes mátrixhoz rendelt szám.

Egy n -edrendű a_{ik} általános elemű mátrix determinánsának számértéke:

$$\det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \sum (-1)^K a_{1k_1} a_{2k_2} \dots a_{nk_n}$$

formulával kiszámítható, ahol k_1, k_2, \dots, k_n az $1, 2, \dots, n$ oszlopindexek valamely permutációja, K pedig e permutációban levő inverziók száma. Az összegezést az $1, 2, \dots, n$ elemek valamennyi permutációjára ki kell terjeszteni [7]. A determinánsnak ez az értelmezése gyakorlatilag nem alkalmas a tényleges számításra.

Az n -edrendű determináns értékének meghatározásához a Laplace-féle kifejtést használva $n!(n-1)$ szorzást és $n!-1$ összevonást kell végezni, amely összesen $n!n-1$ műveletet jelent. Így pl. egy 10-edrendű determináns értékének kiszámításához $10! \cdot 9 = 32\,659\,200$ szorzást és $10! - 1 = 3\,628\,799$ összevonást kellene végezni, vagyis $10! \cdot 10 - 1 = 36\,287\,999$ műveletet.

A gyakorlati számítások követelményeit kielégítő módszerek közül műveletszám szempontjából rövid ismertetését adjuk a Gauss-féle eliminációs eljárásnak és a *determinánsok szorzástételére* alapozott eljárásnak. Az eljárások lépéseihez felhasználjuk a determinánsok alaptulajdonságait [7].

A Gauss-féle eljárást az

$$A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

n -edrendű determinánsra a következő lépésekkel alkalmazzuk: Tegyük fel, hogy a determináns bal felső sarok eleme zérustól különböző, azaz $a_{11} \neq 0$. (Ha $a_{11} = 0$, akkor sor- vagy oszlopcserevel zérustól különböző elemet viszünk a helyére.) Emeljük ki az a_{11} elemet az első sorból, és minden további sorból vonjuk ki az első sornak az egyes sorok első elemével szorzott alakját, ekkor az eredeti n -edrendű determináns értékének meghatározását az a_{11} elem és egy $(n-1)$ -edrendű determináns szorzatára vezettük vissza, mert az első oszlopban az első elem 1 kivételével minden elem 0.

$$A = a_{11} \begin{vmatrix} 1 & b_{12} & \cdots & b_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} 1 & b_{12} & \cdots & b_{1n} \\ 0 & a_{22,1} & \cdots & a_{2n,1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2,1} & \cdots & a_{nn,1} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22,1} & a_{23,1} & \cdots & a_{2n,1} \\ a_{32,1} & a_{33,1} & \cdots & a_{3n,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2,1} & a_{n3,1} & \cdots & a_{nn,1} \end{vmatrix}.$$

Feltéve, hogy $a_{22,1} \neq 0$, megismételjük az előző eljárást erre az $(n-1)$ -edrendű determinánsra. Az eljárást folytatva, a determináns értékét

$$A = a_{11} \cdot a_{22,1} \cdot \cdots \cdot a_{nn,n-1} = \prod_{\substack{i=1 \\ k=1}}^n a_{i,i-k+1}$$

alakban nyerjük.

Az eljárás folyamán végzett szorzások száma:

$$1 + 2^2 + \cdots + (n-1)^2 + n - 1 = \frac{(n-1)n(2n-1)}{6} + n - 1 = \frac{1}{6}(2n^3 - 3n^2 + 7n) - 1;$$

összevonások száma:

$$1 + 2^2 + \cdots + (n-1)^2 = \frac{(n-1)n(2n-1)}{6} + n - 1 = \frac{1}{6}(2n^3 - 3n^2 + n);$$

$$\text{osztások száma: } 1 + 2 + \cdots + (n-1) = \frac{(n-1)n}{2} = \frac{1}{2}(n^2 - n),$$

$$\text{az összes műveletszám: } \frac{1}{6}(4n^3 - 3n^2 + 5n) - 1.$$

A determináns értékének kiszámítását (az egyes lépések végrehajtásakor) a legnagyobb abszolút értékű elem, az ún. főelem kiemelésével is elvégezhetjük. Ha a főelem a determináns i -edik sora j -edik oszlopában van, akkor $(-1)^{i+j}$ tényezővel kell szoroznunk a kiemelt elemet. A Gauss-féle eljárás a főelem kiemelésével a determináns értékét pontosabban adja, mert a hányadosok relatív és abszolút hibái csökkennek.

A **szorzástételre alapozott eljárás** a mátrix két olyan \mathbf{C} és \mathbf{B} mátrix szorzatára való felbontásán alapszik, amelyek közül az egyik alsóháromszög-mátrix, a másik pedig olyan felsőháromszög-mátrix, melynek főátlójában csupa 1-es áll. Ekkor

$\det \mathbf{A} = \det \mathbf{C} \cdot \det \mathbf{B} = \det \mathbf{C}$, azaz

$$A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} c_{11} & 0 & \dots & 0 \\ c_{21} & c_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{vmatrix} = \begin{vmatrix} 1 & b_{12} & b_{13} & \dots & b_{1n} \\ 0 & 1 & b_{23} & \dots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix}.$$

A C és B determináns elemeit a következő formulákkal számíthatjuk ki:

$$c_{i1} = a_{i1}; \quad b_{ij} = \frac{a_{ij}}{a_{11}}; \quad j > 1$$

$$c_{ij} = a_{ij} - \sum_{k=1}^{j-1} c_{ik} b_{kj}; \quad 1 < j \leq i$$

$$b_{ij} = \frac{1}{c_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} c_{ik} b_{kj} \right); \quad j > i > 1, \quad (i, j = 1, 2, \dots, n), \text{ és}$$

$$A = \det \mathbf{A} = c_{11} c_{22} \dots c_{nn} = \prod_{i=1}^n c_{ii}.$$

A szorzások száma:

$$\begin{aligned} & ((n-1) + 2(n-2) + \dots + (n-2) \cdot 2 + (n-1)) + \\ & + ((n-2) + 2(n-3) + \dots + (n-3) \cdot 2) + (n-2) + (n-1) = \\ & = n-1 + \sum_{k=1}^{n-1} k(n-k) + \sum_{k=2}^{n-1} (k-1)(n-k) = n-1 + n \sum_{k=1}^{n-1} k - \sum_{k=1}^{n-1} k^2 + \\ & + n \sum_{k=2}^{n-1} (k-1) - \sum_{k=2}^{n-1} k^2 + n \sum_{k=2}^{n-1} (k-1) - \sum_{k=2}^{n-1} k^2 + \sum_{k=2}^{n-1} k = \frac{1}{6} (2n^3 - 3n^2 + 7n) - 1. \end{aligned}$$

Az összevonások száma $(n-1)$ -gyel kevesebb a szorzások számánál:

$$\frac{1}{6} (2n^3 - 3n^2 + n);$$

az osztások száma:

$$1 + 2 + \dots + (n-2) + (n-1) = \frac{(n-1)n}{2} = \frac{1}{2} (n^2 - n);$$

$$\text{az összes műveletszám: } \frac{1}{6} (4n^3 - 3n^2 + 5n) - 1.$$

A Gauss-féle és a szorzástételre alapozott eljárás műveletszáma egyenlő.

Egy 10-edrendű determináns értékének kiszámítása a két utóbbi eljárás szerint:

$$\frac{1}{6} (4 \cdot 10^3 - 3 \cdot 10^2 + 5 \cdot 10) - 1 = 624 \text{ művelettel végezhető el.}$$

A kerekítési hibák halmozódását a 624 művelet a Laplace-féle eljárás 36 287 999 műveletszámához képest „érzékenyebben” kevésbé befolyásolja.

3. A négyzetes mátrix inverze

Mátrixok körében az osztás általában nem végezhető el. Bizonyos négyzetes mátrixoknak azonban van inverze (reciproka). Ilyen esetben a négyzetes \mathbf{A} mátrixszal való osztást úgy értelmezzük, mint \mathbf{A} inverzével való szorzást. A skalár aritmetikában ennek az felel meg, hogy az $a \neq 0$ számmal való osztás ekvivalens az a reciprok értékével $\frac{1}{a}$ -val való szorzással.

A négyzetes \mathbf{A} mátrix **inverzén** olyan \mathbf{A}^{-1} -gyel jelölt mátrixot értünk, amely kielégíti az

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{E} \quad \text{és} \quad \mathbf{A}^{-1}\mathbf{A} = \mathbf{E}$$

egyenletet.

Igazolható, hogy az \mathbf{A}^{-1} *inverz mátrix* akkor és csak akkor létezik, ha az \mathbf{A} mátrix determinánsa nem nulla, azaz ha

$$\det \mathbf{A} \neq 0.$$

Ha $\det \mathbf{A} \neq 0$, akkor azt mondjuk, hogy az \mathbf{A} mátrix **reguláris**. Ellenkező esetben **szinguláris**. Inverze tehát csak reguláris mátrixnak van.

$$\text{Az } \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

négyzetes **mátrix adjungáltja** (jelölése: $\text{adj } \mathbf{A}$) az

$$\text{adj } \mathbf{A} := \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix}$$

mátrix, ahol A_{ik} az \mathbf{A} mátrix a_{ik} eleméhez tartozó előjeles aldetermináns. Az $\text{adj } \mathbf{A}$ kiszámítását célszerű a transzponált mátrix képzése után végezni.

Az \mathbf{A} mátrix determinánsának és adjungáltjának ismeretében az **inverz mátrix** kiszámítható:

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \text{adj } \mathbf{A},$$

amely részletesebben felírva:

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix}.$$

Innen az is következik, hogy

$$\mathbf{A} \cdot \text{adj } \mathbf{A} = (\det \mathbf{A})\mathbf{E},$$

ahol \mathbf{E} az \mathbf{A} -val azonos rendű egységmátrix.

Példa

$$\text{Számítsuk ki: } \mathbf{A} = \begin{bmatrix} 2 & 0 & 7 \\ -1 & 4 & 5 \\ 3 & 1 & 2 \end{bmatrix} \text{ mátrix inverzét.}$$

Megoldás \mathbf{A} mátrix reguláris, mert determinánása, $\det \mathbf{A} = -85 \neq 0$. Az adjungált mátrix elemei:

$$A_{11} = \begin{vmatrix} 4 & 5 \\ 1 & 2 \end{vmatrix} = 3, \quad A_{21} = -\begin{vmatrix} 0 & 7 \\ 1 & 2 \end{vmatrix} = 7, \quad A_{31} = \begin{vmatrix} 0 & 7 \\ 4 & 5 \end{vmatrix} = -28,$$

$$A_{12} = -\begin{vmatrix} -1 & 5 \\ 3 & 2 \end{vmatrix} = 17, \quad A_{22} = \begin{vmatrix} 2 & 7 \\ 3 & 2 \end{vmatrix} = -17, \quad A_{32} = -\begin{vmatrix} 2 & 7 \\ -1 & 5 \end{vmatrix} = -17,$$

$$A_{13} = \begin{vmatrix} -1 & 4 \\ 3 & 1 \end{vmatrix} = -13, \quad A_{23} = -\begin{vmatrix} 2 & 0 \\ 3 & 1 \end{vmatrix} = -2, \quad A_{33} = \begin{vmatrix} 2 & 0 \\ -1 & 4 \end{vmatrix} = 8.$$

Tehát az inverz mátrix:

$$\mathbf{A}^{-1} = \frac{1}{-85} \begin{bmatrix} 3 & 7 & -28 \\ 17 & -17 & -17 \\ -13 & -2 & 8 \end{bmatrix}.$$

Szorzással érdemes meggyőződni arról, hogy $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{E}$.

A mátrix sajátértékei

Legyen \mathbf{A} egy n -edrendű mátrix, \mathbf{E} egy \mathbf{A} -val azonos rendű egységmátrix, k pedig egy valós vagy komplex szám. A $k\mathbf{E} - \mathbf{A}$ mátrixot az n -edrendű \mathbf{A} mátrix karakterisztikus mátrixának nevezzük. A k -ra n -edfokú

$$p(k) = \det(k\mathbf{E} - \mathbf{A}) = k^n + c_1 k^{n-1} + c_2 k^{n-2} + \dots + c_{n-1} k + c_n$$

polinomot az \mathbf{A} mátrix karakterisztikus polinomjának, a $p(k) = 0$ egyenletet az \mathbf{A} mátrix karakterisztikus egyenletének, a karakterisztikus egyenlet gyökeit pedig az \mathbf{A} mátrix sajátértékeinek mondjuk.

Példa Az $\mathbf{A} = \begin{bmatrix} 2 & 2,000\,01 \\ 2 & 2 \end{bmatrix}$ másodrendű (2×2) mátrix karakterisztikus egyenlete:

$$\det(k\mathbf{E} - \mathbf{A}) = \det \begin{bmatrix} k-2 & -2,000\,01 \\ -2 & k-2 \end{bmatrix} = 0,$$

melynek kifejtett determinánsát k -ra rendezve $k^2 - 4k - 0,000\,02 = 0$ egyenletet kapjuk. Ennek a másodfokú egyenletnek a gyökei:

$$k_1 = 4,000\,005\,000, \quad k_2 = -0,499\,999\,375\,0 \cdot 10^{-5}$$

az \mathbf{A} mátrix sajátértékei. Látni fogjuk a 6. pontban, hogy a mátrix abszolút értékben legnagyobb és legkisebb sajátértékének hányadosa a mátrix jellemzésére ad lehetőséget. A példában szereplő mátrix nagyobb és kisebb sajátértékeiből képzett

$$\left| \frac{k_1}{k_2} \right| = \left| \frac{4,000\,005\,000}{-0,499\,999\,375\,0 \cdot 10^{-5}} \right| = 800\,002,000\,002\,5$$

hányados nagysága gyengén meghatározott egyenletrendszerre utal. A mátrixelmélet számos alkalmazási területén, így a mátrixsorozatok konvergenciafeltételei között is fontos szerepet tölt be a mátrix sajátértékének nagysága.

Az inverz mátrix öröklött hibája

Becsüljük meg az inverz mátrix öröklött hibáját, vagyis az eredeti adatok hibakorlátjának ismeretében a kiszámított inverz mátrix elemeinek hibakorlátját.

Tegyük fel, hogy az $\mathbf{A}\mathbf{Y} = \mathbf{E}$ n -edrendű mátrixegyenlet \mathbf{A} mátrixának a_{ik} elemei egy n -edrendű \mathbf{B} mátrix elemeit Δa_{ik} hibával közelítik. A bevezetés 2. példájában a

$$\left. \begin{array}{l} 2x + 2,000\,01y = 0 \\ 2x + 2y = 1 \end{array} \right\}$$

kétismeretlenes egyenletrendszer $\mathbf{A}_2 = \begin{bmatrix} 2 & 2,000\,01 \\ 2 & 2 \end{bmatrix}$ mátrixának $\det \mathbf{A}_2$ determinánsa, egyetlen elemét $-0,000\,01$ -del megváltoztatva, nullával lesz egyenlő, azaz a mátrix szingulárisává válik. Így a $\det \mathbf{A} \neq 0$ ellenére előfordulhat, hogy a $\det \mathbf{B} = 0$, vagyis a pontos elemekkel felírt mátrix szinguláris, nem létezik az inverze. Az

$$\mathbf{A}_2^{-1} = \begin{bmatrix} -100\,000,00 & 100\,000,50 \\ 100\,000,00 & -100\,000,00 \end{bmatrix}$$

inverz mátrix elemei az \mathbf{A}_2 mátrix elemeihez képest nagyok, ez is az egyenletrendszer gyengén meghatározottságára utal. Ez akkor is gyanút ébreszt, ha a mátrix inverzével szorozva tiszta egységmátrixot ad:

$$\mathbf{A}_2^{-1} \cdot \mathbf{A}_2 = \begin{bmatrix} 1,000\,0 & 0 \\ 0 & 1,000\,0 \end{bmatrix}.$$

A pontos \mathbf{B} mátrix inverzésére feltételt, a pontos és közelítő mátrix inverzeiben előálló elemek eltérésére pedig becslést adhatunk.

Az \mathbf{A} mátrix legyen invertálható, és inverzét jelölje \mathbf{Y} , azaz $\mathbf{A}^{-1} \equiv \mathbf{Y}$.

A

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix}$$

mátrix és az \mathbf{A} mátrix elemeinek különbsége legyen $b_{ik} - a_{ik} = \Delta a_{ik}$ ($i, k = 1, 2, \dots, n$).

$A |\Delta a_{ik}|$ legnagyobb értékét jelölje ε , azaz $\varepsilon = \max_{i, k} |\Delta a_{ik}|$, az \mathbf{Y} mátrix abszolút értékben legnagyobb elemét jelölje y , azaz $y = \max_{i, k} |y_{ik}|$.

\mathbf{A} \mathbf{B} mátrix inverze létezik, ha teljesül az

$$n^2 y \varepsilon < 1$$

feltétel. Legyen $\mathbf{B}^{-1} = \mathbf{Z}$. \mathbf{A} pontos és a közelítő mátrix inverzének eltérését az

$$|y_{ik} - z_{ik}| \leq \frac{n^2 y^2 \varepsilon}{1 - n^2 y \varepsilon} \quad (i, k = 1, 2, \dots, n)$$

egyenlőtlenséggel becsülhetjük.

Most vizsgáljuk meg, hogy egy fizikai folyamatot leíró egyenletrendszer pontos és közelítő mátrixából milyen következtetéseket vonhatunk le. Legyen a pontos mátrix:

$$\mathbf{B} = \begin{bmatrix} -0,610\,959\,808 & 0,336\,037\,2819 & 14\,929,466\,43 \\ 0,407\,306\,5399 & 0,224\,027\,9363 & 9\,952,840\,539 \\ -0,067\,884\,423\,36 & -0,112\,009\,3432 & -4\,976,625\,767 \end{bmatrix}, \quad \det \mathbf{B} = 0,099\,999\,593\,44,$$

és a jobb oldali szabad vektor:

$$\mathbf{b} = [1,004\,325\,65; 1,043\,268\,51; 2,005\,634\,98]^T.$$

\mathbf{A} közelítő mátrix és a szabad vektor:

$$\mathbf{A} = \begin{bmatrix} -0,6110 & 0,3360 & 14\,929,4664 \\ 0,4073 & 0,2240 & 9\,952,8405 \\ -0,0679 & -0,1120 & -4\,976,6258 \end{bmatrix}, \quad \det \mathbf{A} = 0,050\,008\,066\,68 \text{ és}$$

$$\mathbf{a} = [1,0043; 1,0433; 2,0056]^T.$$

A két mátrix különbsége:

$$\mathbf{B} - \mathbf{A} = \begin{bmatrix} 0,000\,040\,192 & 0,000\,037\,2819 & 0,000\,03 \\ 0,653\,99 \cdot 10^{-5} & 0,000\,027\,9363 & 0,000\,039 \\ -0,442\,336 \cdot 10^{-5} & -0,934\,32 \cdot 10^{-5} & 0,000\,033 \end{bmatrix},$$

Képezzük az inverz mátrixokat:

$$\mathbf{A}^{-1} = \begin{bmatrix} -0,920\,711\,4583 & 0,920\,491\,4938 & -0,921\,171\,3841 \\ 27\,023,25774 & 81\,069,53162 & 243\,199,9079 \\ -0,608\,151\,4847 & -1,824\,499\,247 & -5,473\,452\,948 \end{bmatrix},$$

a \mathbf{B} mátrix stabil inverze létezik, ha $n^2 y \varepsilon < 1$. Mivel

$$\varepsilon = \max_{i,k} |b_{ik} - a_{ik}| = \max_{i,k} |\Delta a_{ik}| = 0,000\,040\,192, \quad y = \max_{i,k} |y_{ik}| = 243\,199,907\,9, \text{ így}$$

$$n^2 y \varepsilon = 3^2 \cdot 243\,199,907\,9 \cdot 0,000\,040\,192 \approx 87,97,$$

ez viszont 1-nél nagyobb, tehát stabil \mathbf{B}^{-1} mátrix nem várható.

$$\mathbf{B}^{-1} = \begin{bmatrix} -0,920\,683\,743\,1 & 0,920\,673\,743\,1 & -0,920\,723\,743\,3 \\ 13\,513,748\,75 & 40\,540,130\,25 & 121\,617,043\,0 \\ -0,304\,142\,544\,0 & -0,912\,452\,749\,0 & -2,737\,433\,604 \end{bmatrix}.$$

Ilyen esetben célszerű a $\mathbf{B}^{-1} \cdot \mathbf{B}$, valamint az $\mathbf{A}^{-1} \cdot \mathbf{A}$ szorzatokat is kiszámítani:

$$\mathbf{B}^{-1} \cdot \mathbf{B} = \begin{bmatrix} 1,000\,00 & 0,223\,80 \cdot 10^{-5} & 0,099\,419 \\ 0,3 \cdot 10^{-5} & 1,000\,01 & 0,2 \\ 0 & -0,1 \cdot 10^{-9} & 1,000\,00 \end{bmatrix} = \mathbf{E},$$

$$\mathbf{A}^{-1} \cdot \mathbf{A} = \begin{bmatrix} 1,000\,00 & 0,223\,96 \cdot 10^{-5} & 0,099\,515 \\ 0 & 1,000\,00 & 0 \\ 0 & 0 & 1,000\,00 \end{bmatrix} = \mathbf{E}.$$

Mindkét esetben a tiszta egységmátrix helyett a főátló 1 értékű elemein kívül, a főátló felett és alatt vannak zérust jól közelítő és zérustól lényegesen különböző elemek (szemetes az egységmátrix). Már a szemetes egységmátrixokból is következtethetünk arra, hogy az \mathbf{A} és \mathbf{B} instabil mátrixok.

Az \mathbf{A}^{-1} és \mathbf{B}^{-1} inverz mátrixok elemeinek különbségére érvényes a következő egyenlőtlenség:

$$|y_{ik} - z_{ik}| \leq \left| \frac{n^2 y^2 \varepsilon}{1 - n^2 y \varepsilon} \right| = \left| \frac{3^2 \cdot 243\,199,907\,9^2 \cdot 0,000\,040\,192}{1 - 3^2 \cdot 243\,199,907\,9 \cdot 0,000\,040\,192} \right| = 245\,996,202\,1.$$

A becslés elég durva, a legnagyobb eltérés az inverz mátrixok két azonos indexű elemére:

$$y_{23} - z_{23} = 243\,199,907\,9 - 121\,617,043\,0 = 121\,582,864\,9 \leq 245\,996,202\,1.$$

\mathbf{A} \mathbf{B} mátrixú egyenletrendszer megoldásvektora:

$$\mathbf{B}^{-1} \cdot \mathbf{b} = [-1,810\,792\,121, \quad 299\,785,841\,4, \quad -6,747\,683\,970]^t.$$

Az \mathbf{A} mátrixú egyenletrendszer megoldásvektora:

$$\mathbf{A}^{-1} \cdot \mathbf{b} = [-1,811\,907\,896, \quad 599\,487,682\,8, \quad -13,491\,973\,44]^t.$$

És a kerekítéssel előállított \mathbf{a} vektorral:

$$\mathbf{A}^{-1} \cdot \mathbf{a} = [-1,811\,823\,070, \quad 599\,481,035\,4, \quad -13,491\,823\,83]^t.$$

A jobb oldali vektor kerekített és nem kerekített értékeivel számolva a megoldásvektor koordinátáinak ötödik-hatodik jegyeiben van eltérés.

b) **Megoldhatóság.** A lineáris egyenletrendszernek

- lehet egyetlen megoldása,
- lehet végtelen sok megoldása,
- lehet, hogy nincs megoldása,

függetlenül attól, hogy mennyi az egyenletek száma és mennyi az ismeretlenek száma.

Tétel. A lineáris egyenletrendszer akkor és csak akkor oldható meg egyértelműen, ha mátrixának és bővített mátrixának lineárisan független oszlopvektorszáma azonos, és megegyezik az ismeretlenek számával.

A mátrixhoz jellemző számként hozzárendeljük a lineárisan független oszlopvektorainak, illetve sorvektorainak számát, s ezt a **mátrix rangjának** mondjuk. E fogalmat felhasználva azt is mondhatjuk, hogy a lineáris egyenletrendszer akkor és csak akkor oldható meg egyértelműen, ha mátrixának és bővített mátrixának rangja megegyezik az ismeretlenek számával.

Megoldási módszerek

a) **Inverz mátrix alkalmazása.** Tekintsük az előző pontban szereplő (1) egyenletrendszert, amelynek mátrixos alakja:

$$\mathbf{Ax} = \mathbf{b}.$$

Legyen ebben az egyenletrendszerben az egyenletek száma és az ismeretlenek száma egyenlő, azaz $m = n$. Ekkor az \mathbf{A} együtthatómátrix négyzetes, n -edrendű. Tételezzük fel, hogy \mathbf{A} reguláris, vagyis az **egyenletrendszer determinánsa:**

$$\det \mathbf{A} \neq 0.$$

Ekkor az \mathbf{A} mátrix rangja n . De ugyanennyi a bővített mátrix rangja is, hiszen ebből a mátrixból is csak legfeljebb n lineárisan független oszlopvektor választható ki. Mivel az ismeretlenek száma is n , ezért az előbbi tétel szerint az egyenletrendszernek egyértelmű (tehát egyetlen) megoldása van.

Mivel $\det \mathbf{A} \neq 0$, létezik az \mathbf{A}^{-1} inverz mátrix. Az $\mathbf{Ax} = \mathbf{b}$ egyenlet mindkét oldalát balról szorozva \mathbf{A}^{-1} -gyel, az egyenlet megoldását:

$$\mathbf{x} = \mathbf{A}^{-1} \mathbf{b} \tag{1}$$

alakban kapjuk, mivel

$$\mathbf{A}^{-1} \mathbf{Ax} = \mathbf{A}^{-1} \mathbf{b}, \quad \mathbf{Ex} = \mathbf{A}^{-1} \mathbf{b}$$

melyből (1) következik. Ha a jobb oldali vektor koordinátáit c_i -vel jelöljük, akkor a két vektor (bal oldali és jobb oldali) egyenlőségéből a megfelelő komponensek egyenlősége következik, azaz $x_i = c_i$, vagyis

$$x_1 = c_1, x_2 = c_2, \dots, x_n = c_n.$$

b) **Cramer-szabály.** Az (1) egyenlőség mindkét oldalának részletes felírása alapján, igazolható, hogy az ismeretlenek előállíthatók a következő formulákkal:

$$x_1 = \frac{D_1}{\det \mathbf{A}}, \quad x_2 = \frac{D_2}{\det \mathbf{A}}, \dots, \quad x_n = \frac{D_n}{\det \mathbf{A}}, \quad (\det \mathbf{A} \neq 0),$$

ahol a D_i módosított determináns az \mathbf{A} mátrix determinánsából úgy származtatható, hogy annak i -edik ($i = 1, 2, \dots, n$) oszlopa helyére a jobb oldali \mathbf{b} oszlopot írjuk, azaz

$$D_i = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1i-1} & b_1 & a_{1i+1} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2i-1} & b_2 & a_{2i+1} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{ni-1} & b_n & a_{ni+1} & \dots & a_{nn} \end{vmatrix}.$$

c) **A Gauss-féle módszer.** A *Cramer*-szabály elvileg nagyon egyszerű módszer, alkalmazása azonban nehézkes. Ugyanis ha az egyenletek (és az ismeretlenek) száma nagy, akkor gyakorlatilag nem alkalmazható, mert a determinánsok kiszámítása hosszadalmas, a műveletszám $n!$ -sal arányos. Ezért több módszert is kidolgoztak, amelyek a gyakorlat számára előnyösebbek a *Cramer*-szabálynál. Egyik ilyen a **Gauss-féle módszer**.

Magát a módszert nem részletezzük. A lényege úgy foglalható össze, hogy az egyenletrendszert, és így annak mátrixát is átalakítjuk. Ez egy-egy egyenlet számmal való szorzásával, majd két vagy több egyenlet összeadásával (kivonásával) valósítható meg. Ha ezeket az átalakításokat alkalmas módon végezzük el, akkor elérhető, hogy bizonyos számú lépés után az egyenletrendszer mátrixa és egyúttal az egyenletrendszer is olyan speciális szerkezetű lesz, amelyből az ismeretlenek igen egyszerűen számíthatók ki. Ilyen mátrix például a háromszögmátrix (amelynek minden olyan eleme zérus, amely a főátló alatt van). Ilyen mátrix az átlós mátrix is, de más, speciális mátrixok is keletkezhetnek. Előnye a módszernek, hogy sem az egyenletek, sem az ismeretlenek számára semmilyen megkötést nem kell tenni.

Az

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \dots & \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{aligned} \right\}$$

egyenletrendszert ($\det \mathbf{A} \neq 0$) soronként végzett műveletekkel úgy alakítjuk át, hogy vele ekvivalens:

$$\left. \begin{aligned} x_1 + c_{12}x_2 + c_{13}x_3 + \dots + c_{1n}x_n &= d_1 \\ x_2 + c_{23}x_3 + \dots + c_{2n}x_n &= d_2 \\ \dots & \\ x_{n-1} + c_{n-1n}x_n &= d_{n-1} \\ x_n &= d_n \end{aligned} \right\}$$

egyenletrendszerhez jussunk. Ebből az x_i értékeit az utolsó egyenlettől az első felé haladva behelyettesítésekkel kiszámíthatjuk:

$$\left. \begin{aligned} x_n &= d_n \\ x_{n-1} &= d_{n-1} - c_{n-1n}x_n \\ \dots & \\ x_1 &= d_1 - c_{12}x_2 - c_{13}x_3 - \dots - c_{1n}x_n \end{aligned} \right\}.$$

d) **Iterációs eljárás és hibabecslés.** Az egyenletrendszerek megoldásának előző három megoldási módszerét pontos módszereknek szokás nevezni. A *Neumann* és munkatársai által végzett vizsgálatok azt mutatták, hogy a numerikus közelítő módszerek:

1. a számítógépre írt programok szempontjából kedvezőbbek,
2. a kerekítési hibák halmozódására kevésbé érzékenyek,
3. az előírt pontosságú gyökök előállításához szükséges lépésszám jól becsülhető,
4. az iterációsorozat konvergenciafeltételei általában könnyen teljesíthetők,
5. az előírt pontosság elérése az egymásután kiszámított közelítő értékek különbsége alapján meghatározható.

A numerikus közelítő módszerek szemléltetéséhez röviden **ismertetjük az egyszerű iterációs módszert lineáris egyenletrendszerekre alkalmazva.** Az eljárásban szereplő becslő formulákhoz az alábbi **mátrix-**, illetve **vektornormák** bármelyikét használhatjuk:

$$1. \| \mathbf{A} \|_I = \max_i \sum_j |a_{ij}|; \quad 2. \| \mathbf{A} \|_{II} = \max_j \sum_i |a_{ij}|; \quad 3. \| \mathbf{A} \|_{III} = \sqrt{\sum_{i,j} |a_{ij}|^2}; \quad \text{ill.}$$

$$1. \| \mathbf{x} \|_I = \max_i |x_i|; \quad 2. \| \mathbf{x} \|_{II} = \sum_{j=1} |x_j|; \quad 3. \| \mathbf{x} \|_{III} = \sqrt{\sum_{j=1} |x_j|^2} (=|\mathbf{x}|).$$

Négyzetes mátrixokra a 3. norma $\| \mathbf{A} \| = \sqrt{sp(\mathbf{A}'\mathbf{A})}$ alakban is írható, mely a transzponálttal szorzott mátrix nyomának (a főátló elemösszegének) a gyöke, valamint használatos a

$$\| \mathbf{A} \| = n \max_{i,j} |a_{ij}|$$

képlettel definiált norma is.

Legyen az egyenletrendszer mátrix alakban

$$\mathbf{Ax} = \mathbf{q}. \quad (1)$$

Tegyük fel, hogy az \mathbf{A} főátlójában lévő elemek zérustól különböznek. Az első egyenletet osszuk el a_{11} -gyel, a második egyenletet a_{22} -vel és így tovább, az utolsót pedig a_{nn} -nel. Ha az így előállt egyenletrendszerből kifejezzük az 1 együtthatójú ismeretleneket, akkor

$$\mathbf{x} = \mathbf{c} + \mathbf{Bx} \quad (2)$$

alakú egyenletrendszert kapunk, amely az (1) rendszer iterációs alakja. Az (1) egyenletrendszer megoldásvektorát a fokozatos közelítés módszerével a (2) egyenletrendszer felhasználásával a következő lépésekkel állítjuk elő:

A nulladik közelítő vektor, $\mathbf{x}^{(0)}$ legyen \mathbf{c} , azaz a (2) szabad vektora:

$$\mathbf{x}^{(0)} = \mathbf{c},$$

az első közelítést a 0. közelítő vektor (2) egyenlet jobb oldalába való behelyettesítéssel képezzük:

$$\mathbf{x}^{(1)} = \mathbf{c} + \mathbf{Bx}^{(0)},$$

a második közelítést az 1. közelítő vektor (2) egyenlet jobb oldalába való behelyettesítéssel képezzük:

$$\mathbf{x}^{(2)} = \mathbf{c} + \mathbf{Bx}^{(1)},$$

így tovább, a $(k+1)$ -edik közelítéshez felhasználjuk a k -edik közelítést, és így az egyszerű iterációs eljárás általános formuláját

$$\mathbf{x}^{(k+1)} = \mathbf{c} + \mathbf{B}\mathbf{x}^{(k)}, \quad (k = 0, 1, 2, \dots), \quad (3)$$

alakban kapjuk. Igazolható, hogy az $\mathbf{x}^{(k)}$ sorozatnak van határértéke:

$$\mathbf{x} = \lim_{k \rightarrow \infty} \mathbf{x}^{(k)},$$

és (3)-ból határértékre áttérve $\mathbf{x} = \mathbf{c} + \mathbf{B}\mathbf{x}$ áll elő, vagyis az \mathbf{x} vektorhatárérték a (2) rendszer megoldása. A (2) rendszer ekvivalens az (1) rendszerrel, tehát az \mathbf{x} vektor az (1) rendszernek is megoldása. Az irodalomban a módszer számos javításával találkozhatunk.

Bizonyított, hogy a (3) iterációs sorozat konvergenciájának szükséges és elegendő feltétele, hogy a \mathbf{B} mátrix sajátértékei abszolút értékben 1-nél kisebbek legyenek, azaz a

$$\det(k\mathbf{E} - \mathbf{B}) = 0$$

karakterisztikus egyenlet gyökeire az

$$|\lambda_i| < 1 \quad (i = 1, 2, \dots, n)$$

feltétel teljesüljön. Az $\mathbf{x}^{(0)}$ kezdővektor tetszőleges lehet.

A (3) iterációs sorozat konvergenciájához sokszor kényelmesebb a $\|\mathbf{B}\| < 1$ feltétel teljesülését ellenőrizni, ahol a három mátrixnorma közül bármelyik szerepelhet az egyenlőtlenségben. Gyakran egyszerűbb az iterációs sorozat konvergenciájának megállapítása az

$$|a_{ii}| > \sum_{k=1}^n |a_{ik}| \quad (i = 1, 2, \dots, n; i \neq k)$$

egyenlőtlenség teljesülése alapján. Ez azt kívánja meg, hogy az \mathbf{A} mátrix minden sorában a főátló eleme abszolút értékben nagyobb legyen, mint a kérdéses sor összes többi elemének abszolút értékkel képzett összege.

Az iterációs sorozattal előállított közelítés hibáját

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{1}{1 - \|\mathbf{B}\|} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| \quad (k \geq 1) \quad (*)$$

illetve

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{\|\mathbf{B}\|}{1 - \|\mathbf{B}\|} \|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|$$

formulákkal becsülhetjük. A formulákból látható, hogy a konvergencia annál jobb, minél kisebb a $\|\mathbf{B}\|$ értéke.

Az utóbbi formulából $\|\mathbf{B}\| = \frac{1}{2}$ értékre:

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|.$$

Ha az egymást követő két közelítő vektor különbségének vektornormája 10^{-m} -ediknél kisebb, vagyis

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| < 10^{-m},$$

akkor $\|\mathbf{B}\| < \frac{1}{2}$ esetében a közelítés hibájára is következik, hogy

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| < 10^{-m}.$$

Ha az

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| \leq \frac{1 - \|\mathbf{B}\|}{\|\mathbf{B}\|} \cdot 10^{-m}$$

egyenlőtlenség teljesül, akkor az $\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq 10^{-m}$ egyenlőtlenségből az

$$|x_j - x_j^{(k)}| \leq 10^{-m} \quad (j = 1, 2, \dots, n)$$

egyenlőtlenséget kapjuk, amiből látható, hogy az egyes gyökök közelítő értékének abszolút hibája 10^{-m} -nél nem nagyobb. Ez azt jelenti, hogy a pontos gyöktől a közelítő gyök legfeljebb csak a tizedesvessző utáni m -edik jegyben térhet el.

A (*) formulából, ha felhasználjuk az egymás után következő becsléseket, megkapjuk a

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{\|\mathbf{B}\|^k}{1 - \|\mathbf{B}\|} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| \quad (**)$$

formulát. Ha a nulladik közelítésnek a \mathbf{c} vektort használjuk, akkor

$$\mathbf{x}^{(1)} = \mathbf{B}\mathbf{c} + \mathbf{c}, \text{ és } \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| = \|\mathbf{B}\mathbf{c}\| \leq \|\mathbf{B}\| \|\mathbf{c}\|,$$

tehát a (**)-ből a pontos és a közelítő vektorkülönbség normájának becslésére az

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{\|\mathbf{B}\|^{k+1}}{1 - \|\mathbf{B}\|} \|\mathbf{c}\| \quad (***)$$

formulát kapjuk. Ha a kerekítési hibával terhelt megoldásvektort $\bar{\mathbf{x}}^k$ -val jelöljük, és $\mathbf{e}^{(k)}$ -val jelöljük azt a vektort, amelynek koordinátái egyenlők $\bar{\mathbf{x}}^{k+1}$ vektor megfelelő koordinátáiban jelentkező kerekítési hibával, akkor

$$\|\mathbf{x} - \bar{\mathbf{x}}^{(k)}\| \leq \frac{\|\mathbf{B}\|^{k+1} \|\mathbf{c}\|}{1 - \|\mathbf{B}\|} + \frac{e}{1 - \|\mathbf{B}\|}$$

egyenlőtlenségből azt a következtetést vonhatjuk le, hogy az $\bar{\mathbf{x}}^{(k)}$ közelítő megoldásvektor eltérése a pontos \mathbf{x} megoldásvektortól tetszőleges kicsivé tehető nagy k és elegendő kicsi $e = \max_k \|e^{(k)}\|$ mellett.

Példa

Oldjuk meg az a) inverz mátrix felhasználásával, b) Cramer-szabállyal, c) Gauss-féle eliminációs eljárással a

$$\left. \begin{array}{l} 2x_1 - x_2 + x_3 = 1 \\ -x_1 + 2x_2 - x_3 = 2 \\ 3x_1 - x_2 - x_3 = 3 \end{array} \right\} \text{ lineáris egyenletrendszer.}$$

Megoldás a) Az egyenletrendszer determinánása:

$$D = \det \mathbf{A} = \begin{vmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 3 & -1 & -1 \end{vmatrix} = 2 \cdot (-3) - (-1) \cdot 4 + 1 \cdot (-5) = -7 \neq 0,$$

tehát az egyenletrendszernek egyértelmű megoldása van. Az \mathbf{A} mátrix adjungáltja:

$$\text{adj } \mathbf{A} = \begin{bmatrix} -3 & -2 & -1 \\ -4 & -5 & 1 \\ -5 & -1 & 3 \end{bmatrix},$$

és így az \mathbf{A} mátrix inverze:

$$\mathbf{A}^{-1} = \frac{\text{adj } \mathbf{A}}{\det \mathbf{A}} = \frac{1}{-7} \begin{bmatrix} -3 & -2 & -1 \\ -4 & -5 & 1 \\ -5 & -1 & 3 \end{bmatrix} = \begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{1}{7} \\ \frac{4}{7} & \frac{5}{7} & -\frac{1}{7} \\ \frac{5}{7} & \frac{1}{7} & -\frac{3}{7} \end{bmatrix},$$

$$\text{és a megoldás: } \mathbf{A}^{-1} \mathbf{A} \mathbf{x} = \frac{1}{-7} \begin{bmatrix} -3 & -2 & -1 \\ -4 & -5 & 1 \\ -5 & -1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 3 & -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

$$\mathbf{A}^{-1} \mathbf{b} = \frac{1}{-7} \begin{bmatrix} -3 & -2 & -1 \\ -4 & -5 & 1 \\ -5 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{10}{7} \\ \frac{11}{7} \\ \frac{2}{7} \end{bmatrix}, \quad \text{tehát} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{10}{7} \\ \frac{11}{7} \\ -\frac{2}{7} \end{bmatrix},$$

a két vektor egyenlőségéből pedig felírhatók az ismeretlenek értékei:

$$x_1 = \frac{10}{7}, \quad x_2 = \frac{11}{7}, \quad x_3 = -\frac{2}{7}.$$

b) A módosított determinánsok, D_1 : a D első oszlopát kicseréljük a \mathbf{b} vektor koordinátaival, D_2 : a D második oszlopát kicseréljük a \mathbf{b} vektor koordinátaival, D_3 : a D harmadik oszlopát kicseréljük a \mathbf{b} vektor koordinátaival, majd kiszámítjuk az így kapott determinánsokat:

$$D_1 = \begin{vmatrix} 1 & -1 & 1 \\ 2 & 2 & -1 \\ 3 & -1 & -1 \end{vmatrix} = -10, \quad D_2 = \begin{vmatrix} 2 & 1 & 1 \\ -1 & 2 & -1 \\ 3 & 3 & -1 \end{vmatrix} = -11, \quad D_3 = \begin{vmatrix} 2 & -1 & 1 \\ -1 & 2 & 2 \\ 3 & -1 & 3 \end{vmatrix} = 2.$$

Tehát az egyenletrendszer megoldása Cramer-szabállyal:

$$x_1 = \frac{D_1}{\det \mathbf{A}} = \frac{-10}{-7} = \frac{10}{7}; \quad x_2 = \frac{D_2}{\det \mathbf{A}} = \frac{-11}{-7} = \frac{11}{7}; \quad x_3 = \frac{D_3}{\det \mathbf{A}} = \frac{2}{-7} = -\frac{2}{7}.$$

c) Az egyenletrendszer bővített mátrixa: $\begin{bmatrix} 2 & -1 & 1 & 1 \\ -1 & 2 & -1 & 2 \\ 3 & -1 & -1 & 3 \end{bmatrix}$. Az első sort osszuk el 2-

vel, és az így kapott sort adjuk hozzá a második sorhoz, háromszorosát pedig vonjuk ki a harmadik sorból:

$$\begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{3}{2} & -\frac{1}{2} & \frac{5}{2} \\ 0 & \frac{1}{2} & -\frac{5}{2} & \frac{3}{2} \end{bmatrix},$$

a második sort szorozzuk meg $\frac{2}{3}$ -dal, és $\frac{1}{2}$ -szeresét vonjuk ki a harmadik sorból:

$$\begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -\frac{1}{3} & \frac{5}{3} \\ 0 & 0 & -\frac{7}{3} & \frac{2}{3} \end{bmatrix},$$

a harmadik sort szorozzuk meg $-\frac{3}{7}$ -del:

$$\begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -\frac{1}{3} & \frac{5}{3} \\ 0 & 0 & 1 & -\frac{2}{7} \end{bmatrix}.$$

Ezzel előállítottuk azt a felsőháromszög-mátrixot, amelynek főátlójában csupa 1-es áll, tehát az eredeti egyenletrendszerrel ekvivalens egyenletrendszer:

$$\left. \begin{array}{l} x_1 - \frac{1}{2}x_2 + \frac{1}{2}x_3 = \frac{1}{2} \\ x_2 - \frac{1}{3}x_3 = \frac{5}{3} \\ x_3 = -\frac{2}{7} \end{array} \right\},$$

melynek megoldása:

$$x_3 = -\frac{2}{7}; \quad x_2 = \frac{5}{3} + \frac{1}{3} \cdot \left(-\frac{2}{7}\right) = \frac{11}{7}; \quad x_1 = \frac{1}{2} + \frac{1}{2} \cdot \frac{11}{7} - \frac{1}{2} \cdot \left(-\frac{2}{7}\right) = \frac{10}{7}.$$

Az egyenletrendszer megoldásának hibabecslése

Az n -edrendű \mathbf{B} mátrix és az őt közelítő \mathbf{A} mátrix inverzeinek elemeire felírt egyenlőt-lenséghez hasonló becslőformulát adhatunk az $\mathbf{Ax} = \mathbf{b}$ lineáris egyenletrendszer $\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$ megoldásának öröklött hibájára is.

Ha $\delta = \max_k |\Delta b_k|$, $x = \max_k |x_k|$ jelölést bevezetjük, akkor a lineáris egyenletrendszer megoldásának öröklött hibáját

$$|\Delta x_i| \leq \frac{ny(\delta + nx\varepsilon)}{1 - n^2 y\varepsilon}$$

egyenlőt-lenséggel becsülhetjük.

Példa. Számítsuk ki a

$$\begin{bmatrix} 6,93 & 0,55 & 1,65 \\ 0,55 & 3,74 & 1,10 \\ 1,65 & 1,10 & 7,81 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3,74 \\ 3,52 \\ 4,62 \end{bmatrix} \quad (1)$$

egyenletrendszer megoldását az **1.** inverz mátrix felhasználásával, és becsüljük meg az öröklött hibát **2.** iterációs eljárással, és határozzuk meg 10^{-5} pontossághoz a lépésszámot, valamint a konvergencia feltételének teljesülését.

$$\mathbf{1. A} = \begin{bmatrix} 6,93 & 0,55 & 1,65 \\ 0,55 & 3,74 & 1,10 \\ 1,65 & 1,10 & 7,81 \end{bmatrix}, \quad \det \mathbf{A} = 183,487667, \quad \mathbf{b} = [3,74; 3,52; 4,62]^t,$$

$\det \mathbf{A} \neq 0$, tehát az egyenletrendszernek létezik megoldása, \mathbf{A} -nak van inverze:

$$\mathbf{A}^{-1} = \begin{bmatrix} 0,152\,595\,542\,0 & -0,013\,518\,619\,76 & -0,030\,334\,463\,84 \\ -0,013\,518\,619\,76 & 0,280\,132\,179\,1 & -0,036\,599\,190\,07 \\ -0,030\,334\,463\,84 & -0,036\,599\,190\,07 & 0,139\,604\,478\,2 \end{bmatrix}.$$

$$\text{A megoldásvektor } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{A}^{-1} \mathbf{b} = \begin{bmatrix} 0,382\,976\,562\,6 \\ 0,766\,417\,374\,4 \\ 0,402\,692\,645\,5 \end{bmatrix}.$$

Most becsüljük meg az egyenletrendszer megoldásának öröklött hibáit:

$$\begin{aligned} |\Delta x_i| &\leq \frac{ny(\delta + nx\varepsilon)}{1 - n^2 y\varepsilon} = \frac{3 \cdot 0,280\,132\,179\,1 \cdot (0,005 + 3 \cdot 0,766\,717\,374\,4 \cdot 0,005)}{1 - 3^2 \cdot 0,280\,132\,179\,1 \cdot 0,005} = \\ &= 0,014\,040\,392\,76, \end{aligned}$$

vagyis az egyes gyökök öröklött hibái nem érik el a 0,02 értéket.

Az \mathbf{a} mátrix M és P kondíciószáma:

$$M = \frac{1}{3} \cdot 3 \cdot 0,2801321791 \cdot 3 \cdot 7,81 = 6,563496956,$$

$$P = \frac{9,338723268}{3,455878845} = 2,702271604.$$

Mindkét kondíciósám kicsi, amiből a mátrix stabil voltára következtethetünk.

2. Az (1) lineáris egyenletrendszer $\mathbf{A} = \begin{bmatrix} 6,93 & 0,55 & 1,65 \\ 0,55 & 3,74 & 1,10 \\ 1,65 & 1,10 & 7,81 \end{bmatrix}$ mátrixának minden sorában

a főátlóbeli elem nagyobb, mint a sor többi elemének összege, vagyis teljesül az egyszerű iterációs eljárás konvergenciájának elégséges feltétele:

$$|a_{ii}| > \sum_{k=1}^3 |a_{ik}|, \quad (i = 1, 2, 3; \quad i \neq k).$$

Az (1) rendszerből a főátlóbeli elemmel való soronkénti osztás után $\mathbf{x} = \mathbf{c} + \mathbf{B}\mathbf{x}$ iterálásra alkalmas alakot képezzük:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0,5396825397 \\ 0,9411764704 \\ 0,5915492957 \end{bmatrix} + \begin{bmatrix} 0 & -0,07936507937 & -0,2380952381 \\ -0,1470588235 & 0 & -0,2941176470 \\ -0,2112676056 & -0,1408450704 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Az I. norma szerint: $\|\mathbf{B}\|_I = \max_i \sum_{j=1}^3 |b_{ij}| = 0,4411764705 < 1$ (a sorösszegek közül a maximális érték 1-nél kisebb), tehát az iterációs sorozat konvergens.

A k lépésszám meghatározásához használjuk fel a 4. fejezet (***) formuláját:

$$\|\mathbf{c}\|_{III} = \sqrt{0,5396825397^2 + 0,9411764704^2 + 0,5915492957^2} = 1,235718803, \text{ és}$$

$$\|\mathbf{B}\|_I = 0,4411764705, \text{ tehát}$$

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leq \frac{\|\mathbf{B}\|_I^{k-1} \|\mathbf{c}\|_{III}}{1 - \|\mathbf{B}\|_I} = \frac{0,4411764705^{k+1} \cdot 1,235718803}{1 - 0,4411764705} < 10^{-5}$$

egyenlőtlenség-rendszert kell k -ra megoldani.

$$(k+1) \cdot \lg 0,4411764705 < \lg \frac{10^{-5} \cdot 0,5588235295}{1,235718803},$$

$$-(k+1) \cdot 0,35538765581 < -5,344644971,$$

és így

$$k+1 > 15,038911553,$$

melyből

$$k > 14,038911553.$$

Tehát legfeljebb 15 iterációs lépéssel a gyököket 10^{-5} pontossággal megkaphatjuk. A megoldásvektor kezdőértéke, vagyis a nulladik közelítése legyen a \mathbf{c} vektor,

$$\mathbf{x}^{(0)} = [0,5396825397; 0,9411764704; 0,5915492957]^t,$$

akkor az első közelítés $\mathbf{x}^{(1)} = \mathbf{c} + \mathbf{B}\mathbf{x}^{(0)}$ formula szerint:

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^{(1)} &= \begin{bmatrix} 0,5396825397 \\ 0,9411764704 \\ 0,5915492957 \end{bmatrix} + \\ &+ \begin{bmatrix} 0 & -0,07936507937 & -0,2380952381 \\ -0,1470588235 & 0 & -0,2941176470 \\ -0,2112676056 & -0,1408450704 & 0 \end{bmatrix} \begin{bmatrix} 0,5396825397 \\ 0,9411764704 \\ 0,5915492957 \end{bmatrix} = \\ &= \begin{bmatrix} 0,3241409240 \\ 0,68782630041 \\ 0,3449717916 \end{bmatrix}. \end{aligned}$$

Az $\mathbf{x}^{(1)}$ vektor behelyettesítésével kiszámítjuk az $\mathbf{x}^{(2)}$ közelítést és így tovább. Az $\mathbf{x}^{(11)}$, $\mathbf{x}^{(12)}$ és $\mathbf{x}^{(13)}$ közelítő vektorok:

$$\mathbf{x}^{(11)} = [0,3829743863; 0,7664145637; 0,4026902864]^t,$$

$$\mathbf{x}^{(12)} = [0,3829773474; 0,7664183882; 0,4026935009]^t,$$

$$\mathbf{x}^{(13)} = [0,3829762785; 0,7664170074; 0,4026923367]^t.$$

A 12. és a 11. közelítő vektor különbsége:

$$\mathbf{x}^{(12)} - \mathbf{x}^{(11)} = [0,29311 \cdot 10^{-5}; 0,38245 \cdot 10^{-5}; 0,32145 \cdot 10^{-5}]^t,$$

melyből látható, hogy már a 12. közelítő vektor koordinátái is az ismeretleneket 10^{-5} -nél kisebb hibával közelítik.

$$\text{Próba: } \mathbf{A} \cdot \mathbf{x}^{(12)} = \begin{bmatrix} 3,740007408 \\ 3,520005169 \\ 4,620009092 \end{bmatrix} = \mathbf{b} \text{ (négy tizedesjegyre kerekített értékekre).}$$

5. Gram–Schmidt-féle ortogonalizálási eljárás

A kétismeretlenes egyenletrendszer (1. példa) grafikus megoldásának vizsgálatából azt a következtetést vontuk le, hogy merőleges, illetve közel merőleges egyenesek metszéspontjának koordinátái jegyszámcsökkentő kerekítés esetében sem szenvednek „ugrásszerű” változást. Az $ax + by = d$ kétismeretlenes lineáris egyenletnek megfelelő egyenes irányvektorának koordinátái $-b$ és a , tehát az irányvektor $\mathbf{e} = [-b, a]$. A háromismeretlenes lineáris egyenletrendszer $ax + by + cz = d$ alakú egyenletei a háromdimenziós térben olyan síkokat ábrázolnak, melyekhez tartozó normális vektorok koordinátáit az ismeretlenek

együtthatói adják, $\mathbf{n} = [a, b, c]$. Ha a három sík nem párhuzamos, akkor a grafikus megoldást a három sík metszéspontjaként kapjuk. A közel párhuzamos normális vektorokkal rendelkező síkok közel párhuzamosak, metszéspontjuk koordinátái az együtthatók kerekítésének hatására ugrásszerűen megváltozhatnak. A jelenség a kétismeretlenes egyenletrendszer megoldásakor tapasztaltakkal megegyező. Az ilyen egyenletrendszereket, mint már említettük, **gyengén meghatározott (rosszul kondicionált) egyenletrendszereknek** nevezük.

Általánosítva: n -ismeretlenes lineáris egyenletrendszer ismeretlenjeinek együtthatóit n -dimenziós vektor koordinátáiként felfogva, azt mondhatjuk, hogy eme vektorok által bezárt szögek koszinuszai, illetve kényelmesebb számítás céljából, koszinusznégyzetei az 1 értékhez közel esnek, akkor gyengén meghatározott egyenletrendszerekkel van dolgunk. Az egyenletrendszerek átalakítása a műveletszám növekedéséhez vezet, ez pedig a kerekítési hibák (öröklött hibák) halmozódásával ugyancsak ronthatja a megoldások pontosságát. A nagy sebességű számítógépek kezdeti alkalmazása kikényszerítette e problémák vizsgálatát. *Neumann* és munkatársai számos cikkben a mátrixokkal, determinánsokkal, egyenletrendszerekkel és az arra visszavezethető differenciál- és integrálegyenletekkel kapcsolatos megoldási módszereket a fenti szempontok alapján vizsgálták [1], [2], [3], [4].

A gyengén meghatározott egyenletrendszerekből jól meghatározott rendszert állíthatunk elő a vektorok merőlegességi feltételének felhasználásával. A vektoralgebrából ismert, hogy két vektor skaláris szorzatát a háromdimenziós térben két vektor abszolút értékének és a közbezárt szög koszinuszának szorzataként definiáljuk, azaz

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| \cdot |\mathbf{b}| \cdot \cos \varphi, \text{ és így a két vektor által bezárt szög: } \cos \varphi = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| \cdot |\mathbf{b}|}.$$

Ebből következik, hogy két zérustól különböző vektor merőleges, ha skaláris szorzatuk nulla.

Ha a vektorok koordinátákkal adottak, akkor a skaláris szorzatot a megfelelő koordináták szorzatösszegeként kapjuk, azaz

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1 \cdot b_1 + a_2 \cdot b_2 + a_3 \cdot b_3.$$

Ezt általánosítva n -dimenziós térre:

$$\mathbf{a} = [a_1, a_2, \dots, a_n], \quad \mathbf{b} = [b_1, b_2, \dots, b_n],$$

koordinátákkal adott vektorok esetén:

$$\mathbf{a} \cdot \mathbf{b} = a_1 \cdot b_1 + a_2 \cdot b_2 + \dots + a_n \cdot b_n.$$

Ha két vektor skaláris szorzata zérus, akkor a két vektort ortogonálisnak nevezük. Ezt felhasználva páronként ortogonális vektorokat tudunk előállítani.

Egy lineárisan független $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ vektorrendszerből az ún. **Gram-Schmidt-féle eljárás**sal páronként egymásra merőleges, más szóval **ortogonális** vektorrendszert állíthatunk elő.

Az első vektor legyen $\mathbf{y}_1 = \mathbf{x}_1$, és a rá merőleges második vektort keressük $\mathbf{y}_2 = \mathbf{x}_2 + a\mathbf{y}_1$ alakban. Az a értékét a két zérustól különböző vektor merőlegességi feltételéből határozhatjuk meg:

$$\mathbf{y}_1 \cdot \mathbf{y}_2 = \mathbf{y}_1 \cdot (\mathbf{x}_2 + a\mathbf{y}_1) = \mathbf{y}_1 \cdot \mathbf{x}_2 + a(\mathbf{y}_1 \cdot \mathbf{y}_1) = 0,$$

amelyből

$$a = -\frac{\mathbf{y}_1 \cdot \mathbf{x}_2}{\mathbf{y}_1 \cdot \mathbf{y}_1},$$

tehát

$$\mathbf{y}_2 = \mathbf{x}_2 - \frac{\mathbf{y}_1 \cdot \mathbf{x}_2}{\mathbf{y}_1 \cdot \mathbf{y}_1} \mathbf{y}_1.$$

Ha jól számoltunk, akkor $\mathbf{y}_1 \cdot \mathbf{y}_2 = 0$. Az \mathbf{y}_1 -re és \mathbf{y}_2 -re ortogonális \mathbf{y}_3 vektort

$$\mathbf{y}_3 = \mathbf{x}_3 + a\mathbf{y}_2 + b\mathbf{y}_1$$

alakban keressük.

Az a és b meghatározására két egyenletet kapunk, ha kikötjük, hogy \mathbf{y}_1 és \mathbf{y}_3 , valamint \mathbf{y}_2 és \mathbf{y}_3 ortogonálisak legyenek, azaz

$$\left. \begin{aligned} \mathbf{y}_1 \cdot \mathbf{y}_3 &= \mathbf{y}_1 \cdot \mathbf{x}_3 + a\mathbf{y}_1 \cdot \mathbf{y}_2 + b\mathbf{y}_1 \cdot \mathbf{y}_1 = 0 \\ \mathbf{y}_2 \cdot \mathbf{y}_3 &= \mathbf{y}_2 \cdot \mathbf{x}_3 + a\mathbf{y}_2 \cdot \mathbf{y}_2 + b\mathbf{y}_2 \cdot \mathbf{y}_1 = 0 \end{aligned} \right\},$$

és figyelembe vesszük, hogy $\mathbf{y}_1 \cdot \mathbf{y}_2 = \mathbf{y}_2 \cdot \mathbf{y}_1 = 0$, a következő egyenletrendszert kapjuk:

$$\left. \begin{aligned} \mathbf{y}_1 \cdot \mathbf{x}_3 + b\mathbf{y}_1 \cdot \mathbf{y}_1 &= 0 \\ \mathbf{y}_2 \cdot \mathbf{x}_3 + a\mathbf{y}_2 \cdot \mathbf{y}_2 &= 0 \end{aligned} \right\}.$$

A második egyenletből:

$$a = -\frac{\mathbf{y}_2 \cdot \mathbf{x}_3}{\mathbf{y}_2 \cdot \mathbf{y}_2},$$

az első egyenletből:

$$b = -\frac{\mathbf{y}_1 \cdot \mathbf{x}_3}{\mathbf{y}_1 \cdot \mathbf{y}_1},$$

tehát

$$\mathbf{y}_3 = \mathbf{x}_3 - \frac{\mathbf{y}_2 \cdot \mathbf{x}_3}{\mathbf{y}_2 \cdot \mathbf{y}_2} \mathbf{y}_2 - \frac{\mathbf{y}_1 \cdot \mathbf{x}_3}{\mathbf{y}_1 \cdot \mathbf{y}_1} \mathbf{y}_1.$$

.....

Hasonlóan kapjuk meg az ortogonális vektorrendszer többi tagját, m -edik alakja:

$$\mathbf{y}_m = \mathbf{x}_m - \frac{\mathbf{y}_{m-1} \cdot \mathbf{x}_m}{\mathbf{y}_{m-1} \cdot \mathbf{y}_{m-1}} \mathbf{y}_{m-1} - \dots - \frac{\mathbf{y}_1 \cdot \mathbf{x}_m}{\mathbf{y}_1 \cdot \mathbf{y}_1} \mathbf{y}_1. \quad (1)$$

Ha mindegyik vektort elosztjuk az abszolút értékével, akkor az új páronként ortogonális vektorrendszer mindegyikének az abszolút értéke egységnyi lesz. Ezt a

$$\mathbf{g}_i = \frac{\mathbf{y}_i}{|\mathbf{y}_i|} \quad (i = 1, 2, \dots, m) \quad (2)$$

új vektorrendszert, melynek elemei **páronként ortogonális egységnyi hosszúságú vektorok ortonormált vektorrendszernek** nevezzük.

Példa A háromdimenziós geometriai tér

$$\mathbf{x}_1 = [1, 1, 1]^t, \quad \mathbf{x}_2 = [2, -1, -1]^t, \quad \mathbf{x}_3 = [2, 3, 1]^t$$

lineárisan független vektorai felhasználásával állítsunk elő egy ortogonális és ortonormált vektorrendszert.

Megoldás

Mivel $\mathbf{x}_1 \cdot \mathbf{x}_2 = \mathbf{x}_2 \cdot \mathbf{x}_3 = 0$, de $\mathbf{x}_1 \cdot \mathbf{x}_3 \neq 0$, így a *Gram-Schmidt*-eljárást alkalmazzuk:

$$\mathbf{y}_1 = \mathbf{x}_1 = [1, 1, 1]^t;$$

$$\mathbf{y}_2 = \mathbf{x}_2 - \frac{\mathbf{y}_1 \cdot \mathbf{x}_2}{\mathbf{y}_1 \cdot \mathbf{y}_1} \mathbf{y}_1 = [2, -1, -1]^t - \frac{0}{3} \cdot \mathbf{y}_1 = [2, -1, -1]^t;$$

$$\begin{aligned} \mathbf{y}_3 &= \mathbf{x}_3 - \frac{\mathbf{y}_2 \cdot \mathbf{x}_3}{\mathbf{y}_2 \cdot \mathbf{y}_2} \mathbf{y}_2 - \frac{\mathbf{y}_1 \cdot \mathbf{x}_3}{\mathbf{y}_1 \cdot \mathbf{y}_1} \mathbf{y}_1 = [2, 3, 1]^t - \frac{0}{6} \cdot \mathbf{y}_2 - \frac{6}{3} \cdot [1, 1, 1]^t \\ &= [0, 1, -1]^t. \end{aligned}$$

Mivel $\mathbf{y}_1 \cdot \mathbf{y}_2 = \mathbf{y}_1 \cdot \mathbf{y}_3 = \mathbf{y}_2 \cdot \mathbf{y}_3 = 0$, így $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ ortogonális vektorhármas. Az $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ ortogonális vektorhármasból (2) alapján kapjuk az ortonormált bázist, mivel

$$|\mathbf{y}_1| = \sqrt{3}, \quad |\mathbf{y}_2| = \sqrt{6}, \quad |\mathbf{y}_3| = \sqrt{2}, \quad \text{így}$$

$$\mathbf{g}_1 = \frac{\mathbf{y}_1}{|\mathbf{y}_1|} = \left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right]^t, \quad \mathbf{g}_2 = \frac{\mathbf{y}_2}{|\mathbf{y}_2|} = \left[\frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}} \right]^t,$$

$$\mathbf{g}_3 = \frac{\mathbf{y}_3}{|\mathbf{y}_3|} = \left[0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]^t.$$

A $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$ vektorok mindegyike egységnyi hosszúságú, és mivel

$$\mathbf{g}_1 \cdot \mathbf{g}_2 = \mathbf{g}_1 \cdot \mathbf{g}_3 = \mathbf{g}_2 \cdot \mathbf{g}_3 = 0,$$

ezért ortonormált vektorhármasat alkotnak. A következő pontban a bevezetés 2. példáján szemléltetjük az ortogonalizálás módszerének előnyeit.

6. A gyengén meghatározott lineáris egyenletrendszerek megoldásáról

A kérdést úgy tették fel, hogy vajon a nagyméretű lineáris egyenletrendszer együttható-mátrixából, annak jellemzőiből, miként következtethetünk az egyenletrendszer gyengén meghatározottságára. A *Neumann* és *Goldstine* eredményeit közlő cikkek hatására számos matematikus azzal az elképzeléssel végezte kutatásait e témakörben, hogy a mátrixokhoz tartozik egy olyan valós szám, amelyből következtetni lehet arra, hogy a vizsgált egyenletrendszer gyengén meghatározott, és ekkor az öröklött hibák halmozódását elkerülendő a megoldást például az ortogonalizálással előállított új ekvivalens egyenletrendszerből határozzák meg.

Ortogonalizáljuk a bevezetés 2. példa rendszerét, mely $\mathbf{Ax} = \mathbf{b}$ mátrix alakban:

$$\begin{bmatrix} 2 & 2,000\,01 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \det \mathbf{A} = -0,000\,02,$$

az egyenletrendszer megoldása: $x = 100\,000,00$; $y = -99\,999,50$.

A sorvektorok által közbezárt szög koszinuszának négyzetes kerekítési hibán belül: $\cos^2 \varphi = 1$, vagyis közelítően párhuzamos egyenesek felelnek meg a két egyenletnek. Láttuk, hogy a 2,000 01 együttható 0,000 02-del megváltoztatva a megoldásban ugrásszerű változást okoz. Vizsgáljuk meg, hogy az ortogonalizálással előállítandó egyenletrendszerben a 2,000 01 együttható megváltoztatása, kevesebb jegyre kerekítése mennyire változtatja meg a gyököket.

Az első sorvektor: $\mathbf{y}_1 = \mathbf{x}_1 = [2, 2,000\,01]$, a második sorvektor: $\mathbf{x}_2 = [2, 2]$

$$\mathbf{y}_2 = \mathbf{x}_2 + a\mathbf{y}_1, \text{ ahol } a = -\frac{\mathbf{y}_1 \cdot \mathbf{x}_2}{\mathbf{y}_1 \cdot \mathbf{y}_1} = \frac{8,000\,02}{8,000\,04} = 0,999\,997\,5, \text{ tehát}$$

$$\mathbf{y}_2 = \mathbf{x}_2 - a \cdot \mathbf{y}_1 = [2, 2] - 0,999\,9975 \cdot [2, 2,000\,01] = [0,500\,0 \cdot 10^{-5}, 0,500\,0 \cdot 10^{-5}].$$

Az ortogonalizált egyenletrendszer és mátrixának determinánása:

$$\begin{bmatrix} 2 & 2,000\,01 \\ 0,500\,0 \cdot 10^{-5} & 0,500\,0 \cdot 10^{-5} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$\det \begin{bmatrix} 2 & 2,000\,01 \\ 0,500\,0 \cdot 10^{-5} & 0,500\,0 \cdot 10^{-5} \end{bmatrix} = -0,000\,01.$$

Az ortogonalizált egyenletrendszer megoldása: $x = 100\,001,00$; $y = -100\,000,00$. A két gyök nem váltott előjelet, jó közelítése az eredeti megoldásnak.

$$\mathbf{A} \begin{bmatrix} 2 & 2,000\,0 \\ 0,500\,0 \cdot 10^{-5} & 0,500\,0 \cdot 10^{-5} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

egyenletrendszer determinánása: $-0,000\,02$; megoldása: $x = 100\,000,00$; $y = -100\,000,00$.

$$A \begin{bmatrix} 2 & 1,99999 \\ 0,5000 \cdot 10^{-5} & 0,5000 \cdot 10^{-5} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

egyenletrendszer determinánása: $-0,00002$; megoldása: $x = 100\,000,00$; $y = -100\,000,00$.

$$A \begin{bmatrix} 2 & 1,99 \\ 0,5000 \cdot 10^{-5} & 0,5000 \cdot 10^{-5} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

egyenletrendszer determinánása: $-0,00001995$; megoldása: $x = 99\,749,37374$;
 $y = -100\,250,6266$.

Két vektor által közrezárt szög koszinuszának négyzete: $\cos^2 \varphi = 0,62813 \cdot 10^{-5}$, azaz közel merőleges egyenesek felelnek meg az egyenletrendszernek még a durva jegyelvágás esetén is. Látható, hogy az ortogonalizált egyenletrendszer $2,00001$ együtthatójának durva megváltoztatására sem változik ugrásszerűen a megoldás.

A fentiekből levonhatjuk azt a következtetést, hogy a lineáris egyenletrendszer bármely megoldásmódszerének megválasztása előtt tájékozódni kell a megoldhatóságról ($\det A \neq 0$), valamint arról, hogy az együtthatók kis megváltoztatása (pl. a kerekítési hibakorláttal azonos megváltoztatása) milyen mértékben változtatja meg a $\det A$ értékét, A^{-1} inverz mátrix elemeit és az egyenletrendszer megoldását.

Az A mátrix és inverzének összehasonlítása alapján azt mondjuk, hogy az inverz mátrix stabil, ha az A elemeinek kicsi megváltoztatása az inverz mátrix elemeiben megfelelő kis változást hoz létre, ellenkező esetben pedig instabil. Ha az A^{-1} stabil, akkor A jól kondicionált, ha A^{-1} instabil, akkor A gyengén kondicionált. Az A^{-1} inverz mátrix stabilitásához szükséges, hogy az $|\det A|$ ne legyen túlságosan kicsi az abszolút értékben legnagyobb együtthatóhoz viszonyítva.

A kondicionáltság mérésére többféle eljárást vezettek be. Azt azonnal láthatjuk, hogy a determináns értéke nem jellemezheti egyértelműen a mátrix kondicionáltságát, hiszen az A és kA mátrix kondicionáltságát, ahol k tetszőleges valós szám, azonosnak kell vennünk, pedig $\det(kA) = k^n \det(A)$. A mátrix kondicionáltságának egyetlen számmal való jellemzésére Neumann vizsgálatait figyelembe véve *A. M. Turing*

$$N = \frac{1}{n} N(A)N(A^{-1}), \quad M = \frac{1}{n} M(A)M(A^{-1})$$

képleteket, *J. Todd* pedig

$$P = \frac{\max |\lambda_i|}{\min |\lambda_i|}, \quad H = \sqrt{\frac{\mu_1}{\mu_n}}$$

képleteket javasolta, ahol

$$N(A) = \sqrt{\text{sp}(A^t A)}, \quad M(A) = n \cdot \max_{ij} |a_{ij}|,$$

továbbá λ_i -k az A mátrix sajátértékei, μ_1 és μ_n az $A^t A$ szorzatmátrix legnagyobb és legkisebb sajátértéke, $\text{sp} A^t A$ az $A^t A$ szorzatmátrix főátlójában lévő elemek összegét jelenti.

Az N , M , P és H értékekre fennállnak az

$$N \leq M \leq n^2 N,$$

$$N \leq H \leq nN,$$

$$P \leq H$$

egyenlőtlenségek.

Vizsgáljuk meg a leírtak alapján a

$$\left. \begin{aligned} 4,007575x_1 + 6x_2 + 5x_3 + 4x_4 &= 19 \\ 6x_1 + 9x_2 + 7x_3 + 6x_4 &= 28 \\ 5x_1 + 7x_2 + 9x_3 + 8x_4 &= 29 \\ 4x_1 + 6x_2 + 8x_3 + 3x_4 &= 21 \end{aligned} \right\} \quad (*)$$

egyenletrendszer. Az egyenletrendszer mátrixa, determinánsa és megoldásvektora:

$$\mathbf{A} = \begin{bmatrix} 4,007575 & 6 & 5 & 4 \\ 6 & 9 & 7 & 6 \\ 5 & 7 & 9 & 8 \\ 4 & 6 & 8 & 3 \end{bmatrix}, \quad \det \mathbf{A} = 0,000100; \quad \mathbf{m} = \begin{bmatrix} 9999,551883 \\ -5982,981809 \\ -226,2398161 \\ -756,4660512 \end{bmatrix}.$$

A determináns zérushoz közel eső értéke alapján várható, hogy az inverz mátrix elemei között nagy értékek is előfordulnak, instabil lesz:

$$\mathbf{A}^{-1} = \begin{bmatrix} -1320000,00 & 790000,00 & 30000,00 & 100000,00 \\ 790000,00 & -472802,75 & -17954,75 & -59848,50 \\ 30000,00 & -17954,75 & -681,75 & -2272,50 \\ 100000,00 & -59848,50 & -2272,50 & -7576,00 \end{bmatrix}.$$

Az inverz mátrix elemei abszolút értékben nagyok az \mathbf{A} mátrix elemeihez képest, vagyis feltehető, hogy az \mathbf{A} gyengén kondicionált mátrix.

Ha az \mathbf{A} mátrix első sorvektorának a második, harmadik és negyedik sorvektorral bezárt szögét rendre φ_{12} , φ_{13} , φ_{14} jelöli, akkor ezek koszinuszainak négyzetei:

$$\cos^2 \varphi_{12} = 0,9991064371; \quad \cos^2 \varphi_{13} = 0,9739827574; \quad \cos^2 \varphi_{14} = 0,9304086028,$$

s mivel a φ_{23} , φ_{24} , φ_{34} szögek koszinusz négyzetei is azonos nagyságrendűek, így ezek alapján is feltételezhetjük, hogy \mathbf{A} gyengén kondicionált mátrix.

Számítsuk ki az \mathbf{A} mátrix kondíciós számait *Turing* és *Todd* által javasolt formulákkal.

$$N = 0,1138254459 \cdot 10^8; \quad M = 4752; \quad P = 0,4498701869 \cdot 10^8; \quad H = 395173,3478.$$

A kondíciós számok igen nagy értékei is azt mutatják, hogy az \mathbf{A} mátrix gyengén kondicionált.

Tekintsük az egész értékre kerekített (\mathbf{A}_1) mátrixú egyenletrendszer, determinánsát és megoldásvektorát:

$$\begin{bmatrix} 4 & 6 & 5 & 4 \\ 6 & 9 & 7 & 6 \\ 5 & 7 & 9 & 8 \\ 4 & 6 & 8 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 19 \\ 28 \\ 29 \\ 21 \end{bmatrix}, \quad \det \mathbf{A}_1 = 1.$$

Megoldásvektora: $[1,1,1,1]^t$ oszlopvektor. A (*) egyenletrendszer megoldása két tizedesjegyre kerekítve: $[9\,999,59; -5\,983,01; -226,24; -756,47]^t$ oszlopvektor. A két megoldás lényeges eltérést mutat. Képezzük a bővített mátrix

$$s_1 = [4, 6, 5, 4, 19], \quad s_2 = [6, 9, 7, 6, 28], \quad s_3 = [5, 7, 9, 8, 29], \quad s_4 = [4, 6, 8, 3, 21]$$

sorvektorainak ortogonális vektorrendszerét *Gram-Schmidt*-eljárás alkalmazásával a számítógép által adott maximális jegyszámmal, és az új mátrixot A_m -mel, az új jobb oldali vektort b_m -mel jelöljük:

$$A_m = \begin{bmatrix} 4 & 6 & 5 & 4 \\ 0,105\,726\,872\,2 & 0,158\,590\,308\,4 & -0,367\,841\,409\,7 & 0,105\,726\,872\,2 \\ -0,638\,554\,216\,9 & -1,457\,831\,325 & -0,132\,530\,120\,5 & 2,361\,445\,783 \\ 0,069\,526\,627\,22 & -0,048\,816\,568\,05 & -0,004\,437\,869\,822 & -0,011\,834\,319\,53 \end{bmatrix}$$

$$b_m = [19,0,002202643172,0,1325301205,0,004437869822]^t.$$

Az ortogonalizált egyenletrendszer megoldásvektora:

$$[0,999\,999\,999\,1; 1,000\,000\,001; 0,999\,999\,999\,8; 0,999\,999\,999\,7]^t.$$

Kerekítve a (**) egyenletrendszerrel megegyező megoldást kaptunk.

Most az A_m mátrix $a_{11} = 4$ elemét cseréljük 4,007 575-re. Az ehhez tartozó determináns: 1,000 383 755, és ezzel a mátrixszal képzett egyenletrendszer megoldásvektora:

$$[0,999\,616\,391\,2; 0,999\,583\,035\,1; 0,999\,599\,712\,8; 0,999\,616\,391\,9]^t,$$

nem szenved ugrásszerű változást. Ellenőrzésül a két értékes jegyre kerekített mátrixszal is végezzük el a megoldást, azaz

$$A_k = \begin{bmatrix} 4,01 & 6,00 & 5,00 & 4,00 \\ 0,11 & 0,16 & -0,37 & 0,11 \\ -0,64 & -1,46 & -0,13 & 2,36 \\ 0,070 & -0,049 & -0,004\,4 & -0,012 \end{bmatrix},$$

melynek determinánsa: 1,017 186 662, és megoldásvektora:

$$[0,993\,178\,971\,9; 0,993\,240\,759\,1; 1,014\,892\,667; 0,995\,861\,108\,6]^t.$$

Ha összehasonlítjuk az ortogonalizált egyenletrendszer három megoldásvektorát a (**) egyenletrendszer $[1,1,1,1]^t$ megoldásvektorával, azt látjuk, hogy egész jegyre kerekítéssel azonosak.

Megjegyzés. *Neumann János* tudományos munkásságának egy vékonyka szelete, amelyről írtunk. Igen jelentősek a matematikai logika, a halmazelmélet, az operátoralgebrák elmélete, a majdnem periodikus függvények, az operációkutatás, a játékelmélet, a nemlineáris parciális differenciálegyenletek területein elért eredményei. *Lánczos Kornéltól* származik az a mondás, hogy **Neumann János nem csupán az egyik legkiválóbb matematikus volt, hanem a legkiválóbb matematikus a 20. században.**

Irodalom

[1] John von Neumann collected works, general editor: A. H. Taub. Volume V. Design of computers, theory of automata and numerical analysis. Pergamon Press, Oxford–London–New York–Paris, 1963

[2] V. Bargmann, D. Montgomery, J. von Neumann: Solution of linear systems of high order. 1946. 421–477. oldal.

[3] J. von Neumann, H. H. Goldstine: Numerical inverting of matrices of high order, 1947. 479–557. oldal.

[4] H. H. Goldstine, J. von Neumann: Numerical inverting of matrices of high order II. 1950. 558–572. oldal.

A témakör alapismereteit az alábbi könyvek tartalmazzák:

[5] Obádovics J. Gyula: Numerikus módszerek és programozásuk. 2. kiadás. Tankönyvkiadó, Budapest, 1977

[6] Obádovics J. Gyula–Szarka Zoltán: Felsőbb matematika. 2. kiadás. SCOLAR Kiadó, Budapest, 2002

[7] Obádovics J. Gyula: Lineáris algebra példákkal. SCOLAR Kiadó, Budapest, 2001

Dávid Gyula

Rejtett paraméterek, avagy a statisztikus fizika és a kvantummechanika matematikai megalapozása

1. Neumann János és a fizika

Neumann János – más tudományokban elért eredményei mellett – a fizika sok területén is maradandót alkotott. Érdeklődése a legmodernebb elméleti kérdésektől a gyakorlati, technikai problémáig terjedt. Összegyűjtött műveinek csaknem negyede fizikai jellegű cikkeket tartalmaz. Persze még hozzászámíthatjuk a határterületeket is, hiszen hidrodinamikai, áramlástan vizsgálati vezették el a meteorológiai kérdések tanulmányozásához, és a légköri jelenségek vizsgálata során is fizikai modelleket használt, fizikai egyenletek megoldásával foglalkozott.

Fizikai érdeklődése igen sokrétű volt. Szorosan együtt dolgozott (előbb Európában, majd az Egyesült Államokban) kora legkiválóbb fizikusaival, többükkel (pl. Wigner Jenővel) közös cikkeket is publikált. Egy-egy újonnan felmerült fizikai kérdésben általában a matematikai kihívást látta meg – az új problémával küszködő, a matematikai eszközöket nehezen megtaláló fizikusok segítségére sietett egy-egy részletkérdés megoldásával, vagy (több esetben is) átfogó tanulmányosorozat, illetve könyv formájában az egész kérdéskör matematikai alapjainak világos tisztázásával. Az előző esetre jó példa találkozása a kozmológiával és az asztrofizikával: néhány publikálatlan kéziratban foglalkozott az általános relativitáselmélet (ma is minden fizikus rémálmát jelentő, igen bonyolult) egyenletei által leírt bizonyos téridőkkel, illetve a csillagok szerkezetére vonatkozó egyenletrendszer megoldásaival. Emellett egy csillagász kollégájával közösen két hosszú és alapos cikket írt a gravitációsan kölcsönható csillagrendszerek dinamikájára, statisztikus viselkedésére vonatkozóan. A második eset egyik példája a többelektronos atomok kvantummechanikájával kapcsolatban Wignerrel közösen végzett munkája. A kvantummechanika alapító atyái a legegyszerűbb atom, a hidrogénatom sikeres leírása után egymásnak ellentmondó eredményekhez jutottak a bonyolultabb, több elektront tartalmazó rendszerek tanulmányozása során. Ami nem volt csoda, hiszen nem egységes matematikai módszert, hanem ad hoc ötleteket használtak. Neumann és Wigner közös cikksorozata használta először a csoportelmélet matematikai módszerét e kérdések tárgyalására, és így sikerült tisztáznuk az atomszerkezet kapcsolatát a téridő szimmetriáival, valamint az elektronok felcserélhetőségével kapcsolatos szimmetriával. Ezt az itt megkezdett kutatási irányt aztán fizikusok és matematikusok egész csoportja vitte tovább a következő években, végleg megalapozva a téridő szimmetriái, a kvantumelmélet alapelvei és az elemi részecskék tulajdonságai közti szoros összefonódást, ami később, a XX. század második felében kibontakozó modern részecskefizika alapjává vált.

Ugyancsak a második eset példája Neumann cikksorozata, majd alapvető könyve, amely a néhány évvel korábban megszületett kvantummechanika szilárd matematikai megalapozását nyújtotta. E művet és a benne felvetett kérdéseket a későbbiekben részletesen tárgyaljuk.

Elméleti fizikus vagy matematikus?

Neumann magát elsősorban matematikusnak tartotta (ezzel kezdi utolsó könyve, „*A számítógép és az agy*” előszavát is: „*Matematikus vagyok, nem pedig neurológus vagy elmeorvos, ezért némi magyarázatra és igazolásra szorul, amire itt vállalkozom.*”). Fizikus kortársai körében azonban ez nem volt magától értetődő. Ők Neumannt a legnagyobb kortárs elméleti fizikusok egyikének tartották – fizikushoz képest szokatlan nagy matematikai képzettséggel, ismeretanyaggal és intuícióval. Amikor 1929-ben Ortvyay Rudolf, a budapesti tudományegyetem frissen kinevezett elméleti fizika professzora saját korábbi helyére Neumann (vagy annak barátját és kollégáját, gyakori szerzőtársát, Wigner Jenőt) ajánlja a szegedi egyetem elméleti fizika katedrájára, természetesnek veszi, hogy a jelöltek elméleti fizikusok: „*Tudományos működésük az elméleti fizikai kutatások előterében álló ún. kvantummechanikára vonatkozik. E nehéz és centrális problémák tisztázásában jelentékeny eredményeket értek el, amiben nagy része volt mély matematikai műveltségüknek...*”. Neumann később, az amerikai szenátus Atomenergia Bizottsága előtti meghallgatásakor így kezdi bemutatkozását: „*Matematikus vagyok és matematikai fizikus.*” Láthatóan tisztában volt azzal a mély hatással, amit működése kora fizikájának fejlődésére gyakorolt.

Igazi elméleti fizikusként viselkedik akkor, amikor – számos kortársához hasonlóan – a XX. század első felében formálódó két nagy elméleti építmény, az általános relativitáselmélet és a kvantummechanika összekapcsolására tesz – inkább jelképes, próbálkozászerű – kísérletet. E cikkek, kéziratok nem tekinthetők komoly elméletnek, a szerzők inkább csak próbálgatják az új fizika adta lehetőségeket. Háttha a korábban távoli területek újszerű összeillesztése adja meg a kulcsot az egyik vagy másik elméletben fellépő furcsa, nehezen érthető jelenségek igazi magyarázatához... Így Neumann is megpróbálkozik az általános relativitáselmélet által leírt egyik szokatlan tér-időben megfogalmazni és megoldani az elektron Dirac-féle hullámegyenletének megfelelőjét. A furcsa matematikai eredményt (a téridő „kvantálódását”) később lelkes levélben meséli el magyarországi levelezőpartnerének.

Neumann János fizikai cikkeiben és nagy könyvében nem elefántcsonttoronyban ülő matematikusként jelenik meg, akinek a fizika csak arra való, hogy érdekes, megoldandó matematikai feladatokat szállítson neki. Neumann nemcsak a matematikai kérdések izgatják, hanem a fizikai elmélet mögöttes tartalma is, az is, hogy egyáltalán miért, milyen fizikai motiváció alapján fordul az elmélet a matematika egyik vagy másik ágához vagy fogalmához. Persze ha kiderül, hogy a matematika adott része (korábbi motiváció, alkalmazói „megrendelés” hiányában) nincs elég részletesen, alaposan kidolgozva, az eszközök hiányosak vagy gyengék, akkor Neumann nem habozik ezt a hibát a tiszta matematika eszközeivel pótolni. Fizikai cikkei ilyenkor a szokásos matematikus definíció–tétel–bizonyítás külsőt veszik fel, és ha a fizikai motivációról nem tudnánk, tiszta matematikai szaklapba

valónak is találhatnánk őket. És valóban, e fizika felvetette kérdéseket Neumann később gyakran továbbfejlesztette, matematikai jellegű műveiben általánosította (pl. az operátor-algebrák és operátorgyűrűk elméletét). De a fizikai cikkekben a matematikai alapozást hamarosan követi a fizikai rész. Itt Neumann kora fizikáját, a formálódó elmélet körüli szenvedélyes vitákat jól és alaposan ismerő fizikusnak mutatkozik. Olykor a technikai részletekbe menően elemzi pl. a kvantummechanikai mérés problematikáját, amely Einstein és Bohr között annyi vita tárgya volt. Neumann a precíz matematikai megalapozás birtokában határozott kijelentéseket tudott tenni, amelyek egyes kérdéseket végleg eldöntöttek, más problémák fejlődését pedig alapvetően új utakra terelték.

Érdekes megfigyelni e cikkekben Neumann viszonyát a kvantumelmélet többi úttörőjéhez, és az azóta is sokat emlegetett „fizikus pongyolaság – matematikusi precizitás” kérdéskörhöz. Mint később részletesen szó lesz róla, Neumann matematikusként mereven elutasította a Dirac által bevezetett „általánosított függvényeket”, a nevezetes Dirac-féle delta-függvényt, matematikailag megalapozatlan, jogtalan extrapolációnak nevezve azt. Alkalomadtán azonban – pl. az atombomba lökéshullámáról szóló egyik cikkében – vérbeli fizikus módjára gátlás nélkül alkalmazta a Dirac-deltát (precíz matematikusként mellékesen megjegyezve, hogy „...*ez tisztán formális eszköz, és könnyen kikerülhető – a szükséges transzformáció és az azt követő számítás azonban természetessé teszi használatát*”). Ugyanakkor sohasem feledkezett meg a kvantummechanika Dirac-féle formalizmusa és a szerző dicséretéről sem. Nagy könyvének előszavában¹ pl. a következőképpen ír erről: „*Dirac mind nemrég kiadott könyvében, mind pedig számos cikkében a kvantummechanikát felülmúlhatatlanul elegánsan, tömören és ugyanakkor invariáns formában fogalmazta meg.*” Ez az a módszer, az a formalizmus, amelyről Neumann néhány oldal múlva megmutatja, hogy matematikailag megalapozatlan! Ő viszont – számos matematikus társától eltérően – érti és érzi az elmélet fizikai lényegét és szépségét is, látja, hogy Dirac módszere győzte meg az akkor még vonakodó fizikusok többségét az elmélet helyességéről és hasznosságáról. Neumann is értékeli e szépséget és heurisztikus erőt. Más matematikus a Dirac-elmélet formális hibáit felfedezvén gögös fölényvel, ellentmondást nem tűrve lesöpörné azt az asztalról. Neumann ehelyett szinte szabadkozik: „*Ezért talán helyénvaló, hogy a mi módszerünk mellett – ez Diracétól lényegesen eltér – néhány érvet sorakoztassunk fel.*” Saját módszerét az elmélet megalapozásának szánja, nem lép fel azzal az ígérennyel, hogy a fizikusok térjenek át Dirac formalizmusáról az övére (nem is tértek át). Könyve későbbi részében viszont részletesen elemez egy mély problémát (a rejtett paraméterek kérdését), amely állandó viták tárgya volt a kvantumelmélet alapító atyái és az elméleti fizika más nagyjai (pl. Einstein) körében. Dirac elegáns formalizmusa ebben a kérdésben nem segít, nem foglal állást. Neumann viszont a saját módszere alapján meg tudja válaszolni a rejtett paraméterek kérdését. Ez a megfelelő érv, a döntő bizonyíték módszere mellett: a fizikai hatékonyság, nem a hamis matematikusi fölény...

Neumann tisztában van a fizika és a matematika bonyolult kapcsolatával, a tudománytörténet azon gyakran ismétlődő eseményeivel, amikor egy új fizikai terület (pl. a mechanika vagy az elektrodinamika) követelte meg bizonyos új matematikai eszközök kifejlesztését.

¹ Az idézetek a könyv magyar kiadásából valók: Neumann János: A kvantummechanika matematikai alapjai. Akadémiai Kiadó, Budapest, 1980. Fordította Sebestyén Ákos.

Az első lépéseket általában fizikusok tették meg – persze pongyolán –, és csak később jött néhány precíz matematikus, hogy elegyengesse a terepet, tisztázza a matematikai alapokat. Fel is teszi a kérdést, vajon ez-e a helyzet a kvantummechanikával kapcsolatban is:

„Arra gondolhatnánk, hogy amint Newton mechanikája maga után vonta az eredeti formájában kétségtelenül ellentmondásos infinitézimális kalkulus fejlődését, úgy a kvantummechanika is a „végtelen sok változó analízisének” új formáját sugallja, vagyis hogy a matematika apparátusát kell megváltoztatni, nem a fizikai elméletet.”

Neumann erre is készen állna, azonban könyvében megmutatja, hogy nincs szükség ilyen forradalmi lépésre, a meglévő matematikai fogalomrendszer (a hiányzó részletek megfelelő kidolgozásával) alkalmas az új fizikai gondolatok precíz kifejezésére. A későbbiekben kitűnt, hogy a kvantummechanikán túllépve, a mezők kvantumelméletében már nem ez a helyzet – élete vége felé Neumann maga is sokat dolgozott e bővebb fizikai elméletek gondolatait matematikailag pontosan és ellentmondásmentesen visszatükröző matematikai eszközök kifejlesztésén.

A kvantumelmélet születésétől kezdve filozófiai jellegű viták kísérték. Ez nem csoda, hiszen ez az elmélet olyan gyökeresen megváltoztatta a valóságról, az elemi objektumokról, a térről és időről, a mozgásról, a fizikai jelenségek megfigyeléséről, leírásáról, a mérésről, a kauzalitásról és még sok egyébről kialakult és mélyen belénk gyökerezett fogalmakat, mint korábban még egyetlen tudományos elmélet sem tette. (Talán csak az evolúciós biológia térhódításának forradalmi hatásával hasonlítható össze.) A fizika a huszadik század elején már produkált egy hasonló szemléleti forradalmat, paradigmaváltást: a speciális relativitáselmélet térről és időről alkotott fogalmai annyira újszerűek voltak, hogy sokan még ma, száz esztendő múltán is idegenkedve és elutasítóan fogadják őket. Ekkoriban, a huszadik század hajnalán vált szokásossá, majd a kvantumelmélet megszületésével egészen általánossá, hogy az elmélet élvonalában dolgozó fizikusok azonnal filozófiailag is értelmezni és kritizálni kezdték a maguk és kollégáik eredményeit. Neumann élénk figyelemmel kísérte ezeket a vitákat, de némi kívülrőlállással is szemlélte. Amikor hozzászólt, nem a filozófus szempontjából tette. Megmutatta, hogy bizonyos (látszat)viták forrása a fogalmak nem pontos definiálása, ezért a vitázó felek ugyanazt a fogalmat különböző értelemben használják. A pontos definícióra pedig leginkább a matematika képes. A kvantummechanika valószínűségi értelmezésével kapcsolatos neumann-i analízis ennek a hozzáállásnak a mesterei példája: anélkül, hogy explicit módon bírálná valamelyik szembenálló álláspontot, saját fogalomrendszerének párhuzamos szóbeli és matematikai definiálására alapozva olyan brilliáns konstrukciót mutat be, amely egyértelműen elvezeti a régóta vitatott – és filozófiailag is sokat diszkutált – „rejtett paraméter” probléma véglegesnek szánt megoldásához.

A tudománytörténet fintora, hogy Neumann mégsem zárhatta le véglegesen ezt a problémakört. Természetesen felmerül a kérdés, hogy ha Neumann János matematikailag precízen bebizonyította: nincsenek rejtett paraméterek – akkor hogy lehet az, hogy a huszadik század második felében még annyian és olyan sokszor elővették ezt a problémát, csiszoltak, finomítottak rajta, és még napjainkban is vitatják. Csak ennyi ereje van egy matematikai bizonyításnak, csak ekkora a tekintélye Neumann Jánosnak? Cikkünk későbbi részében megkísérelünk válaszolni erre a kérdésre.

Neumann a fizika területén működött. Alapvető, mély és maradandó hatást két területen gyakorolt a XX. századi fizika fejlődésére. Az egyik terület a statisztikus fizika megalapozása, „az ergodikus tétel” bizonyítása, a másik pedig az új tudományág, a kvantummechanika szilárd matematikai megalapozása. E két témával a későbbiekben részletesen foglalkozunk.

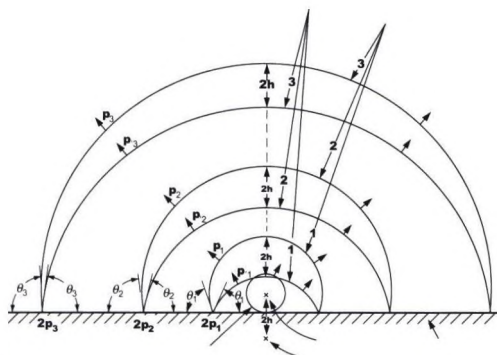
Mielőtt azonban erre rátérnénk, meg kell említenünk Neumann János és a fizika, mi több, a technika egy kevésbé elvont, vaskosan gyakorlati kapcsolatát is.

2. Lökéshullámok és az atombomba

Neumann részt vett az atombomba kifejlesztésén dolgozó titkos Manhattan-tervben is. Ennek kapcsán kezdte meg a hidrodinamikai jelenségek, ezen belül a lökéshullámok keletkezésének, terjedésének és visszaverődésének vizsgálatát. Eredményeit titkos belső jelentésekben publikálta, és csak a háború után, a titkosítás feloldása után tette közzé tudományos folyóiratokban. Jó néhány tanulmánya, jelentése még hat évvel halála után, összegyűjtött műveinek kiadása idején is titkosítva volt!

A lökéshullámok terjedésének problémája két helyen is kapcsolódott az atombomba-programhoz. A nukleáris láncreakció megindításához a hasadásra képes anyag megfelelő mennyiségét (ez az ún. kritikus tömeg) egy helyre kell hozni, és megfelelő ideig együtt kell tartani. Az utóbbi a nehezebb – az ismeretterjesztő könyvekben látható szemléletes modell, amikor két uránból készült félgömböt hagyományos robbanóanyaggal egymásnak lönek, biztosan nem működik. A meginduló láncreakció, a felszabaduló hő és a keletkező nyomás azonnal szétdobná a két félgömböt, és leállna a reakció. Ezért Los Alamosban azal kísérleteztek, hogy gömb alakban rendezzék el a hasadóanyag kis adagjait, és köréje szintén gömb alakban helyezték el a hagyományos robbanóanyagot. A „berobbanás” (implózió) megfelelő mértékben összesűríti a hasadóanyagot, és a nyomás megfelelő ideig együtt is tartja, amíg a láncreakció ténylegesen bekövetkezik. Ez volt az ötlet, a végrehajtás sokkal nehezebb. Több Los Alamosban dolgozó kutató emlékirataiban olvasható, hogy időnként robbanást lehetett hallani, majd valahol az utcán krumpli formára deformálódott, eredetileg gömb alakú tartályokra bukkantak. A robbanás nem sikerült pontosan gömbszimmetrikusra – és a tartalom összesűritése helyett a tartály deformálására, no meg messzire repítésére használandó el a löpor energiája. Neumannnak az implózió fizikájára vonatkozó elméleti számításai és a kísérleti csoport kitarató munkája végül meghozta az eredményt.

A lökéshullámok terjedésének és hatásainak elmélete egy másik fontos alkalmazását az atombomba hatásainak elemzése (és ezeknek még a bomba tényleges kipróbálása előtti megjósálása) területén találta meg. Neumann részletes számításokat végzett a várható lökéshullám paramétereinek meghatározására, valamint a lökéshullámok sík felületről (pl. a talajról, házak faláról vagy a robbanás közelében levő hegyoldalról) történő visszaverődésének elemzésére is. Ez a téma igen kritikus a bomba alkalmazhatósága tekintetében – végül is ezen múlik, eléri-e az atombomba a tervezett pusztító hatást.



1. ábra. Az atombomba lökéshullámának terjedése és visszaverődése. Ábra Neumann János egyik Los Alamos-i kutatási jelentéséből 1947-ből. A jelentés nagy része még 1963-ban, Neumann összes műveinek amerikai kiadása idején is titkosítva volt. Az ábrán szereplő feliratok fordítása és magyarázata: A bomba a levegőben robban, h magasságban (real explosion – valódi robbanás). A talaj (rigid ground: szilárd talaj) elérő lökéshullám (emerging wave – beeső hullám) visszaverődik. A visszavert hullám úgy modellezhető, mintha a föld alatt, a bomba tükröképének helyén egy másik robbanásból (virtual explosion – látszólagos robbanás) indult volna ki. A távoli megfigyelőt (vagy áldozatot) a kétféle hullám kombinációja éri el.

azt a kritikus kérdést, hogy milyen körülmények között fog a bombarobbanás okozta lökéshullám a szokásos, „akusztikai”, illetve a Neumann által levezetett különleges formában visszaverődni egy szilárd felületről vagy egy másik lökéshullámról. Részletes elemzést készített a Föld felszíne fölött magasan felrobbantott atombomba kezdetben gömb alakú lökéshullámának torzulásáról, amint az összetalálkozik a felszínről visszaverődött másik hullámmal.

E cikkek és jelentések némelyike konkrét numerikus számításokat is tartalmaz. A cikkek végén táblázatok, milliméterpapírra rajzolt függvénygrafikonok foglalják össze az eredményeket. E számításokat részben „kézzel”, részben az akkoriban rendelkezésre álló mechanikus számológépekkel végezték. Bizonyára ez a hatalmas emberi erőfeszítést igénylő, ám teljesen mechanikus jellegű munka is hozzájárult ahhoz, hogy Neumann figyelme később a számítások automatizálása, az elektronikus számítógépek irányába fordult.

A katonai céllal végzett fizikai számítások, a hidrodinamika egyenleteivel, modelljeivel, numerikus módszereivel való többéves bensőséges kapcsolat a háború után arra ösztönözte Neumannt, hogy e tapasztalatokat, módszereket és eredményeket a békés, polgári célú kutatás, a meteorológia, az időjárás-előrejelzés szolgálatába állítsa. Későbbi hidrodinamikai és meteorológiai tárgyú cikkeiben könnyű megtalálni a folytonosságot, a lökéshullámok fizikájának tanulmányozása során összegyűlt tapasztalatokat, olykor szó szerint átvett idézeteket titkos tanulmányaiból. Meteorológiai munkásságának jelentős része olyan numerikus módszerek, számítási algoritmusok kidolgozására vonatkozott, amelyek lehetővé tennék a bonyolult hidrodinamikai egyenletek ésszerű időn belüli (ma úgy mondanánk:

Neumann lökéshullámokra vonatkozó cikkei nem elméleti matematikai, hanem inkább mérnöki jellegűek. A fizikai alapok felvázolása, az egyenletek megkonstruálása, illetve a (korábban nem túlságosan részletesen kidolgozott) fizikai modellek ismertetése után konkrét modellekre vonatkozó részletes számítások következnek. Neumann világosan felvázolja az adott modell fizikai előfeltevéseit, elemzi a megengedhető matematikai közelítéseket, és eredményeit összehasonlítja az irodalomban korábban publikált, „a fizikusi józan észre” alapozott következtetésekkel. E korábbi sejtések sok esetben hibásak voltak, olykor pont az ellenkezőjük bizonyult igaznak. Neumann e meglepő eredmények elemzése során nem habozott kilépni a matematikus elefántcsonttoronyából: szorosban együttműködött az amerikai haditengerészet és a légierő, valamint a Los Alamos-i nukleáris kutatóközpont munkatársai által folytatott robbantási kísérletek tervezőivel, végrehajtóival és elemzőivel. Együtt vizsgálták meg

real time, azaz a vizsgált jelenség lefolyásával egyidejű vagy azt megelőző) numerikus megoldását, így az egyenletek által leírt folyamat előrejelzését. E numerikus algoritmusok első képviselői a lökéshullámok terjedésére vonatkozó cikkekben bukkantak fel.

Mondhatjuk tehát, hogy Neumann János, az igen absztrakt matematikai fogalmak embere adott esetben földönjáró, gyakorlati problémákkal is megbirkózó mérnökként is megállta a helyét. Ebben hasonlított barátjára, a fizikus Wigner Jenőre, aki a kvantummechanika és az elméleti magfizika eredményes művelése mellett az első atombombák plutóniumát legyártó első ipari atomreaktorok tervezőjeként, mérnökeként is beírta nevét az atomkorszak történetébe.

3. Az ergodikus tétel

Az „*ergodikus hipotézis*”, majd bizonyítása után „*ergodikus tétel*” nevet viselő állítás a fizika egyik gyorsan fejlődő, számtalan alkalmazásban hasznosnak bizonyult ágának, a statisztikus fizikának sarkköve. Története Maxwellnek és Boltzmann-nak, a kinetikus gázelmélet alapítóinak munkásságával kezdődött. Az i-re a pontot Neumann János tette fel.

A kinetikus gázelmélet

A XIX. század közepén kialakult kinetikus gázelmélet a fizika egy teljes ágát, a hőjelenségek elméletét, a termodinamikát próbálta meg visszavezetni egy másik, sokak által alapvetőbbnek érzett, matematikailag alaposan kidolgozott és jól bevált tudományterületre, a (klasszikus) mechanikára. Először csak az ideális gázok iskolából is jól ismert törvényeit magyarázták meg, majd áttértek a bonyolultabb rendszerek, folyadékok és szilárd anyagok termodinamikájának megalapozására is. Ezt a visszavezetést a kortársak megnyugtónak találták – hiszen a mechanikát ismerjük és értjük! – ugyanakkor a kritikus szemű fizikusok matematikai hiányosságokat találtak a kínált magyarázatban.

Az alapötlet igen egyszerű. Tekintsünk egy tartályba zárt gázt. A termodinamika kidefinitte, hogy milyen fizikai mennyiségekkel (termodinamikai szakzsargonban „állapotjelzőkkel”: nyomás, hőmérséklet, térfogat, tömeg) írhatjuk le tulajdonságait, és ezek között milyen kapcsolat áll fenn (ezek az ún. állapotegyenletek). Emellett szabályokat és képleteket ad annak leírására, mi történik, ha egyik vagy másik paramétert meg akarjuk változtatni, mennyi hőt vagy munkát kell ehhez közölnünk a rendszerrel.

A *kinetikus gázelmélet* a tartályba zárt gázt nem folytonos közegnek, hanem apró részecskék, molekulák, illetve atomok együttesének tekinti – e részecskék között vákuum van. A gáz makroszkopikus tulajdonságai, viselkedése e mikroszkopikus szerkezet és az elemi objektumok mozgása alapján érthetők meg (innen az elmélet nevében a „kinetikus”, azaz „mozgási” jelző). Ma ez a kép triviálisnak tűnik – a XIX. század közepén azonban csak igen közvetett, jórészt kémiai és (Faradaynek köszönhetően) elektrokémiai bizonyí-

tékok voltak az atomok létezéséről. Sokan az atomhipotézist pusztán játéknak, matematikai feltevésnek tekintették, ami megkönnyíti a kémiai számításokat, de nem szabad neki realitást tulajdonítani. (Ez a helyzet kísértetiesen megismétlődött az 1960-as években, amikor a proton és nukleon építőkövei, a kvarkok realitását nem akarta elfogadni a fizikus tudományos közösség.) Éppen a kinetikus gázelmélet, majd a statisztikus fizika sikere és sokrétű alkalmazhatósága lett a legjobb bizonyíték az atomok létezése mellett – eme elméletek módszereinek finom és ötletes alkalmazásával mérte meg a XX. század elején Jean Perrin francia fizikus az Avogadro-számot, és határozta meg ezáltal az addig csak hipotetikus atomok pontos méretét.

A kinetikus gázelmélet első tárgya az ideális gáz volt. E gáz részecskéit apró rugalmas golyóknak tekinthetjük, amelyek távolról nem fejtenek ki egymásra vonzó- vagy taszítóerőt (épp ettől lesz „ideális” a gáz, szemben a „reális” gázokkal, ahol már ilyen vonzó- vagy taszítóerőkkel is számolni kell). E golyók tehát Newton első törvénye, azaz a tehetetlenség törvénye értelmében egyenes vonalú egyenletes mozgást végeznek – mindaddig, amíg egymással vagy az edény falával nem ütköznek. (Persze az edény fala is atomokból áll, de az egyszerűség kedvéért ettől szokás eltekinteni.) Két golyó ütközése tökéletesen rugalmas – ezért megmarad a két golyóból álló rendszer teljes mozgási energiája és lendülete (idegen szóval impulzusa), azaz a két golyó ütközés utáni mozgási energiájának összege megegyezik az ütközés előtti összeggel – csak épp másként oszlik el a közös vagyon (a termodinamika egyik legsikeresebb és leghatékonyabb, az ismeretterjesztésben jól bevált modelljéhez érkezünk: az energiára úgy gondolhatunk, mint a pénzre, ami számtalan formában létezhet, ezek átválthatók egymásra, az egyes szereplők között cserélgethetők, de a pénz teljes mennyisége a folyamatok során megmarad). Hasonló igaz az impulzusra is. A középiskolai fizikaórákról tudjuk, hogy rugalmas ütközés esetén az energia és az impulzus megmaradásának tétele elegendő arra, hogy az ütközés előtti sebességekből kiszámítsuk az ütközés utáni sebességeket. Elegendő – de csak ha a mozgás egy dimenzióban, egy egyenes mentén, sínen vagy a laboratóriumi ütközőpadon zajlik le! Csakhogy a molekulák térben, három dimenzióban mozognak, sebességük irányított mennyiséggel, vektorral írható le, és számtalan olyan végső sebességvektor-kombináció van, amely kielégíti az energia- és az impulzusmegmaradás követelményeit. Az ütközés pontos leírásához ennél többre lenne szükség – például pontosan meg kellene mondani, mekkorák és milyen alakúak a molekulák, hiszen az ütközés részletei, a kialakuló sebességvektorok attól függenek, hogy a két molekula melyik része ütközik! Gömb alakúnak feltételezett molekulák esetén csak a gömbök sugarát és a középpontok pályaegyeneseit kellene megadni, ebből már kiszámítható lenne az ütközés eredménye. (A kinetikus gázelmélet és a statisztikus fizika nagy sikere épp az lett, hogy ilyen részletekbe menő elemzés nélkül is képes leírni a molekulákból álló rendszer állapotát.)

A fallal történő ütközést hasonlóképp írhatjuk le. A falat formálisan végtelen nagy tömegűnek feltételezve az ütközés törvényei arra vezetnek, hogy a ferdén a falnak ütköző molekula úgy pattan vissza, mint fénysugár a tükrőről: sebességének a falra merőleges komponense előjelet vált, a fallal párhuzamos komponense megmarad.

Az elemi folyamatok tehát viszonylag egyszerűen leírhatók a klasszikus mechanika eszközeivel. A newtoni mechanika szellemében a mozgástörvényeken kívül még egy dolog kell a rendszer történetének kiszámításához: a kezdeti feltételek ismerete. Azaz egy adott időpillanatban ismerni kell a rendszer minden részecskéjének pontos helyzetét és sebességét

(az utóbbi helyett a mechanikában az impulzust szokás megadni, ami a részecske tömegének és sebességének szorzata). Meg kell tehát mondanunk a rendszert alkotó összes golyó helyzetét és sebességét a kezdő pillanatban, a többit a fizika törvényei elintézik helyettünk.

Manapság – viszonylag kevés golyó esetén – ez valóban megtehető. Aki látta már a számítógépen folyó hokimeccset, és megpróbálta követni a korong mozgását, vagy megfigyelte a „Jezball” nevű, néhány éve divatban volt egyszerű számítógépes játékban a falról és egymásról visszapattanó golyócskák mozgását, rájöhetett, hogy a – jelentős részben Neumann gondolatai alapján kifejlesztett – számítógép épp a kinetikus gázelmélet fentebb felvázolt programját valósítja meg: adott kezdeti helyzetből adott sebességgel indítva a golyókat, kiszámítja mozgásukat. A mozgás állandó sebességű és egyenes vonalú, a falról történő visszapattanás a mechanika törvényeit követi, a golyók találkozásakor pedig relatív helyzetük függvényében változik meg a sebesség. Nem nehéz elképzelni ugyanezt sík helyett térben, három dimenzióban, öt vagy tíz golyó helyett pedig annyival, ahány molekula a dobozba zárt gázt alkotja.

Azaz – hány molekuláról is van szó? Ma már tudjuk, hogy a köznapi mennyiségű gáz molekuláinak száma az Avogadro-szám nagyságrendjébe esik, azaz kb. 10^{23} molekula szaladgál a tartályban. Nincs (és még nagyon sokáig nem is lesz!) olyan nagy sebességű és nagy memóriakapacitású számítógép, amely ennyi molekula mozgásának követéséhez szükséges számításokat *real time* el tudná végezni...

A helyzet tehát a következő: fizikailag egyszerű modellünk a nagy részecskeszám miatt matematikailag kezelhetetlen. Zsákutcába kerültünk, nem tudjuk leírni a gáz tulajdonságait.

Mikro- és makrojellemzők

De tulajdonképpen miért is lenne szükségünk a gáz összes részecskéje helyzetének és sebességének ismeretére? A tapasztalattal való összehasonlításhoz? Ez biztosan nem igaz, hiszen mérőműszereink sem képesek meghatározni az egyes golyók helyzetét és sebességét. Végül is mi az, amit mérünk a gázon? A termodinamikai állapotjelzőket: térfogatot, nyomást és hőmérsékletet. Hogy függnek ezek össze az egyes molekulák adataival?

A kinetikus gázelmélet megalapítóinak nagy érdeme, hogy erre a kérdésre választ adtak. Számításaik nyomán kiderült, hogy a gáz hőmérséklete egyszerűen az egy részecskére jutó (mozgási) energiával arányos. (A hőmérséklet mértékegységeinek bevezetésekor ezt még nem tudták, ezért a hőmérsékletet nem energiaegységekben mérjük. Így szükségünk van egy átváltási tényezőre: ez a nevezetes Boltzmann-állandó. A termodinamikai és statisztikus fizikai számítások során a hőmérséklet mindig a Boltzmann-állandóval megszorozva, energiára átszámítva jelenik meg.)

De az egyes részecskék energiája az ütközések során változik! Semmi baj: az átlagos energiát kell tekinteni. És mivel az egyes ütközések során az energia megmarad, ezért a rendszer összenergiája is állandó – így elegendő a tartályba zárt gáz molekuláinak összes energiáját összeadni, és elosztani a részecskék számával. Ezzel megkapjuk a gáz hőmérsékletét, energiában kifejezve.

A nyomás pedig nem más, mint a felületegységre jutó erő. Miért gyakorol a gáz nyomást a falra? Azért, mert az egyes részecskék a falnak ütközve visszapattannak, sebességük – így

impulzusuk – falra merőleges komponense ellenkezőjére változik. Az erő pedig nem más, mint az időegységenkénti impulzusváltozás. Most már csak azt kell tudni, hogy időegységenként hány molekula ütközik a fal egységnyi területének. Ez pedig a gáz sűrűsége, azaz a térfogategységenkénti molekulák száma alapján könnyen meghatározható.

A fenti néhány bekezdés fizikai megfontolásait képletekre lefordítva a középiskolainál nem nehezebb matematikával levezethető a híres „általános gáztörvény”, azaz az ideális gáz nyomása, hőmérséklete, térfogata és részecskeszáma közti ismert összefüggés.

Ez valóban egyszerű – mit kellett ehhez Neumann Jánosnak hozzátennie? Az ördög per sze a részletekben lakik.

Időátlag

A gáztörvény legegyszerűbb levezetésénél szokás feltenni, hogy minden molekula merőlegesen ütközik a falnak, és ugyanakkora sebességgel mozog. Ennyiből ki is jön a tétel – csak hogy a feltevések nem igazak! A molekulák egymással folyamatosan ütköznek, mindegyik más irányban, más sebességgel mozog. A fallal történő ütközésük során egyszer több, másszor kevesebb molekula érkezik, hol nagyobb, hol kisebb impulzust adnak át a falnak. Azaz a falra gyakorolt erőhatás pontról pontra és időről időre gyors ütemben változik, ingadozik, ahogy mondják: *fluktuál*.

Miért érzünk – sőt mérünk – mégis állandó nyomást? Mert érzékszerveink és műszereink nem képesek követni ezt a gyors változást! Nem az egyes molekulák falra gyakorolt lökéseit érezzük, hanem ennek átlagát. Mintha az autópálya forgalmát hosszú távon követve nem az egyes kocsikat detektálnánk, hanem az óránként áthaladt acél mennyiségét mérnénk meg valami (bizonyára mágneses) mérőműszerrel. A hasonlat annyiban sántít, hogy sok nagyságrenddel több molekula egyedi eseményeire kell átlagolnunk, mint az ember által előidézett, statisztikusan vizsgálható jelenségek esetében.

Mekkora ez az átlagolási idő? Az emberi érzékszervek felbontása kb. egy tized másodperc – egymást ennél gyorsabban követő jelenségeket már nem tudunk megkülönböztetni. Műszereink ennél jobb felbontásúak, de a klasszikus fizika legtöbb területén meg sem közelítik a molekuláris jelenségek vizsgálatához szükséges érzékenységet.

Hasonló a helyzet a statisztikus fizika más alkalmazási területein is. Minden makroszkopikusan mérhető adat mikroszkopikus, térben és időben gyorsan változó jelenségek átlagaként áll elő. A valódi érték az átlag körül gyorsan változik, fluktuál. A legtöbb esetben csak az átlagot mérjük. Nagyon speciális esetekben a fluktuációk is észrevehetőek. Ez a helyzet a rádióadóról elhangolt, csak a magában a berendezésben „céltalanul” mozgó elektronok által képviselt zajt felerősítő rádiókészülék, vagy a tévéadás megszüntét követően a képernyőn megjelenő „hangyafoci” esetén: ekkor magát a (nulla) átlagjel körüli ingadozást, fluktuációt halljuk vagy látjuk. Egy másik érdekes példa az ún. Brown-mozgás: itt a molekulák mozgásának fluktuációi – ha nem is makroszkopikusan, de – mikroszkóp alatt közvetlenül szemügyre vehetők.

Térjünk vissza a dobozba zárt gázhoz. Ha ezt környezetétől hőszigetelő fallal zárjuk el, akkor nem cserélhet vele energiát, így a benne levő energia állandó lesz. Ha hőszigetelő doboz helyett pl. egy nagyobb, állandó hőmérsékletű edénybe, ún. hőtartályba helyezzük a

gáztartályt, akkor a gáz energiája sem lesz állandó, hanem a környezettel történő energiacsere miatt az átlagérték körül fluktuál. A makroszkopikusan észlelt nyomás viszont mindkét esetben az időben gyorsan fluktuáló erőhatások időbeli átlaga.

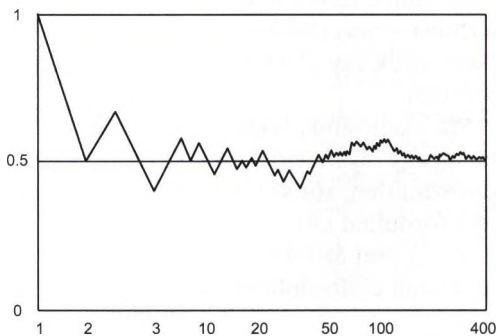
Hogy kell ezt az időbeli átlagot képezni? Képzelnék el, hogy vannak olyan pontos műszereink, amikkel az egyes molekulák sorsát is követni tudjuk. Mérjük meg tehát igen hosszú időn át (pl. egy tized másodpercig, ami a molekulák ütközései közt eltelt átlagos időhöz képest maga az örökkévalóság!) az egyes molekulák által a falnak átadott impulzust, adjuk ezeket össze, majd osszuk el az eltelt idővel. Ha ezt a procedúrát csak rövid ideig folytatnánk, az eredmény erősen ingadozna. Ahogy azonban az átlagolási idő nő, és jelentősen meghaladja az elemi események nagyságrendjét, a görbe kisimul, és egy határozott átlagértékhez tart. Ez lesz a keresett időátlag, a makroszkopikusan észlelhető nyomás.

A helyzet hasonlít a „fej vagy írás” játékhoz. Dobjunk sokszor egy érmével, és jegyezzük fel a fejek arányát. Az eleinte hektikusan ingadozó görbe mind közelebb kerül az 1/2 értékhez, ami az elméletileg várható eredmény, a „fej” dobás valószínűsége.

Valóban így van ez? Mit szólnánk, ha mondjuk harmincezer dobás után a fejek aránya csak 42% körül ingadozna? Bizonyára arra gondolnánk, hogy csalás történt, az érmét preparálták, valaki be akar csapni bennünket. A „fej vagy írás” játék előfeltétele az elemi események (a „fej” és az „írás”) egyforma valószínűsége. Hogyan tudjuk ellenőrizni ennek fennállását? Csak a játék tényleges lefolytatásával, az eredmények statisztikai elemzésével. (Persze ha kiderül, hogy cinkelt érmével van dolgunk, még mindig játszhatunk igazságos játékot: a „fej” és „írás” tényleges előfordulásának arányában módosítanunk kell a tétet és a nyereményeket.)

Hasonló helyzet előfordulhat a molekulák világában is. Képzelnék el, hogy kezdetben az összes molekula az edény bal oldali falánál van, mindegyikük sebességvektora párhuzamos egymással, és a jobb oldali fal irányába mutat. (Valószínűtlen? Az, de nem lehetetlen.) A molekulák végig párhuzamosan mozognak, egyszerre adják át impulzusukat, és egyszerre fordulnak vissza. Aztán az edényt odavissza bejárva ismét egyszerre löknek egy nagyot a jobb oldali falon. A falnál ülő megfigyelő által regisztrált időátlag nagyon speciális fűrészfogas szerkezetet fog mutatni – esze ágában sincs simán hozzásimulnia az átlagértékhez.

Az ilyen valószínűtlen, de nem lehetetlen események leírására a matematikusok egy speciális kifejezést használnak: *nullmértékű halmaz*. Ennek megértéséhez képzeljünk el egy hasonló szituációt: nyíllal lövünk egy kör alakú céltáblára, különösebb célzás és céllövői érzék nélkül. Tegyük fel, hogy csak azokat az eseményeket vesszük figyelembe, amikor egyáltalán eltaláljuk a céltáblát. Mi annak a valószínűsége, hogy a tábla jobb felébe találunk? Mindenki rávágja, hogy 50%. Itt is vannak ki nem mondott feltevéseink, mint a pénzfeladobásnál (nem húz félre az íj, nem fordulunk el lövés közben, nem fúj a szél) – ezek lényegé-



2. ábra. „Fej vagy írás” játék. A görbe a „fej” dobások relatív gyakoriságát mutatja a dobásszám függvényében (a vízszintes tengely beosztása logaritmikus). Az ábra nem számítógépes szimuláció, hanem valódi érmédobások sorozatán alapul.

ben azt fejezik ki, hogy az elemi események egyformán valószínűek. Csakhogy a pénzfeldobással vagy a kockadobással ellentétben itt végtelen sok „elemi esemény” van, hiszen a tábla végtelen sok pontja közül bármelyiket eltalálhatjuk! A lehetőségek egyszerű, kombinatorikai számbavétele most nem segít. Gömbérzékünk viszont azt súgja, hogy a valószínűség a vizsgált tartomány *mértékével*, területével arányos.

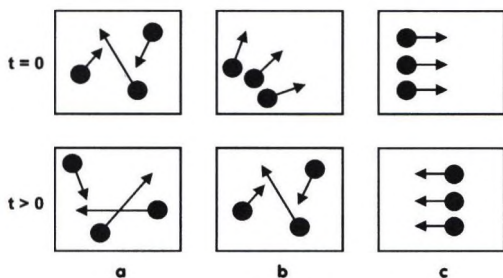
Mekkora a valószínűsége a tábla közepén levő, fele akkora sugarú kör eltalálásának? Nyilvánvalóan 25%, hiszen ennek területe pont negyede a teljes tábláénak.

És mekkora annak a valószínűsége, hogy pontosan a céltábla középpontját találjuk el? Természetesen nulla, hiszen a pont „területe” nulla. Ám ez nem lehetetlen esemény!

A diszkrét események valószínűség-elméletében megszoktuk, hogy a nulla valószínűség a lehetetlen esemény jellemzője. (Például két kockával dobva nulla annak esélye, hogy a pontok összege 13 legyen – ilyen esemény egyszerűen nem fordul elő a lehetőségek halmazában.) Folytonosan sok lehetőség esetén más a helyzet. A céltábla közepének eltalálása nem lehetetlen, de nulla valószínűségű esemény. Ennek a célhalmaznak a területe, „mértéke” nulla. Ebből a példából kiindulva – de persze pontos matematikai definíciókkal alátámasztva – vezették be a matematikusok a *nullmértékű halmaz* fogalmát. Ez a lehetséges események egy olyan részhalmaza, amelynek „eltalálása” nulla valószínűségű, ám nem lehetetlen.

Az a sejtésünk, hogy a fenti, igen mesterkéltnek tűnő kezdőfeltétel a molekulák mozgásának egy olyannyira speciális esetét valósítja meg, ami – bár nem *ab ovo* lehetetlen – igen valószínűtlen, sőt szeretnénk azt mondani: nulla valószínűségű. Azaz gyakorlatilag soha sem fordulhat elő.

Ez az eset azonban arra figyelmeztet, hogy nem magától értetődő az időátlagok képzése, hiszen előfordulhatnak olyan esetek, amikor az átlag nem követi várakozásunkat.



3. a b c ábra. Gázmolekulák mozgása különböző kezdőállapotok esetén:

a) átlagos kezdőállapot, átlagos folytatás,

b) speciális kezdőállapot, átlagos folytatás,

c) speciális kezdőállapot, speciális folytatás.

A c) esetnek megfelelő, túlságosan szabályos kezdőállapotok nullmértékű halmazt alkotnak a fázis térben.

A fenti példában a gáz által a falra gyakorolt nyomást vizsgáltuk. Nyilvánvaló, hogy bármilyen más fizikai információ, amit a gázzal szerezhetünk, kiszámítható (elvileg legalábbis) a molekulák helyzete és sebessége (vagy impulzusa) alapján. A makroszkopikus mérések eredményét pedig mindig a fentihez hasonló időátlag adja.

Mi van akkor, ha nincs – mint fentebb feltételeztük – olyan szuperműszerünk, amely az egyes molekulák mozgását is detektálja? Kénytelenek leszünk visszatérni korábbi, már kivihetetlennek bizonyult terünelhöz: matematikai eszközökkel kell követnünk az egyes molekulák mozgását, ezután – minden pillanatban ismerve az összes molekula helyzetét és sebességét – ki

tudjuk számítani a keresett fizikai mennyiséget, majd (ismét nem mérésrel, hanem számítással) elkészíthetjük az időátlagot is. E program végrehajtása sajnos reménytelen. Van-e valami olyan trükk, amivel e procedúra elkerülhető, és mégis megkaphatjuk a helyes eredményt?

Eloszlásfüggvény

Két zseniális gondolat segít ki a csávából, egyben konkrét számítási módszert is nyújtva a fizikusoknak. Az egyik gondolat a molekulák egymástól való megkülönböztethetetlenségét hangsúlyozza. Nem az számít, hogy adott pillanatban *melyik* molekula löki meg adott erővel az edény falát, csak az, hogy adott idő alatt *hány* egyforma sebességű molekula érkezik a falhoz. Ezért elegendő a molekulák helyett *eloszlásukkal* foglalkozni.

Soroljuk a gáz molekuláit osztályokba (pl. x koordinátája 1 és 1,01 között, y koordinátája 0,42 és 0,43 között, z koordinátája 1,36 és 1,37 között, sebességének irányszöge 10 és 11 fok között, nagysága pedig 555 és 556 m/s között van stb.), és mondjuk meg, hogy egy adott pillanatban melyik kategóriában éppen hány molekula van – ezzel definiáltunk egy eloszlást (vagy eloszlásfüggvényt).

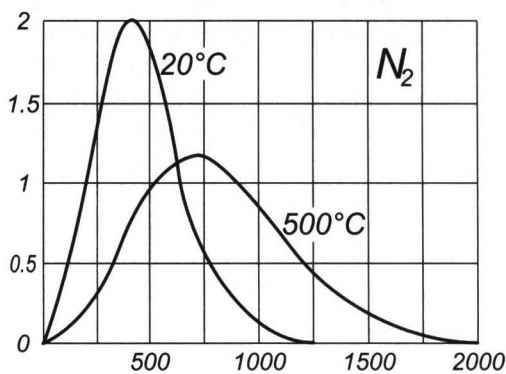
Az eloszlás ismeretében ki tudjuk számítani a makroszkopikus mennyiségeket – hiszen azok úgysem érzékenyek a molekulák egyediségére, csak az egyes kategóriákba került molekulák száma számít.

Vigyázni kell azonban az időátlag képzésénél, hiszen az eloszlásfüggvény időben változik. Hogy hogyan változik, az természetesen a molekuláris dinamika alapján vezethető le. Az eloszlás általában nem marad állandó, mert az egyes molekulák a másodperc töredéke múlva bekövetkező ütközésük során átkerülnek egy másik kategóriába, hiszen sebességük iránya és nagysága is megváltozik. Előfordulhat azonban, hogy egy másik molekula épp az előző helyére kerül, azaz sebességének nagysága és iránya egy ütközés következtében épp az előző címkének megfelelő lesz. Boltzmann felállított egy igen bonyolult egyenletet (a róla elnevezett transzportegyenletet), amely leírja a mozgás és ütközés következtében felépülő kategóriaváltásokat.

Láttuk, hogy van olyan eset, amikor a molekulák mozgása, eloszlása – és ennek következtében az adataiból készült statisztika – „nem tipikus”. De mit nevezhetünk ebben az esetben tipikusnak?

Maxwell és Boltzmann, a kinetikus gázelmélet alapító atyái azt felelték: az *önreprodukáló* eloszlásokat tekinthetjük tipikusnak. Ez a másik fentebb említett zseniális gondolat. A Boltzmann-féle transzportegyenlet segítségével válaszolni lehet arra, milyen az az eloszlás, amely reprodukálja önmagát, azaz időben nem változik: az egyes kategóriákból kilépő molekulák helyébe ugyanannyi másik lép. Így kapjuk a nevezetes Maxwell-eloszlást. Az ábra a sebesség nagysága szerinti eloszlásfüggvényt mutatja be – a sebességvektor minden iránya, valamint az edényen belüli tetszőleges pontban való tartózkodás egyformán valószínű.

Ha a gáz molekuláinak eloszlásfüggvénye egy adott pillanatban Maxwell függvé-



4. ábra. A gázmolekulák sebességének egyensúlyi Maxwell-Boltzmann-eloszlása. Az ábrán a nitrogéngáz két különböző hőmérséklethez tartozó eloszlásfüggvényét tüntettük fel. A vízszintes tengely a molekulák sebességét jelöli, m/s egységekben. A görbe magassága az adott sebességű molekulák relatív gyakoriságával arányos.

nyének megfelelő alakú, akkor az ütközések ellenére később is ilyen alakú marad. A makroszkopikus mennyiségeket pedig az egyes részecskékre vett átlag segítségével számíthatjuk ki. Ha az eloszlás időben nem változik, e mennyiségek is állandóak maradnak. Sikerült végre leírunk egyszerűnek látszó fizikai objektumunkat, az edényben békésen kuksoló, időbeli változást nem mutató gázt!

A transzportegyenlet módszere jól bevált az ideális gáz esetén, de nem általánosítható minden sok részecskéből álló rendszerre. Részben azért, mert a szabad mozgás (ami az ideális gáz esetében a molekulák két ütközés közötti egyenes pályájú, állandó sebességű repülése volt) leírása is nehéz, másrészt az egyes alkotórészek közti kölcsönhatás is sokkal bonyolultabb az ideális gáz részecskéinek ütközésénél.

Fázistér és fázispályák

Gibbs amerikai fizikus volt az, aki megtalálta azt a kulcsot, amely a statisztikus fizikát bonyolult rendszerek leírására is alkalmas hatékony fizikai elméletté tette. Ez a kulcs a statisztikus sokaságok módszere volt, sarkköve pedig az ergodikus hipotézis.

Illusztrációképpen térjünk vissza ismét az ideális gázhoz, és nézzük meg, hogy írná le ezt a rendszert egy matematikai fizikus, aki még sohasem hallott a termodinamikáról, viszont alaposan megtanulta Newton klasszikus mechanikájának a XIX. század folyamán kifundált továbbfejlesztését és általánosítását, a Hamilton-féle mechanikát!

Tudjuk, hogy a több részecskéből álló rendszer állapotát egyértelműen meghatározza összes részecskéje helyvektorának 3-3, és impulzusvektorának 3-3 komponense. Ha ezeket a mennyiségeket az idő függvényében ismerjük, mindent tudunk a rendszerről. Bármely más fizikai mennyiség e változók függvényeként állítható elő.

A Hamilton-féle mechanikában szokás az N részecskéből álló rendszer helyvektorainak és impulzusvektorainak összesen $6N$ komponensét egy $6N$ dimenziós tér, a *fázistér* koordinátáinak tekinteni. E tér minden pontja a teljes rendszer egy lehetséges állapotának felel meg. A rendszer pillanatnyi állapotát a tér egy pontja jellemzi. A rendszer időbeli fejlődését e pont elmozdulása jelzi a fázistérben (e mozgásra vonatkoznak Hamilton ún. kanonikus egyenletei). Ha a rendszer mozgása során időátlagot akarunk képezni, akkor a fázistérben bejárt pályát kell követnünk, és a vizsgált függvényeknek az egyes fázispontokban felvett értékeire kell átlagolnunk.

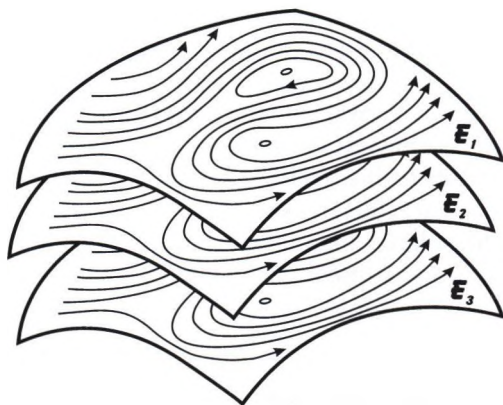
A rendszer mozgását bizonyos feltételek korlátozzák. Ezek legfontosabbika az energiamegmaradás: a mozgás során a rendszer teljes energiája állandó marad. Az energia is fizikai mennyiség, ennek értéke is kiszámítható az egyes tömegpontok helyzete és impulzusa, azaz a fázistérbeli pont koordinátái alapján. Tehát a fázistér minden pontjához meghatározott energiaérték tartozik. Megfordítva: az energia értékének rögzítése kijelöli a fázistér pontjainak egy részhalmazát, az ún. energiafelületet. (Szemléletesen nem tudjuk elképzelni a sokdimenziós tereket. Ezért szemünk előtt mindig egy háromdimenziós ábra legyen. Ha a fázistér háromdimenziós lenne, akkor az állandó energiához tartozó részhalmazok kétdimenziós – általában bonyolult módon görbülő, kanyargó – felületek lennének. Innen az energiafelület neve.)

Az energiacsökkentések nem metszik egymást, hiszen a metszéspont egyszerre két energiaértékhez tartozna. Úgy képzelhetjük őket, mint a hagyma egymást követő rétegeit.

A rendszer mozgását leíró görbe mindig egy energiacsökkentésen kanyarog. E fázispálya önmagát sem metszi, hiszen akkor a fázistér egy pontjából, azaz a rendszer egy meghatározott állapotából kétfelé is vezetne a jövőbeli fejlődés útja. Ezt pedig a newtoni fizika determinizmusa (a kezdőállapot egyértelműen meghatározza a rendszer további sorsát) kizárja. Úgy gondolhatunk a fázispályákra, mint az energiacsökkentésen lefektetett sínekre. A kezdőpillanatban a rendszert ráhelyezzük a fázispálya egy pontjára, és az ott áthaladó sínen, a Hamilton-egyenletek által meghatározott sebességgel mozgó vonat viszi tovább a rendszert.

Bár eredetileg a sok tömegpontból álló rendszer leírására vezettük be, a fázistér és a vele kapcsolatos fogalomrendszer tetszőlegesen klasszikus mechanikai rendszer leírására is alkalmas. Az energiacsökkentések és az azon kanyargó fázispályák képe általános, minden mechanikai rendszerre érvényes (amennyiben a rendszer zárt és konzervatív, azaz a mozgás során az energia megmarad).

Ez a geometriai jellegű kép (amely az eredeti newtoni mechanika ekvivalens matematikai átfogalmazása) rendkívül impresszívnek és termékenynek bizonyult, és sok tudományos gondolat kiindulópontjává szolgált.



5. ábra. Energiacsökkentések és fázispályák. A rendszer fázistérben hagymahéjszerűen helyezkednek el a különböző energiaértékekhez tartozó energiacsökkentések. Az adott energiájú rendszer fázispályája a megfelelő energiacsökkentésen kanyarog. A különböző fázispályák nem metszik egymást, hiszen a mozgástörvények egyértelműen meghatározzák a rendszer további fejlődését, ezért elágazás nem lehetséges. Ha az ergodikusan hipotézis igaz lenne, egyetlen (igen kanyargós) fázisgörbe bejárná az egész energiacsökkentést. Az ábrán vázolt eset közelebb áll a valósághoz (Neumann kváziergodikus tétele által leírt) helyzethez.

Az ergodikusan hipotézis

Többek között annak a(z első pillantásra egészen meglepítő) sejtésnek a kimondásához is elvezetett, mely szerint egy eléggé bonyolult rendszer fázispályája a megfelelő energiacsökkentés minden pontján áthalad. Ez volt az *ergodikusan hipotézis* eredeti formája (Boltzmann 1871).

Ez a hipotézis könnyen cáfolható. Tegyük fel, hogy az energián kívül van még egy megmaradó mennyiség, egy mozgásállandó. Ennek megadott értéke szintén kijelöli a fázistér egy részhalmazát. A rendszer fázispályájának mindkét részhalmazon (az energia és a másik mozgásállandó által kijelölt felületen) rajta kell maradnia, ezért csak e két felület közös részén mozoghat (ami általában még alacsonyabb dimenziójú halmaz: pl. két felület közös része általában egy vagy több görbe) – így nem járhatja be a teljes energiacsökkentést.

Egy másik ellenvetés így szól: ha egyetlen görbe bejárna a teljes felületet, akkor lényegében csak egyetlen (adott energiájú) fázispálya lenne – a rendszer különböző kezdőfeltételű (de azonos energiájú) mozgásai ugyanannak az egyetlen mozgásnak egymástól időben eltolt példányai lennének. Ez elképzelhető lenne, de hamar születtek konkrét ellenpéldák.

Ha ennyi és ilyen súlyos ellenvetés szól ellene, miért ragaszkodtak a fizikusok mégis az ergodikus hipotézishez (és miért gyúrták-gyömöszölték később olyan alakra, hogy a fenti egyszerű ellenérvekkel már ne lehessen belekötni)? Azért, mert egyszerű és frappáns megoldást kínál a korábban részletesen elemzett problémára, a bonyolult rendszerek fejlődése során definiált időátlagok kiszámítására!

Statisztikus sokaságok

Fogadjuk el egy pillanatra igaznak az ergodikus hipotézist! Ha a rendszer fázispályája bejárja a teljes energiafelületet, akkor a pálya menti időátlagolás helyettesíthető az energiafelület pontjaira vett átlagolással. Csak azt kell tudnunk, hogy az energiafelület egyes pontjait milyen súllyal vegyük figyelembe. Azaz azt, hogy az egyes pontok környékén milyen sokáig tartózkodik a rendszer. Ez pedig a mozgásegyenletek alapján könnyen megtudható.

Miért jobb, miért egyszerűbb az energiafelület pontjaira átlagolni, mint a rendszer tényleges mozgására? Mert az utóbbit általában nem ismerjük! Még a legegyszerűbbnél kicsit bonyolultabb klasszikus mechanikai rendszerekre is reménytelen a mozgásegyenletek pontos megoldásának felírása. Például míg az egymás gravitációs terében mozgó két tömegpont mozgásegyenleteinek egzakt megoldását évszázadok óta ismerjük (ez az ún. Kepler-probléma, megoldásai a bolygók jól ismert ellipszispályái), addig a háromtest-probléma, azaz három, egymást kölcsönösen vonzó test mozgásegyenleteinek megoldása a matematika legnehezebb feladatai közé tartozik. Évszázadok óta bíbelődnek fizikusok és csillagászok a háromtest-probléma egy-egy speciális esetének vizsgálatával – az általános megoldás megtalálására nincs remény. Amikor viszont a mozgás egyszerű, akkor a részecskék nagy száma állíthat leküzdhetetlen technikai nehézségeket a megoldás elé – ilyen az ideális gáz korábban részletesen megvizsgált problémája.

Tehát az általános mechanikai rendszer mozgásegyenleteit fel tudjuk írni, de megoldani nem tudjuk. Így hiába tudjuk, hogy az energiafelületen valahol ott kanyarog a rendszer fázispályája, sohasem fogjuk megtalálni, pontosan merre is jár. A pálya mentén időátlagot számolni ezért teljesen reménytelen vállalkozás. Ha viszont az ergodikus hipotézis igaz, akkor nem kell törődni a fázispályával (amely úgyis felkeresi az energiafelület összes pontját, de a fizikust – és az átlagot – nem érdekli, hogy milyen sorrendben), és elegendő az ismert struktúrájú energiafelületre átlagolni.

Gibbs ugyanezt kissé másképp fogalmazta meg, de a fizikai (és matematikai) következmények azonosak. Képzeljük el a vizsgált fizikai rendszert nagyon sok példányban. A különböző kényszerek által rögzített makroszkopikus kényszerfeltételek (pl. az ideális gázt tartalmazó edény térfogata, a rendszer teljes energiája stb.) legyenek azonosak, de a rendszer belső szerkezete tetszőleges lehet. Az ideális gáz példáján: ugyanakkora térfogatot, nyomást és hőmérsékletet a molekulák nagyon sokféle eloszlása megvalósíthat. Így is fogal-

mazzák: azonos makroállapothoz nagyon sok mikroállapot tartozhat. (Eme mikroállapotok száma a makroállapot rendezetlenségének mértékéül szolgálhat: e szám logaritmusát azonosította Boltzmann az *entrópia* nevű híres – vagy inkább hírhedt – mennyiséggel, amely a termodinamika második főtétele szerint minden fizikai folyamatban nő, időbeli aszimmetriát, irreverzibilitást hozván a makrovilágba.) Gibbs nyomán a fizikai rendszer eme kópiáinak, sok-sok példányának összességét *statisztikus sokaságnak* nevezzük, a rendszer egyes példányai a sokaság elemei. (Vegyük észre az érdekes szemléletváltást: az eredeti kinetikus gázelméletben a molekulák nagy száma miatt merült fel a statisztikai, valószínűség-számítási módszerek alkalmazásának igénye. Gibbsnél a teljes, sokatomos rendszer már csak egy példány a statisztikus sokaságban – a sok molekula „csak arra való”, hogy a rendszernek elég sok lehetséges belső állapota legyen, ezért érdemes legyen statisztikus sokaságot készíteni. Ez nem szükségszerű, ezért a statisztikus fizika alkalmazható bonyolult, de nem atomos-molekuláris struktúrájú rendszerek, pl. az elektromágneses mező leírására is.)

A fizikai mennyiségek átlagának kiszámítására Gibbs ezekre a sokaságokra alapozva dolgozott ki eljárást. A fő kérdés az, hogy melyik Gibbs-egyedet milyen súllyal kell figyelembe venni az átlagolásnál. Erre a kérdésre Gibbs nagyon egyszerű választ adott: mind-egyiket ugyanakkora súllyal. (Gondoljunk a pénzfeldobás vagy a kockadobás példájára, ahol *a priori* egyformának gondoljuk az egyes elemi események valószínűségeit, vagy a fizikai esetekhez közelebb álló másik példára, a nyilakkal való vaktában lövöldözésre: semmi sem tünteti ki eleve a céltábla egyes pontjait, ezért minden tartomány eltalálásának valószínűségét a tartomány nagyságával, mértékével arányosnak tételezzük fel.) A Gibbs-sokaság elemei megfelelnek a fázistér energiafelülete egyes pontjainak. Időbeli változásról, fejlődésről nincs szó, az e módszerben nem szerepel. Viszont Gibbs feltételezése az azonos statisztikai súlyokról csak akkor helytálló, ha igaz az ergodikus hipotézis, tehát a fizikai rendszer időbeli fejlődése során valóban végiglátogatja a fázistér energiafelületének minden pontját, azaz minden Gibbs-egyedet.

Az ergodikus probléma

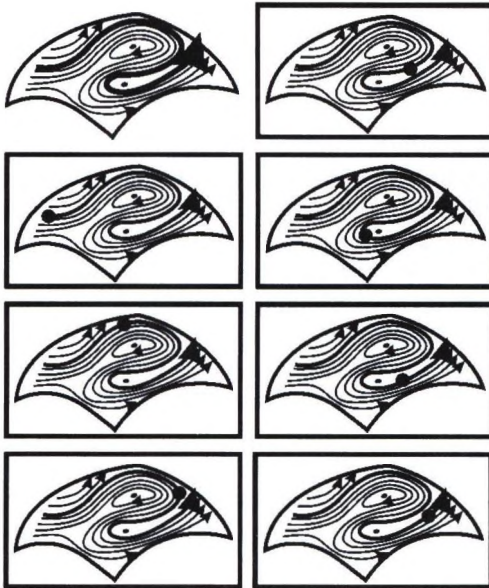
Az *ergodikus kérdés* tehát teljes általánosságában így szól: megegyezik-e tetszőleges fizikai mennyiség Gibbs-féle sokaságátlaga a rendszer időfejlődése által definiált időátlagával?

Ha igaz lenne az ergodikus hipotézis, azonnal igennel felelhetnénk az ergodikus kérdésre. És jó lenne igennel felelni, mert a Gibbs-módszert felhasználva termékeny és szerteágazó statisztikus fizikai elmélet fejlődött ki a XIX. század utolsó és a XX. század első évtizedeiben. De ez az egész hatalmas és szépséges elmélet homokra épült, amíg az ergodikus kérdésre pozitív válasz nem érkezett.

Az ergodikus hipotézis túl erős, sok ellenérv szól ellene. Lehet-e úgy gyengíteni, hogy igaz legyen, és mégis pozitív válasz következzen belőle az ergodikus kérdésre? Ezen dolgozott a XX. század elején számos fizikus és matematikus.

Viszonylag egyszerűen kivédhető az első említett ellenérv, a másik megmaradó mennyiség kérdése. Ha ez a mennyiség egyszerűen megadható, akkor elég a fázistér azon tarto-

mányára korlátoznunk magunkat, ahol ez a mennyiség adott értéket vesz fel. Vizsgáljuk meg ismét a tartályba zárt ideális gázt! Ha az egész tartály áll hozzánk képest, akkor nyilvánvaló, hogy a gázmolekulák impulzusának (vektori) összege zérus. De ha áttérünk egy olyan megfigyelőre, aki szép lassan elsétál a tartály mellett, akkor az ő koordináta-rendszeréből nézve minden molekula impulzusvektora megváltozik, a rendszer teljes impulzusa pedig nem lesz nulla. Világos, hogy egy ilyen leírásbeli különbség semmiféle hatással nem lehet a vizsgált rendszer belső viszonyaira, termodinamikai tulajdonságaira. Egyszerű tehát a megoldás: kössük ki, hogy a molekulák összipulzusa legyen nulla. Ezzel a rendszer eredeti fázisterét (amelyben mindenféle összipulzusérték előfordulhatott) korlátoztuk, de



6. ábra. Időátlag és sokaságátlag. Az első ábra azt mutatja be, hogy az időátlag képzésekor végig követjük a rendszer mozgását a (vastagon rajzolt) fázispályán, és ennek során számítjuk ki a fizikai mennyiségek átlagát. A többi ábra a statisztikus sokaságot jelképezi: a rendszer egyes kópiái azonos energiájú állapotban vannak, de mikroállapotuk különböző. Az egyes kópiák állapotát a fekete golyók jelzik. Az ergodikus hipotézis szerint ezek a kópiák ugyanannak a fázispályának más-más pontján járnak, ezért az egyes kópiákra történő átlagolás eredménye megegyezik az időátlaggal. Figyeljük meg, hogy az ábrán vannak olyan tartományok, ahova a fázispont nem jut el (pl. a kis zárt görbék belsejébe). Neumann részletesen foglalkozott ezzel a problémával is.

nem változtattunk a rendszer lényeges tulajdonságain. Hasonlóképp kezelhetők azok az univerzális megmaradási tételek, amelyek – Emmy Noether tétele alapján – a téridő szimmetriatulajdonságaiból származtathatók, és amelyek kvantummechanikai implementálásával épp Wigner Jenő, Neumann barátja és munkatársa ért el átütő sikereket, amikor (Neumann tanácsára) a szimmetriák matematikai elméletét, a csoportelméletet alkalmazta leírásukra.

Bonyolultabb a helyzet, ha a másik megmaradási tételt nem általános szimmetriaelv kényszeríti ki, hanem a rendszer saját dinamikája. Ismert például a Kepler-probléma esetén egy sajátos, csak erre a fizikai feladatra jellemző megmaradó mennyiség, a Runge–Lenz-vektor. A XIX. század folyamán (főleg égimechanikai kérdések vizsgálata nyomán) a fizikusok alaposan megvizsgálták, mikor bukkanak fel egy rendszer mozgása során sajátos mozgásállandók (ez a szög- és hatásváltozók módszere). Ezeknek az állandóknak a léte ellentmond az ergodikus hipotézisnek, de nem küszöbölhető ki az előbb vázolt, az összipulzusra alkalmazott triviális módon. A statisztikus fizika kutatói nagyot sóhajtottak tehát, és visszavonultak: csak olyan rendszerekre tartottak ki a hipotézis mellett, amelyekben az energián kívül nincs más megmaradó mennyiség.

A kváziergodikus hipotézis

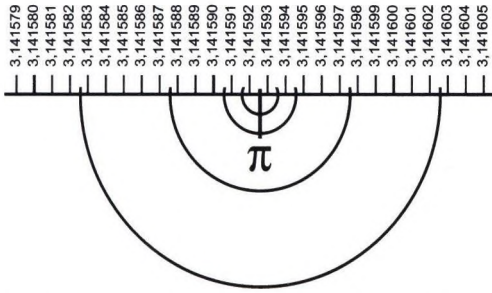
A másik, korábban említett nehézség csak jóval ravaszabb módon volt kiküszöbölhető. Ha a fázispálya az energiafelület minden pontján áthaladna, akkor csak egyetlen fázispálya lenne. Amúgy elég nehéz elképzelni ilyen alakzatot... A probléma abban áll, hogy egy egydimenziós alakzattal, egy vonallal akarunk lefedni egy sokdimenziós alakzatot, az energiafelületet. Miért is van erre szükség? Mert így szerepelhet az energiafelület minden pontja a sokaságátlagban. De ennek van más módja is – és a matematikusok erre is kitaláltak egy alkalmas fogalmat, a mindenütt sűrű halmaz fogalmát.

Tekintsük a valós számegyenest, és vizsgáljuk a racionális koordinátájú pontokat, azaz azon pontokat, amelyek koordinátája két egész szám hányadosaként felírható. E pontokból nyilvánvalóan végtelen sok van, azonban ennél sokkal erősebb állítást is kimondhatunk elhelyezkedésükről: az egyenes tetszőleges pontjának tetszőleges közelségében is találunk ilyen racionális koordinátájú pontot.

Első pillantásra azt gondolhatnánk, hogy a racionális számok teljesen kitöltik a számegyenest. Válasszunk két racionális koordinátájú pontot, és keressük meg a felezőpontjukat. E pont koordinátája nyilvánvalóan a két eredeti szám számtani közepe lesz, ami – a törtek összeadásának szabálya alapján – szintén racionális szám lesz. Felezzük meg ezután az egyik eredeti szám és a most kapott új szám közti intervallumot – újabb racionális számot kapunk. Az eljárás vég nélkül folytatható. Így tetszőleges, akármilyen rövid intervallum belsejében is találhatunk végtelen sok racionális koordinátájú pontot.

De – a szemléletes várakozástól eltérően – az egyenes nem minden pontja racionális koordinátájú! Elemi bizonyítással belátható, hogy pl. $\sqrt{2}$ nem írható fel két egész szám hányadosaként, ún. irracionális szám. (Annak idején ez a felfedezés nagy sokkot váltott ki az ókori görög matematikusok körében – addig ugyanis nyilvánvalónak vették, hogy bármely szakasz hossza kifejezhető két egész szám hányadosaként, arányaként. Egy legenda szerint e döbbenetes felfedezést megünneplendő száz ökröt áldoztak az isteneknek. Egy másik legenda szerint azóta félnek az ökrök a tudományos haladástól.)

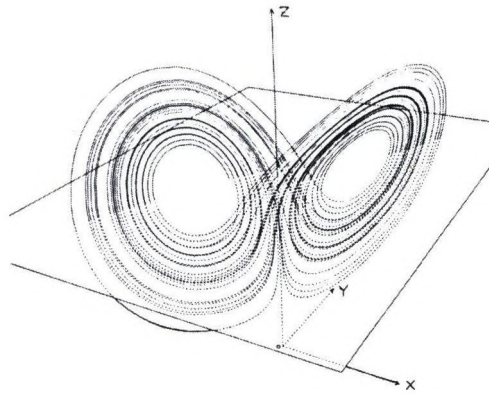
A matematika későbbi fejlődésével, a XIX. század végén Cantor által kifejlesztett halmazelmélet kereteiben az is kiderült, hogy az irracionális számok jóval többen vannak, mint a racionális számok (szakkifejezéssel: nagyobb a számosságuk). Mégis: bármely irracionális számhoz tetszőlegesen közel találunk racionális számot. Ehhez elég arra gondolni, hogy az irracionális számok felírhatók végtelen tizedes tört formájában. Ha azt kérdezzük, van-e racionális szám mondjuk a π (amely szintén irracionális) számtól egymilliomodnyi távolságon belül, egyszerű a válasz: írjuk fel a π tizedestört-alakjának első néhány jegyét: $\pi \approx 3,1415926\dots$, kerekítsük ezt hat tizedesre, majd növeljük meg eggyel az utolsó számjegyet: 3,141592 és 3,141593 egymástól pontosan 0,000001 egységnyi, azaz egymilliomodnyi távolságban van, köztük ott a π , amelyet tehát mindkettőtől a keresetnél kisebb távolság választ el. A felírt két szám pedig – lévén véges tizedes tört – racionális. Nyilvánvaló, hogy a fenti konstrukció tetszőleges valós (azaz racionális vagy irracionális) számból kiindulva elvégezhető, egymilliomod helyett tetszőleges pontossággal. Így hát tetszőleges valós szám tetszőlegesen kicsiny környezetében (azaz tőle tetszőlegesen kis távolságban) van legalább egy racionális szám (ténylegesen végtelen sok van, de most egy is elég). Ezt a tulajdonságot a matematikusok a következő formában fogalmazzák meg: a racionális számok *mindenütt sűrűn* helyezkednek el a valós számok között.



7. ábra. Racionális számok a π környezetében. Az ábrán a számegyenes egy részletének kinagyított képe látható. A körök a π -nek megfelelő pont körüli 1, 2, 5 és 10 millióod egységnyi sugarú környezeteket jelölik. Látható, hogy a π tetszőleges környezetében vannak racionális számok.

rendszer fázispályája nem megy át az energiafelület minden pontján, de mindenütt sűrű az energiafelületen. Azaz a felület tetszőleges pontjának tetszőleges közelségében elhalad, mintegy belemetsz a pont körüli tetszőlegesen kicsiny sugarú gömbbe.

Nehéz persze ezt elképzelni. Aki látott az utóbbi évtizedekben divatba jött fizikai fogalomról, a *káoszról* (amely maga persze szoros kapcsolatban van az ergodikus problémakörrel)



8. ábra. Kaotikus fázisgörbék: a Lorenz-attraktor. Az ábrán egy egyszerűsített, három szabadsági fokú (azaz háromdimenziós fázistérű) hidrodinamikai rendszer időbeli fejlődését leíró differenciálegyenletek számítógépes megoldása látható. A modellt Lorenz 1963-ban dolgozta ki meteorológiai jelenségek szimulációja és számítógépes előrejelzése céljából (ezt a kutatási irányt egyébként éppen Neumann János kezdeményezte). Az egyenletek viszonylagos egyszerűsége és a fázistér alacsony dimenziója ellenére a mozgás rendkívül bonyolult, kaotikus, lényegében előrejelezhetetlen. A fázisgörbe egy igen komplex struktúrájú halmazon, az ún. különös attraktoron kanyarog.

Nézzük most a két-, illetve háromdimenziós esetet! Tekintsük azokat a pontokat, amelyeknek minden koordinátája racionális. A fentihez hasonlóan belátható, hogy a sík, illetve a tér tetszőleges pontjának tetszőlegesen kicsiny környezetében van racionális koordinátájú pont. (Környezeten most a pont körüli kicsiny gömböt kell érteni.) Az állítás nyilvánvalóan kiterjeszthető tetszőleges dimenziójú térre is. E terekben tehát a racionális koordinátájú pontok mindenütt sűrűn helyezkednek el.

E matematikai fogalom segítségével gyengítették az eredeti ergodikus hipotézist „kváziergodikus hipotézissé”. Eszerint a

szóló, számítógépes grafikákkal illusztrált könyvet, az halvány fogalmat alkothat arról, mire is képes néhány kaotikusan összegabalyodott tartomány két és három dimenzióban. És ezek a legegyszerűbb esetek messze nem tükrözik az általános, sokdimenziós eset elképesztően bonyolult struktúráját és határtalan geometriai lehetőségeit.

Tény, hogy a kváziergodikus hipotézis bevezetése két legyet ütött egy csapásra. Először is megengedte több különböző fázispálya létezését. Úgy képzelhetjük el ezeket, mint végtelenül kanyargós drótkötegeket: mindegyik egyedi drót közel halad el minden ponthoz, egymást mégsem metszik. A rendszer különböző kezdőfeltételekkel indított példányai más-más drót mentén haladnak. A drótkötegek később aztán szét is válhatnak, az egyes drótok a fázistér más-más tartományában kalandoznak tovább. (A káosz problémaköre épp ezzel kapcsolatos: egymáshoz tetszőlegesen közeli kezdőfeltételek is vezethetnek lényegesen más fejleményekhez. Ironikus módon épp ez, a

Neumann munkásságához is kapcsolódó tudományos eredmény mutatott rá arra, hogy munkásságának egy másik célja, az időjárás hosszú távú, pontos előrejelzése lényegében megoldhatatlan feladat, hiszen a földi légkör kaotikus rendszer, és a kezdeti feltételek még oly pontos ismerete, meg az igen hatékony numerikus algoritmusok sem tudják kiküszöbölni a fázispályák szét tartásából szükségszerűen következő előrejelzési hibákat. Gondoljunk arra, hogy a fázistér egyik tartománya napsütéses időnek, a másik pedig zivatarnak felelhet meg. Az egymás közelében indult fázispályák egyike ide, a másika oda tart. És ezen a még pontosabb mérés és a még nagyobb számítógépek sem segíthetnek.)

○ A kváziergodikus hipotézis érvényessége tehát fizikailag jóval realiztikusabb lenne, hiszen lehetővé tenné a rendszer lényegesen különböző fejlődési pályáinak létezését. Ugyanakkor nem rontaná el azt az előnyt, amiért a fizikusok az ergodikus hipotézisre szavaztak! A kváziergodikus fázispálya éppen úgy közel jár az energiafelület minden fontjához, mint azt az ergodikus pálya tenné – ezért a Gibbs-féle statisztikai sokaságban minden pontot, azaz a rendszer minden állapotát figyelembe kell venni.

Neumann kváziergodikus tétele

A kváziergodikus hipotézis még ebben a formában sem bizonyult helyesnek. Ha azonban megengedjük, hogy azon fázispályák összessége, amelyekre nem teljesül a kváziergodikus hipotézis, az energiafelületen nullmértékű részhalmazt alkossanak (ez persze nem befolyásolja a statisztikai tulajdonságokat – emlékezzünk vissza a céltábla közepére!) – nos, akkor megkapjuk Neumann János 1932-ben bizonyított tételének állítását.

Neumann ekkor már túl volt a kvantummechanika alapjainak matematikai vizsgálatán, amelynek kapcsán számos tételt bizonyított és számos módszert kifejlesztett a függvények terén ható operátorokkal kapcsolatban. Matematikus kollégáival való beszélgetések során merült fel benne, hogy ezeket a matematikai eszközöket az ergodikus (pontosabban a kváziergodikus) hipotézis bizonyítására is be lehetne vetni. Az 1932-ben publikált cikksorozat első, szinte teljesen matematikai része közli a tétel pontos bizonyítását, az ezt követő cikk már „Az ergodikus hipotézis fizikai alkalmazásairól” szól, a harmadik cikk megmutatja, hogy az ergodicitáshoz nem szükséges igen sok elemi objektumból álló rendszer, azaz a fázistérnek nem kell nagyon nagy dimenziójúnak lenni – ezzel Neumann mintegy a fél évszázaddal későbbi kaoszfizika matematikai alapjait teremti meg.

Neumann említett cikkei nem csak az itt körülírt tételt bizonyítják, számos más kérdésről is szólnak. A bizonyítás módszere alapján például becslést lehet adni arra, hogy milyen hosszú is az a bizonyos időszakasz, amin túl a kiszámított időátlag már nem fluktuál, hanem megállapodik, és így értéke megegyezik a sokaságátlag alapján számított értékkel. Konkrét fizikai rendszerre ennek az időtartamnak a becslése meglehetősen nehézségekbe ütközhet.

○ Igen érdekes, hogy Neumann a bizonyítás matematikai részleteit közlő alapvető cikket speciális, kéttételes „fizikus” okoskodással zárja. Először is rámutat, hogy tisztán matematikai szempontból az extra mozgásállandók léte nem okozna gondot: meg kell szorítani a fázistert az extra mozgásállandó konkrét értékének megfelelő részhalmazra (ahogy azt a gáz összipulzusa esetében láttuk). Fizikai szempontból azonban a helyzet problematikus,

hiszen e mozgásállandók meghatározása egyáltalán nem triviális (sőt, ez a klasszikus mechanika egyik legtöbbet vitatott kérdése), a fázistér megfelelő részalmazának megtalálása pedig gyakorlatilag kivihetetlen.

A másik „fizikus” problémafelvetés arról szól, hogy mit kell érteni a „mindenütt sűrű” fázisgörbén, mit kezdjen a fizikus azzal a kérdéssel, vajon egy pont rajta van-e a görbén vagy nincs. A szokásos választ, mely szerint gyakorlati szempontból e kérdés értelmetlen, hiszen a tényleges mérés mindig csak véges pontosságú, tehát az eredmény nem egy pont, hanem egy kis „folt” a fázistérben, ezért szükségszerűen tartalmazza a mindenütt sűrű halmaz pontjait is – Neumann elutasítja, és egy matematikailag korrekt gondolatmenetet ajánl a fizikusok figyelmébe.

Neumann néhány évvel később előadást tartott az ergodikus tételről a budapesti tudományegyetemen, az Ortway Rudolf által szervezett szemináriumsorozaton. Fennmaradt egy fogalmazvány, amelyben Neumann röviden összefoglalja a téma jelentőségét:²

Az „Ergod-hypothézis” először Boltzmann és Maxwell kinetikai gázelméletében szerepelt: általa foglalta össze Boltzmann egzaktan az ezen elméletben szükséges „rendetlenségi” feltevéseket. Az „Ergod-hypothézis” egy tisztára matematikai feltevés a klasszikus mechanika rendszerében, tehát egy oly tétel, melynek minden fizikai vonatkozástól eltekintve is vagy bebizonyíthatónak vagy megcáfolhatónak kellene lennie, és pedig *tisztára analitikai eszközökkel*. És csak akkor vezethető vissza a kinetikai gáz- és hőelmélet a klasszikus mechanikára, ha ez a matematikai tétel [...] igaz. A statisztikai mechanika minden ismeretes formájára áll ez, mert mindenütt valahol fellépnek „rendetlenségi” feltevések ([...] azáltal, hogy – Gibbs eljárása szerint – az „időátlag” indokolás nélkül a „mikrokanonikus sokaságátlaggal” pótoltatik [...])

Az „E. h.” fogalmazása: Minden pálya az energiafelületen mindenütt sűrű, kivéve esetleg egy 0 területű kivételes halmazt. Sőt a pálya mentén csinált statisztikája bármely paraméternek megegyezik a „mikrokanonikus sokaságban” csinált statisztikájával.

[...]

Röviddel ezután J. v. Neumann ezen módszerrel az „Ergod-hypothézis”-t be is bizonyította. Az „E. h.” ekkor 42 éve volt ismert és megoldatlan. Azaz: bebizonyította, hogy az „E. h.” érvényességének szükséges és elegendő feltétele, hogy az energián kívül további „mozgásintegrálok” ne létezzenek (ez lévén Boltzmann és Maxwell sejtése).

Neumann-nak ez a bizonyítása tehát lezárt egy rég vajúdo kérdést, és szilárd alapokra helyezte a statisztikus fizika egyre terebélyesedő elméletét. Az ergodicitás kérdése mindazonáltal továbbra is fennáll: konkrét rendszerekre meg kell vizsgálni, teljesülnek-e a Neumann által megadott matematikai feltételek. Ez persze sokkal könnyebb feladat, mintha minden rendszerre közvetlenül próbálnák bizonyítani vagy cáfolni az ergodicitást. A dina-

² A dokumentum megjelent: Nagy F.: Neumann János és a „magyar titok”, OMIKK 1987, p. 205–207. Néhány Neumann által németül írt fizikai szakkifejezést magyar megfelelőjével pótoltam.

mikai rendszerek elméletének (melyet magyar fizikusok és matematikusok is intenzíven művelnek) ma is egyik központi kérdése a vizsgált rendszer ergodikus volta.

Neumann tétele hamar közismert lett a fizikusok körében. Ennek egyik jele, hogy amikor a vele amúgy is állandó személyes és levelezési kapcsolatban álló Ortway Rudolf 1934-ben a Magyar Tudományos Akadémia tagjának javasolta Neumannt, e szavakkal méltatta ergodelméleti munkásságát:³

Talán még fontosabbak ama vizsgálatait, melyek a mechanika egy klasszikus problémájára, az ergodhypotézisre vonatkoznak, és melyek eme, a mechanika thermodynamikai alkalmazásának előfeltételét képező probléma teljes tisztázásához vezettek.

Ennél többet ma sem mondhatunk.

4. A kvantummechanika matematikai alapjai

Neumann János a fizika fejlődésére a legnagyobb, máig tartó hatást a kvantummechanika precíz matematikai megalapozásával gyakorolta. Ahhoz, hogy ezt a teljesítményt értékelni tudjuk, át kell tekintenünk – természetesen csak vázlatosan – a kvantumelmélet, ezen belül a kvantummechanika fejlődésének történetét, valamint az ezzel kapcsolatos fizikai és matematikai problémákat.

A klasszikus mechanika csődje

A klasszikus mechanika, ahogy azt (Galilei gondolatai alapján) Isaac Newton kidolgozta, és a maga (és tőle függetlenül Leibniz) feltalálta matematikai módszerrel (a „kalkulusal”, mai nevén a differenciál- és integrálszámítással) összeötvözve hatékony, magyarázó és egyben prediktív erejű elméletté fejlesztette, majd ahogy azt utódai (többek között Euler, Lagrange, Laplace, Hamilton, Jacobi) hajlékony, elegáns, a földi és csillagászati gyakorlatban előforduló problémák kezelésére alkalmas koncepció- és módszeregyüttessé fejlesztették, mi több, világmagyarázó filozófiai rendszert építettek rá (a Laplace-féle mechanikai determinizmust) – tehát ez a klasszikus mechanika univerzális jellegű és igényű elmélet volt. Semmiféle megkötést nem ismert, saját alkalmazhatósági határaitól nem tudott, és nem is akart tudni. A földi és égi jelenségek leírása és megmagyarázása során elért hatalmas sikerei arra indították a kor tudósait, hogy a mechanika által még nem leírt jelenségekre is sorra mechanikai magyarázatokat keressenek – és olykor találjanak is.

³ Ua. p. 209.

Az eredetileg pontrészcskék mozgásának leírására kidolgozott matematikai módszereket hamarosan általánosították a merev testek, a rugalmas közegek, valamint a folyadékok és a gázok mechanikájának tárgyalására. Ehhez új matematikai eszközöket kellett bevezetni, pontosabban megfelelően általánosítani kellett a newtoni kalkulusot. Így dolgozták ki pl. a parciális differenciálegyenletek fogalmát és elméletét. Ebben a korban kart karba öltve haladt a fizika (értsd: mechanika) és a matematika fejlődése, a két tudomány művelői is nagyrészt ugyanazok a tudósok voltak.

A mechanika azonban a fent említett – nyilvánvalóan mechanikai jellegű – jelenségek mellett más fizikai területek magyarázatára is igényt tartott. Ilyen irányú első nagy sikere a hangjelenségek mechanikára történő visszavezetése volt. E jelenségeket egészen más-képp érzékeljük, mint a mechanikaiakat (pl. nem látjuk a részecskék mozgását), mégis, a hangnak a közegben tovaterjedő mechanikai nyomáshullámként való interpretálása egy-csapásra helyére tette az egész jelenségkört.

A következő nagy sikert az előző fejezetben részletesen elemzett kinetikus gázelmélet (majd továbbfejlesztése, a statisztikus mechanika) jelentette. Ezzel ismét a jelenségek egy fontos típusát, a hőjelenségek sok fontos kérdését sikerült bevonni a mechanikai magyarázat és megalapozás ernyője alá (bár láttuk, hogy ennek az eredménynek precíz matematikai megalapozásával Neumann Jánosig kellett várni).

Nyilvánvalónak tűnt, hogy a fizikai jelenségek „maradéka” esetében is előbb-utóbb megtalálják a mechanikai magyarázatot, a megfelelő mechanikai modellt, és a kezelésére szolgáló matematikai eszközöket. A fizikusok már nem sokkal Newton után arról ábrándoztak, hogy egyszer majd felírják – a newtoni gravitációs törvény mintájára – az egymáshoz „vonzódó” kémiai anyagok közti „affinitás”, vagy az akkoriban megismert elektromos és mágneses jelenségek pontos erőtvényét. Sőt egyesek még a szerelmesek közt ható vonzóerő képzetének megtalálásában is reménykedtek.

E remények némelyike valóra vált. Coulomb valóban kimérte és felírta az elektromos töltések közt ható Coulomb-erő törvényét, később a mágneses mező által a mozgó elektromos töltésre gyakorolt Lorentz-erő képlete is megszületett. Hiába reménykedtek azonban a kémiai (és persze a társadalmi) erők mechanikai leírásában. A XIX. század folyamán aztán az is kiderült, hogy a mechanika nem fedi le a fizika teljes területét. Az elektromos és mágneses jelenségkörnek csak egyik aspektusát jelentik a töltésekre gyakorolt erők – a másik (és fontosabb) rész csak egy új fizikai létező, az elektromágneses mező (Faradaytól származó) fogalmának bevezetésével vált megfoghatóvá. Amikor Maxwell az 1860-as években felírta az elektromágneses mező differenciálegyenleteit, ezzel – a mechanika kinetikus gázelméletben elért, egyebek közt neki is köszönhető diadalával lényegében egy időben – megkongatta a lélekharangot a mechanika és a rá épülő mechanikus világkép egyeduralmának korszaka fölött. (Meg kell jegyeznünk azonban, hogy maga Maxwell saját egyenleteit nem az önállóan létező, nemmechanikai jellegű elektromágneses mező mozgástörvényeinek tekintette – mint ahogy ma gondolunk rájuk –, hanem egy mélyebb, mechanikai jellegű mozgás, az éter rezgéseinek és csavarodásainak visszatükröződését látta bennük. Élete végéig kísérletezett azzal, hogy mechanikai modellt konstruáljon egyenletei mögé – az éter kis csavarocskáit, megfeszülő rúdait és örvényeit kereste. Ma azt mondhatnánk: *rejtett paraméteres* elméletet keresett saját, jól működő egyenletei mögé – mint ahogy az utolsó fejezetben látni fogjuk, a szintén jól működő kvantumelmélet mögé is kerestek és keresnek sokan. Maxwell „rejtett paraméteres” elmélete természetesen mechanikai jellegű lett volna. Nagy úr a megszokás, még a legnagyobb tudósok esetében is.)

A klasszikus mechanikát újabb csapás érte a XX. század elején, abból az irányból, ahonnan senki sem számított rá. A speciális relativitáselmélet (amely az elektromágneses fényelmélet és a fényvel végzett kísérletek nem várt eredményeinek értelmezésére született meg) kiderítette, hogy az egymáshoz képest gyorsan (azaz a fény sebességével összehasonlítható sebességgel) mozgó objektumok viselkedésének leírására már nem érvényes a klasszikus mechanika. Pontosabban a klasszikus mechanika alapjául szolgáló tér- és időmodell (amely nagyjából megegyezett a köznapi „józan ész” alapján elképzelt tér- és időfogalommal) nem helyes, hanem egy másik, kissé bonyolultabb tér-idő fogalommal kell pótolni. Ez a paradigmaváltás akkora hatást váltott ki a művelt és a fizika elvi kérdései iránt érdeklődő közvéleményből, hogy hullámai még mára sem csitulnak el egészen.

A relativitáselmélet nem azt mondta ki, hogy a klasszikus mechanika egészében hibás, hanem csak azt, hogy érvényességi köre korlátozott. Új fizikai tartományokra – jelen esetben a korábban nem vizsgált nagy sebességek tartományára – kiterjesztve vizsgálatainkat, új törvényeket tapasztalunk. E törvények *határeseteként* kapjuk meg a klasszikus mechanika törvényeit – jelesül akkor, ha a mozgó testek relatív sebessége sokkal kisebb a fény sebességénél. Ezek az új törvények történetileg oly módon bukkantak fel, hogy az új tartományba eső jelenségekre a fizikusok (jobb híján, és persze azért, mert meg voltak győződve általános érvényességükről) megpróbálták alkalmazni a klasszikus mechanika ismert törvényeit – és a kísérleti tapasztalat ellentmondott az így kiszámolt jóslatoknak.

Nagyon hasonló szituáció jellemzi a klasszikus mechanika más jellegű korlátait felfedező, és az új terület új törvényeit feltérképező kvantumelmélet születését is. Új fizikai objektumok – ezúttal az atomok és alkotórészeik – bukkantak fel a fizikusok szemhatárán, és a kísérletezők laboratóriumaiban gyűlni kezdtek a rájuk vonatkozó (persze igencsak közvetett) adatok. Így első lépésben az derült ki, hogy az atomok – amelyeket korábban többen csak a vegyszerek elemkonstrukcióinak tartottak – valóságosan léteznek, sőt hamarosan sikerült méretüket és tömegüket is meghatározni. Nemsokára már azt is tudták, hogy az atomok – görög eredetű nevükkel ellentétben – nem oszthatatlanok, hiszen sikerült kisebb alkotórészeiket (az elektronokat) „kipiszkálni” belőlük. Az élesebb eszű fizikusok máris elkezdtek tűnődni egy furcsaságon, a feltáruuló mikrovilág első, még egyáltalán nem szembeötlő paradoxonán, egy olyan tényen, amivel a klasszikus mechanika (sőt az egész klasszikus fizika) egyáltalán nem tudott mit kezdeni.

Már Demokritosz, az akkor még teljesen spekulatív atomhipotézis első ókori megalkotója is úgy gondolta, hogy egy homogén anyagtömb egyforma atomokból áll. (Lásd a gye-rekdalt: „*Vízből van a víz, sóból van a só – én meg gyerekből vagyok!*”) A kémiai atomfogalom kidolgozói, Dalton, Berzelius és követőik is magától értetődőnek tekintették, hogy egy kémiai elem atomjai teljesen egyformák. Más elem atomjaitól persze különböznek, egymástól azonban abszolút megkülönböztethetetlenek. A XIX. század végére azonban kiderült, hogy az atom kisebb alkotórészekből áll, azok stabil kötött állapota. A XX. század elején Rutherford felfedezte az atommagot, és megalkotta az atom Naprendszer-jellegű modelljét: az elektronok a nehéz atommag körül keringenek, mint a bolygók a Nap körül (sok ember atomokról szóló ismeretei ma sem mennek ennél tovább).

Ha a legközelebbi csillaghoz érkező majdani űrszondánk azt találná, hogy annak is pont 9 bolygója van, méretük és keringési adataik pedig megegyeznek a mi Naprendszerünk bolygóival – igencsak csodálkoznánk (vagy azt hinnénk, hogy a szonda eltévedt, és tévedésből visszaérkezett). A newtoni mechanika azt tanítja, hogy a gravitációs erőhatások ál-

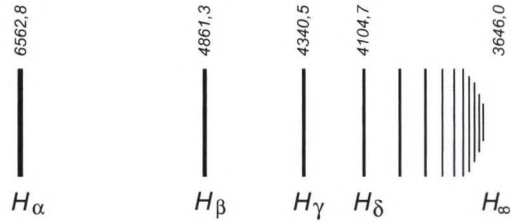
tal összetartott összetett képződmény *mozgástörvényeit* az általános fizikai elvek határozzák meg, konkrét szerkezetüket viszont az esetleges, történeti véletlenektől függő *kezdeti feltételek*. Ezért a csillagok körül keringő bolygók száma és elhelyezkedése igen változatos lehet. Ezzel szemben az atomok alkotórészeit összetartó erőhatás „filozófiája” egészen más (noha – mint kiderült – a mag és az elektronok között a jól ismert elektrosztatikus vonzóerő hat, amelynek képlete azonos a tömegvonzás képletével) – hiszen a hatására kialakuló összetett képződmények valamennyien teljesen egyformák... Mi akadályozza meg például azt, hogy az atomnak egy szomszédjával történő ütközése az elektron pályáját egy hajszálnyival megváltoztassa, ezzel kicsit módosítva az atom átmérőjét? A klasszikus fizika értetlenül áll ez előtt a jelenség előtt: vajon miért különbözik ennyire az atomokat összetartó hatás a megszokottól? (Pontosabban: a klasszikus fizikát művelő tudósok jelentős részének fel sem tűnt ez a furcsaság, csak akkor vettek róla tudomást, amikor a kvantummechanika megoldotta a rejtélyt.)

A Rutherford-féle atommodellel szemben felvethető még egy magától értetődő ellenérv (és a kortársak azonnal fel is vetették): hogy lehetséges az, hogy az atommag körül keringő elektron tartósan pályáján marad? A keringő, azaz gyorsuló elektromos töltésnek a Maxwell-egyenletek szerint elektromágneses hullámokat kellene kisugároznia (az elméletnek ezt a jóslatát akkoriban közvetett módon, később közvetlenül is igazolták, persze makroszkopikus mozgások esetén) – a sugárzás pedig energiát és perdületet visz el, így az elektron egyre közelebb spirálozik az atommaghoz, végül belezuhan, a számítások szerint a másodperc igen kis törtrésze alatt. Az így felépülő atomok az érvényes fizikai elmélet szerint nem lehetnének tartósan fennmaradó, stabil képződmények – pedig mégis azok.

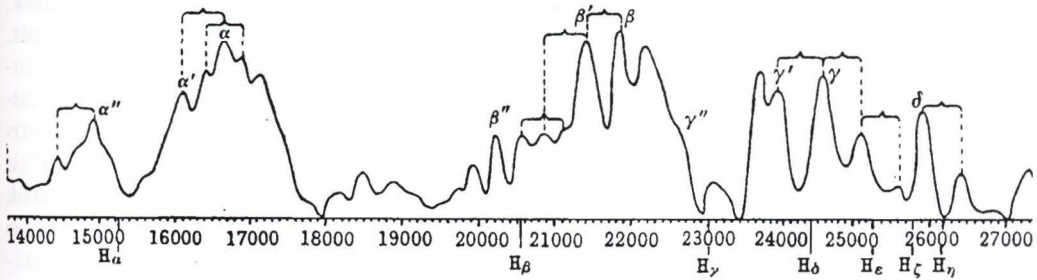
További nehézséget okoztak az atomfizika úttörői számára a kémiai jelenségek. A vegyészek a „vegyérték” fogalmával szimbolizálták az atomoknak azt a tulajdonságát, hogy csak igen válogatós módon hajlandók egymással bonyolultabb kötött állapotot, még összetettebb struktúrát, molekulát alkotni: egy oxigénatom csak két hidrogénatommal alkot kötött rendszert (a közismert vízmolekulát), eggyel, hárommal vagy negyvenkettővel nem. Ez a tulajdonság teljesen idegen a klasszikus fizikától: ha a kilencbolygós Naprendszerbe behelyeznénk egy tizedik bolygót, akkor egy kicsit megváltozna a többi bolygó pályája, de semmi drasztikus nem történne, nem lépne fel semmi olyan erő, amely kilökné a rendszerből a „felesleges” bolygót, megakadályozná annak csatlakozását. Ugyanígy semmi sem korlátozza egy elektromosan feltöltött test töltésének nagyságát, az általa kifejezhető erőt vagy az általa odavonzható kicsiny bodzabél golyócskák számát sem. A vegyérték fizikai mibenlétéről senkinek sem volt semmiféle fogalma. Az atom összetett voltának felismerésével sejthetővé vált, hogy a „vegyértéknek” valami köze lehet az atom belső szerkezetéhez, de a kapcsolat mibenléte rejtély maradt.

Hasonló, de még vadabb furcsasággal szembesültek az atomok által kibocsátott fény spektrumát vizsgáló fizikusok. A XIX. század közepén fedezték fel, hogy minden elem jellegzetes, meghatározott hullámhosszú hullámok kombinációjából álló sugárzást bocsát ki (az elem gőze pedig a folytonos spektrumú sugárzásból, pl. egy izzó test fényéből pontosan ugyanezen hullámhosszú komponenseket nyeli el). Az elem spektruma mintegy az illető elem atomjainak ujjlenyomatául szolgál – a sugárzás alapján messziről is meg lehet állapítani az adott elem jelenlétét, sőt mennyiségét is. Ezen a jelenségen alapul a távoli objektumok, pl. a csillagok atomi összetételéről szerzett összes tudásunk. Egy adott elem (pl.

a sokat vizsgált hidrogén vagy az alkálifémek) spektrumvonalainak hullámhosszai között érdekes és furcsa matematikai kapcsolatokat állapítottak meg a spektroszkópusok. A dolog a titkosírások megfejtéséhez hasonlít: az adathalmaz túl szabályos, rendszer van benne, ezért bizonyosak vagyunk benne, hogy valamit közölni akarnak vele – csak épp a kódot, és így a tartalmat sem értjük. Egy-egy részsiker, pl. a hidrogén vonalaira vonatkozó nevezetes Balmer-formula annak felel meg, amikor a titkosírás kutatói rájönnek, hogy egy bizonyos jel a király nevét kódolja – de még messze vannak az írás értelmének megfejtésétől.



9a) ábra. Részlet a hidrogén spektrumából. Az a) ábra a prizmával vagy spektroszkóppal létrehozott színekép egy részét, a látható fény frekvenciatartományát mutatja be. A fényes színekvonalak (a negatív felvételen sötét vonalak) alkotják az ún. Balmer-sorozatot. A vonalakhoz írt számok a hullámhosszal arányosak. Ezek között a számadatok között Balmer egy egyszerű képlettel leírható összefüggést talált, amit a Bohr-modell, majd később a kvantummechanika alapján sikerült megmagyarázni.



9b) ábra. Az ábra egy modernebb, nagyobb felbontású spektroszkóppal készült, és a sugárzás egyes összetevőinek intenzitását is jelzi. A vízszintes tengelyre írt számok a hullámhosszal arányosak. A függőleges tengely a sugárzás intenzitását méri. A görög betűkkel jelölt csúcsok a hidrogén korábban is ismert jellegzetes színekvonalai. Az ábrán látható, hogy a nagyobb felbontású spektroszkópban a már korábban elnevezett vonalak egész vonalrendszerre bomlanak fel, ezeket a görög betűk mellé tett vesszőkkel jelölik. A spektrumnak ez a „finomszerkezete” okozta a legnagyobb nehézséget a korai kvantumelméletben. A finomszerkezet pontos magyarázata végül a Neumann és Wigner által kifejlesztett csoportelméleti módszerrel sikerült (lásd pl. a 15. ábrát).

A hőmérsékleti sugárzás

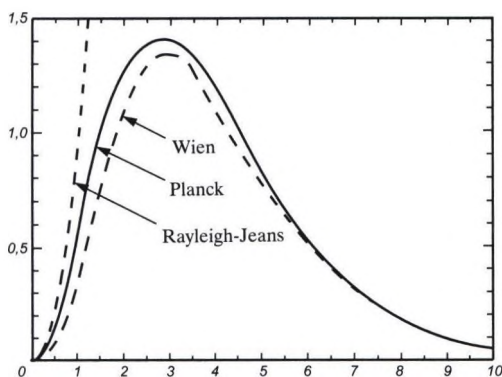
A XIX. század vége még egy klasszikus tudományág, a (statisztikus mechanikával alátámasztott) termodinamika hagyományos fogalom- és feltevérendszerének csődjét is elhozta: a kísérletek eredménye durva ellentmondásba került az elmélet előrejelzéseivel.

Az elektromágneses mező egyenleteinek kidolgozása után (ezt Maxwell végezte el az 1860-as években) hamarosan kitűnt, hogy ez a mező is az anyag egyik új, nem hagyományos formája. A mező is rendelkezik az anyag alapvető attribútumaival, energiája, impulzusa van, erőt képes kifejteni, energiát tud átadni. Ezért alanya lehet az energiaátadási jelenségek általános tudományának, a termodinamikának.

A legegyszerűbb ilyen vizsgálható jelenség az üregbe zárt elektromágneses sugárzás. Ha a zárt üreg fala nincs az abszolút nulla fokra lehűtve, sugárzást bocsát ki. Ennek teljes energiasűrűsége a kísérletileg talált Stefan–Boltzmann-törvény szerint az abszolút hőmérséklet negyedik hatványával arányos. A sugárzás nem tud távozni az üregből, azt a fal egy másik része elnyeli, majd újból kisugározza. Emlékezzünk vissza a molekulák „önreprodukáló” energiaeloszlására, ami a hőmérsékleti egyensúlyra jellemző! Előbb-utóbb itt is kialakul az egyensúly, azaz az üregben mindenütt ugyanolyan sugárzási energiasűrűség lesz, és a fal minden darabkája időegység alatt ugyanannyi energiát nyel el, mint amennyit kisugároz: beáll a stacionárius állapot.

Felmerül a kérdés, hogy milyen lesz a sugárzás spektrális eloszlása, azaz az energia mekkora része jut az egyes frekvenciatartományokra. A XIX. század végének fizikusai úgy érezték, hogy minden eszköz a rendelkezésükre áll a kérdés elméleti megválaszolására: ismerik az elektromágneses sugárzás törvényeit, a termodinamikát, és a termikus egyensúly statisztikus fizikai megalapozását. Neki is láttak a számításnak, és döbbenetes eredmény jött ki: az üregben a sugárzás összenergiája végtelen nagyra adódott (közvetlen ellentmondásban a határozott, véges értéket adó kísérleti eredménnyel).

Természetesen sokan és sokféleképpen keresték a hibát a levezetésben, de nem találták. A számítás kulcsa a statisztikus fizika egyik fontos eredménye, az ekvipartíció tétele volt. Ez a tétel azt mondja ki, hogy egy termodinamikai egyensúlyban levő fizikai rendszer minden szabadsági fokára átlagosan ugyanannyi energia jut – ez a mennyiség arányos a hőmérséklettel. („Szabadsági fokon” a rendszer teljes leírására alkalmas adatok minimális számát értjük. A legegyszerűbb esetben, az N golyóval modellezhető ideális gáz esetén ez épp a golyók összes hely- és impulzusvektorának 3-3 komponense, azaz összesen $6N$ adat, ami épp a fázis tér dimenziószáma.) A probléma magját az jelenti, hogy az elektromágneses sugárzás végtelen szabadságfokú rendszer! (Egyszerű belátni: ahhoz, hogy az üregben pontosan megismerjük az elektromágneses mező állapotát, a tér minden pontjában meg kell adnunk az elektromos és a mágneses térerősség vektorának irányát és nagyságát. A térnek pedig – bármilyen kicsiny is az üreg térfogata – végtelen sok pontja van.)



10. ábra. A Planck-görbe. Az ábrán az üreg falával termikus egyensúlyban levő elektromágneses sugárzás energiasűrűsége látható a frekvencia függvényében. A vízszintes tengely beosztása a frekvencia és a hőmérséklet hányadosával arányos, a függőleges tengely az energiasűrűséget méri. A görbe alatti terület végtelen lenne, azaz végtelen nagy teljes energiasűrűséget jelentene, ami nyilvánvalóan nonszensz eredmény. (A másik szaggatott vonal az elektromágneses fényelmélet alapján számított közelítő Wien-törvény görbéje: ez nagy frekvencián helyes, kis frekvencián viszont jelentősen eltér a mérési adatoktól.)

Természetesen sokan és sokféleképpen keresték a hibát a levezetésben, de nem találták. A számítás kulcsa a statisztikus fizika egyik fontos eredménye, az ekvipartíció tétele volt. Ez a tétel azt mondja ki, hogy egy termodinamikai egyensúlyban levő fizikai rendszer minden szabadsági fokára átlagosan ugyanannyi energia jut – ez a mennyiség arányos a hőmérséklettel. („Szabadsági fokon” a rendszer teljes leírására alkalmas adatok minimális számát értjük. A legegyszerűbb esetben, az N golyóval modellezhető ideális gáz esetén ez épp a golyók összes hely- és impulzusvektorának 3-3 komponense, azaz összesen $6N$ adat, ami épp a fázis tér dimenziószáma.) A probléma magját az jelenti, hogy az elektromágneses sugárzás végtelen szabadságfokú rendszer! (Egyszerű belátni: ahhoz, hogy az üregben pontosan megismerjük az elektromágneses mező állapotát, a tér minden pontjában meg kell adnunk az elektromos és a mágneses térerősség vektorának irányát és nagyságát. A térnek pedig – bármilyen kicsiny is az üreg térfogata – végtelen sok pontja van.)

Max Planck, a termodinamika egyik vezető tudósa 1900 novemberében kapta meg a sugárzás spektrális eloszlására vonatkozó addigi legpontosabb mérések adatait. A kísérleti görbe a kis frekvenciák tartományában megegyezett az elméletileg számítottal, nagyobb frekvenciáknál viszont lekonyult,

és a görbe alatti terület, ami a teljes energiasűrűséggel arányos, a Stefan–Boltzmann-törvénnyel megegyező értékűnek adódott.

Planck pusztán matematikai játékként megpróbált olyan függvényt találni, amely a kis és a nagy frekvenciák tartományában is pontosan írja le a görbe menetét. Ő maga döbönt meg leginkább, amikor ez sikerült. A viszonylag egyszerű függvény ma a Planck-görbe nevet viseli. Ennek a képletben bukkant fel először a XX. század meghatározó fizikai állandója, a Planck-féle hatáskvantum.

Planck bemutatta eredményeit kollégáinak, majd megpróbálta úgy módosítani a korábbi termodinamikai számításokat, hogy a most már ismert képletű függvényt kapja eredményül. Hamarosan kiderült, ehhez egyetlen feltevésre van szükség: adott frekvenciájú sugárzás energiája nem vehet fel tetszőleges értéket, csak a frekvenciával arányos „*energiakvantum*” egész számú többszöröse lehet. Az energia és a frekvencia közti arányossági tényező épp a Planck-állandó. Ebből a feltevésből (meg persze a statisztikus fizika és a termodinamika szokásos apparátusának bevetésével) már levezethető a Planck-görbe alakja, a kísérletekkel teljes egyezésben.

A XX. század első perceivel tehát a „kvantum” fogalma is megjelent a fizikában, hogy aztán a róla elnevezett tudományág a század meghatározó (és az atombombától a számítógépig terjedő alkalmazásokban is megjelenő) elméletévé nője ki magát.

Planck maga is érezte, hogy kényszer szülte feltevése ellentmondásban van saját egész korábbi munkásságával, minden fizikai ismeretével. Mi köze lehet egymáshoz a sugárzás frekvenciájának és energiájának? Egy hullám energiáját frekvenciája mellett amplitúdója is meghatározza – és ez a két adat a klasszikus mechanika hullámtana, valamint az elektromágnesség Maxwell-elmélete szerint is egymástól teljesen független, szabadon megválasztható mennyiség volt. Ha az amplitúdót egy kicsit megváltoztatjuk, a sugárzás energiája megnő – de mitől lenne ez a növekedés ugrásszerű, „kvantált”? A klasszikus elmélet szerint az energia folytonosan változna...

A kérdésre ezen a szinten nem volt válasz. Előállt tehát a XX. század első évtizedei fizikájának tipikus alaphelyzete: a klasszikus elmélet nem működik, helytelen vagy abszurd eredményekre vezet – a kísérleti eredmények helyes leírásához pedig az addigi elmélet szellemétől teljesen idegen, ahhoz szervesen nem illeszkedő, kiegészítő „*kvantumhipotézisek*” bevezetése szükséges. Az így előállt „öszvér” elmélet jól írja le a tényeket, csak épp senki sem érti – még maga az elmélet megalkotója sem.

Az elmélet feldozgatása

Hamarosan hasonló helyzet állott elő az atomfizikában is: megszületett az oda vonatkozó „öszvér” elmélet, az atomok Bohr-modellje.

A fordulatot az hozta meg, amikor 1913-ban Niels Bohr – támaszkodva Plancknak a sugárzás energiakvantumaira vonatkozó hipotézisére – egyszerű, ám teljesen érthetetlen modellt adott az atomok szerkezetére, amivel a korábban említett furcsaságok jó része magyarázható volt. Bohr modellje a klasszikus mechanika és néhány attól teljesen idegen, ún. „kvantumszabály” együttese. Azért nem mondjuk, hogy ötvözet, mert az össze nem illő komponensek továbbra is idegenek maradtak, az elmélet számos sikere ellenére sem sikerült szervülniük.

Bohr modellje elfogadja a klasszikus mechanikát, annak alapján számítja ki az elektron pályáját az atomban, azonban egy extra szabály alapján megtiltja, nem engedélyezi a klasszikusan lehetséges ellipszispályák nagy részét. A szabály szerint bizonyos mechanikai mennyiségek (az ún. hatásváltozók, amelyeket az égimechanikával foglalkozó csillagászok már a XIX. század közepén bevezettek, és amelyeket a keringő részecske pályájának adataiból lehet kiszámítani) értéke csak egy alapvető természeti állandó, a *Planck-féle hatáskvantum egész számú többszöröse* lehet. Ez a szabály alaposan megrostálja a lehetséges pályákat, és – ami a legfontosabb – a klasszikus mechanika szerint folytonos pályasokaságot diszkrétté, „kvantálttá” teszi.

A hatásváltozók a pálya adataiból számíthatók ki, azokat pedig a kezdőfeltételek határozzák meg. Ha a kezdőfeltételeket egy parányit megváltoztatjuk, a mozgás egy kicsit megváltozik, és persze a hatásváltozók értéke is. Csakhogy ha az eredeti mozgás Bohr szerint megengedett volt, az új, attól kissé különböző mozgás már nem lesz az, hiszen a hatásváltozó csak a Planck-féle hatáskvantum törtrészével módosult. Kissé durvábban meg kell tehát változtatnunk a mozgást, hogy a hatásváltozó egy egész kvantumnyit ugorjon – a megengedett pályák klasszikus folytonos sokasága helyett a meg nem engedett pályák folytonos halmazából itt-ott felbukkanó, egymástól véges távolságban, diszkrétan elhelyezkedő megengedett pályákkal találkozunk. Ezeket akár meg is számozhatjuk, hiszen a helyzet ahhoz hasonlít, mint amikor a számegyenes folytonosan elhelyezkedő valós számai helyett csak az egymástól elkülönülten elhelyezkedő egész számokat vesszük figyelembe.

Bohr egyik posztulátuma szerint a megengedett pályán mozgó elektron nem sugároz ki elektromágneses hullámokat. Maxwell forog a sírjában, de a modell jól működik.

Más fizikai mennyiségek is kiszámíthatók a pálya alapján, például a rendszer energiája. És ha a pályák diszkrét lépésekben változnak, az energia sem vehet fel akármilyen értéket. A klasszikus mechanikából származó energiaképleteket és a diszkrét, kvantált pályákat figyelembe véve Bohrnak sikerült kiszámítania az egyik legegyszerűbb, de a valóságban is előforduló rendszer, a hidrogénatom (azaz a magányos proton körül keringő egyetlen elektron) lehetséges energiáit.

És persze a különböző pályák energiáinak különbségét. Hiszen a Bohr-modell következő feltevése épp az, hogy a különböző energiájú pályák közt van átmenet – nem folytonos, hanem diszkrét, pillanatszerű: az elektron mintegy „átugrik” egyik pályáról a másikra. Ha a nagyobb energiájú pályáról ugrik a kisebb energiájúra, akkor közben pedig kisugározza a felesleges energiát, azaz a két pályához tartozó energiák különbségét. A fordított irányú ugráshoz pedig a megfelelő energiakülönbség befektetésére van szükség – ezt külső forrás, általában az atom által elnyelt sugárzás biztosítja. Planck képlete alapján viszont a sugárzás elemi adagjának (a fotonnak) energiája arányos a sugárzás frekvenciájával. Lássunk csodát: Bohrnak ezzel az összetákolt (és belső ellentmondásokat tartalmazó) modellel sikerült kimagyaráznia az atomok által kibocsátott fény diszkrét spektrumvonalait. Mi több, ez az egyszerű számítás a hidrogénatom esetén pontosan megadta a spektroszkópusok által kísérletileg talált Balmer-formulát! Az abban szereplő két, addig misztikusnak tűnő egész szám hirtelen egyszerű magyarázatot kapott: az „ugrás” előtti és utáni elektronpálya sorszáma jelent meg a formulában.

Az atomok titokzatos stabilitása is érthetővé vált. A Bohr-modell feltevései szerint az elektron „önként” csak a nagyobb energiájú állapotból ugorhat át a kisebb energiájúba, hiszen a fordított irányú ugráshoz valahonnan energiát kell szereznie. A modell szerint vi-

szont a rendszernek van egy legalacsonyabb energiájú, ún. *alapállapota* – innen nem tud hova ugrani az elektron, mert már nincs kisebb energiájú állapot, következésképpen tartósan ebben az állapotban marad. Így a rendszer stabil.

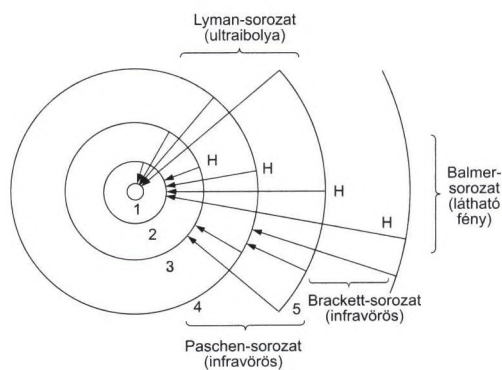
Azt is megmagyarázhatjuk, miért nyelik el az atomok épp azokat a frekvenciájú sugárzásokat, amiket kibocsátani is szoktak: az adott frekvenciának pontosan megfelel két megengedett atomi pálya energiájának különbsége, így az alacsonyabb energiájú pályán mozgó elektron a sugárzást elnyelve feljebb, magasabb energiaszintre kerül. Az elem gőzén áthaladó, folytonos spektrumú fényből a nekik kedves frekvenciájú sugárzást elnyelik az atomok, és ún. gerjesztett állapotba kerülnek – a továbbhaladó fényből pedig hiányzik a megfelelő frekvenciájú komponens. (Persze az atom később visszaugrik az alacsonyabb energiájú állapotba, újra kibocsátva a fényt – csak hogy ez a kisugárzás általában nem az eredeti fénysugár terjedési irányába történik.)

A Bohr-modell alapján azonnal megoldhatóvá, sőt triviálissá válik a régi rejtély is: miért is egyformák az atomok. A válasz egyszerű: a megengedett állapotok halmazából mindegyik ugyanabban az állapotban, a legalacsonyabb energiájú alapállapotban van. Így minden fizikai tulajdonságuk megegyezik. Külső energia befektetésével az atom kibillenthető ebből az állapotból (gerjeszthető), de hamarosan visszatér az alapállapotba. Ekkor pedig már nem különböztethető meg attól a rokonától, amely ezalatt végig az alapállapotban maradt.

A közönséges, atomos anyag létét és stabilitását tehát sikeresen megmagyarázta a Bohr-modell. Hasonlóképpen bonyolultabb és fejlettebb utódelméletei is. Ma is a kvantumelmélet egyik legnagyobb sikerének tartjuk, hogy megoldotta ezt a sokak számára korábban fel sem merült alapvető problémát: miért stabil a bennünket körülvevő anyag.

(A Bohr-modellben egyszerű feltevésnéként szerepelt, hogy az elektron a kisebb energiájú állapotból csak külső energiabefektetés, „gerjesztés” hatására megy át nagyobb energiájú állapotba, a fordított folyamat viszont „önként” végbemegy – ez az ún. spontán emisszió. A későbbi kvantummechanikában épp ez az önkéntesség okozott gondot: a Schrödinger-egyenlet szerint az atom gerjesztett állapotai is stabilak, az elektronnak esze ágában sincs leugrani az alapállapotba. Erre a rejtélyre csak az elmélet következő lépcsőfoka, a kvantumelektrodinamika adott választ. Ez az elmélet, amely már az elektromágneses mező kvantumtulajdonságait is figyelembe veszi, a spontán emissziót az elektronnak az elektromágneses mező kvantumfluktuációival történő kölcsönhatásával magyarázza.)

A klasszikus mechanikából, elektrodinamikából, valamint a mindkettőnek ellentmondó kvantumposztulátumokból felépített Bohr-modell tehát – legalábbis kvalitatíve – jól leírta a tapasztalt atomfizikai jelenségeket. A legegyszerűbb atom, a hidrogénatom esetében pe-



11. ábra. A hidrogénatom spektrumának (lásd 9. ábra) magyarázata a Bohr-modell alapján. A spektrumban korábban felismert színekvonalak a lehetséges elektronpályák közti átmeneteknek felelnek meg. A felfedezőikről elnevezett vonalsorozatok a különböző szintekről induló, de azonos elektronpályán végződő ugrások során kibocsátott színekvonalakat tartalmazzák.

dig, amikor a szükséges számításokat ténylegesen is el lehetett végezni, az eredmény pontosan egyezett a kísérleti tapasztalattal.

A következő évtizedben hihetetlen pezsgés indult meg: fizikusok hada vetette magát az új elméletre. Egyik részük megpróbálta megalapozni, egyszerűbb és érthetőbb hipotézisekre visszavezetni Bohr feltevéseit (erre vonatkozott Bohr híres *bon mot*-ja: *Az Ön elmélete kétségtelenül örült. De nem eléggé örült ahhoz, hogy igaz lehessen.*). Egyes fizikusok a kor másik „örült”, de igaznak bizonyult elméletét, a relativitáselméletet próbálták bevetni, és a kvantumfeltételeket a téridő nemtriviális tulajdonságaira akarták visszavezetni. (Később, a harmincas évek elején Neumann is próbálkozott egy ilyen elmélettel.)

A tudósok másik része pedig – nem kutatva a Bohr-posztulátumok „eredetét”, de látva azok működését a hidrogénatom esetében – más, összetettebb rendszerekre, bonyolultabb atomokra, illetve molekulákra is megpróbálta alkalmazni. E rendszerek spektrumában is látszik valami szabályosság, csak nem olyan egyszerű, mint a hidrogénatoméban. A kutatók igen nagy matematikai bonyodalmakba ütköztek. Már az előző fejezetben is volt szó arról, hogy míg a kéttestprobléma a klasszikus mechanikában egyszerűen tárgyalható, három test esetében már leküzdhetetlen matematikai nehézséggel jár a lehetséges pályák kiszámítása. A hidrogén után következő legegyszerűbb atom, a héliumatom pedig három testből: atommagból és két elektronból áll. A pályák meghatározása után ki kellene választani közülük a Bohr által megengedettet – csakhogy a pályákat sem lehet pontosan kiszámítani. Érdekes módon az alkálifémek (a periódusos rendszer első oszlopának elemei, pl. a kálium és a nátrium) esetén viszonylag jól működött Bohr módszere, más atomokkal nem sikerült dűlőre jutni.

A hullámmechanika

Újabb fordulatot 1924 hozott, amikor egy fiatal francia fizikus, Louis de Broglie egyszerű, szemléletes magyarázatot talált Bohr kvantumszabályaira – legalábbis a megengedett pályákat kiválasztó szabályra.

Gondoljuk meg, a fizika mely klasszikus területén találkozunk a Bohr-féle szabályokhoz hasonló feltételekkel: a lehetséges mozgások folytonos halmazából bizonyos ésszerű feltételek csak néhány, diszkrétan elhelyezkedő mozgás megvalósulását engedik meg? A válasz: a hullámtanban.

Vizsgáljuk meg a két végén leszorított húr, azaz egy transzverzális rezgésekre képes, kifeszített rugalmas szál mozgását! A mozgásra vonatkozó ún. hullámegyenlet legegyszerűbb formális megoldásai a tetszőleges frekvenciájú harmonikus (azaz időben és térben is szinuszos függvény szerint változó) rezgések. Ha csak határozott frekvenciájú, szinuszos időfüggésű mozgásformákat keresünk, akkor a határfeltétel (az, hogy a húr két végét rögzítettük) csak eme rezgések egy bizonyos részhalmazának megvalósulását teszi lehetővé: azokat, amelyek esetén a húr megadott hosszán éppen egész számú félhullám fér el. A hullámhossz tehát egy rögzített hosszúság és egy (pozitív) egész szám hányadosa lehet. A hullámhossz alapján kiszámítható a rezgésszám is (fordítva arányosak): a rezgésszám épp egy alapfrekvencia egész számú többszörösének adódik. A húr egyes rezgési lehetőségei, „módusai” megszámozhatók, egész számmal jellemezhetők, akárcsak az atomi energiaszintek.

A húr természetesen bonyolultabb mozgásokra is képes. Az általános mozgás az egyes a határozott frekvenciájú egyszerű mozgások (ún. normálrezgések) keverékének, *szuperpozíciójának* tekinthető. Ilyenkor a húr egy adott pontjának kitérését úgy kapjuk, hogy az egyes normálrezgések alapján kiszámolt kitérésüket összeadjuk. Eme összeadódási, szuperponálódási lehetőség a mozgást leíró hullámegyenlet lineáris volta teszi lehetővé.

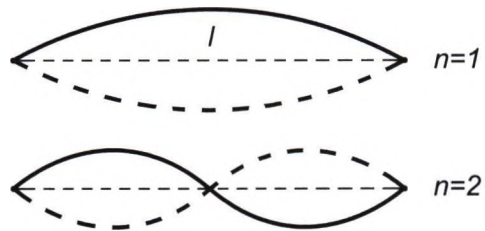
Hasonló a helyzet bonyolultabb rezgő rendszerek, pl. egy kifeszített membrán vagy egy üregben rezgő levegő esetén. Itt is felbukkannak egész számok, ún. kvantumszámok a rezgési lehetőségek indexelésére.

A lehetséges frekvenciákat most bonyolultabb képlet állítja elő, azok nem többszörösei egy alapfrekvenciának, nincsenek „harmonikus” viszonyban. (Ezért ad ki harmonikus zenei hangot az egydimenziós húr vagy a fuvola csövében rezgő levegő, és ezért ad diszharmonikus hangot, zajt a kétdimenziós membránú dob.) E bonyolultabb rezgő rendszerek esetén az egyes normálrezgések szuperpozíciójaként kapható meg a hullámegyenlet általános megoldása.

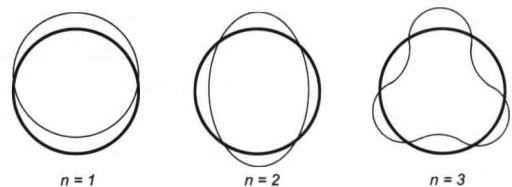
De Broglie ötlete szerint valami hasonló hullámjelenségre kellene visszavezetni a Bohr-féle posztulátumot is. De mi adja a határfeltételt? A zseniális megoldás szerint az, hogy nincs határ: ha az elektron az atommag körüli pályán hullámszerűen viselkedik, akkor a pálya köre (vagy ellipszise) mentén éppen egész számú hullámnak kell elférnie, hogy a hullámgörbe – körbeérve – akadálytalanul folytatható legyen.

Ez eddig természetesen csak – kissé bonyolult – átfogalmazása a Bohr-posztulátumnak: minden Bohr-pályára kiszámíthatjuk a hullámhosszat, de ez nem alapvetően új információ. De Broglie azonban megmutatta, hogy minden ismert (egzaktul számítható) esetben az így kapott hullámhossz fordítva arányos a megfelelő klasszikus mozgás során kiszámítható impulzussal, az arányossági tényező éppen a Planck-állandó.

Ez viszont olyan általános törvénynek tűnik, amely mélyebb, mint a Bohr-posztulátum! Ha ezt a tételt állítjuk az elmélet élére, akkor Bohr eredményei ebből levezethetővé válnak (és a tudósok reménye szerint a Bohr módszerével kezelhetetlen bonyolultabb esetek is vizsgálhatóak lesznek).

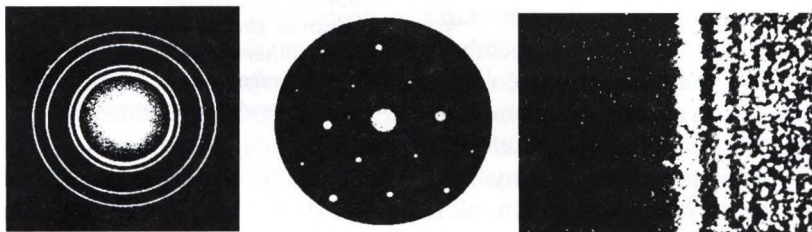


12. ábra. A húr normálrezgései. A két végén rögzített húr (pl. a gitár húrja) akkor végezhet tiszta harmonikus (azaz fix frekvenciájú, tiszta szinuszos) rezgéseket, ha a húr hossza épp a félhullámhossz egész számú többszöröse. Az ábrán az első három normálrezgés látható. Adott l hosszúságú húr esetén az n -dik normálrezgés hullámhosszával fordítva arányos a rezgés frekvenciája, azaz a húr által kibocsátott hang magassága.



13. ábra. De Broglie-hullámok az atommag körül. De Broglie szerint csak azok a pályák valósulhatnak meg, amelyekben éppen „elfér” az elektronhoz rendelt hullám, azaz a körpálya kerülete egész számú többszöröse a hullámhossznak. A hullámhossz de Broglie képlete alapján fordítva arányos az impulzussal, az impulzust viszont a klasszikus mechanika alapján kell kiszámítani. Ez az eljárás a hidrogénatom esetében a Bohr-féle kvantálással azonos eredményre vezet: azonos pályasugarakat és azonos energiaszinteket határoz meg. Az ábrán az első három Broglie-féle hullám szimbolikus ábrázolása látható (valójában különböző sugarú körökre kellene elhelyezni a hullámokat).

Bohr elmélete nem mond semmit az elektron szabad mozgásáról, azaz arról, hogyan viselkedik akkor, amikor nincs atomban kötve. Ha viszont elfogadjuk de Broglie hipotézisét, az elektronhoz ekkor is hozzárendelhető egy hullám. A hullámok alapvető tulajdonsága a diffrakció és az interferencia: megfelelő rácsra ejtve a hullám elhajlik, ernyőre érve pedig két hullámhegy egymást erősítő, illetve hullámhegy és hullámvölgy egymást gyengítő hatása következtében jellegzetes interferenciacsíkok kialakulása várható. De Broglie képlete alapján kiszámítható, hogy milyen sűrű rácsra van szükség a mérhető elhajlási kép létrehozásához: a kristályrácsban elrendezett atomok alkotta diffrakciós rács éppen megfelel. És valóban: a hipotézis után nem sokkal már ki is mérték az elektronhullámok kristályrácsra történő elhajlása okozta interferenciaképet.



14. ábra. Elektronhullámok diffrakciós képei. Az első fényképen polikristályos szerkezetű alumíniumfólián, a másodikon nátrium-klorid egykristályon létrejött diffrakció látható. Korábban hasonló képeket csak röntgensugarakkal (melyek az elektromágneses hullámok egyik fajtáját alkotják) tudtak létrehozni. A harmadik felvétel a látható fény optikájából is jól ismert jelenséget, a hullámok él mentén történő elhajlása során létrejövő interferenciaképet mutat be – csak hogy ezúttal nem fény, hanem elektronhullám hajlott el egy alumínium-oxid kristály éles szélén. A felvételek nem sokkal az elektronok hullámtermészetének felfedezése után, az 1920-as évek közepén készültek.

Innen már csak egy lépés hiányzott a modern kvantummechanika megszületéséhez: a hullámokhoz hullámegyenlet kell, aminek a hullámfüggvény a megoldása. Két *ad hoc* hipotézist (Planck tételét az energia és a frekvencia arányosságáról, de Broglie képletét a hullámhossz és az impulzus fordított arányosságáról), valamint a részecske energiája és impulzusa között a klasszikus mechanikában fennálló kapcsolatot felhasználva Erwin Schrödinger osztrák fizikus felírta az elektronhullámok hullámegyenletét, először csak szabad elektronra, majd ezt általánosítva az atommag elektromos terében mozgó elektronra is. Megszületett a Schrödinger-egyenlet, a kvantummechanika alapegyenlete.

Az elméleti fizikusok most ismét elemükben érezték magukat. Fél évszázada, a Maxwell-elmélet elfogadása óta mást sem csináltak, mint hullámegyenleteket oldottak meg furcsa mellékfeltételekkel (az egyik neves elméleti fizikus a századforduló táján kiszámította egy afrikai sós tó mellé telepített rádióantenna hullámainak terjedését, a sós víz elektromos vezetőképességének figyelembevételével). Számos jól kidolgozott matematikai módszer állt rendelkezésükre, és ezeket most be is vetették. Hamarosan már a különböző rendszerekre alkalmazott Schrödinger-egyenlet számos megoldását ismerték.

Emlékezzünk vissza a húr esetére: a hullámegyenlet megoldásához határfeltételek is kellenek. A hidrogénatom Schrödinger-egyenletének „természetes” határfeltétele az, hogy a hullámfüggvény (bármilyen is legyen annak fizikai jelentése) az atommagtól végtelen távol tartson nullához. Ezzel az egyszerű határfeltétellel és az elméleti fizikusok differenciálegyenlet-megoldó rutinjával hamarosan sikerült is meghatározni a hidrogénatom-problé-

ma lehetséges energiaszintjeit. És természetesen megkapták a kísérletekből ismert Balmer-formulát.

Hogy kerül ide az energia? Gondoljunk Planck képletére, amely a részecske energiáját és a hullám frekvenciáját kapcsolja össze: ahogy a húr sem rezeghet – az adott határfeltétel mellett – tetszőleges frekvenciával, úgy a Schrödinger-egyenletben paraméterként szereplő energia tetszőleges értékére sincs az egyenletnek a végtelenbeli határfeltételt kielégítő megoldása. Csak bizonyos energiák esetén – és ezek megegyeznek a Bohr-modell „megengedett” energiaszintjeivel.

Schrödinger egyenlete – a húr hullámegyenletéhez hasonlóan – lineáris differenciálegyenlet. Megoldásai tehát összeadhatók, szuperponálhatók. Ahogy a húr esetében a szuperponált megoldás több különböző frekvenciájú normálrezgés keveréke, úgy a Schrödinger-egyenlet általános megoldásához sem tartozik határozott energiaérték – a megoldás különböző energiájú állapotokat leíró hullámfüggvények szuperpozíciója. Mennyi hát az energia értéke ebben az állapotban? Határozatlan – ez egyik megnyilvánulása a kvantummechanika valószínűségi, statisztikai jellegének, aminek részletes elemzéséről hamarosan szó lesz, hiszen Neumann János e téren is alapvetően hozzájárult az elmélet fejlődéséhez.

Adósok vagyunk a válasszal egy igen fontos kérdésre. Ez a kérdés azonnal felmerült a kortársakban is, amint megismerték de Broglie hipotézisét, Schrödinger hullámegyenletét, és még inkább, amikor a hullámtermészet kísérleti bizonyítéka, az elektronhullámok diffrakciója köztudomásúvá vált: *mi is az, ami hullámszik?* Mit jelent a Schrödinger egyenlet megoldása, a nevezetes hullámfüggvény? Vagy ahogy az egyik kortárs fogalmazta: *mi is a „hullámszani” ige alanya?*

Ezen a kérdésem azóta is vitatkoznak a fizikusok. Később részletesen visszatérünk a problémára.

Szimmetriák

A Schrödinger-egyenlet könnyen általánosítható többelektronos atomok esetére. Akkor is differenciálegyenlet marad, csak kicsit bonyolultabb. A hullámegyenleteken edzett fizikusok mégis több reménnyel láttak neki eme egyenlet – legalábbis közelítő – megoldásának, mint a teljesen reménytelen mechanikai soktestprobléma Bohr-féle kvantálásának. Hamarosan megszülettek az első részeredmények.

Ahogy az elektronok száma növekszik, úgy nehezedik a Schrödinger-egyenlet megoldása is. Ugyanakkor felbukkan egy új jelenség: a *szimmetria*. Ha a sokelektronos atom két elektronját felcseréljük, a rendszer nem változik. Hasonló a helyzet az elektronok tetszőleges permutációja esetén is. Ezt a szimmetriát nyilvánvalóan figyelembe kell venni az atom elektronrendszerének leírása során. Kétféleképpen is: egyrészt meg kell követelni, hogy az egyenlet megoldása ne legyen érzékeny két elektron felcserélésére. Másrészt a szimmetria segítséget is jelent: kevesebb esetet kell végigbogarászunk a megoldás során, hiszen ha két eset csak az elektronok felcserélésében különbözik, akkor fizikailag ekvivalensek, felesleges mindegyikkel külön-külön foglalkozni.

Az atomok leírásakor egy másik fajta szimmetriának is fontos szerepe van. Ez pedig az atom saját középpontja körüli elforgatása iránti érzéketlensége, forgásszimmetriája. Az atommag elektrosztatikus tere, ami az elektronokat keringésre kényszeríti, nyilvánvalóan gömbszimmetrikus: a vonzóerő sugárirányú, nagysága csak a középponttól való távolságtól függ, az iránytól nem. Ha egy atom bonyolult elektronrendszerét a középponton átmenő valamelyik tengely körül tetszőleges szöggel elforgatjuk, a rendszer az eredetivel fizikailag ekvivalens marad, egyebek közt energiája sem változik meg. Ennek a ténynek tükröződnie kell a Schrödinger-egyenlet megoldása során is.

A helyzetet némileg bonyolította az az időközben felfedezett tény, hogy az elektronoknak spinjük, saját belső impulzuszórájuk (és ehhez kapcsolódó mágneses momentumuk) is van, valamint a Pauli-elv, mely szerint két elektron nem lehet azonos állapotban. Ezek a fizikai fogalmak lefordíthatók a Schrödinger-egyenlet szimmetriaviszonyainak nyelvére.

A sokelektronos atomokkal foglalkozó fizikusok átértékelték a szimmetriatulajdonságok fontosságát, és megpróbálták őket figyelembe venni a Schrödinger-egyenlet megoldása során. Ám különböző módszereik igen eltérő eredményekhez vezettek.

Ez az a pont, ahol Neumann János először jelenik meg a kvantummechanika történetében.

Neumann barátja, Wigner Jenő is egyike volt azoknak, akik a sokelektronos atomok energiaszintjeinek meghatározásával foglalkoztak. 1926-ban Neumann hívta fel Wigner figyelmét arra, hogy a szimmetriák leírására és szisztematikus tárgyalására létezik egy jól kidolgozott matematikai módszer: a *csoportelmélet*, ezen belül is a csoportok lineáris ábrázolásainak elmélete. Wigner ezt követően olyan alaposan elsajátította (majd később alkotó módon továbbfejlesztette) a csoportelméletet és kvantummechanikai alkalmazásait, hogy hamarosan a téma vezető szakértőjévé vált. A következő években publikált cikkeiben csoportelméleti alapon tisztázta az elektronrendszer szimmetriasajátosságainak hatását a Schrödinger-egyenlet megoldásaira. Később, 1932-ben megírta alapvető jelentőségű, ma is sokat idézett tankönyvét (*Csoportelméleti módszer a kvantummechanikában*), amely az eltelt évtizedek alatt sem avult el, élményszerű és kedvet adó, ugyanakkor mélyre hatoló bevezetés a témakörbe. Későbbi kutatásai során az atommagok, majd az elemi részecskék fizikájára alkalmazta a csoportelméletet, és számos ismeretterjesztő jellegű esszében is fel-

hívta a figyelmet a szimmetriáknak a fizikában és általában a természettudományban játszott fontos szerepére (esszéi magyarul is megjelentek *Szimmetriák és reflexiók* címmel).

1928-ban jelent meg a kor vezető fizikai folyóiratában, a Zeitschrift für Physik-ben Neumann és Wigner háromrészes cikksorozata, amelyben a csoportelmélet alapján szisztematikus tárgyalták az atomi spektrumok elméletét, az elektron spinjének és az elektronrendszerek szimmetriatulajdonságainak figyelembevételével. Ez a cikksorozat a következő évek hasonló jellegű ku-

Im ganzen haben wir also folgende Terme:

$\mu =$	-3	-2	-1	0	1	2	3
Antisymmetrisch	0	1	2	3	2	1	0
Entartet	1	3	6	7	6	3	1
Symmetrisch	1	2	4	4	4	2	1

Das sind 81 linear unabhängige Eigenfunktionen, wie man auch direkt abzählen kann. Wenn wir die durch die Drehgruppe geforderten Zusammenfassungen vornehmen, so erhalten wir:

$l =$	3 (7fache Terme)	2 (5fache Terme)	1 (3fache Terme)	0 (einfache Terme)
Antisymmetrisch	0	1	1	1
Entartet	1	2	3	1
Symmetrisch	1	1	2	0

15. ábra. Részlet Neumann és Wigner cikkéből, amelyben az atomok elektronrendszerét és spektrumát tárgyalják a csoportelmélet alapján. A táblázatban szereplő számok a szinképponalak finomszerkezetét írják le a tapasztalattal (lásd 9. ábra) jó egyezésben.

tatásaiban példaként, vezérfonalként szolgált, számtalanszor hivatkoztak rá. Az atomok és molekulák elektronrendszerének leírása során ma is a Neumann és Wigner által kijelölt úton halad a kutatás.

A kvantummechanika kialakulása

Időközben (pontosabban a de Broglie–Schrödinger-féle fejleményekkel párhuzamosan) a kvantummechanika alapjainak kiépítése során egy nem várt, absztraktabb jellegű fordulat is bekövetkezett. Egy fiatal német fizikus, Werner Heisenberg meglepő és újszerű matematikai sémával állt elő, amivel ugyancsak sikerült reprodukálnia a Bohr-modell addigi eredményeit, és további általánosításra is lehetőséget adott. Ez volt a *mátrixmechanika* kezdete. Míg Schrödinger egyenlete mintegy megnyugvással töltötte el az idősebb fizikusnemzedéket, hiszen az új és érthetetlen fizikát jól bevált, általuk ismert matematikai eszközökkel, differenciálegyenletekkel próbálta leírni, addig Heisenberg próbálkozása inkább a fiatalabb, a hagyományos matematikai ágakban kevésbé képzett, és (tán ezért is) újításokra mindig kész fizikusok tetszését nyerte el. Ehhez hozzájárult az elmélet matematikai kifejtéséhez kapcsolódó, lázadó jellegű „filozófiai” interpretáció is.

Heisenberg ugyanis – mintegy Einstein híres tettét megismétlendő, amellyel száműzte a fizikából a mindent betöltő, ám fizikailag észlelhetetlen éter hagyományos, sok nehézséget okozó fogalmát – most a Bohr-pályák és egyéb kimutathatatlan „boszorkányok” végleges kiseprűzésére szólította fel kollégáit. Pozitívista filozófiai hozzáállása szerint e nem észlelhető objektumok nem szerepelhetnek az alapvető elméletben – csakis olyan mennyiségek, amelyeket ténylegesen mérni lehet, például a különböző energiaszintek közti átmenetek során kibocsátott sugárzás intenzitása. Ezeket a valóban mérhető adatokat táblázatokba (később úgy mondták: *mátrixokba*) rendezte, majd matematikai összefüggéseket keresett – és talált – köztük. Ezek alapján pedig újabb, kísérletileg is meghatározó adatokat tudott (sikeresen) kiszámolni. Talán már ennyiből is látszik az eljárás absztrakt volta – szemben a de Broglie–Schrödinger-féle hullámok megszokott, otthonos, szemléletes mivoltával (eltekintve persze attól, hogy senki sem tudta: ugyan mi is hullámmik).

A mátrixok elmélete sem volt teljesen új – a közgazdászok és a mérnökök már évtizedek óta használták. A fizikusoknak azonban nem tanították, hiszen a mátrixok – egy-két perifériális területtől eltekintve – csak elvétve szerepeltek a fizikában. Amikor Heisenberg kidolgozta elméletét, egyik professzora figyelmeztette, hogy az általa bevezetett táblázatok és a rájuk értelmezett műveleteket más tudományágakban már régóta ismerik, és mátrixnak nevezik. Persze van egy kis különbség, hiszen a mérnöki és közgazdaságtani gyakorlatban csak véges méretű számtáblázatok szerepeltek, míg Heisenberg mátrixai végtelen nagyok voltak.

Természetesen felmerül a kérdés, mennyiben jogosult a véges mátrixokra értelmezett műveleteket (összeadás, mátrixszorzás stb.) végtelen sok elemet tartalmazó számtáblázatokra kiterjeszteni. Végtelen sok szám összeadása értelmezhető, de nem mindig vezet egyértelmű eredményre, sőt olykor végtelen értéket ad. Az új matematikai eszközzel ismerkedő fizikusokat ez nem nagyon izgatta, Neumann és néhány más matematikus viszont elkezdett foglalkozni a kérdéssel: Vajon milyen feltételeket kielégítő végtelen mátrixokkal lehet

a véges mátrixokkal analóg műveleteket végezni, és az ezekre vonatkozó szabályok, tételek közül melyek maradnak érvényben, melyek módosulnak? A kérdést néhány év múlva Dirac transzformációelmélete, majd Neumann operátorokra vonatkozó általános tételei tisztázták megnyugtatóan.

Heisenberg és ifjú kollégái a következő hónapokban és években megvizsgálták a kvantumelmélet konkrét feladatait, és alkalmazták rájuk a mátrixmechanika módszereit. Ez – a módszer újdonsága, szokatlan volta miatt – olykor csak nagyon körülményesen sikerült. Pauli dolgozata, amelyben a hidrogénatom energiaszintjének problémáját tárgyalta a mátrixmechanika alapján, sokkal terjedelmesebb, bonyolultabb, és nehezebben követhető, mint a Schrödinger-egyenlet alapján történő analóg levezetés. De a két módszer végeredménye megegyezett!

Vajon véletlen ez az egybeesés? Két elkeseredett ellenfél versengése gyakran végződik furcsa kiegyezésben, összeborulásban. Schrödinger nemsokára publikált egy cikket „*A Heisenberg–Jordan-féle kvantummechanika viszonya az enyémhez*” címmel, amelyben bebizonyította, hogy a két elmélet matematikailag ekvivalens. Más szóval ha egy problémát mindkét elméletben meg lehet oldani, az eredmény azonos lesz. Az elméletek egyes állításai is „lefordíthatók”, matematikailag áttranszformálhatók egymásra.

Ettől kezdve nincs értelme a szembeállításnak. Illetve... Bár a fizikai eredmények azonosak, az elérésükhöz vezető út nehézsége, illetve szépsége a két elméletben különbözhet. Mivel a fizikusok többsége a „régimatematikát”, a differenciálegyenleteket tanulta egyetemi éveiben, érthető, hogy a következő években, amikor a kvantummechanika megkezdte behatolását a fizika különböző szaktudományába (pl. kialakult a molekulafizika, a kémiai kötés elmélete, a szilárd testek és a szuperfolyékony hélium kvantumelmélete stb.), a legtöbben a Schrödinger-féle hullámmechanikát alkalmazták, illetve abból kiindulva dolgozták ki a különböző számítási és közelítő módszereket. A mátrixmechanika viszont verhetetlen volt azokon a speciális területeken, ahol az eredeti Heisenberg-féle végtelen táblázatok helyett véges mátrixokkal lehetett dolgozni. Ilyen volt az elektronspin nemrelativisztikus elmélete, majd néhány év múlva a Paul Dirac által kidolgozott relativisztikus kvantumelmélet (ehhez a tudományterülethez Neumann János is hozzájárult egy cikkel), még később az atommag újonnan kialakuló fizikája.

Dirac és a transzformációelmélet

A kvantummechanika e két, matematikailag ekvivalens alakja úgy tekinthető, mintha egy fizikai jelenséget egyik vagy másik koordináta-rendszerre vonatkoztatva íránk le. A számadatok, a konkrét részeredmények mások, de a háttérben a mennyiségek közti összefüggéseknek, alapvető kapcsolatoknak ugyanaz a rendszere húzódik meg.

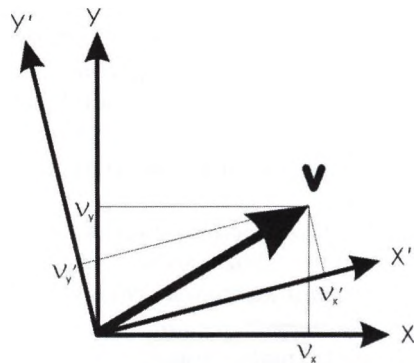
Ez a szituáció emlékeztet például a *vektor* matematikai fogalmának bevezetésére. E fogalom az egyszerűségéhez képest – ma már az általános iskolában is tanítják – meglepően későn, a XIX. század végén született meg. Egy vektort egy adott koordináta-rendszerben három számmal adhatunk meg (N dimenziós térben persze N számmal), és a vektoron értelmezett műveleteket (összeadás, számmal való szorzás, skaláris és vektoriális szorzás stb.) e koordináták vagy komponensek segítségével fejezhetjük ki. Egy másik koordináta-

rendszerre átvérve általában minden szereplő vektor minden komponense megváltozik, de a köztük fennálló összefüggések – mint pl. az, hogy két vektor merőleges egymásra, vagy hogy egy vektor két másik vektor összege –, mindkét rendszerben egyaránt fennállnak. Ezért a két koordináta-rendszerbeli két számhármast nem két önálló matematikai objektumnak tekintjük, hanem ugyanannak az absztrakt matematikai entitásnak (magának „ a ” vektornak) két különböző reprezentációjának. E két számhármásban ugyanaz az információ rejtőzik, ezért kölcsönösen egyértelműen át is számíthatók egymásra. Ezzel foglalkozik a lineáris algebra transzformációelmélet nevű ága.

Érdeemes bepillantani egy száz évnél régebbi, kb. a XIX. század derekán írt fizikai, pl. hidrodinamikai tankönyvbe. A mai olvasó meglepve veszi észre, hogy akkor még nem használtak vektorokat, azokat a mennyiségeket, amiket ma egy vektor komponenseinek nevezünk, önálló betűvel jelölték. Ennek következtében az e könyvekben szereplő egyenletek rengeteg különböző betűjelet tartalmaztak, és ijesztően terjedelmesek voltak, pl. a ma egy sorban elférő Navier–Stokes-egyenlet részletesen kiírva egy egész könyvdalt betöltött. Az absztrakt vektorfogalom (és a vele kapcsolatos műveletek) bevezetése tehát nemcsak „magasabb szintre” emelte a matematikai fogalmat, hanem jelentősen egyszerűsítette az egyenletek kezelését, áttekinthetőségét, megjegyezhetőségét, segítette a fizikai lényeg kihamozását a matematikai szimbólumok dzsungeléből.

Hasonló program véghezvitelére vállalkozott Paul Dirac angol fizikus a kvantummechanika két alakjának megszületését, majd ekvivalenciájuk bizonyítását követően. Ha két, első látásra ennyire eltérő elméletnek, mint a mátrixmechanika és a hullámmechanika, azonos a matematikai és fizikai tartalma, és azonos következtetések vonhatók le belőlük, akkor mindkettő úgy tekinthető, mint egy absztraktabb, általánosabb „kvantummechanika” két reprezentációja. Ez a szemlélet segít lehámozni az egyes elméletek esetlegességeit, sallangjait, és megkeresni a közös mögöttes tartalmat. (Senki sem adna egy lyukas garast sem egy olyan fizikai elméletért, amelyben egy sebességvektor komponensei csakis prímszámok lehetnek – és a szerző később ebből vonna le messzemenő következtetéseket. Miért nem fogadnánk el ezt az „elméletet”? Mert tudjuk, hogy egy megengedett transzformáció, a koordináta-rendszer elforgatása során a vektorok komponensei úgy változnak meg, hogy az új rendszerben nem lesznek – nemhogy prímszámok, hanem még egész számok sem. Egy ilyen „elmélet” csak egyetlen koordináta-rendszerben lehetne helytálló, nem lenne általánosan igaz, eltranszformálható, szakszóval *kovariáns*.)

Dirac elegáns formalizmusa megtalálta a kvantuma két verziójának közös magját, és konkrét recepteket adott arra, hogy lehet az absztrakt formalizmusról egyik vagy másik korábbi verzióra áttérni. Az egyszerűbb kérdések, mint pl. a harmonikus oszcillátor kvantumelmélete közvetlenül az absztrakt formalizmusban is tárgyalhatók. A kvantummechanikával ismerkedő egyetemi hallgatókat mindig elbűvöli, amikor az első órák egyikén, szinte



16. ábra. Vektor és komponensei két különböző koordináta-rendszerben

közvetlenül a Dirac-formalizmus bevezetése után az előadó „szinte a semmiből” levezeti a harmonikus oszcillátor lehetséges energiaszintjeit. E formalizmus szintjén nincsenek mátrixok, nincsenek hullámfüggvények és differenciálegyenletek, csak a kvantummechanika lepárolt lényege működik.

A bonyolultabb feladatok esetén természetesen már általában nem elég az absztrakt formalizmus, „le kell szállni a fellegekből”, és a Schrödinger- vagy Heisenberg-féle reprezentációk valamelyikében szokás tovább számolni. Elsősorban azért, mert e formalizmusokban nagy számú közelítő módszert dolgoztak ki az egzaktul nem megoldható problémák kezelésére. (Hasonlít ez ahhoz, amikor egy bonyolult vektoregyenlet megoldása során kénytelenek vagyunk koordináta-rendszert választani, az egyenletet a komponensekre mint számokra vonatkozó egyenletekre átírni, és az így kapott egyenletrendszert számítógéppel megoldani.) Mindazonáltal egy-egy nagyobb lélegzetű problémakör részletes tárgyalása után szokás az általános jellegű eredményeket visszaírni a Dirac-féle absztrakt és így könnyebben megjegyezhető formába (pl. a szórásjelenségek Lippmann–Schwinger-féle elméletét).

Ez az a Dirac-féle formalizmus, amelyről Neumann a bevezetőben is idézett módon igen elismerően nyilatkozik könyvében: „*Dirac mind nemrég kiadott könyvében, mind pedig számos cikkében a kvantummechanikát felülmúlhatatlanul elegánsan, tömören és ugyanakkor invariáns formában fogalmazta meg.*” – hogy aztán néhány oldal múlva megmutassa: a Dirac-elmélet matematikailag nem kielégítő, egyszerűbben szólva helytelen. És amely elmélet helyett Neumann – több más szerzővel, köztük a matematika akkori fejedelmével, Hilberttel együttműködve – több cikkben kidolgozta, majd végül 1932-ben megjelent nagy könyvében (*A kvantummechanika matematikai alapjai*) részletesen is közreadta a kvantummechanika alternatív, ám matematikailag abszolút precíz elméletét.

De tulajdonképpen mi kifogásolni valót talált Neumann a kvantummechanika Dirac-féle felépítésében? Ahhoz, hogy erre választ tudjunk adni, meg kell ismerkednünk a kvantumelmélet interpretációs problémáival is.

Mi hullámzik?

A de Broglie-féle hullámhipotézis és a Schrödinger-egyenlet megszületése után mindenki arra gondolt, hogy a hullámzó valami egyszerűen az elektron anyaga: az elektron nem pontszerű objektum, nem is golyó, hanem a térben szétkent, hullámzó felhőcske. Ezt a nézetet alátámasztotta, hogy a Schrödinger-egyenletből viszonylag könnyen sikerült az elektron töltéssűrűségére és áramsűrűségére vonatkozó, a teljes töltés megmaradását kifejező ún. kontinuitási egyenlet levezetése.

Ennek a nézetnek a hátulütői azonban hamarosan megmutatkoztak. Ha az elektron el- kent felhő, akkor egyes részei – lévén mind negatív elektromos töltésük – nyilvánvalóan taszítják egymást. Mi tartja mégis össze? Hát persze: a pozitív töltésű atommag vonzóereje. És mi a helyzet a számítógép monitorának katódsugárcsővében most éppen velem szemben repülő, atommagoktól való szabad elektronnal – azt miért nem szórja szét saját részeinek kölcsönös taszítása?

Természetesen feltételezhetünk valami furcsa belső szerkezetet az elektron belsejében, rugókat vagy más, nem elektromos jellegű összetartó mezőket – ezt persze igen nehéz elképzelni, hiszen a Schrödinger-egyenlet megoldása szerint az elektronfelhő az egész atomot betölti –, ám van egy súlyosabb ellenérv is.

A szétkent elektronfelhő egyes részeinek tasztításából származó helyzeti energiát be kellene számítani a rendszer egyéb energiái közé, és be kellene írni a Schrödinger-egyenletbe. Ezt el is végezték, csakhogy az így módosított egyenlet már nem írta le helyesen, a kísérletekkel egyezően a hidrogénatom energiaszintjeit!

A helyzet ismét paradox: a pontszerű proton és pontszerű elektron feltételezésével felírt egyenlet helyesen adja meg a rendszer energiaszintjeit, de elkent elektronfelhőt ír le – az elkent elektronfelhő elektromos terének figyelembevétele viszont elrontja a helyes energiaértékeket...

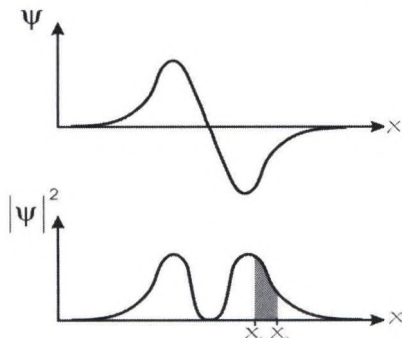
Még súlyosabb a következő ellenérv: ha az elektron elkent felhő, akkor miért észlelünk a kísérletek során mindig csak egész elektront? Soha senki sem mért még fél- vagy negyedelektronnyi tömeget vagy elektromos töltést! Az elektron fura jószág: amikor mozog, akkor hullám, elkent paca, amikor becsapódik, akkor határozott tömegű és töltésű, az észlelő ernyő egyetlen pontjára koncentrálódott részecske.

Hasonló a helyzet a fotonnal, a Planck feltételezte fénykvantummal is: a terjedés során a Maxwell-egyenletek leírta hullámtulajdonságokat mutatja, keletkezése és elnyelődése során viszont határozott energiájú és helyzetű, pontszerű részecskeként viselkedik.

Bohr ismerte fel, hogy ezek az ellentétesnek látszó (*komplementer*) tulajdonságok egyszerre jellemzik a valódi fizikai részecskéket. Bizonyos körülmények között az egyik, más körülmények között pedig a másik tulajdonság dominál, és csak az egyik képből kiindulva nem lehet helyesen leírni az objektum viselkedését. Heisenberg pedig levezette ennek az állításnak a matematikai megfelelőjét, a nevezetes *határozatlansági relációt*. Eszerint az objektum (pl. elektron) bizonyos alapvető fizikai jellemzői komplementer párokba rendezhetők: ilyen például a helyzet és az impulzus. Két mennyiség értékét nem lehet egyszerre pontosan megmérni. Ez nem a mérőeszközünk hibája: a két mennyiségnek egyszerűen *nincs* egyszerre határozott értéke. Az elektronnak pl. nincs egyszerre határozott helykoordinátája és impulzusa. Ha az egyik mennyiséget kísérjük megmérni (pl. egy detektorral vagy fénysugárral meg akarjuk határozni az elektron helyzetét), akkor a mérőeszközzel való elkerülhetetlen kölcsönhatás, szóródás következtében nem tudhatjuk meg pontosan az elektron impulzusát. Ha viszont mérésünkkel az impulzusra kérdezzük rá (azaz de Broglie képletének megfelelően a hullámhosszat akarjuk meghatározni, pl. interferenciakísérletekkel), akkor a helykoordináta mosódik el – hiszen hol is van *pontosan* egy hullám? A hely és az impulzus bizonytalanságának szorzata nem lehet kisebb a Planck-állandónál. Ez a tétel következik a kvantummechanika alapegyenleteiből, de a levezetésénél a mérés fogalmára vonatkozó számos (kimondott vagy ki nem mondott) feltevéssel éltek – ezért a határozatlansági relációt évtizedeken át értelmezték, vitatták és magyarázták a fizikusok és az ismeretelmélettel foglalkozó filozófusok.

Bohr és iskolája dolgozta ki a kvantummechanika ún. koppenhágai értelmezését is. Ez választ ad arra a kérdésre, mi hullámszerű, mit is jelent a Schrödinger-egyenlet megoldásaként adódó hullámfüggvény. A válasz az, hogy *valószínűséget*. Pontosabban szólva a hullámfüggvény abszolút értéke négyzetének és a tartomány térfogatának szorzata adja meg annak valószínűségét, hogy az elektront a vizsgált kis tartományban találjuk meg. (A hul-

lámfüggvény értéke nem valós szám, hanem komplex, ezért kell ilyen bonyolultan fogalmazni, ezért nem emlegethetjük egyszerűen a hullámfüggvény négyzetét.)



17. ábra. A hullámfüggvény és az elektron megtalálásának valószínűsége. A felső ábra a de Broglie–Schrödinger-féle hullámfüggvény egy lehetséges alakját ábrázolja a helykoordináta függvényében, az alsó ábra a hullámfüggvény abszolút értékének négyzetét. Annak a valószínűsége, hogy a részecskét az x_1 – x_2 intervallumban találjuk, a besatírozott területtel arányos. A második görbe alatti teljes terület (azaz annak a valószínűsége, hogy a részecskét valahol megtaláljuk) éppen 1 egységnyi.

képp az energia, az impulzusmomentum, és más fizikai mennyiségek mérésére is. A kvantumelmélet matematikai apparátusa – fejlődésének ebben a fázisban legtisztábban a Dirac-féle formalizmus alapján – pontos képleteket ad e mennyiségek várható értékére, feltéve, ha a rendszer egy adott hullámfüggvénnyel leírható állapotban van.

Emellett az elmélet módszert szolgáltat annak meghatározására is, hogy egyáltalán milyen számérték lehet e mennyiségek mérésének eredménye. Sok fizikai mennyiség értéke nem vehet fel tetszőleges értéket, hanem csak bizonyos diszkrét, kvantált értékek valamelyikét. Ilyen mennyiség – kötött állapotok esetén – a rendszer energiája is. Így kapjuk meg a kvantumelméletből pl. a hidrogénatom energiaszintjeit. Más fizikai mennyiségek – pl. a helykoordináta és az impulzus – értéke egy folytonos sokaság eleme lehet, azaz tipikusan akármelyik valós szám. (Nem igaz tehát az a sokszor hallott leegyszerűsítő állítás, mely szerint „a kvantumelméletben minden mennyiség értéke kvantált”. Mennyisége és rendszere válogatja.)

Neumann-nak épp a folytonos értékészletű mennyiségek Dirac-féle leírásával szemben voltak matematikai kifogásai.

A Dirac-delta

A kvantummechanika szerint a rendszer állapotából kiszámítható minden fizikai mennyiség mérésének valószínűség-eloszlása. Vannak-e olyan állapotok, amikor ez az eloszlás egyetlen értékre koncentrálódik, azaz pl. egy adott energiaérték mérésének valószí-

A valószínűségi értelmezés megoldja az elkent felhő – pontszerű részecske ellentmondást. Az elektron sohasem elkent felhő, töltése nem áll össze egymástól térben elkülönült, egymást taszító darabokból, mindig pontszerű – csak éppen nem tudni, hol van. Nem csak mi nem tudjuk, maga a természet sem!

Adott állapotú elektron tehát itt is lehet, meg ott is. Ugyanabban az állapotban sok különböző helyen megtalálhatjuk, igaz, különböző valószínűséggel. De hol is van *valójában*? A koppenhágai interpretáció azt feleli, hogy a kérdésnek nincs értelme. Akik mégis értelmet akartak neki adni, azok a *rejtett paraméterek* feltételezéséhez fordultak. Erről szól cikkünk utolsó fejezete.

A hely és az impulzus fizikai mennyiség, amelyek mérésére határozott mérési utasításunk, mérőberendezésünk van. Hasonló-

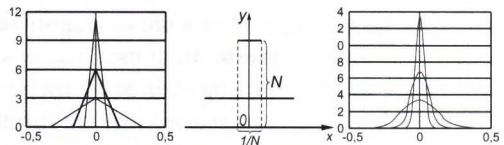
núsége 100%, az összes többié pedig nulla? Igen, vannak – ezeket hívják az adott mennyiség *sajátállapotainak*. A klasszikus mechanikában csak ilyen állapotok vannak – a rendszer bármilyen állapotában bármelyik fizikai mennyiség teljesen határozott értéket vesz fel. A kvantummechanika matematikájának egy nevezetes eredménye (ennek másfajta kifejezési módja a már említett határozatlansági reláció is), hogy nincs olyan állapot, amelyben minden fizikai mennyiségnek határozott értéke lenne (ennek legfrappánsabb bizonyítása éppen Neumann nagy könyvében található meg). Azaz ha az egyik fizikai mennyiség értéke határozott, akkor számos más mennyiségé nem. Például előfordulhat olyan állapot, amelyben a rendszer impulzusmomentumának értéke pontosan a Planck-állandó kétszerese (ez az egyik lehetséges értéke), de sok mérést végezve kiderül, hogy a rendszer energiája 5% valószínűséggel az alapállapot, 53% valószínűséggel az első, 42% valószínűséggel pedig a harmadik gerjesztett állapot energiájának felel meg. (A gyakorlatban előforduló esetek ennél sokkal bonyolultabbak.) Az energia sajátállapotai bizonyos szempontból kitüntetett szerepet játszanak: ha a rendszer egy időpillanatban energia-sajátállapotban van, akkor továbbra is abban marad (egészen addig, amíg egy másik rendszerrel való kölcsönhatás ki nem billenti onnan). Ezek az energiaértékek tehát tartósan fennmaradnak – valószínű, hogy egy (a rendszer saját időskálájához képest hosszan tartó) makroszkopikus mérés tartósan ugyanabban az energia-sajátállapotban találja a rendszert. Ezek között is van egy még inkább kitüntetett állapot: a legkisebb energiájú, az *alapállapot*. Külső gerjesztés nélkül a rendszer nagy valószínűséggel ebben található meg. Más fizikai mennyiségek értéke természetesen az alapállapotban is szórást mutathat.

Van-e *helysajátállapot*? Azaz olyan állapot, amikor a részecskén helymérést végezve, mindig ugyanazt a számértéket kapjuk? Emlékezzünk vissza arra, hogy a hely spektruma (lehetséges értékeinek halmaza) folytonos, ezért egy adott pontban találás valószínűsége mindig nulla. (Gondoljunk a céltábla középet eltaláló nyíllövés valószínűségére és a nullmértékű halmazokra!)

Milyen lehet(ne) a helysajátállapot hullámfüggvénye? A hullámfüggvény négyzete arányos a valószínűséggel, ezért ha a függvény egy adott szakaszon kívül nulla, akkor ott biztosan nem találjuk meg a részecskét. A szakaszon belül a függvény értékét úgy kell megválasztani, hogy a megtalálás teljes valószínűsége (azaz a függvény négyzetének görbéje alatti terület) pontosan 100%, azaz 1 legyen.

Próbáljuk meg „koncentrálni” a hullámfüggvényt a kiválasztott pontra! Húzzuk össze egyre rövidebb szakaszokra – és ennek megfelelően növeljük a magasságát! „Határesetben” olyan függvényt kapunk, amely az egyenes mentén mindenütt nulla, kivéve egyetlen pontot, itt értéke végtelen, és az alatta levő terület pontosan 1.

– Ez a *delta-függvény*, a hely sajátfüggvénye! – mondja Dirac, a pragmatikus fizikus.
– Ilyen függvény pedig nincs! – mondja Neumann, a matematikus, avagy a matematikai szigorúságra is érzékeny elméleti fizikus.



18. a b c ábra. A Dirac-féle delta-függvény közelítő függvényei:

- a) háromszögimpulzusok,
- b) négyzetgimpulzusok,
- c) Gauss-görbék.

Minél keskenyebbek a közelítő görbék, tulajdonságaik annál pontosabban közelítik a Dirac-féle delta-függvény megkívánt jellemzőit. Sajnos az elképzelt határfüggvény, a „végtelenül keskeny” impulzus nem tartozik a matematikailag szigorúan definiálható függvények közé.

A Dirac-féle delta-függvény nemcsak az itt leírt módon építhető fel (közelíthető például Gauss-görbékkel is), de a konstrukció mindig tartalmaz egy olyan határátmenetet, ami nem állja ki a szigorúbb matematikai vizsgálódás próbáját.

Dirac elegáns kvantumformalizmusa szükségképpen tartalmaz delta-függvényeket, hiszen a folytonos spektrumú fizikai mennyiségek (köztük a legfontosabbak, a hely és az impulzus) sajátállapotainak leírására mindenképpen ezeket kell használnia. Enélkül (pl. a diszkrét és a folytonos spektrumú mennyiségek külön-külön való kezelésével) csorbulna az elmélet általánossága, eleganciája, és nem utolsósorban használhatósága.

Megjegyezzük, hogy a Dirac-féle delta-függvény nem csak a kvantummechanikában bukkan fel. Bevezetése után sok más fizikai elméletben alkalmazták, korábban bonyolult módon megfogalmazható tételeket vagy definíciókat lehetett vele jelentősen egyszerűbb alakra átírni (pl. az elektrosztatika potenciálméletének Green-tételét). Mit jelenthet a Dirac-delta alakú töltéseloszlás az elektrosztatikában? Egyszerűen régi ismerősünket, a ponttöltést, amelynek elektromos töltése egyetlen pontba koncentrálódik, azon kívül mindenütt nulla, de az össztöltés mégis véges érték.

Neumann könyve és programja

Neumann-nak természetesen igaza volt: hasznossága ellenére a Dirac-féle delta-függvény nem korrekt matematikai objektum.

Miért baj az, ha egy fizikai elméletben matematikailag kétes, megalapozatlan részletek maradnak? A fizikusok körében elterjedt meggyőződés szerint ez nem baj – az ilyen esetekben segít a híres fizikus szimat, a „gömbérezék”, és megmutatja, mikor szabad használni a kétes tisztaságú matematikai eszközt, mikor nem. A tudománytörténet ugyanakkor számos példával igazolja, hogy bár az új, még nem kellően megalapozott matematikai eszközt bevezető fizikus nagymesterek (pl. az infinitézimális kalkulus kezdeményező Newton, a harmonikus analízist megalapító Fourier, Heaviside, az operátorszámítás atyja, a funkcionálintegrálokat bevezető Feynman vagy épp maga Dirac) birtokában voltak e sokat emlegetett gömbérezéknek, és új eszközeikkel később is helyesnek bizonyult eredményekhez jutottak – lelkes követőikről ez korántsem mondható el. Gyakran előfordult, hogy a mesterek tanítványai lejáratták a mestertől kapott eszközt, abszurd, sőt nevetséges eredményeket vezettek le. Ez a helyzet addig tartott, amíg (olykor csak néhány tudósgenerációval később) precíz matematikusok egy csoportja meg nem vizsgálta alaposan a fizikus intuíción alapján bevezetett matematikai fogalmakat, és meg nem határozta érvényességi határokat, használhatóságuk tartományát. Rendre kiderült, hogy míg az atyamester ösztönösen bent maradt ebben a tartományban, az óvatlan tanítványok általában kiléptek belőle. A matematikai megalapozás után viszont már mindenki használhatja az új eszközt, hiszen tudhatja, meddig mehet el vele. A legjobb persze az, ha a fizikai ihletettséggű új matematikai eszközöket már megszületésük pillanatában tesztelik a matematikusok, és kijelölik a határokat. (Ennél már csak az lenne jobb, ha a fizikusok matematikai igényessége emelkedne arra a szintre, hogy a matematikai számítások során csak matematikailag megengedett eszközöket használnának – ha angolul beszélünk a fizikáról, használjuk a korrekt an-

gol nyelvet, ha a matematika nyelvén, akkor használjuk a korrekt matematikát – de erre sajnos belátható időn belül még nem számíthatunk.)

Neumann ismerte ezt a helyzetet, hiszen gyakran publikált fizikai folyóiratokban, és folyamatosan követte a fizikus szakirodalmat, tapasztalta a híres fizikus pongyolaságot. Ezért is féltette a kor nagy eredményét, sok zseniális tudós közös szellemi gyermekét, a kvantummechanikát attól, hogy alkotói – matematikai pongyolaságokba tévedve – homokra alapozzák az új fizika szilárdnak szánt épületét. Ahogy egy pohár szennyvíz is tönkre tehet egy egész hordónyi nemes bort – elég egyetlen megalapozatlan matematikai állítás, és a nemes konstrukció porrá omlik össze. (Jó példa erre az utóbbi évekből a „sztochasztikus elektrodinamika” nevű, jobb sorsra érdemes elmélet szomorú sorsa.) A logika szerint ugyanis egy hamis állításból bármit le lehet vezetni... Ezért érezte kötelességének, hogy elválassa az ocsút a tiszta búzától, a kvantummechanikát kidolgozó fizikusok magasaróptú gondolatait szilárd alapra helyezze.

Neumann könyve előszavában ugyan felveti annak lehetőségét, hogy a matematikai módszerek megfelelő általánosításával kidolgozzák a Dirac-delta, azaz a helysajátfüggvények precíz elméletét (hiszen Newton idejében az általa kidolgozott „kalkulus” is éppen olyan pongyola, matematikailag megalapozatlan konstrukció volt, mint Neumann korában a delta-függvény – hogy később, sok fizikus és matematikus működése nyomán a mai kiterjedt és jól megalapozott matematikai analízissé fejlődjön), de aztán elveti ezt a lehetőséget, és könyve főszövegében megmutatja, hogy a meglévő matematikai eszközökkel is pontosan meg lehet fogalmazni a folytonos spektrumú fizikai mennyiségek kvantumelméletét. Ehhez persze az egész kvantummechanikát a kezdetektől fogva szigorú és pontos matematikai alapokra kell helyeznie. Könyvében (és az azt előkészítő, társszerzőkkel írt cikksorozatban) ezt megteszi.

A tárgyalás alapja a végtelen dimenziós tereken ható operátorok elmélete. Ezt az elméletet néhány évvel korábban Neumann és néhány matematikus kollégája dolgozta ki, illetve általánosította a korábbi eredményeket a végtelen dimenziós esetre. Az elmélet véges dimenziós változata már régóta ismert volt, és viszonylag egyszerű. A végtelen dimenziós változat számos matematikai buktatót rejteget, ide tartozik a folytonos spektrumú fizikai mennyiségek esete is. A kulcslépés a végtelen dimenziós tér bizonyos részalmazaira, altereire vetítő operátorok (projektorok) precíz elmélete. Ugyanezek az operátorok és a rájuk vonatkozó tételek bukkantak fel az ergodikus hipotézis bizonyítása során is.

Neumann az egyik ide vonatkozó cikkben megjegyzi, mennyire örül annak, hogy kora matematikájának a kívülállók – köztük a fizikusok – számára absztraktnak, túl precíznek, feleslegesen szőrszálhasogatónak tűnő fogalomalkotásai (mint a Lebesgue-féle mérték vagy a Stieltjes-integrál fogalma) munkássága nyomán ennyire fontos és magától értetődő szerepet kaptak néhány kulcsfontosságú fizikai elmélet precíz megalapozásában.

A kvantummechanikának ez a Neumann által adott matematikai alapozása később további érdekes fejlemények kiindulópontja lett. Neumann módszere ugyanis rámutatott, hogy a kvantumelmélet lényegi mondanivalóját, „matematikai magját” nem a Schrödinger-féle hullámfüggvény, nem Heisenberg végtelen mátrixai, de még csak nem is Dirac absztrakt állapotvektorai hordozzák. Hanem mi? Mint a későbbi részletes kifejtés során kiderült, a végtelen dimenziós tér alterei (amelyekre Neumann projektor-operátorai vetítenek), és a köztük fennálló kapcsolatok, relációk (tartalmazás, metszet, lineáris burok). Ezek az alterek megfeleltethetők egy-egy (absztrakt módon elképzelt) fizikai mérés külön-

bőző eredményeinek, és segítségükkel a kvantumelmélet minden méréssel ellenőrizhető eredménye rekonstruálható. Az alterek közti matematikai relációk pedig analógiába állíthatók a klasszikus valószínűség-számítás eseményalgebrájával, amelyben pl. a kockadobás elemi eseményeiből új, összetett eseményeket képezünk (páros szám dobása, két dobásból legalább 9 pont dobása stb.), majd az elemi események valószínűségének ismeretében ki tudjuk számítani az összetett események valószínűségét. További analóg matematikai rendszer a formális logika: itt egyes elemi állítások kapcsolhatók össze összetett állításokká (A és B, A vagy B, nem A stb.). Az állításokat összekapcsoló műveletek meghatározott szabályoknak tesznek eleget (pl. az „A és B” állítás azonos a „B és A” állítással).

Az analógia alapján a kvantumelmélet alterekre és ezek relációira alapozott felépítését *kvantumlogikának* nevezik. Ennek szabályai nem a klasszikus eseményalgebrából vagy formális logikából, hanem a végtelen dimenziós tér Neumann által tanulmányozott matematikai törvényszerűségeiből származtathatók le. Hamarosan kiderült, hogy a kvantumlogika eseményeket összekapcsoló szabályai eltérnek a „köznapi józan ész” formalizálásából született klasszikus logikáétól. Ezért a rá alapozódó valószínűségi-statisztikai elmélet is a hagyományostól eltérő eredményeket szolgáltat. Mint később részletesen látni fogjuk, ez az eltérés áll a kvantumelmélet valószínűségi kijelentései és a köznapi szemlélet közti végletes ellentmondások háttérében. A kvantumlogika, amely Neumann vizsgálatai nyomán született, majd önállósult, ma is fejlődő tudományág – művelői megpróbálják a kvantummechanikánál absztraktabb, matematikailag még nehezebben kezelhető diszciplínát, a mézők kvantumelméletét is szigorú matematikai alapokra helyezni.

Neumann munkája ugyanazt a területet fedi le, mint Dirac kalkulusa, csak sokkal precízebb, mélyebb – és ennek megfelelően a napi fizikusi munkában nehezebben kezelhető. Ez az általános formalizmus a Dirac-féléhez hasonlóan a kvantummechanika Schrödinger-és Heisenberg-féle felépítése fölött helyezkedik el, azokat speciális esetként tartalmazza, egyben módszereket is nyújt a konkrét reprezentációk felépítéséhez, illetve az egyikről a másikra történő áttéréshez. Ugyanakkor általános volta miatt az alapvető kérdések (pl. a később említendő rejtett paraméterek problémája) tárgyalására a konkrét reprezentációknál sokkal alkalmasabbnak bizonyult.

A matematikai részletektől e helyütt természetesen el kell tekintenünk. A felépítés precíz és alapos, ugyanakkor látszik, hogy nem öncélú: a fizikus tudományos közösség által a fizika legnagyobb szemléleti forradalma, a kvantumelmélet robbanásszerű kifejlődésének néhány éve alatt elért eredményeket, a kutatás és a szemléletváltás megszenvedett gyümölcseit kell megmentenie, biztonságba helyezni, megfelelő tartósítással átadnia a következő kutatónemzedéknek, hogy az – immár az alapok matematikai megkérdőjelezése nélkül – bátran használhassa új tudományos problémák megoldására a XX. század fizikájának legkiemelkedőbb elméletét, a kvantummechanikát.

Ma, hetven évvel e nagy mű (melyet gyakran hasonlítanak egy másik „...*matematikai alapjai*” című könyvhöz, Newton *Principia*-jához) megszületése (és sajnos csak alig húsz évvel magyar kiadása) után is érezni a könyv olvasása során ezt az alkotó kettősséget: a precíz matematikus szigorúságát, és az új fizika eredményeit, gondolatait értő és óvó fizikus gondoskodó keze nyomát.

A tudománytörténet újabb fintora (úgy látszik, ez az istennő Neumann-nal kapcsolatban gyakran és kedvtelve fintorgott), hogy azt a matematikai programot, amit Neumann könyve előszavában megemlített és elvetett (azaz a Dirac-delta és a hozzá kapcsolódó fogalmak

szigorú matematikai megalapozását), néhány évvel később más matematikusok végrehajtották. Így született a matematika disztribúcióelméletnek nevezett ága. Ma tehát egy fizikus különösebb aggály nélkül számolhat a Dirac-deltával, tudván, hogy ha szükség lenne rá, akkor néhány kézikönyv áttanulmányozása és néhány tucat papírlap képletekkel történő teleírása után precízen is kiszámolhatná azt, amit így a Dirac-formalizmussal néhány sorban megkaphat. És minden gyakorló elméleti fizikus meg van róla győződve, hogy ugyanazt az eredményt kapná így is, úgy is.

Mindez természetesen mit sem von le Neumann János érdemeiből. Neumann megtette azt, amit – mély és széles körű matematikai ismeretei, ugyanakkor az új fizika kidolgozásában való közvetlen részvétele, az „alapító atyák” és alapgondolataik közvetlen megismérése és olykor alakítása révén – csakis ő tehetett meg: szigorú matematikai alapokra helyezte a kvantummechanikát. Az utódok most már bátran használhatják ezt a csodálatos eszközt, az emberi szellem e kiemelkedő alkotását.

5. Rejtett paraméterek

Neumann nagy könyvének második része a kvantummechanika valószínűségi jellegének matematikai megalapozásával és értelmezésével foglalkozik, egyben kitér a statisztikus mechanika kvantumos kiterjesztésének kérdéseire is. Neumann János e témakörökben is alapvető fontosságú eredményeket ért el.

Isten nem kockajátékos

Láttuk, hogy az általánosan elfogadott koppenhágai interpretáció szerint a kvantummechanika valószínűségi kijelentésekhez vezet. Ez az állítás nem pontosan írja le a helyzetet: van olyan kérdés, amire az elmélet egyértelmű választ ad – pl. arra, hogy mekkora energiájúak a hidrogénatom lehetséges energiaszintjei. Valószínűségi jellegű választ kapunk viszont az olyan kérdésekre, hogy mi lesz egy fizikai mennyiség mérésének eredménye, ha a vizsgált rendszer egy adott állapotban van.

Mi ennek a statisztikai jellegnek, ennek a bizonytalanságnak az oka? A tudomány történetében már korábban is fordultak elő valószínűségi megfontolások, sőt épp ilyen megfontolásokból eredt maga a valószínűség-számítás is, mint a matematika egyik ága. Az elterjedt interpretáció szerint olyan jelenségek leírására alkalmazzuk a valószínűség-számítást, amikor a vizsgált rendszer viselkedését túl sok, szerteágazó, nehezen követhető, nehezen figyelembe vehető körülmény határozza meg – és feltételezhetjük, hogy ez a sok hatás egymást mintegy kiegészítve, egyben egymás ellen dolgozva valamilyen átlagos viselkedéshez vezet.

Gondoljunk csak a kockadobásra. Senki sem kételkedik abban, hogy ha pontosan figyelembe vennénk a kockára elhajításakor ható erőket és forgatónyomatékokat, a levegő el-

lenállását, az esetleges légmozgások hatását, a kocka gurulása során az asztal egyenetlenségei által okozott erőket és még számtalan más körülményt, akkor pontosan ki tudnánk számítani (esetleg évek munkájával!), hogy melyik oldalára esik a kocka. Sok hűhó semmiért – ez az eredmény, és főleg a számítás sok részlete ugyanis senkit sem érdekel. Hallgatólagosan feltesszük, hogy a sokféle hatás egymástól független, egyik az egyik, a másik irányba befolyásolja a kocka mozgását, és végeredményben a kocka egyforma, $1/6$ valószínűséggel esik mindegyik oldalára.

Mit is jelent az utóbbi kijelentés? Képzeljük el, hogy ugyanazzal a kockával, ugyanazon körülmények között nagyon sokszor dobunk. Azaz az általunk tudatosan befolyásolható körülményeket (pl. a kéztartás, a dobás magassága stb.) igyekszünk azonosan tartani – csak hogy számtalan más körülmény is van, amelyekre nincs befolyásunk, és ezek is beleszólnak a kocka mozgásába. Ennek következtében a dobás kimenetele nem lesz azonos: hol ez, hol meg amaz a szám jön ki, ha elég sokáig dobálunk, nagyjából egyforma gyakorisággal.

Mire emlékeztet ez a helyzet? Természetesen a Gibbs-féle statisztikai sokaságokra. Ezek is a vizsgált rendszer igen sok párhuzamos kópiáját tartalmazzák, amelyek az általunk kontrollálható (makroszkopikus) paraméterekben (pl. térfogat, nyomás, hőmérséklet, részecskeszám) megegyeznek, a mikrodinamikát meghatározó belső változók azonban példányról példányra különböznek. (Az nem számít, hogy a sok kísérlet egymás után zajlik, mint a tényleges kockadobásnál, vagy párhuzamosan, a rendszer valóban elkészített sok példányán.) Az atomfizikában a Gibbs-sokaság gyakran tényleges kísérletekben is realizálható, a természet ugyanis bőkezűen ellátott bennünket ugyanazon fizikai rendszer azonos példányaival: ilyenek az atomok vagy az elemi részecskék. (Láttuk már, hogy – legalábbis az atomok esetében – ezt az azonosságot a kvantummechanika meg tudta magyarázni.) A modern részecskegyorsítóknál azonos állapotú elemi részecskékből álló nyalábok ütköztetnek egy céltárgyba. A másodpercenként sok száz milliárd elemi ütközési folyamat a Gibbs-sokaság legjobb realizációja – és valóban, a részecskefizikusok elméleteiket az eme ütközések során végbemenő folyamatok statisztikai analízisére alapozzák.

Másik példaként vizsgáljuk meg a kinetikus gázelmélet egyik sokat tanulmányozott alanyát, a tartályba zárt ideális gázt. Tudjuk, hogy az egyes molekulák össze-vissza mozgása miatt a tartály egyik falának egyszer több, egyszer kevesebb molekula ütközik, az észlelt erőhatás időben gyorsan ingadozik, fluktuál. Első közelítésben természetesen csak az átlagértéket, az állandó nyomást érzékeljük. Finomabb mérésekkel – illetve a fluktuációkat más termodinamikai mennyiségekkel összekapcsoló statisztikus fizikai tételek alapján – maguknak a fluktuációknak a nagyságát is meghatározhatjuk.

Egy ilyen elmélet nem mondja meg, hogy adott pillanatban pontosan mekkora nyomás hat az edény falára. Csak ennek a gyorsan változó, véletlen jellegű folyamatnak a statisztikus jellemzőiről (várható érték, szórás, különböző momentumok stb.) szolgáltat adatokat – egyébként épp az ergodikus tételre épülő Gibbs-féle statisztikus fizika alapján. Ennek ellenére senki sem kételkedik abban, hogy az említett mennyiség (azaz az adott pillanatban a falra ható erő) *létezik*, sőt: ha vennénk a fáradságot (és persze a megfelelő kapacitású számítógépet), akkor az egyes molekulák mozgás egyenetlenségének megoldásával, az ütközések pontos követésével – legalábbis elvben – pontosan ki is tudnánk számítani. Csak éppen az eredmény teljesen érdektelen lenne, nem érné meg a fáradságot. (Azt sem éri meg igen sok adatgyűjtő, kérdezőbiztos, adatfeldolgozó és számítógép bevetésével megtudakolni,

hogy 2003. április 26-án 20:28-kor Budapesten hány ember fogyasztott vajás zsömlét. Pedig meg lehetne tudni, csak aránytalanul drága – és felesleges lenne. A közgazdászok és a kereskedők ettől függetlenül – statisztikai alapon – pontosan ismerik a város zsömle- és vajszükségletét, sőt azt is tudják, hogy ez szombaton mennyivel szokott emelkedni.) A statisztikai viselkedés ismerete teljesen elegendő a gáz makroszkopikus viselkedésének leírásához. Mindenesetre jó tudni, hogy ha akarnánk, megismerhetnénk a mikroszkopikus részleteket is – legalábbis a klasszikus mechanika tanítása szerint.

Ha a gáz egy adott makroállapotában megmérnénk az egy pillanatban a falra ható (amúgy gyorsan fluktuáló) erőt, akkor hol ezt, hol meg amazt az értéket kapnánk. Ennek az az oka, hogy egy makroállapothoz igen sok mikroállapot tartozik, és a keresett mennyiség pontos értéke mikroállapotról mikroállapotra változik. Ha a makroszkopikus szintről leszállnánk a mikroszkopikusra, és egy adott mikroállapotban végeznénk el a mérést, akkor – szinte definíció szerint – csakis egyetlen értéket kaphatnánk. Hiszen a vizsgált fizikai mennyiség az egyes molekulák helyzete és impulzusa alapján számítható ki, azokat pedig az adott mikroállapotban mind pontosan ismerjük.

Ezeket a triviálisnak látszó, közismert állításokat azért taglaltuk ilyen részletesen, mert – éppen közismertségük okán – azonnal hasonló gondolatok bukkantak fel a legtöbb fizikus fejében, amikor először hallott a kvantummechanika valószínűségi jellegéről. A klasszikus statisztikus fizika nagyon derék, jól működő, a gyakorlati igényeket teljesen kielégítő elmélet – de *nem teljes*: nem szolgál a vizsgált rendszer részletekbe menő, aprólékos leírásával. És persze éppen ebből adódik statisztikus jellege. Ha minden molekula mozgását pontosan ismernénk, nem kellene valószínűségi kijelentésekhez folyamodnunk. De nem ismerjük – és persze nem is akarjuk ismerni, megelégszünk a nem teljes elmélet nyújtotta információkkal. Mindazonáltal az elmélet statisztikus kijelentései mögött ott rejlik a valóság mélyebb, determinisztikus törvényekkel irányított, elvileg megismerhető, és számításokkal követhető tartománya (legalábbis a klasszikus mechanikára épülő statisztikus fizika szerint – később ugyanis, Neumann kezdeményezésére kifejlődött a kvantummechanikán alapuló statisztikus fizika is).

Ha a kvantummechanikára a klasszikus statisztikus mechanikával analóg módon tekintünk, sok hasonlóságot találunk. A kvantummechanika is jól működő, a gyakorlati igényeket kielégítő, széles körben alkalmazható elmélet, következtetései minden vizsgált esetben megegyeznek a kísérleti tapasztalattal. De csak valószínűségi kijelentéseket tesz, ezért *nyilvánvalónak tűnik*, hogy nem teljes. Valahol alatta, mögötte ott rejlik a valóságnak az a tartománya, ahol a mozgásokat hagyományos, a klasszikus fizikában megszokotthoz hasonló determinisztikus törvények irányítják, ahol az ok egyértelműen meghatározza az okozatot, ahol a rendszer mikroállapotának ismeretében egyértelműen megmondhatjuk az összes fizikai mennyiség értékét, és ahol a valóban elvégzett mérések 100% bizonyossággal ezeket az értékeket szolgáltatják. Azaz a rend, a kauzalitás, a determinizmus, egyszóval: a klasszikus fizika világa. A XX. század fizikusainak elveszett paradicsoma.

Eme hipotetikus mélyebb szint változóit (az úgynevezett *rejtett paramétereket*) és változásuk törvényei megismerve elvileg – átlagolással, a statisztikus fizikával analóg módon – reprodukálni lehetne a kvantummechanika minden eredményét. De természetesen ennél sokkal többet is: amikor a rendszer egyik kvantumállapotában a kvantummechanikai elmélet csak a vállát vonogatja, és nem tudja (nem is akarja) megmondani, hol is van pontosan az a fránya elektron, csak a megtalálásának valószínűségéről papol – akkor a mélyebb

szint, a rejtett paraméterek értékei alapján egyértelműen ki tudnánk számítani az elektron helyzetét. Meg persze az összes egyéb fizikai mennyiség értékét is. Egy kvantumállapot – a statisztikus fizika makroállapotaihoz hasonlóan – nagyon sok, egymástól megkülönböztethető mikroállapot közös fedőneve lenne. Egy-egy ilyen mikroállapotban minden fizikai mennyiség határozott értéket venne fel. Csak (kísérleti és matematikai) eszközeink korlátozott voltán múlik, hogy ezeket a mikroállapotokat (még) nem tudjuk megkülönböztetni egymástól.

Ebben a képben a kvantummechanika érvényességi tartománya mintegy közbenső zónát képezne a klasszikus, determinisztikus fizika két tartománya: a makroszkopikus világ és az atomi jelenségeknél mélyebben fekvő, de azokat meghatározó szint, a rejtett paraméterek világa között.

A kvantummechanika az atomfizika szintjén természetesen jól működik, minden gyakorlati esetben elegendő információt szolgáltat. Az észlelt jelenségek értelmezéséhez és előrejelzéséhez nincs szükség a rejtett paraméterek világának ismeretére, elegendő a kvantummechanika. (A Gay–Lussac-féle gáztörvények felfedezéséhez és alkalmazásához sem kellett ismerni a kinetikus gázelméletet.)

Ennek a feltételezett mélyebb szintnek a keresését sem gyakorlati szükséglet, sem pedig az elmélet és a kísérlet ellentmondása nem indokolja (ezen a XIX–XX. század fordulója környékén megszokott állapoton a kvantummechanika megszületése és széles körben történt alkalmazásai után kb. egy évtizeddel már szerencsésen túljutottunk). A mélyebb szint keresésének egyetlen indoka az a filozófiai vagy esztétikai igény, amely az emberek jelentős részével nem engedi elfogadtatni azt a gondolatot, hogy a valóság legmélyebb szintjén nincs szigorú determinizmus, ott valóban statisztikai jellegű törvények uralkodnak. Vagy durvábban és egyértelműbben fogalmazva: a klasszikus fizika megszokott világa iránti *nosztalgia*.

Nem véletlen, hogy az idősebb fizikus generáció tagjai körében (ők azok, akik fiatalon még láthatták a boldog békeidőket, a klasszikus fizika diadalmas sikereit a XIX. század második felében) ezek a gondolatok szinte természetesnek számítottak. Közéjük tartoztak a kvantumelmélet alapítói közül is sokan: Planck, Einstein, Schrödinger, sőt a fiatalabbak közül de Broglie is. Ezek a nagy fizikusok borzadva szemlélték az általuk (is) szabadjára engedett szellemet, az indeterminizmus kísértetét, amint nagyra növe romba döntötte a klasszikus fizika gyönyörűséges épületét. Életük végéig nem fogadták el a kvantummechanikát mint a valóság „végső” elméletét, és legfőképpen minden bűnök gyökerével, a koppenhágai értelmezéssel álltak szemben.

A fizikusok ifjabb generációja – melynek képviselői (élükön Heisenberggel) már az új fizika bővületében, a mindent meg-, sőt újramagyarázás forradalmi hevében és állandó sikerei között nőttek fel, és lettek vezető tudósokká – nem osztozott a nagy elődök nosztalgijában, pesszimizmusában és szkepticizmusában. Közéjük kell számítani Niels Bohrt is, aki – bár a korábbi generációhoz tartozott – tanítványaival együtt az új fizika és annak filozófiája lelelkesebb apostolai, programjának kidolgozói közé számított. Ez a tudós társaság sajnálkozás nélkül vetette le és rúgta félre a klasszikus fizika megunt gúnyját, forradalmi (sőt Bohr *szavaival* „*elegendően örült*”) elméletek és teóriatöredékek kiötlésével és kidolgozásával kereste a megoldást a napi tudományos munka során felmerült apró kérdésekre is. Ellenállás nélkül elfogadták és magukévá tették az új fizika (pontosabban a koppenhágai interpretáció) eretnek gondolatait a valóság „végső” szintjének indeterminisz-

tikus voltáról, és gyanús restaurációs kísérletekként utasítottak vissza minden, a rejtett paraméterek létezésére utaló megfontolást.

A kvantummechanika ún. második generációja, azok a tudósok, akik még középiskolások voltak a nagy áttörés idején, de néhány év múlva az új elmélet magától értetődő alkalmazóiként teremtették meg a fizika sok „kvantumoz” részterületét (a kvantumkémia, molekulafizikát, a statisztikus kvantumfizikát, a szilárd testek kvantumelméletét, a mezők kvantumelméletét, a magfizikát, a csillagok energiatermelésének elméletét, az elemi részecskék fizikáját stb.), majd az ezekre épülő technológiát (nukleáris technika és energetika, félvezetők, lézerek stb.) mind az ifjú forradalmárok tanítványaiként nőttek fel, és fel sem merült bennük a kétség a kvantumelmélet (valamint annak koppenhágai interpretációja) iránt. Napi tudományos munkájukban folyamatosan tapasztalták, hogy az új elmélet *működik*, és igen termékenynek bizonyul. Holmi rejtett paraméterek keresése szemükben öreges szőrszálhasogatásnak tűnt volna.

A két irányzat képviselői a kor tudományos kongresszusain, konferenciáin csaptak össze. Közmondásossá váltak Bohr és Einstein vitái a kvantumelmélet alapkérdéseiről. Einstein ravasz gondolat kísérletek segítségével próbált fogást találni a koppenhágai értelmezésen, látszólagos ellentmondásokat, paradoxonokat vezetve le belőle. (Hasonlóan ahhoz, ahogy néhány évvel korábban az ő speciális és általános relativitáselméletét támadta paradoxonok kreálásával az értetlen és konzervatív fizikus közvélemény.) Ezek a viták, amelyekből – a kortársak visszaemlékezései szerint – rendre Bohr került ki győztesen, tovább finomították a koppenhágai interpretáció híveinek érvelését, gazdagították elméletüket. Einstein és társai természetesen nem adták fel – a további fejleményekre még visszatérünk.

A kísérleti technika a 80-as, 90-es évekre érkezett el oda, hogy a nagy fizikusok kiagyalta gondolat kísérletek némelyikét ténylegesen megvalósítsa. Az eddigi eredmények kivétel nélkül a kvantumelmélet előrejelzéseit igazolták – bármilyen meghökkentőek is voltak azok.

„*Isten nem kockajátékos*” – mondta Einstein, akinek ez a mondása mintegy programjává vált a kvantummechanika koppenhágai irányzatát elutasító, nagy nevekkel fémjelzett irányzatnak. Képviselői a kvantummechanikát, koruk fizikájának fő vonulatát átmeneti korszaknak, a klasszikus, determinisztikus fizika két – volt és eljövő – korszaka közé ékelődött csúf szakasznak, barbár, kultúrát és szépséget romboló középkornak tekintették.

E fizikusok egyetlen reményét a rejtett paraméterek feltételezett világának ígérete, a fentebb vázolt, a kvantummechanika és a statisztikus fizika közti analógiára építő gondolatmenet jelentette. Az ő reményeiket zúzta porrá Neumann Jánosnak a rejtett paraméterek nem létezéséről szóló, nagy könyve második részében közölt bizonyítása.

A sűrűségoperátor

A valószínűség-számítás a matematika viszonylag régi fejezete, a szerencsejátékok elemzéséből fejlődött ki néhány évszázaddal ezelőtt. Precíz matematikai alapjait mégis csak 1929-ben teremtette meg – a Neumann-nal egy évben, 1903-ban született – Andrej Kolmogorov. Ez az elmélet is (akárcsak Neumann ergodikus tétele és operátorelmélete)

Lebesgue-nek a század elején kidolgozott általános mértékelméletén alapul. (Kolmogorov elmélete tette helyére egyebek között a céltábla közepét eltaláló nyíl nulla valószínűségű találatának paradoxonát is – a nullmértékű halmaz Lebesgue-től származó fogalma segítségével.) A valószínűség-számítás eme forradalmi megújítása lényegében a kvantumelmélet forradalmával párhuzamosan folyt, és hamarosan számos új matematikai eredmény és irányzat megszületéséhez vezetett. Neumann jól ismerte ezeket a fejleményeket. Annál inkább figyelemre méltó, hogy a kvantummechanika valószínűség-fogalmának elemzése során nem az absztrakt, a matematikai analízisre támaszkodó kolmogorovi valószínűség-fogalmat, hanem a sokkal szemléletesebb, a fizikusok gondolatvilágához és gyakorlatához is közelebb álló Gibbs-féle statisztikus sokaságokat tekintette kiindulópontul.

El kell tehát képzelnünk a vizsgált kvantummechanikai rendszer igen sok példányából álló sokaságot, amelyek a kvantummechanika szempontjából azonos makroállapotot képviselnek, mikroszkopikusan azonban különbözhetnek. Azaz...

A fenti mondat két szempontból is revideálásra szorul. Neumann precíz matematikai vizsgálatai derítették ki, hogy még ebbe az egyszerű állításba is igen sok klasszikus fizikai „csökevény”, rejtett feltevés szorult.

Az első módosítás az „állapot” fogalmára vonatkozik. Ezt a szót egyrészt „a rendszert meghatározó adatok összessége” értelemben használjuk, másrészt létezik a „kvantummechanikai állapot” fogalma is, amit Neumann a könyv első részében pontos matematikai fogalmak segítségével definiál. A statisztikai sokaság fogalmának elemzése megmutatja, hogy ez a két jelentés jelen esetben nem esik egybe, ezért a fogalom további finomítására van szükség.

Az egyes fizikai mennyiségek mérési eredményeinek valószínűség-eloszlása két dologtól függ: attól, hogy milyen mennyiséget kívánunk mérni, és a rendszert jellemző adatoktól (a rendszer „állapotától”). Neumann részletesen megvizsgálja ezt a kérdést, és kimutatja, hogy tetszőleges fizikai mennyiség valószínűség-eloszlása kifejezhető néhány jól definiált, egyszerű fizikai mennyiség mérésének eredményei segítségével (ezt a kapcsolatot a kvantumelmélet lineáris struktúrája teszi lehetővé, ami köznapi nyelven az „anyag hullámok” szuperponálhatóságában, interferenciaképességében nyilvánul meg). Más szóval a rendszer teljes jellemzésére elegendő néhány mennyiség mérési statisztikájának ismerete. Ezeket az adatokat Neumann az ún. *sűrűségoperátorban* egyesíti, amely ettől kezdve a rendszer állapotának egyetlen legitim képviselője. Ennek az operátornak a számításokra is alkalmas konkrét megjelenési formája egy végtelen mátrix, a sűrűségmátrix. Minden lehetséges mérés eredményeinek eloszlása, várható értéke, szórása stb. kiszámítható a kérdéses fizikai mennyiség és a rendszer sűrűségmátrixa ismeretében. Neumann meg is adja a kiszámításhoz szükséges képletet.

Az így bevezetett sűrűségoperátor a statisztikus sokaságra jellemző, és általánosabb fogalom az addig használt „kvantummechanikai állapot” fogalmánál, amelyet Schrödinger elméletében a hullámfüggvény, Heisenberg és Dirac elméletében a végtelen dimenziós állapotvektor képvisel. Az utóbbiak speciális esetét jelentik az előbbinek – ebben az esetben az állapotot képviselő sokaságot „*tiszta sokaságnak*” nevezik, míg az általános esetben „*kevert sokaságról*” beszélünk. A kevert sokaságokra – kissé pongyola fogalmazásban – úgy gondolhatunk, mint különböző tiszta sokasággal képviselt állapotú kvantummechanikai rendszerek speciális „keverékeire”.

A másik említett probléma szorosabban kapcsolódik klasszikus fizikán nevelkedett gondolatvilágunkhoz. Mindenki úgy képzei, hogy ha szükséges, egyenként megvizsgálhatjuk a statisztikai sokaság egyes elemeit, és külön-külön méréseket hajthatunk végre rajtuk.

Holott erről szó sincs! A statisztikai sokaság épp arra való, hogy segítségével az egyetlen, *valódi* fizikai rendszeren végrehajtott mérések elméletét vizsgáljuk. Erre pedig a kvantumelmélet határozott matematikai eszközöket szolgáltat: az egyes fizikai mennyiségeket leíró operátorokat (melyek általános elméletét Neumann dolgozta ki, és alkalmazásukat a könyv első részében részletesen bemutatta). A sokaság egyes elemeiről tehát nem tudhatunk meg más információt, mint amit a kvantummechanika méréselmélete megenged.

Természetesen feltevésekkel élhetünk a sokaság tulajdonságaira vonatkozóan (de ezt ne úgy képzeljük, mint ha módunkban állna külön-külön megvizsgálni a sokaság egyes elemeit). Az egyik ilyen – értelmesnek és hasznosnak tűnő, a klasszikus fizikában pedig magától értetődő – lehetőség, ha olyan sokaságot keresünk, amelyen méréseket végezve minden fizikai mennyiség egyértelmű, határozott értéket vesz fel (ez lenne az ún. *szórásmentes sokaság*).

A klasszikus fizika Gibbs-sokaságaival ez elvileg megvalósítható lenne: legyen a sokaság minden eleme (a közös makroállapot mellett, pontosabban azon túlmenően) *ugyanabban a mikroállapotban*. Ekkor minden mérés eredménye kiszámítható e mikroállapot adatai alapján.

Égészen más a helyzet a kvantummechanikában. Neumann a hosszas és precíz matematikai előkészítés után lecsapja a labdát, és alig néhány soros egyszerű levezetésben megmutatja, hogy ilyen *szórásmentes sokaság nem létezik*. Létezése ellentmondásban lenne a kvantummechanikai leírás korábban már elfogadott, axiomatizált matematikai feltevéseivel.

Ez az alapvető eredmény búvik meg Heisenberg híres és sokat idézett határozatlansági relációjának hátterében is. Mivel nem vehet fel minden fizikai mennyiség egyszerre határozott értéket, mindig találhatunk olyan mennyiségeket, amelyek közül a rendszer – adott sűrűségoperátorral jellemezhető – állapotában legalább az egyiknek szórása van. A határozatlansági reláció e szórások között teremt kapcsolatot.

A következő lépésben Neumann matematikai kritériumot ad a tiszta és a kevert sokaságok megkülönböztetésére. Ez a kritérium pontosan és egyértelműen definiálja, mikor lesz az adott sűrűségoperátorral jellemezhető sokaság tiszta, azaz a hagyományos állapotvektorral vagy hullámfüggvénnyel is leírható sokaság. (Érdekességképpen megemlítjük, hogy a kritérium így szól: a tiszta állapot sűrűségoperátora épp a végtelen dimenziós állapotter egy egydimenziós alterére vetítő operátorral egyezik meg. Ez az egydimenziós alter azonosítható Dirac állapotvektorával.)

A sűrűségoperátor látszólag ártatlan matematikai definíciója döntő fontosságúnak bizonyult a rejtett paraméterek létezésének vagy nemlétezésének kérdéskörében, amire pedig a fizika nagyjai oly sok szót, tintát és nyomdafestéket pazaroltak.

Rejtett paraméterek pedig nincsenek

Gondoljuk végig még egyszer, mit is kellene tudnia egy – a klasszikus fizika iránt nosztalgiját érzők által vágyott – rejtett paraméteres elméletnek!

Mindenekelőtt ismételjük meg, hogy a rejtett paraméteres elméletre nem azért van szükségünk, mert a kvantummechanika hibás – éppen ellenkezőleg, egy jól működő, helyes

eredményeket szolgáltató elmélet filozófiailag vagy esztétikailag kielégítőbb megalapozását szeretnék elérni.

Ezért elvárjuk, hogy a kiépítendő rejtett paraméteres elmélet vezessen ugyanazokra az eredményekre, mint a kvantummechanika – minden olyan esetben, amikor az utóbbi (saját hatáskörén belül) értelmes kijelentéseket tesz. A rejtett paraméteres elmélet extra jó tulajdonságainak olyankor kell megmutatkoznia, amikor a kvantummechanika nem kíván nyilatkozni (pl.: hol is van *valójában* az elektron?).

Másképp szólva: a nem statisztikai jellegű, determinisztikus rejtett paraméteres elmélet alapján készített valószínűség-eloszlások (amikor is a mélyebb elmélet közös makroállapothoz tartozó mikroállapotait „egybeösszuk”, átlagolunk rájuk) egyezzenek meg a kvantummechanika alapján számítható eloszlásokkal. Neumann statisztikai vizsgálatai szerint az utóbbi eloszlásokat kimerítően jellemzi a sűrűségoperátor – ennek alapján az előző követelmény úgy fogalmazható meg, hogy a rejtett paraméteres számolás köteles reprodukálni a rendszer sűrűségoperátorát, annak összes matematikai sajátosságával együtt.

A kvantummechanikai szint tiszta állapota a mélyebb, rejtett paraméteres szinten mikroállapotok keveréke, a neki megfelelő altér már többdimenziós lenne. Márpedig a vetítő operátor dimenziószáma egyértelmű, matematikailag pontosan definiált mennyiség – nincs mód tehát arra, hogy a kvantummechanikai szint valamennyi statisztikus tulajdonságát visszatükrözze a mélyebb szint leírásán végzett átlagolás! Nem lehetséges tehát olyan rejtett paraméteres leírás, ami a kvantummechanika valamennyi eredményét visszaadná.

Neumann lépésről lépésre, egyszerű és mindenki által elfogadható definíciók és levezetési lépések láncolatán át vezet el olvasóját ehhez a valóban meghökkentő végeredményhez. És persze rögtön válaszol néhány lehetséges ellenérvre is.

Először is felhívja a figyelmet arra, hogy az utóbbi lépések tulajdonképpen feleslegesek voltak. Hiszen a rejtett paraméteres szintet klasszikusnak tételezik fel, azaz a megfelelő sokaság szórásmentes lenne. Ilyen pedig – mint korábban megmutatta – nem létezik. Ezzel az állítással a következő ellenérvet lehetne szembeszegezni: egyáltalán nem szükségszerű az, hogy a mélyebb, rejtett paraméteres szinten használt matematika megegyezzen a kvantummechanikáéval, vagy akár csak hasonlítson arra. E szinten tehát esetleg nem releváns a szórásmentes sokaság nemlétezéséről szóló tétel (akárcsak a makroszkopikus klasszikus fizikában sem). Az utóbbi bizonyítási lépések viszont már függetlenek voltak a rejtett paraméteres szint feltételezett matematikai leírásától, csak a kvantummechanikai szint – elfogadott, és jól működő – matematikáját használták. E bizonyítási lépések tehát tetszőleges rejtett paraméteres elmélet létezését megtiltják – feltéve, ha a jelen alfejezet elején leírt követelményt elfogadjuk, azaz reprodukálni kívánjuk a kvantummechanika helyes eredményeit. Persze ha a kvantummechanika hibázna, azaz rosszul írná le a kísérletek eredményeit, akkor ismét megnyílna a tér sokféle alternatív elmélet – köztük a rejtett paramétereket tartalmazók – előtt.

Érdemes e helyen kissé hosszabban idézni Neumann könyvének idevágó eszmefuttatásából. (Ahol Neumann „kauzalitást” ír, mai szóhasználattal inkább „*determinizmust*” mondanánk.) Megfigyelhetjük az egzaktul bizonyított eredményére büszke matematikus mellett a mindig szkeptikus és óvatos természettudóst (hiszen senki sem jelentheti ki egy fizikai elmületről, hogy az lesz a végső szó a világ megismerésében, az elméletek igazságtartalmát mindig a konkurens elméletekhez viszonyítják, és végső soron a kísérletek dönté-

nek), valamint a „nagyok” filozófiai vitáiból általában kimaradó, de most e vitához döntő tudományos érvet szolgáltató csendes forradalmárt is, aki a klasszikus fizika iránti nosztalgia helyett még „*az emberiség ősi gondolkodásmódjával*” is kész szakítani egy sikeres és szép fizikai elmélet érdekében:

Az sem lenne elegendő, ha a kvantummechanika operátorainak megfelelő fizikai mennyiségek mellett még más, eddig még fel nem fedezett mennyiségek is léteznének, mert a kvantummechanika által felvetett összefüggések a már ismert mennyiségekre érvényüket vesztenék. Ezért – noha azt gyakran feltételezik – az elemi folyamatok statisztikustól különböző leírásának lehetősége nem a kvantummechanika átértelmezésének kérdése, mert ez csak úgy lenne lehetséges, ha a kvantummechanika jelenlegi rendszere hibás eredményeket adna. [...]

Természetesen túlzás lenne azt állítani, hogy a kauzalitást ezennel elintéztük: a kvantummechanikának a jelenlegi formájában számos komoly hiányossága van, és még az is előfordulhat, hogy hibás, bár ez az utóbbi lehetőség nagyon valószínűtlen annak fényében, hogy az általános problémák kvalitatív magyarázatában és a speciális problémák kvantitatív megoldásában milyen átütő sikerhez vezetett. Annak ellenére, hogy a kísérletekkel a kvantummechanika jól egyezik, és a világ egy minőségileg új oldalát tárta fel, azt nem mondhatjuk az elmélet-ről, hogy a tapasztalat bebizonyította, hanem csak annyit, hogy a tapasztalat legjobb ismert összegzése. Ám, ha még oly elővigyázatosak is vagyunk, s e kételyeket szem előtt is tartjuk, mégis azt mondhatjuk, hogy jelenleg nincs alkalom és ok arra, hogy a kauzalitás kérdését a természetben feszegezzük, erre ugyanis semmilyen kísérlet nem utal. A makroszkopikus kísérletek a kérdés eldöntésére elvileg alkalmatlanok, az elemi folyamatokkal kapcsolatos tapasztalatokkal összeférő egyetlen elmélet, a kvantummechanika viszont a kauzalitásnak ellene szól.

Felhívjuk a figyelmet arra, hogy itt az emberiség ősi gondolkodásmódjával és nem logikai szükségszerűségekkel állunk szemben (erre utal az is, hogy egyáltalán lehetséges volt statisztikus elméletet kiépíteni), s a kérdés vizsgálatakor nincs okunk arra, hogy a begyökerezett gondolkodásmódhoz ragaszkodjunk. Ilyen körülmények között bölcs dolog-e a kauzalitás kedvéért egy ésszerű fizikai elmélettről lemondani?

... vagy mégis?

A kvantummechanika koppenhágai értelmezésének ellenfelei természetesen Neumann bizonyítása ellenére sem zárták le a vitát. Einstein és két kollégája a harmincas években egy híres cikkben hívták fel a figyelmet az ún. EPR-paradoxonra (melyet részletesen elemezni itt nincs helyünk, de az utóbbi időben a kvantumteleportációs kísérletek kapcsán amúgy is az érdeklődés – és az ismeretterjesztő irodalom – középpontjába került). E paradoxonnal azt igyekeztek bebizonyítani, hogy a kvantumelmélet valóságleírása nem teljes, bizonyos létező valóságalelemek annak horizontján kívülre kerülnek. Néhány éve végrehaj-

tották az eredeti EPR-kísérlet egy változatát, és a látszólagos paradoxon ellenére a kvantummechanikai leírás bizonyult helyesnek.

Más fizikusok (pl. Bohm és de Broglie, majd Vigier) a kvantummechanika módosításával próbálkoztak, hogy filozófiai és természetleírási igényeinek jobban megfelelő elméletet kapjanak. De Broglie nemlineáris elméletében (szemben az eredeti kvantumelmélettel) nem érvényes a szuperpozíció elve, így a Neumann és mások által kidolgozott lineáris operátorelméletet és vele a kvantummechanika matematikai eszköztárának jelentős részét ki kellene dobni az ablakon. A fizikusok nagy része ezt túl nagy árnak tartaná egy még semmit sem bizonyított „realisztikus” elmélet elfogadásáért.

A hatvanas években nagy feltűnést keltett, hogy John Bell publikált egy valóban működő rejtett paraméteres elméletet. Igaz, csak a legegyszerűbb, kétállapotú kvantummechanikai rendszer (pl. az elektronspin fel- és lefelé mutató állapota) esetére. Az igen egyszerű modellben explicit, determinisztikus képlet adja meg, hogyan kell kiszámítani a rendszerre jellemző, a megfigyelő által is ismert (makro-)paraméterek és a megfigyelő által nem ismert (rejtett) paraméterek alapján az egyetlen mérhető mennyiség értékét. A rejtett paraméter dinamikájáról csak annyit kell feltételezni, hogy fázisterét egyenletes eloszlással, ergodikusan járja be. E modellben a kétállapotú rendszer kvantumelméletében minden mérhető mennyisége egzaktul reprodukálható a determinisztikus mikroállapot alapján.

Mi történt? Vajon hibásnak bizonyult Neumann bizonyítása? A részletes vizsgálat azt mutatja, hogy a vizsgált rendszer túl egyszerű. Neumann levezetésének elején teljesen kézenfekvő feltevéseket fogad el (és fogadtat el az olvasóval) a statisztikus függvények elemi tulajdonságaira vonatkozóan. E feltevések némelyike e túlságosan egyszerű rendszerre nem értelmezhető, ezért megnyílik a tér a rejtett paraméteres konstrukció előtt. Bell később maga mutatta meg, hogy modellje egy kicsit is bonyolultabb rendszerre már nem általánosítható.

A Bell-egyenlőtlenségek

Bell a következő években Neumann-nál egyszerűbb bizonyítást keresett (és talált) a rejtett paraméterek kérdéskörében. Módszerének nagy előnye, hogy még Neumannnál is kevesebb, lényegében semmilyen kikötést sem tesz a rejtett paraméterek mibenlétére, a mélyebb szint törvényszerűségeire nézve. Az egyedüli feltétel a *lokálitás*, amit a speciális relativitáselmélet létrejötté óta minden fizikus elfogad: nem lehetséges fénynél gyorsabb információátvitel. E teljesen általános feltételek alapján Bell levezetett egy egyenlőtlenséget bizonyos egyszerű kísérletek lehetséges eredményei közti kölcsönös kapcsolatok (ún. korrelációk) értékeire. Ennek az egyenlőtlenségnek minden rejtett paraméteres lokális elméletben teljesülnie kell, a rejtett paraméterek konkrét tulajdonságaitól és a feltételezett dinamikától teljesen függetlenül.

Már csak egy lépés volt hátra: ki kellett számítani ugyanezen korrelációk értékét a kvantummechanika standard módszerei alapján, és behelyettesítéssel megvizsgálni, igaz-e rájuk a Bell-egyenlőtlenség. Könnyű volt olyan eseteket találni, amikor az egyenlőtlenség nem teljesül.

Mit jelent ez az eredmény? Logikája hasonló Neumann bizonyításához (csak a levezés sokkal elemibb, ezért jobban alkalmas a matematikailag kevésbé képzettek meggyőzésére): a rejtett paraméterek pusztán létezése bizonyos összefüggéseket követel meg a mérési eredmények között – de a megfelelő kvantummechanikai számítás ezeket az összefüggéseket nem teljesítő eredményhez vezet. Következésképpen nem létezik olyan rejtett paraméteres modell, amely minden részletében reprodukálná a kvantummechanika eredményeit, és ezzel mintegy az elmélet felszínét nem borzoló, de a mélységet átalakító „háttérelméletként” lapulhatna a kvantummechanika alatt vagy mögött.

A későbbi vizsgálatok arra utaltak, hogy a Bell-egyenlőtlenség (és különböző későbbi változatai) a kvantummechanika valószínűségi modelljének egy nagyon fontos tulajdonságával: *nem-kolmogorovi* voltával kapcsolatosak.

Említettük, hogy nagyjából a kvantummechanika születésével egyidejűleg Kolmogorov dolgozta ki a valószínűség-számítás matematikailag precíz, a Lebesgue-féle mértékelméltre támaszkodó alapjait. A kolmogorovi valószínűség-elméletben az események rendszerre („eseménytér”) egyes elemeihez egy 0 és 1 közti számot, valószínűséget rendelnek. A bonyolultabb események (A és B, A vagy B) valószínűsége az összetevő események ismert valószínűségeiből meghatározott szabályok alapján számítható ki. Vannak ún. „*elemi*” *események* (kockadobásnál pl. az, hogy egy kockával egyest, kettést... hatost dobunk). Ezek ismert (jelen esetben 1/6 értékű) valószínűségei alapján tetszőleges bonyolultabb, összetett esemény (pl. az, hogy 137 kockadobásból pontosan 42-szer dobunk hatost) valószínűsége kiszámítható. Ezt a valószínűségekkel ellátott eseményrendszert nevezik a matematikusok kolmogorovi eseményhálónak. Ez a matematikai konstrukció nagyjából megfelel a „köznapi józan ész” valószínűség-fogalmának, amit a mindennapi életben, kockázatbecslés és lotóztás során alkalmazunk, csak persze a kolmogorovi elmélet szigorú matematikai fogalmakkal operál.

A kvantummechanika valószínűség-elméletének logikája egészen más. A Schrödinger-egyenlet megoldásával vagy más módon meghatározott hullámfüggvény (vagy Neumann absztrakt felépítésében a sűrűségoperátor) ismeretében tetszőleges esemény valószínűsége közvetlenül megkapható. „Eseményen” itt azt értjük, hogy valamilyen fizikai mennyiség mérése meghatározott értéket ad. Az egyes események valószínűségeit így a kvantummechanikában közvetlen úton kapjuk meg. De vajon érvényesek-e az így kapott, „valószínűségnek” nevezett mennyiségekre a kolmogorovi valószínűség-számítás szabályai, axiómái, tételei?

A viszonylag egyszerű, ténylegesen végigszámolható modellek vizsgálata arra az eredményre vezetett, hogy a kolmogorovi szabályok érvényessége az egyes eseményekre kapott valószínűség-értékek közti speciális összefüggések fennállását, bizonyos egyenlőtlenségek teljesülését követeli meg. És ezek az összefüggések éppen a Bell-egyenlőtlenségek!

A Bell-egyenlőtlenségek tehát úgy tekinthetők, mint az eseménytér és a rajta értelmezett valószínűség-függvény „kolmogorovitásának” feltételei. A Bell és mások által vizsgált példák tanulsága pedig úgy összegezhető, *hogy a kvantummechanikai eseménytér és valószínűség-elmélet nem kolmogorovi!* (Ez az állítás szoros kapcsolatban van a Neumanntól származó kvantumlogikai eseményalgebra klasszikus logikától eltérő szabályaival.) Ezért érezzük a kvantummechanika eredményeit furcsának, olykor abszurdnak – mert nem a köznapi józan észnek megfelelő valószínűség-elméleti modellt használja. De ugyan mit tudhatnának az elektronok a köznapi józan észről...?

Utólag, több évtized után kiderült, hogy Neumann János – valószínűleg fizikusi ösztönére hallgatva – igen jól döntött, amikor nem a kora matematikájának élvonalában álló, frissen született kolmogorovi valószínűség-elméletet választotta a kvantummechanikai statisztikus elmélet kiindulópontjául, nem követelte meg a kolmogorovi valószínűség-számítás szabályainak fennállását, hanem elfogadta a Bohr és iskolája által kidolgozott koppenhágai értelmezést, a hullámfüggvény alapján kiszámítható „valószínűség” fogalmát. Ezen a nyelven ugyanis meg tudta fogalmazni (bár explicit módon nem mondta ki), mennyire eltér a kvantumelmélet valószínűségi modellje a mindennapjainkban alkalmazott modelltől. A döntő szót persze a kísérletek mondják ki...

A kvantummechanika működik!

Bell eredménye – legalábbis elvileg – arra is alkalmas, hogy döntson a két rivális fél (egyik oldalon a kvantummechanika, másik oldalon az összes rejtett paraméteres lokális elmélet) között. Hiszen az egyik küzdő fél egy egyenlőtlenség fennállását követeli meg, a másik pedig éppen a fordított egyenlőtlenségét. E mérés technikai végrehajtása sokáig reménytelennek tűnt, míg végül a nyolcvanas években (lézerfényvel végrehajtott precíziós mérésekkel) végre sikerült. A mérés egyértelműen a kvantummechanika által adott előrejelzést igazolta.

Bell egyenlőtlenségének azóta nagyon sok verziója napvilágot látott, különböző ravasznál ravaszabb fizikai rendszerre alkalmazva a fenti logikát. Néhány esetben a vonatkozó mérés is megtörtént: minden esetben a kvantumelmélet jóslata vált be.

Napjaink mérései, a hajdani nagyok gondolat kísérleteinek modern megvalósulásai is igazolják Neumann János fentebb idézett szavait: a kvantummechanika „*a tapasztalat legjobb ismert összegezése*”. A Neumann könyvének megjelenése óta eltelt hetven év a kvantumelmélet számos új diadalát hozta el, emellett sok gyakorlati alkalmazást – köztük Neumann másik fő műve, a számítógép globális elterjedését és behatolását mindennapjainkba. A mai számítógépek chipjeinek villámnál is gyorsabb működése, a CD-nket olvasó lézergyár – mind-mind a kvantumelmélet egy-egy gyakorlati megvalósulása. Ez a gyönyörű, termékeny, sokszor igazolt elmélet, a XX. század egyik legnagyobb szellemi teljesítménye igazán megérdemli, hogy miatta lemondjunk a rejtett paraméterek kereséséről, hiszen „*nincs okunk arra, hogy e begyökerezett gondolkodásmódhoz ragaszkodjunk*”.

Neumann János – matematikusként elért hatalmas eredményei és a számítógép megalkotásában szerzett érdemei mellett – a statisztikus fizika és a kvantummechanika szilárd matematikai alapokra helyezésével olyan máig ható, sok korábbi kutatást összegezõ és sok további kutatást megindító fizikusi életművet hagyott ránk, amelynek alapján nevét mindenképpen a XX. század legnagyobb elméleti fizikusai között kell számon tartanunk.

Forgó Ferenc–Zalai Ernő

Neumann János hozzájárulása a játékelmélethez és a matematikai közgazdaságtanhoz

Neumann János a 20. századra számukban erősen megfigyelt polihisztorok egyike volt, aki több tudományterületen is kiemelkedőt alkotott. Csak egyetlen közgazdasági munkája volt, „*Az általános gazdasági egyensúly egy modellje*”, amelyet először 1932-ben a princetoni egyetemen adott elő, majd 1937-ben publikált Ausztriában. (Angol nyelvű fordítását 1946-ban közölték, magyar nyelven 1965-ben jelent meg válogatott munkái között. A Neumann-idézeteket rendre az utóbbiból vesszük át.)

Rögtön hozzátesszük, Neumann János a modern *játékelmélet* megalapozásával is jelentősen hozzájárult a közgazdaságtan fejlődéséhez. A játékelmélet azonban egy több tudományágat (matematika, közgazdaságtan, pszichológia, politológia) átfogó önálló terület, amely az ésszerű emberi viselkedés szabályaival foglalkozik, és az esetenként *praxeológiának* nevezett (lásd Lange, 1964) interdiszciplináris tudományágba sorolható.

Neumann Jánosnak a közgazdaságtan alakulására gyakorolt jelentős hatását még a Nobel-díjas közgazdász, *Samuelson* is kénytelen volt elismerni, aki egyébként igen kevésre értékelte közgazdasági szempontból magát a Neumann-modellt: „betoppant egy rövid időre a területünkre (mármint a közgazdaságtanba), és azóta már nem lesz ugyanaz, mint ami volt” (*Samuelson*, 1989. 121. o.). *Roy Weintraub* (1983), a matematikai közgazdaságtan történetének egyik legavatottabb kutatója szerint „Neumann dolgozata... a matematikai közgazdaságtan kiemelkedően legfontosabb (*single most important*) cikke” (uo. 13. o.). Neumann játékelméleti és gazdaságnövekedési modellje számos korábbi fontos fejlemény szintézise, később napvilágot látott eredmények zseniális előrejelzései, és egyidejűleg eltérő szemléletű gazdaságelméleti iskolának közös találkozási, illetve kritikus elágazási pontja is. A legtöbb közgazdasági irányzat igyekszik magáénak vallani.

Neumann modelljeinek megjelenése időben egybeesik a *kvantitatív közgazdaságtan* mint önálló tudományág megszerveződésével. A jelentős részben önállósult kvantitatív közgazdaságtudományi irányzatok kialakulásának kezdeteként ugyanis az *Econometric Society* megalakulását (1930), és lapjának, az *Econometrica*-nak a megjelenését (1933), tehát az 1930-as évek elejét jelölhetjük meg. Az erőteljes matematikai és statisztikai módszertannal dolgozó kvantitatív közgazdaságtani irányzat ekkor még egy név alatt (*econometrics*) jelent meg, de később maga is részben önálló részdiszciplínákra tagozódott. Manapság *matematikai közgazdaságtan*, *operációkutatás* és *ökonometria* (gazdaságstatisztika) elnevezések alatt találkozhatunk vele.

A matematika közgazdaságtani alkalmazásáról *Cournot* (1838) munkájának megjelené-

se óta beszélhetünk, s noha alkalmazhatósága körül váltakozó hevességgel folyik a vita, ma már a közgazdaságtan elválaszthatatlan része. A legtöbb módszertani eszközt és tételt a fizikából kölcsönözték a közgazdászok. Különösen igaz ez az *egyensúly* fogalmára, amely Neumannnál is központi fogalom. Neumann János ugyanakkor a fennmaradt megjegyzések szerint (lásd például *Morgenstern*, 1976) visszatérően hangsúlyozta, hogy a közgazdászoknak a klasszikus fizikából átvett matematikai eszköztárnál jóval korszerűbb és adekvátabb matematikai módszereket kellene felhasználniuk. Minden bizonnyal ez sarkallta a játékelmélet általános elméletének kidolgozására az absztrakt általános egyensúlyelméleti modellek további kutatása helyett. Pedig viszonylag korán szoros kapcsolatba került azokkal az amerikai kutatásokkal, amelyek eredményeként – jelentős részben az ő módszertani újításait követve és továbbfejlesztve – megszülettek az általános egyensúlyelmélet modern modelljei és egzisztenciabizonyításai.

Az új módszertan iránti igényt nemcsak mások felé támasztotta követelményként, hanem követésre méltó példát is mutatott rá. Neumann János volt az első, aki világosan és matematikailag korrekt módon definiálta az absztrakt játék fogalmát, és felhívta a figyelmet az új elmélet szintetizáló jellegére. Olyan modelleket állított fel és olyan alapvető tételeket bizonyított, amelyek önmagukban, de különösen a később hozzá kapcsolódó kutatások eredményeként döntően meghatározták a mai modern játékelmélet arculatát. Külön tudományelméleti ritkaságnak számít, hogy a kooperatív játékok elméletét, egy teljesen új tudományágat, egy olyan monumentális könyvvel indította el 1944-ben az osztrák *Oskar Morgensternnel* együtt, amely csupa új eredményt tartalmazott. Ilyen műveket általában a tudományos cikkekben megjelenő részeredmények összegezeként és rendszerezéseként szoktak megjelentetni.

A tanulmány szűkre szabott kerete nem teszi lehetővé, hogy az egyes önmagukban is roppant összetett és izgalmas területek világával, minden fontos részletre kiterjedő módon, megismertessük az Olvasót. Dolgozatunkban Neumann János eredményeit időrendben mutatjuk be, mivel az időrendi sorrend hívebben tükrözi Neumann tudományos felfogásának alakulását, mint egy tematikus szerkezet. Először a kétszemélyes, zérusösszegű játékok egyensúlyának létezésére adott bizonyítását, majd a hasonló módszertanra épülő általános egyensúlyelméleti modelljét mutatjuk be, s végül – a játékelmélethez visszatérve – a kooperatív játékok elméletéhez kapcsolódó munkásságát ismertetjük.

A két téma, a játékelmélet és az általános gazdasági egyensúlyelmélet bemutatása és megértése korántsem jelent azonos nehézségű feladatot. Azok számára, akik nem jártasak a közgazdaságtanban, igyekeztünk megkönnyíteni az egyensúlyelméleti részek tanulmányozását egy rövid általános közgazdaságtani bevezetővel is. Bemutatjuk továbbá a legfontosabb korabeli általános egyensúlyelméleti modelleket, és összevetjük velük Neumann modelljét. Neumann közgazdaságtani hozzájárulásának jelentősége ugyanis nem értékelhető önmagában, még kevésbé pusztán csak a modell közgazdasági mondanivalója alapján. Neumann modellje – a maga nemében és idejében – egyike volt azoknak a modelleknek, amelyek a legtisztábban és a legszebben testesítették meg azt a szemléletbeli és módszertani változást, szinte robbanásszerű fejlődést, amely a matematika közgazdaságtani alkalmazásában az 1930–1950-es évek folyamán végbement. Röviden érintjük ezt a fejleményt is. Bízunk abban, hogy a témákkal először találkozó olvasó, ha nem is fog első olvasásra mindent megérteni és átlátni, kedvet fog kapni ahhoz, hogy mélyebben megismerkedjen az érintett témákkal. A csatolt irodalomjegyzékben bőven fog tanulmányozásra alkalmas forrásokat találni.

1. A kétszemélyes, zérusösszegű játékok egyensúlya létezésének bizonyítása

Neumann János előtt elszórtan ugyan, de foglalkoztak olyan stratégiai jellegű problémákkal, amelyeket manapság legjobban játékelméleti eszközökkel lehet vizsgálni. Legjobb példa erre *Cournot* (1838), aki olyan oligopolisztikus piacokat vizsgált, ahol néhány nagy szereplő alakítja, termelési döntésein keresztül, az árakat. Neumann János volt ugyanakkor az első, aki világosan, matematikailag korrekt módon definiálta az absztrakt játék fogalmát. Ahhoz, hogy Neumann János munkájának jelentőségét megértsük, szükségünk van néhány játékelméleti alapfogalom tisztázására.

Néhány játékelméleti alapfogalom

Mivel is foglalkozik a játékelmélet? Ha a legáltalánosabban akarjuk megfogalmazni, akkor azt mondhatjuk, hogy a játékelmélet olyan matematikai modellek összessége, amelyeket olyan konfliktushelyzetek tanulmányozása és elemzése céljából állítunk fel, amelyekben az egyes személyek vagy csoportok elért eredménye nemcsak a saját, hanem mások döntéseitől is függ. A konfliktus szót itt nem úgy kell értelmezni, hogy a szereplők (játékosok) érdekei szükségképpen abszolút ellentétesek (ez is lehet speciális esetként), bizonyos mértékig egybe is eshetnek, sőt bizonyos célok elérése érdekében együtt is működhetnek. Az élet minden területén adódnak konfliktushelyzetek és több tudományág is foglalkozik ezek bizonyos vonatkozásaival. Gondoljunk csak a jogra vagy a pszichológiára. A játékelmélet azt a célt tűzte ki, hogy ezeket a helyzeteket matematikai modellekkel elemzi, és így próbálja megmagyarázni, hogy a játékosok miért úgy viselkednek, ahogy az megfigyelhető, hogyan kellene viselkedniük, ha bizonyos ésszerűségi követelményeket tartanak szem előtt, előre tudjuk-e jelezni, hogy a játékosok együttes cselekvésének eredményeképpen mi fog kialakulni, és mennyire tekinthető a kialakult helyzet stabilnak.

Két dolgot fontos itt kiemelni. Noha a köznyelv játéknak nevezi, például a kockajátékot, mi nem tekintjük annak, mivel a játékosok passzívak, nem hoznak a végkimenetelt befolyásoló döntéseket, szerepük mechanikus (a kocka eldobását hajtják végre). Szükséges az is, hogy legalább két tudatos játékos legyen, és így kizárjuk azokat a döntési helyzeteket, amikor egy döntéshozó választ különböző lehetőségek közül, noha ez, mint speciális eset érdekes bizonyos játékelméleti modellek működésének tesztelésére. A vizsgálat módszere is speciális. Mindenekelőtt azt vizsgáljuk, hogy bizonyos helyzetekben a játékosoknak hogyan kellene viselkedniük, és ennek következtében mi alakulna ki, és nem pedig azt, hogy valójában mit csinálnak az emberek. A *kísérleti játékelmélet* foglalkozik az utóbbival, joggal tekinthetjük a klasszikus játékelmélet közeli rokonának, és fontos szerepe van a játékelméleti modellek igazolásában. Ha nagy eltérés mutatkozik egy játékelméleti modell által előírt és a valóságban megfigyelt viselkedési forma között, akkor ez komoly ösztönzés a modellek finomítására, netán gyökeres átalakítására.

Már a kezdetektől fogva két nagy területre különült el a játékelmélet: a *kooperatív* és a

nem kooperatív játékok elméletére. A kooperatív játékok esetében a játék szabályai megengedik, hogy a játékosok cselekvésüket összehangolják, és ezt kötelező erejű szerződésekben rögzítsék. Például egy parlamentáris demokráciában két párt (ezek a játékosok) köthet egy kötelező érvényű koalíciós szerződést, amelynek az előírásait különböző szavazásoknál követniük kell. Nem kooperatív játékok esetében ezt a játék szabályai nem engedik meg, ami persze nem zárja ki azt, hogy egyéb módokon ne hangolják össze cselekvésüket (ha ezt saját érdekük úgy diktálja).

Példaként említjük meg az *oligopolisztikus piaci versenyt*. Képzeljünk el egy olyan piacot, ahol n számú termelő van, mindegyikük ugyanazt a terméket állítja elő, oly módon, hogy a fogyasztó számára teljesen lényegtelen, hogy kinek a termékét veszi meg. Ilyen termék lehet például a cukor, mivel általában az ember, amikor cukrot vásárol, nem szokta megnézni, hogy melyik vállalat állította elő. A vállalatok egymástól függetlenül hozzák meg a termelési döntéseiket. (A versenyhivatal nem engedi meg a termelési volumenek összehangolását, a kartelleket!) A fogyasztók a piacon így megjelenő összes termékmennyiséget egy bizonyos áron hajlandók mind megvenni. Ha sok termék van, akkor ez az ár kisebb, ha kevés, akkor nagyobb. Így az egy vállalat által elérhető nyereség nagysága nemcsak attól függ, hogy ő maga mekkora mennyiséget állít elő és mekkora költséggel, hanem attól is, hogy a többi vállalat milyen termelési döntést hozott.

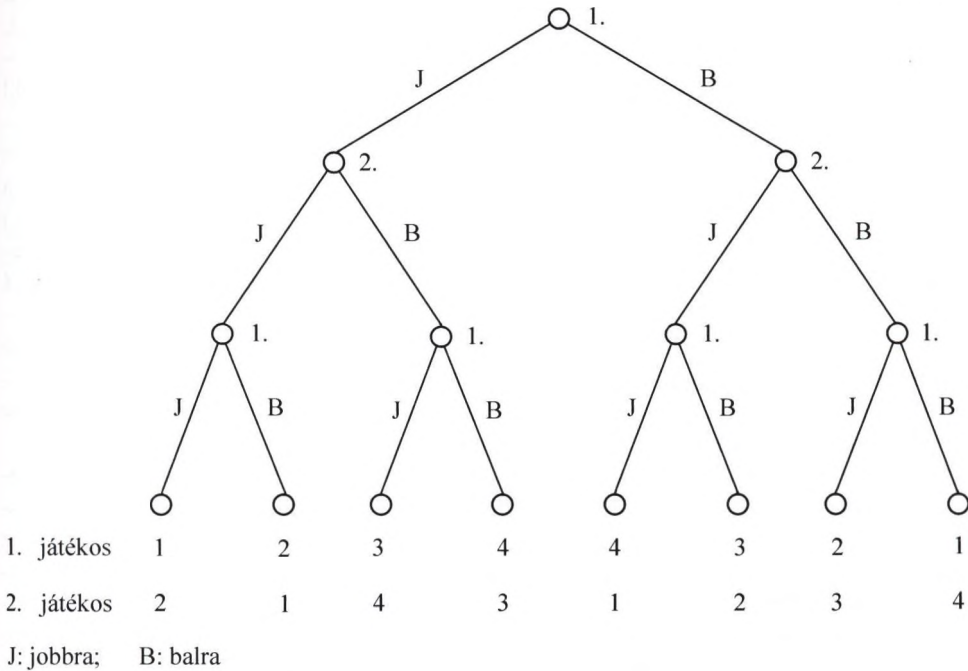
Az előbbi modellben elég természetesen adódott, hogy az egyes játékosoknak milyen döntéseket kellett hozniuk és aztán, miután mindenki döntött a saját termelési szintjéről, mi lesz ennek a következménye, kinek mekkora nyeresége keletkezett. Vannak olyan helyzetek, ahol a lehetséges cselekvések következményei nem ilyen egyértelműek. Ezért feltételezzük, hogy minden játékosnak van egy *hasznossági függvénye*, amely értékeli számára a játék különböző kimeneteleit, vagyis a játékosok együttes cselekvése eredményeként létrejött helyzetet. Egyelőre csak annyit teszünk fel, hogy a játékosok egyenként racionálisak: mindenki azt szeretné, ha a saját hasznossági függvénye minél nagyobb értéket venne fel. Ennek egyik következménye, hogy ha valamelyik játékos tudja azt, hogy a többiek mit fognak tenni, akkor a saját lehetőségei közül olyat fog választani, amely maximalizálja a saját hasznosságát. Ha nem tudja pontosan, hogy mit fognak cselekedni a többiek, de (objektív vagy szubjektív) valószínűségeket tud a többiek lehetséges cselekvéseikhez rendelni, akkor azt tesszük fel a játékosról, hogy a saját *várható hasznosságát* (az egyes hasznosságoknak a valószínűségekkal súlyozott átlagát) maximalizálja.

A nem kooperatív játékok elmélete

A következőkben először a nem kooperatív játékokkal foglalkozunk kicsit részletesebben. Neumann János tette először világossá az átmenetet a nem kooperatív játékok két *alapvető ábrázolási módja* között. Az egyik az extenzív forma, a másik pedig a stratégiai vagy normál forma. Az *extenzív formában* a lehető legnagyobb részletességgel, szinte forgatókönyvszerűen írjuk le, hogy az illető játékot hogyan játsszák. Az ehhez használt eszközt a gráfelméletből vesszük, és *gyökérrel rendelkező véges fának* nevezzük.

Ezt úgy kell elképzelni, mint egy valódi fát, amelynek gyökereiből *ágak* indulnak ki, majd ezekből újabb ágak, és a fa végén vannak a *levelek*, amelyből már nem indulnak ki újabb ágak.

Egy ilyen fát a csomópontjai és az ágai adnak meg, és az jellemzi, hogy bármely csomópontból el lehet jutni ágak mentén bármely másik csomópontba, és sehol sem alakulnak ki körök, tehát ágak mentén nem tudunk visszajutni a kiinduló pontunkba. Az 1. ábrán látunk egy ilyen fát.



1. ábra

A fa minden csomópontjához hozzárendelünk egy játékost, aki arról dönt, hogy a csomópontból kiinduló ágak közül melyikben menjen tovább a fa levelei felé. Az 1. ábrán lévő játéknál előbb az 1. játékos dönt, hogy jobbra vagy balra megy-e, majd a 2. játékos teszi ugyanezt, végül ismét az 1. játékos jön. Megengedett, hogy egyes csomópontokban ne egy játékos, hanem egy véletlen mechanizmus döntsön arról bizonyos ismert (elvből ismert) valószínűség-eloszlás szerint, hogy melyik ágon menjünk tovább. Ez az eset például a kártyajátékokban, ahol a játék általában keveréssel és osztással kezdődik. A fa minden leveléhez hozzárendelünk n számot (az egyes játékosok hasznosságait, amit a játékelméleti irodalomban kifizetésnek is szokás nevezni). Az 1. ábrán a fa végpontjaiban két szám látható, az 1. és a 2. játékos kifizetései.

Egy *játszmának* nevezünk egy ösvényt, amely a fa gyökerétől egy levélhez vezet, ahol a kifizetések megtörténnek. Ezt a játékot a teljes és tökéletes információ jellemzi. Minden játékos ismeri (legalábbis elvből) az egész fát a kifizetésekkel együtt, és tudja azt, hogy a többiek is ismerik és tudja mindenkiről, hogy a (várható) hasznosságát igyekszik maximalizálni.

Jó példa a fentiekre a *sakkjáték*. A sakkjátékot szabályai egyértelműen meghatározzák. A fa gyökeréhez világos van hozzárendelve, és innen 20-féle ágon indulhat el (minden gyaloggal léphet egyet vagy kettőt, és mindkét húszárral két különböző helyre). A következő 20 csomó-

pont mindegyikéhez sötét van hozzárendelve, aki szintén 20-féle ágon indulhat el. A játék így folytatódik tovább, amíg véget nem ér valamelyik fél győzelmével vagy döntetlennel. Feltételezzük, hogy a végességet kikényszerítő szabályok vannak érvényben (például a háromszori túrkörkép azonnali döntetlen, és ha 50 lépéspáron keresztül nincs ütés vagy gyalogtolás, akkor is automatikusan döntetlen a játszma). Kifizetésnek választhatjuk a szokásos 1 a győzelemért, 1/2 a döntetlenért és 0 a vereségért pontozást. Ezt a fát a lehetőségek csillagászati száma miatt természetesen csak elvben tudjuk felrajzolni. Már két lépés után is 421 pontja van a fának!

Nézzük meg most, hogy hogyan adunk meg egy játékot *stratégiai formában*. Feltételezzük, hogy minden játékosnak van egy *stratégiahalmaza*, vagyis adva van n nem üres halmaz S_1, \dots, S_n . A játékot úgy játsszák, hogy minden játékos, a többiektől függetlenül, választ a saját stratégiahalmazából egy elemet. Ennek eredményeképpen létrejön egy $s = (s_1, \dots, s_n)$ *stratégiaprofil*. Ugyancsak feltesszük, hogy minden játékosnak van egy hasznossági (kifizető) függvénye f_1, \dots, f_n , amely minden s stratégiaprofilhoz hozzárendel egy kifizetést. A *stratégiahalmazok és kifizetőfüggvények együttesét nevezzük a játék stratégiai (normál) formájának*. Mint jeleztük, Neumann János fogalmazta meg először világosan az átmenetet az extenzív forma és a stratégiai forma között. A fával ábrázolt extenzív formában adott játékban az i -edik játékos egy *stratégiája egy teljes magatartásterv*, amely megmondja azt, hogy amikor az i -edik játékosnak kell lépnie, akkor melyik ágon fog továbbhaladni.

Ez még olyan szituációkra is vonatkozik, amelyek az i -edik játékos korábbi lépései miatt nem fordulhatnak elő. A sakkjátékban például a világos egy stratégiája egy utasításrendszer, amely megmondja azt, hogy amikor a világosnak kell lépni, mit fog tenni. Ha ez az utasításrendszer egy esetben is valami mást ír elő, akkor az egy másik stratégia. Ugyanez vonatkozik a sötétre is. Nyilvánvaló, hogy noha nagyon sok stratégiája van mindkét játékosnak, ezek száma véges. A játékot most úgy lehet lejátszani, hogy mindkét játékos, egymástól függetlenül, választ egy-egy stratégiát, amely egyértelműen meghatározza a játszmát és ennek kimenetelét is. A játék lejátszása most egy olyan feladattá válik, ami még a játék lejátszása előtt eldönti a játék kimenetelét: igazi intellektuális feladat a stratégia kiválasztása, ami utána következik, csak a bábuk tologatása, és ezt a játékos helyett egy megbízott is csinálhatja. Nem kell mondani, hogy a valóságos versenygyakorlatban a stratégia nem ezt jelenti, és természetesen a megbízottak nem tudják a játszmákat lejátszani a nagymesterek helyett. Ennek az az oka, hogy a stratégiahalmazok elemszáma óriási, és nemcsak az ember, hanem a gép sem tudja valamennyit számon tartani.

Az 1. ábra játékában egy lehetséges stratégiapáros az alábbi magatartásterv:

Az 1. játékos előbb J-t választja, majd ha a 2. játékos J-t választotta, akkor B-t, ha pedig B-t választotta a 2. játékos, akkor J-t. Ha az 1. játékos B-t választaná (ezt ezen stratégia szerint nem fogja megtenni, de ha mégis ez történné), akkor a 2. játékos minden választására az 1. játékos B-t felel. A 2. játékos egy stratégiája: ha az 1. játékos J-t választotta, akkor J, ha B-t, akkor B. Láthatjuk, hogy a stratégiák minden lehetséges esetre adnak egy utasítást, és a J–J–B ágak által meghatározott ösvényhez (játszmához) vezetnek, ami az 1. játékosnak 2, a 2. játékosnak 1 kifizetést eredményez.

Mi az előnye a stratégiai formának? Elsősorban a matematikai kezelhetősége, valamint az, hogy olyan játékok is definiálhatók, amelyek extenzív formában soha nem jelennek meg, illetve csak igen mesterséges módon ábrázolhatók extenzív formában. Neumann Jánosnak is az volt a célja, hogy a normál formában felírt játékok elemzéséhez a matematikai legerőteljesebb eszközeit használja fel.

A kétszemélyes, zérusösszegű, véges játékok egyensúlya

Az 1920-as években egy speciális játék kötötte le a matematikusok egy részének érdeklődését. Bizonyos értelemben a legegyszerűbbnek tekinthető esetet, a kétszemélyes, zérusösszegű, véges játékokat elemezték: a játékosok száma 2, a stratégiahalmazok végesek, és a játékosok kifizetései összege 0. Ez utóbbi feltételezi a teljes antagonizmust: amit az egyik játékos nyer, azt a másiktól nyeri, az érdekek teljesen ellentétesek. Így elég is csak például az első játékos kifizetéseit megadni, a másodiké automatikusan adódik, mint ennek a -1 -szerese. Minthogy a stratégiák száma véges, ezért az egész játékot (normál formában) megadhatjuk egy A számtáblázattal (szokásos nevén egy mátrixszal), amelynek m sora van, ha az 1. játékosnak m stratégiája van és n oszlopa, ha a 2. játékosnak n stratégiája van. A táblázat i -edik sorának j -edik eleme a_{ij} azt a kifizetést mutatja, amit az 1. játékos kap a 2. játékostól, ha ő az i -edik, ellenfele pedig a j -edik stratégiáját követi. Példaként nézzük az alábbi táblázatot, ahol az 1. játékos F és A közül választhat, a 2. játékos pedig B és J közül.

		2. játékos	
		B	J
1. játékos	F	1	2
	A	3	4

Nézzük, hogyan gondolkozhatnak a racionálisnak feltételezett játékosok. Képzeljük magunkat az 1. játékos helyébe. Ha F -et választja, akkor 1 kifizetésre számíthat, hiszen a 2. játékos a B stratégiáját választja ez ellen (ez az érvelés még meggyőzőbb, ha sokszor játsszák a játékot). Ha A -t választja, akkor 3 kifizetésre számíthat. Érdekes ezért az A stratégiát választani, ami számára mindenféleképpen biztosít 3-at, történjék bármi. Az 1. játékos ún. *maximin stratégiát* választott, mivel a választott stratégia maximalizálja a minimális kifizetéseket.

Ugyanezt az okoskodást meg lehet ismételni a 2. játékosra, figyelembe véve azt, hogy számára ezek a kifizetések veszteségeket jelentenek. Ha a B oszlopot választja, akkor számíthat arra, hogy 3-at veszít, míg ha a J oszlopot, akkor 4-et, így érdemes a B -t választania. Ez a stratégia egy *minimax stratégia*, mivel a maximális veszteségeket minimalizálja. Megfigyelhetjük, hogy végeredményben ez a stratégia is a 3 kifizetést eredményezte. Ez a 3 a fenti táblázatban egy *nyeregpontra*, amely a saját oszlopában *maximális*, míg a saját sorában *minimális*. Az adott ésszerűségi feltételek mellett ez a megoldás jól jelzi előre a játékosok viselkedését és a játék kimenetelét. A kialakuló nyeregponthoz tartozó stratégiákat *egyensúlyi megoldásnak* nevezzük. Nem mindig ilyen egyszerű azonban a helyzet.

Nézzük a következő példát. Egy játékos készülődik a tizenegyesrúgás elvégzéséhez, a kapus pedig a kivédéséhez. Közismert, hogy a kapusnak akkor van a legtöbb esélye a hártásra, ha a rúgás pillanatában elhatározza, hogy merre mozdul el. A jó lövéshez is el kell határozni, hogy merre rúgja a játékos a labdát. Az egyszerűség kedvéért tegyük fel, hogy a rúgó játékosnak három stratégiája van: **J**obbra, **K**özépre vagy **B**alra rúgja a büntetőt. A kapusnak is három lehetősége van: **J**obbra vagy **B**alra mozdul, vagy **K**özépen marad. (A **K** stratégiába tartozik az is, amikor csak akkor mozdul el, amikor már látja, hogy merre megy a labda.) Kifizetésnek vegyük azt, hogy adott stratégiapáros mellett 10 büntetőből hány gól

lesz. Az alábbi táblázat mutatja a kifizetéseket (a számok nem objektív statisztikán nyugszanak, de nem is teljesen légből kapottak):

		Kapus		
		J	K	B
Rúgó	J	5	8	9
	K	8	3	8
	B	9	8	5

Nyilván a Rúgó a maximalizáló, a Kapus a minimalizáló játékos. Ha az előbbi logikát követjük, akkor a Rúgó maximin stratégiája vagy J, vagy B, ami legalább 50%-os sikert biztosít neki. A kapus minimax stratégiája K, és ekkor számíthat arra, hogy a büntetők 80%-ánál több nem megy be. Ezek a *stratégiaprofilok nem stabilak*. Ha sokszor ismétlődik ez az eset, akkor nem érdemes mindig jobbra löni a büntetőt, mert a Kapus, ezt látván, felkészül erre. Ebben a mátrixban nincs nyeregpont, és így az a fajta egyensúly sem létezik, ami az előző példában megvolt.

Hogyan kezeljük ezt a szituációt? Itt is segít a megfigyelés. Aki néz labdarúgó mérkőzéseket, az láthatja, hogy mind a Rúgó, mind a Kapus igyekszik bizonytalanságban tartani a másikat és „keverni” a stratégiáit. *Keverésen* itt azt értjük, hogy bizonyos valószínűséggel játssza az egyes stratégiáit. Feltételezve, hogy a másik is ezt teszi, egy új, származtatott játékhoz jutunk, amelyet *kevert bővítésnek* nevezünk. Ha ezt a játékot meg szeretnénk adni normál formában, akkor kifizetőfüggvényeket is kell definiálnunk. Természetesnek tűnik kifizetésnek a várható kifizetéseket tekinteni, ahol az eredeti játék a_{ij} kifizetését az x_i és y_j valószínűségekkel szorozzuk meg, majd összegezzük minden i -re és j -re. Itt x_i jelöli annak a valószínűségét, hogy az 1. játékos az i -edik stratégiáját játssza, míg y_j annak a valószínűségét, hogy a 2. játékos a j -edik stratégiáját játssza. A két valószínűséget azért szoroztuk össze, mert a játékosok egymástól függetlenül választották meg a stratégiáikat. Így a *kevert bővítés* egy olyan játék, ahol a játékosok stratégiahalmazai most már az összes lehetséges valószínűség-eloszlások, a kifizetőfüggvény pedig a kifizetések várható értéke. Az eredeti stratégiák a kevert bővítésben tovább élnek speciális esetként (1 valószínűséggel játszunk egy eredeti stratégiát, és 0-val az összes többi), és *tiszta stratégiáknak* nevezzük őket.

Mint láttuk, csak tiszta stratégiákat tekintve nem jön létre egyensúly, vagyis nem fog megegyezni a maximin érték a minimax értékkel. Vajon mi a helyzet kevert stratégiák esetén? A mátrixjátékok esetében a híres francia matematikus, *Emile Borel* ismerte fel a tiszta és kevert stratégiák jelentőségét 1924-ben, és megpróbálta a maxmin = minmax egyenlőséget bizonyítani a kevert bővítésre. Miután ez nem sikerült neki, elkezdett kételkedni abban, hogy vajon ez az egyenlőség egyáltalán igaz-e. Itt kapcsolódott be 1928-ban Neumann János, aki zseniális matematikusként megtalálta azt az eszközt, egy új fixponttételt, amelynek segítségével bebizonyította a mátrixjátékok kevert bővítésének alapvető tételét, amit azóta is *Neumann-tételnek*, sokszor egyszerűen csak *minimax tételnek* neveznek.

Meg kell jegyezni, hogy Neumann János egy általánosabb tételt bizonyított, és tételét a későbbiekben a legkülönbözőbb irányokban tovább általánosították (lásd *Hegedűs és Zalai*, 1978). A matematika egy egészen új területévé fejlődött a minimax tételek vizsgálata, és igen sok tudományágban lehet alkalmazni a játékelméleten kívül is. A későbbiekben derült ki, hogy a minimax tétel bizonyításához egyszerűbb eszközök is elégségesek, nem kell

fixponttételeket használni, de számos általánosításban a fixponttételek alkalmazása elengedhetetlen.

Térjünk vissza a tizenegyesrúgó játékhoz. Ennek a játéknak a kevert bővítésében a Rúgó játékos maximin stratégiája (az alkalmazott számítási eljárást itt nem részletezzük):

0,416 valószínűséggel jobbra rúgni

0,168 valószínűséggel középre

0,416 valószínűséggel balra rúgni,

a Kapus minimax stratégiája:

0,416 valószínűséggel jobbra mozdulni

0,168 valószínűséggel középen maradni

0,416 valószínűséggel balra mozdulni.

Ennek eredményeként annak a valószínűsége, hogy egy büntetőből gól lesz: 0,717.

2. Az egyensúlyi gazdasági növekedés Neumann-féle modellje

Egy rövid bevezető a modern közgazdaságtanba

Kezdjük egy közkeletű, mondhatni banális megállapítással: az ember társas, társadalmi lény (*homo politicus*). Lét- és fajfenntartását összetett – biológiailag, társadalmilag és történetileg meghatározott – tényezők alakítják, irányítják. *Tevékenységei, döntései* részben ösztönösek, részben hagyományokat-mintákat követnek, részben tudatosak. Az európai civilizációban, különösen a piacgazdaság kibontakozása nyomán, ha korlátozott módon is, de fokozatosan előtérbe került a *racionalitás*. A modern közgazdaságtan, amelynek a kezdetét a 18. századtól, leginkább Adam Smith neves művének (*The Wealth of Nations*) megjelenésétől számíthatjuk, ezért erőteljesen épít a *racionalitás* fogalmára, a sokat vitatott *homo oeconomicus* (az okosan gazdálkodó ember) feltételezésére.

Az emberi *tevékenységek* különböző – biológiai, pszichológiai, társadalmi – *szükségletek kielégítésére* irányulnak. A szükségletek kielégítéséhez emberi munkára, valamint anyagi és szellemi javakra van szükség. Az anyagi javakat az ember eredendően a természetből nyeri, de céltudatos munkával fokozatosan átalakítja azokat, hogy minél alkalmasabbak legyenek az emberi szükségletek kielégítésére. Azokat a *tevékenységeket*, amelyek a különböző *javak* céltudatos átalakítására, végső soron emberi szükségletek kielégítésére (*fogyasztásra*) való alkalmassá tételére (beleértve a felhasználókhöz való eljuttatását) irányulnak *termelő- vagy szolgáltatótevékenységeknek* nevezzük. (Gyakran csak termelést mondunk, de mindkettőre gondolunk.)

Mind az emberi munkaerő, mind a természeti erőforrások, következképpen az általuk előállítható egyéb *javak* is *viszonylag szűkös* mennyiségben állnak az emberek, a társadalom

rendelkezésére, ezért *takarékoskodni, gazdálkodni* kell velük. A *viszonylagos* jelző magyarázatra szorul, mivel több mindenre utalunk vele. Viszonylagos a rendelkezésre álló javak szükségossége a kielégíthető szükségletek tekintetében, relatív az egyes javak szükségossége egymáshoz képest, s viszonylagos a különböző javak és erőforrások szükségossége az időtől, a műszaki-szervezési lehetőségektől függően is.

A *javak szükségossége* a modern közgazdaságtan egyik sarkalatos alapfeltevése és alapfogalma. Egyes uralkodó közgazdasági irányzatok egyenesen úgy definiálják a közgazdaságtant, mint „*az alternatív módon felhasználható szűkös erőforrásokkal való ésszerű gazdálkodás*” tudományát. A gazdasági döntések és cselekvések azonban, a változó társadalmi közegebe beágyazva jelennek meg, ezért jelentős mértékben függenek a *történeti, társadalmi* viszonyoktól is. A közgazdászok fokozatosan törekedtek arra, hogy – gyakran nehezen elfogadható egyszerűsítések árán – elhatárolják a közgazdaságtan tárgyát a más tudományok által tanulmányozott jelenségektől. Ez az egyik magyarázata annak, hogy általánossá vált az egymással rendszerint ideológiai ellentétben álló iskolák párhuzamos jelenléte. Ennek mélyebb okaiba itt nem kívánunk elmélyedni, de szükségesnek tartottuk megjegyezni.

A modern közgazdaságtan a piacgazdaságok működésének *idealizált alapelveit* kellően precíz fogalmakkal és összefüggésekkel igyekeznek leírni, a matematikában és a természettudományokban megszokott *axiomatikus* kifejtést alkalmazva. Ez egyidejűleg erénye és hátránya ennek a megközelítésnek: segíti a megértést, de ugyanakkor lényeges problémákat eltakar. A továbbiakban fő kontúrjaiban felrajzoljuk azt a gazdaságképet, és jelezzük azokat a főbb jelenségeket, amelyekkel a közgazdaságtan általában foglalkozik.

A közgazdaságtan egyik alapelve, hogy „minden mindennel összefügg”. Ezért van némi szabadság abban, honnan kezdjük kifejteni vizsgálati tárgyát. Legtöbbször a *szükségletekkel*, az emberi tevékenységek végső okaival és mozgatórugóival kezdjük a tárgyalást. A közgazdaságtan nem sokat foglalkozik a szükségletekkel önmagukban és általában. A szükségletek csak annyiban érdekesek, amennyiben azok a *gazdasági javakat hasznossá és értékesé* teszik az emberek számára. Az uralkodó gyakorlat szerint a közgazdászok a *gazdasági tevékenységek* két alapvető csoportját különböztetik meg: a *fogyasztást* és a *termelést*. Az első csoportba tartozó tevékenységek végzésének színterei a *háztartások*, a másodiké a *vállalatok*. A *fogyasztás közvetlenül* valamilyen emberi *szükséglet kielégítésére* irányul, a *termelés* célja az egyéni vagy társas *jövedelemszerzés*. A megszerzett jövedelmek azonban végső soron a háztartásokhoz kerülnek, és felhasználásukról ott döntenek.

Az egyéni és/vagy kollektív (kormányzati) *háztartások*, az adott ár és egyéb piaci feltételek mellett, döntenek egyrészt arról, hogy a rendelkezésükre álló erőforrásokból (munka, tőkeként felhalmozott termékkészletek, föld stb.) mennyit bocsátanak a vállalatok rendelkezésére (*a termelési tényezők kínálata*). Másrészt arról, hogy a megszerzett jövedelmekből mennyit fordítanak jelenbeli és mennyit jövőbeli szükségleteik kielégítésére: *fogyasztásra* és *megtakarításra*. Egyensúly esetén az utóbbit teljes egészében a termelőkapacitások bővítésére, *felhalmozásra* (beruházásra) fordítják. A fogyasztás és felhalmozás együtt generálja a *termékek végső keresletét*, ami viszont a termékek termelő felhasználásán keresztül további, *közbenső keresletet* indukál.

Ez utóbbit már a *vállalatok* vezetői határozzák meg, akik egyidejűleg eldöntik, hogy milyen termékeket állítanak elő (*termékek kínálata*), és hogy ezek megtermeléséhez mennyi erőforrást vesznek igénybe (*a termelési tényezők kereslete*). Az áruk cseréjét az anonim *piac* közvetíti, ahol minden jószágnak minden meghatározott időszakban, a vevőtől függetlenül,

egy és ugyanaz az ára. Ha egy adott időszakban a kereslet és kínálat minden jószág piacán megegyezik egymással, akkor azt mondjuk, hogy a gazdaság *általános egyensúlyban* van.

Összefoglalva és kiegészítve a leírtakat: a modern közgazdaságtan a gazdaságot egymástól független, önálló döntési jogokkal rendelkező *gazdasági egységek* együtteseként ábrázolja. Ezek a gazdasági egységek különböző *gazdasági javakat* (árúkat) használnak fel, illetve állítanak elő, és egymással *árukereslet* folytatnak, éspedig a *piac* közvetítésével. A kereslet az *árak* szabályozzák, ezek határozzák meg a *cserearányokat*. A *vállalatok* döntési lehetőségeit alapvetően korlátozzák a *technológiai* (műszaki-szervezési) *lehetőségek*, s a technológia által megszabott lehetőségek közül azt a megoldást választják, amely az adott árak mellett számukra a legnagyobb *nyereséget* nyújtja. A keletkező *jövedelmek elosztását* meghatározott szabályok irányítják, amelyek révén kialakul a háztartások rendelkezésére álló *jövedelem*. A háztartások eldöntik, hogy mennyit fordítanak a jelen- és jövőbeli szükségleteik kielégítésére. A gazdaság hatékony működése megköveteli, hogy a különböző gazdasági egységek döntései, az egyes áruk kínálata és kereslete egymással összhangban legyen. Az összhang ideális esete az *egyensúly*, amikor is a különböző áruk kínálata és kereslete megegyezik egymással. Az uralkodó felfogás feltételezi, hogy a piac működési mechanizmusa (a *kereslet-kínálat törvénye*) képes a gazdaságot ebbe a kitüntetett, egyensúlyi állapotba vinni, illetve annak közelében tartani.

Neumann általános gazdasági egyensúlyi modellje

Kezdjük mindenképp a modell egyszerű és könnyen áttekinthető gazdaságképének felvázolásával. Hasznos lehet az *1. táblázat* megértése. Az első kimutatás egy gazdaság adott időszakban megfigyelt *termelését* tartalmazza, termékenként és termelési tevékenységekként elkülönítve (Y_{ij}). A második az ugyanabban az időszakban felmerült *ráfordításokat* részletezi azonos felbontásban (X_{ij}), ahol i a termékek, j a termelési tevékenységek (folyamatok) indexe.

1. táblázat: A gazdasági tevékenységek eredményének ábrázolása Neumann modelljében

Kibocsátások

	1. folyamat	2. folyamat	...	j . folyamat	...	m . folyamat	egységárak
1. termék	Y_{11}	Y_{12}		Y_{1j}		Y_{1m}	p_1
2. termék	Y_{21}	Y_{22}		Y_{2j}		Y_{2m}	p_2
⋮							
i . termék	Y_{i1}	Y_{i2}		Y_{ij}		Y_{im}	p_i
⋮							
n . termék	Y_{n1}	Y_{n2}		Y_{nj}		Y_{nm}	p_n
a tevékenységek szintje	x_1	x_2		x_j		x_m	

Ráfordítások

	1. folyamat	2. folyamat	...	j. folyamat	...	m. folyamat	egységárak
1. termék	X_{11}	Y_{12}		X_{1j}		X_{1m}	p_1
2. termék	X_{21}	X_{22}		X_{2j}		X_{2m}	p_2
⋮							
i. termék	X_{i1}	X_{i2}		X_{ij}		X_{im}	p_i
⋮							
n. termék	X_{n1}	X_{n2}		X_{nj}		X_{nm}	p_n
a tevékenység szintje	x_1	x_2		x_j		x_m	

Neumann feltételezi, hogy az X_{ij} elemekben megjelenő ráfordítások tartalmazzák a háztartások létfenntartásához szükséges fogyasztást is, és hogy a kibocsátások mindig csak a következő időszakban használhatók fel (*éves megtérülés*). Felteszi továbbá („hogya további bonyodalmak elejét vegyük” – 162. o.), hogy a) „a hozam a termeléssel arányos”, és b) „a termelés természetes tényezői, a munkát beleértve” (az elsődleges erőforrások) a termelt termelési tényezőkhöz, a felhalmozott tőkékhez képest, „korlátlan mennyiségben bővíthetők”, azaz hosszabb távon nem képezhetnek szűkös keresztmetszetet.

Neumann egy olyan elvont gazdaságot vizsgált, amelyben feltevése szerint nem változnak a fogyasztási szokások, nincs műszaki haladás, és ezek következtében az egyszer kialakult egyensúlyi termelési szerkezet és árarányok időben változatlanok tekinthetők. Egyensúly esetén tehát a termelés és a termékek felhasználása az egyik időszakra a másikra *egyenletesen*, λ ütemben változik: λ előjelétől függően nő, stagnál vagy csökken. Az ilyen állapotot *stacionárius egyensúlynak* nevezzük.

A feltevéseknek megfelelően a vizsgált gazdaságban a *termékek keresletének és kínálatának egyensúlyi feltételei* az alábbi egyszerű formát öltik:

$$Y_{i1} + Y_{i2} + \dots + Y_{im} = (1 + \lambda) \cdot (X_{i1} + X_{i2} + \dots + X_{im}), \quad i = 1, 2, \dots, n. \quad (1)$$

Az időben változatlan *egyensúlyi árak* (p_i , $i = 1, 2, \dots, n$) pedig olyanok, amelyek esetén az egységnyi értékű tőkebefektetésre minden tevékenység esetében azonos mértékű, π *megtérülés* (kamat vagy profit) jut. Az ilyen árakat (voltaképpen, mint láthatjuk, csak *árarányokat*) az alábbi egyenletekkel definiálhatjuk:

$$p_1 \cdot Y_{1j} + p_2 \cdot Y_{2j} + \dots + p_n \cdot Y_{nj} = (1 + \pi) \cdot (p_1 \cdot X_{1j} + p_2 \cdot X_{2j} + \dots + p_n \cdot X_{nj}), \quad j = 1, 2, \dots, m. \quad (2)$$

A *tevékenység szintek* (x_j), valamint egységnyi tevékenységszintre vonatkozó, állandónak feltételezett *kibocsátási* ($k_{ij} = Y_{ij}/x_j$) és *ráfordítási* ($r_{ij} = X_{ij}/x_j$) *együtthatók* bevezetésével a fenti összefüggések egyenértékűen átírhatók az alábbi szimmetrikus formákba (a két összefüggés formái szimmetriáját *dualitásnak* nevezzük):

$$k_{i1} \cdot x_1 + k_{i2} \cdot x_2 + \dots + k_{im} \cdot x_m = (1 + \lambda) \cdot (r_{i1} \cdot x_1 + r_{i2} \cdot x_2 + \dots + r_{im} \cdot x_m), \quad i = 1, 2, \dots, n. \quad (3)$$

$$p_1 \cdot k_{1j} + p_2 \cdot k_{2j} + \dots + p_n \cdot k_{nj} = (1 + \pi) \cdot (p_1 \cdot r_{1j} + p_2 \cdot r_{2j} + \dots + p_n \cdot r_{nj}), \quad j = 1, 2, \dots, m. \quad (4)$$

A felírt egyensúlyi feltételekben nincs semmi olyan új közgazdasági tartalom, ami a közgazdászoknál korábban már meg nem jelent volna. Elsősorban az elemzés módszertana volt új, több szempontból is. Így például Neumann a termelési lehetőségek ábrázolásakor feltette, hogy a technológia tartalmazza egyrészt az *ikertermelés* lehetőségét (egy-egy tevékenységgel egyidejűleg több terméket is elő lehet állítani), másrészt *technológiai választékot* (egy-egy terméket egyidejűleg több tevékenységgel is elő lehet állítani). Ez újdonság volt, mert – mint látni fogjuk – a korabeli modellek sokkal egyszerűbben ábrázolták a gazdasági javakat és a technológiai lehetőségeket.

Neumann gondosan kerülte még a látszatát is annak, hogy modelljét a gazdaságot teljesszerte leíró elméletként értelmezze. A modern árutermelés néhány kiemelt jelenségére (az árutermelés „körköröségére”, a technológiák közötti választás és az egyensúlyi árrendszerek meghatározottságára) összpontosítja figyelmét. Az alkalmazott absztrakciók kapcsán érdemes idézni magát Neumannt:

„Hogy teljesen szabadon tárgyalhassuk (a fenti kérdéseket – a szerzők), idealizálni fogjuk a helyzet többi elemét... Ennek az idealizálásnak legnagyobb része irreleváns, de ezt a kérdést itt nem tárgyaljuk” (160–161. o.) – tette abszolút világossá álláspontját. Máshol (102. o.): „minden tudomány így indult, és a közgazdaságtan mint tudomány csak néhány száz éves”, vizsgálatának tárgya pedig olyannyira bonyolult és összetett, hogy „még igen sok kutatásra van szükségünk, hogy kifejlesszük a lényeges koncepciókat – a valóban használható eszméket”.

Mivel Neumann a gazdaságot jellemző ráfordítási és kibocsátási *együtthatókat állandóknak* tekintette, ezért szembesülnie kellett egy problémával: nem várható el ugyanis általában, tetszőleges nem negatív együtthatók mellett, hogy legyen olyan tevékenységi szerkezet, amely esetén a kibocsátások és a ráfordítások szerkezete egybeesik, és így a (3) egyenlőségek egyidejűleg teljesülnek minden termék esetén. Hasonló a helyzet az egyensúlyi árakat meghatározó (4) egyenletek tekintetében, amelyek a különböző tevékenységek esetében a kibocsátások és a ráfordítások értékének (a bevételeknek és a költségeknek) azonos arányát feltételezik.

Neumann ezért lazított az egyenlőségi feltételeken, és helyettük alkalmas irányú gyenge egyenlőtlenségeket vezetett be, amelyekkel megengedte, hogy egyensúly esetén is lehessenek olyan termékek, amelyek kínálata meghaladja keresletüket:

$$\sum_j k_{ij} \cdot x_j \geq \sum_j r_{ij} \cdot x_j, \quad i = 1, 2, \dots, n, \quad (5/a)$$

illetve olyan tevékenységek, amelyek tökemegtérülési rátája nem éri el az egyensúlyi árak által lehetővé tett legnagyobb rátát (π):

$$\sum_i p_i \cdot k_{ij} \leq (1 + \pi) \cdot \sum_i p_i \cdot r_{ij}, \quad j = 1, 2, \dots, m. \quad (6/a)$$

Ha viszont feltesszük, hogy egyes termékek többletkínálata nem okoz zavart a gazdaságban (az ún. „*díjmentes lomtalanítás*” feltevése), akkor nemcsak a lehetséges tevékenység-szintek, de az árak esetében is elő lehet írni, hogy *csak nem negatív értékeket* vehetnek fel ($x_j, p_i \geq 0$). Negatív ár ugyanis azt jelentené, hogy a termelő hajlandó lenne fizetni azért, hogy az előállított termékétől megszabaduljon, amire díjmentes lomtalanítás lehetősége miatt nincs szüksége.

A modell közgazdasági értelmezése ugyanakkor megkívánja, hogy a fenti feltételeket kiegészítsük az alábbi megkötésekkel:

$$p_i \cdot \sum_j k_{ij} \cdot x_j = (1 + \lambda) \cdot p_i \cdot \sum_j r_{ij} \cdot x_j, \quad i = 1, 2, \dots, n, \quad (5/k)$$

azaz, az olyan termék ára, amelynek a kínálata egyensúlyban is meghaladja a keresletét, szükségképpen nulla kell, hogy legyen (szabad jószág), illetve

$$x_j \cdot \sum_i p_i \cdot k_{ij} = (1 + \pi) \cdot x_j \cdot \sum_i p_i \cdot r_{ij}, \quad j = 1, 2, \dots, m, \quad (6/k)$$

az olyan tevékenységeket, amelyek tőke megtérülési rátája nem az adott árak mellett elérhető lehető legnagyobb ráta, azokat egyensúlyban nem használják.

Ez a megoldás nem más, mint a matematikai programozás, illetve a matematikai közgazdaságtan irodalmából ma már jól ismert *komplementaritási elv* alkalmazása. Ennek szükségességére egyébként a Walras–Cassel-modell kapcsán – nagyjából Neumann-nal egy időben és tőle függetlenül – mások is (például *Zeuthen*, 1933 és *Schlesinger*, 1935) felhívták a figyelmet, amelyre még később visszatérünk. Mint látni fogjuk, Neumann egy olyan feltevést alkalmazott, ami miatt nála a komplementaritási elv követelményei automatikusan teljesültek, ezért az (5/k) és a (6/k) egyenlőségek nem jelennek meg modelljének végző leírásában.

Vegyük észre, hogy valahányszor az (5/a), illetve a (6/a) egyenlőtlenségek teljesülnek az x_j , illetve a p_i változók valamely adott értéke mellett, akkor teljesülni fognak akkor is, ha a fenti változókat rendre megszorozzuk ugyanazzal a pozitív számmal. Az egyensúlyi feltételek tehát *csak a változók arányait* határozzák meg, a szintjüket nem. A triviális, azaz az olyan megoldásokat, amelyekben az x_j vagy a p_i változók mindegyike nulla, eleve kizárjuk a vizsgálatból. Mindezek miatt az egyes változó csoportok szintjét tetszőlegesen megkötjük, feltehetjük például, hogy *összegük 1*, azaz

$$\sum_j x_j = \sum_i p_i = 1.$$

Szorozzuk be rendre az (5/a) és a (6/a) egyenlőtlenségek mindkét oldalát a hozzájuk tartozó p_i és x_j (kiegészítő) változókkal, és adjuk össze őket külön-külön a két egyenlőtlenségcsoport esetében. A kapott két összevont egyenlőtlenségből egyszerű átalakítások után az alábbi összefüggést kapjuk:

$$(1 + \lambda) \cdot \sum_{ij} p_i \cdot r_{ij} \cdot x_j \leq \sum_{ij} p_i \cdot k_{ij} \cdot x_j \leq (1 + \pi) \cdot \sum_{ij} p_i \cdot r_{ij} \cdot x_j. \quad (7)$$

Ebből nyilvánvalóan adódik két fontos megfigyelés. Az *egyik*: ha az előállított összes termék értéke pozitív, azaz $\sum_{ij} p_i \cdot k_{ij} \cdot x_j > 0$, amit a közgazdasági szempontból értelmes megoldásoktól joggal elvárunk, akkor $\lambda = \pi$, azaz egyensúlyban a *növekedési ütem és a tőke megtérülési ráta* szükségképpen *megegyezik egymással*. (Ez az összefüggés nyilván csak akkor állhat fenn, ha minden megtermelt többletet felhalmozznak a gazdaságban. Általánosabban: a növekedési ütem – ami itt nem más, mint a felhalmozási ráta – nem lehet nagyobb, mint a megtérülési ráta, $\lambda \leq \pi$.) A *másik* lényeges következtetés: ha $\lambda = \pi$, akkor automatikusan teljesülnek a *komplementaritási elv* követelményei.

Neumann feltette, hogy $k_{ij} + r_{ij} > 0$ minden i és j esetén, azaz minden termék megjelenik minden tevékenységben vagy kibocsátásként, vagy ráfordításként. Ez a csak matematikai alapon védhető feltevés (tegyünk a nullák helyébe igen kis pozitív számokat, mint Neumann érvelt) nemcsak azt eredményezte, hogy az egyensúlyi megoldásban az előállított

összes termék értéke szükségképpen pozitív, és ezért $\lambda^* = \pi^*$, ami miatt viszont automatikusan teljesülnek a *komplementaritási* feltételek, hanem azt is, hogy *csak egy egyensúlyi tényező létezik*.

A fenti kiegészítő megjegyzések után most már megadhatjuk pontosan annak a modellnek a feltételeit, amely keretében Neumann János először adott precíz bizonyítást az általános gazdasági egyensúly létezésére:

$$x_j, p_i \geq 0, \quad \sum_j x_j = \sum_i p_i = 1, \quad \alpha > 0,$$

$$\sum_j k_{ij} x_j \geq \alpha \cdot \sum_j r_{ij} x_j, \quad \text{minden } i\text{-re} \quad (5/a)$$

$$\sum_i p_i \cdot k_{ij} \leq \alpha \cdot \sum_i p_i \cdot r_{ij}, \quad \text{minden } j\text{-re} \quad (6/a)$$

ahol $k_{ij}, r_{ij} \geq 0$ és $k_{ij} + r_{ij} > 0$.

Megjegyzések és általánosítások

1) A megoldás *unicitása* az egyensúlyi növekedési és tőke megtérülési tényező tekintetében (mint maga is kiemeli a 165. oldal alján) fontos szempont volt Neumann számára. A két tényező közös egyensúlyi értéke ugyanis egyik oldalról nem más, mint az adott termelési-felhasználási együtthatók által lehetővé tett *legnagyobb növekedési ráta*, másik oldalról pedig, a fenti együtthatók és a lehetséges árak mellett adódó legkisebb *megtérülési ráta*:

$$\lambda^* = \max \left\{ \lambda: \text{van olyan } x_j \geq 0, \sum_j x_j = 1, \text{ hogy } \sum_j k_{ij} x_j \geq (1 + \lambda) \cdot \sum_j r_{ij} x_j, \text{ minden } i\text{-re} \right\},$$

$$\pi^* = \min \left\{ \pi: \text{van olyan } p_i \geq 0, \sum_i p_i = 1, \text{ hogy } \sum_i p_i \cdot k_{ij} \leq (1 + \pi) \cdot \sum_i p_i r_{ij}, \text{ minden } j\text{-re} \right\}.$$

Ez egy fontos megfigyelés, mert ez kapcsolja össze a növekedési modell egyensúlyi megoldását a kétszemélyes játékok egyensúlyával. Megjegyezzük még, hogy ha nem vezetünk be olyan feltevést, amely szavatolja, hogy a fenti tényezőket tekintve csak egy egyensúlyi megoldás létezik, akkor általában csak azt tudjuk igazolni, hogy λ^* is lehetséges megtérülési ráta, azaz $\lambda^* \geq \pi^*$.

2) A fentiekből egy újabb fontos megállapítás következik. Definiáljuk az F függvényt az alábbi törtkifejezéssel:

$$F(\mathbf{x}, \mathbf{p}) = F(x_1, x_2, \dots, x_m; p_1, p_2, \dots, p_n) = \frac{\sum_{ij} p_i \cdot k_{ij} \cdot x_j}{\sum_{ij} p_i \cdot r_{ij} \cdot x_j},$$

ahol $x_j, p_i \geq 0, \sum_j x_j = \sum_i p_i = 1, \mathbf{x} = (x_1, x_2, \dots, x_m)$ és $\mathbf{p} = (p_1, p_2, \dots, p_n)$.

Egyszerűen beláthatók az alábbi megállapítások: ha a Neumann-modellnek létezik egyensúlyi megoldása, $(\alpha^*, \mathbf{x}^*, \mathbf{p}^*)$, akkor az \mathbf{x} változók \mathbf{x}^* szinten rögzített értékei mellett az $F(\mathbf{x}, \mathbf{p})$ függvény értéke a \mathbf{p}^* pontban minimális, \mathbf{p}^* szinten rögzített \mathbf{p} értékek esetében viszont az \mathbf{x}^* pontban maximális (és pedig mindkét esetben α^*) értéket vesz fel, azaz

$$F(\mathbf{x}, \mathbf{p}^*) \leq F(\mathbf{x}^*, \mathbf{p}^*) \leq F(\mathbf{x}^*, \mathbf{p}),$$

minden $x_j, p_i \geq 0, \sum_j x_j = \sum_i p_i = 1$ esetén.

A Neumann-modell feltételei nem mások, mint ennek a *minimax*, más szóval *nyeregpon-ti megoldásnak* a szükséges feltételei. Neumann tehát ugyanazt a minimax (nyeregpon-t) megközelítést használta most is, mint amelyet a kétszemélyes zérusösszegű játékok egyen-súlyával foglalkozó dolgozatában alkalmazott. A minimax megoldás létezésének bizonyítá-sát itt tovább általánosította. A kétszemélyes, zérusösszegű játék kifizetőfüggvénye – amelyben az x és p változók elemei a tiszta stratégiák keverési arányai – a fenti $F(x, p)$ függvény egy sajátos esete.

3) Egy további fontos megfigyelés a fentebb bevezetett F függvény kapcsán. Legyen a te-vékenységszintek és az árak egy adott (x, p) értéke mellett a ráfordítások értéke pozitív. Az F függvény értéke ilyen esetben a profittényező nagyságát adja meg. Neumann ezért a fen-ti függvényt *profitfüggvénynek* nevezte, s rámutatott arra, hogy a profitfüggvény formai szempontból analóg a *termodinamika* potenciálfüggvényeivel, és úgy vélte: „...feltehető, hogy a hasonlóság fennáll teljes fenomenológiai általánosságában...” (161. o.).

A klasszikus termodinamika matematikai módszertana Hicks (1939) és Samuelson (1947) alapvető művei nyomán a neoklasszikus közgazdaságtan standard eszközévé vált. Azóta is számos kísérlet történt a termodinamikai fogalmak közgazdaságtani alkalmazásá-ra (lásd bővebben Bródy, Martinás és Sajó, 1986, Bródy, 1986). Neumann Jánosnak később az a véleménye erősödött meg, mint már jeleztük, hogy a közgazdászoknak a klasszikus fi-zikából átvett matematikai eszköztárnál jóval korszerűbb és adekvátabb matematikai mód-szereket kellene felhasználniuk. Minden bizonnyal ez sarkallta a játékelmélet általános el-méletének kidolgozására.

4) Érdemes közelebbről is megvizsgálni a *növekedési ütem* és a *kamatráta* egybeesésének feltételét. Neumann *explicité* feltette, hogy a *szükséges fogyasztáson* felül minden jövedel-met megtakarítanak és visszaforgatnak a termelésbe. A növekedési és a kamattényező egyenlősége nála egyenesen következik ebből a feltevésből. Neumann ugyan kamatról be-szél, de vélhetően úgy gondolkozott, hogy egy „pénzügyi bonyodalmaktól mentes” gazda-ságban az egyensúlyi kamatláb megegyezik az egyensúlyi árak által lehetővé tett legna-gyobb profitrátával. Egy helyütt ezt írja például: „...egyensúlyi helyzetben semelyik eljárás nem lehet nyereséges (máskülönben vagy az árak vagy a kamatláb növekedne – világos, hogy ezt az absztrakciót hogyan kell érteni)” (164. o.). A nyereség az ő értelmezésében ka-matköltségen felül jelentkező többletet jelentett.

A modell lehetséges interpretációi megengedik, hogy a *kamat* helyett a *klasszikus érte-lemben vett profitról* beszéljünk, és a *növekedési ütemet* és a *profitrátát elválasszuk egymástól*. A termelés költségeit meghatározó ráfordítási együtthatók (r_{ij}) ugyanis, mint jeleztük, nemcsak a termelő ráfordításokat, hanem az ún. létszükségleti fogyasztást is tartalmazzák. Ha az utóbbi költségét, a klasszikus közgazdászokhoz hasonlóan, munkabéreként értelmez-zük, akkor π egyértelműen megfeleltethető a profitrátának, a kapott egyensúlyi árak pedig közvetlen rokonságba hozhatók a klasszikusok (mindenekelőtt Ricardo és Marx) által használt, *termelési árak* nevezett fogalommal.

Neumann modelljének fenti átértelmezése és egyszerű kibővítése révén megmutatjuk, hogy nem a növekedési ütem határozza meg a profitrátát, mint egyesek ezt feltételezték, hanem fordítva, a profitráta a potenciális növekedési ütemet. Profit keletkezhet ugyanis akkor is, ha a gazdaság egyszerű újratermelést folytat, azaz a növekedési ütem nulla. Ve-zessük be a *luxusfogyasztás* együtthatóit (f_{ij}), és tekintsük továbbra is adottnak a termelés költségeit meghatározó ráfordítási együtthatókat (r_{ij}). A felhasználási ($r_{ij} + f_{ij}$) és a ráfordí-

tási (r_{ij}) együttthatók most már eltérnek egymástól, és ennek megfelelően megváltoznak a modell alapegyenlőtlenségei:

$$\sum_j k_{ij} \cdot x_j \geq (1 + \lambda) \cdot \sum_j (r_{ij} + f_{ij}) \cdot x_j, \quad (5/b)$$

$$\sum_i p_i \cdot k_{ij} \leq (1 + \pi) \cdot \sum_i p_i \cdot r_{ij}. \quad (6/a)$$

Normális feltételek esetén a luxusfogyasztási együttthatók (f_{ij}) növekedésével csökken a növekedés üteme, s lesz olyan értékük is, amelyek esetén $\lambda = 0$, azaz a gazdaság ún. *egyszerű újratermelést* folytat. A luxusfogyasztási együttthatók növekedése ugyanakkor nem érinti a profitráta nagyságát.

A modell fenti általánosítása megszünteti annak szép, szimmetrikus dualitását, és egyúttal nehezebbé teszi a modell matematikai elemzését is. A Neumann-modell ehhez hasonló aszimmetrikus kiterjesztései matematikai szempontból feltétlenül érdekes és izgalmas feladatot jelentettek, és az 1970-es években több érdekes eredmény is született ezen a területen (*Morishima 1964, Łoś et al., 1974*).

5) Neumann, mint jeleztük, feltette, hogy $k_{ij} + r_{ij} > 0$ minden i és j esetén, azaz minden termék megjelenik minden tevékenységben vagy kibocsátásként, vagy ráfordításként. Ez a feltevés biztosította, hogy létezik egyensúlyi megoldás, és csak egyetlen (közös) egyensúlyi tényező létezik. Azt is jeleztük, hogy a fenti feltevés nehezen védhető közgazdasági szempontból. Ezért is figyelemre méltó a *Kemeny, Morgenstern és Thompson (KMT)* szerzőhármas 1956-os tanulmánya, amelyben egyrészt enyhítették ezt a feltevést, másrészt egy Neumannnál jóval egyszerűbb bizonyítást adtak az egyensúly létezésére (lásd *Zalai, 2000*). Itt jegyezzük meg, hogy *Morgenstern és Thompson (1976)* monografikus igényvel megírt könyve a Neumann-modell különböző általánosítási lehetőségeit mutatja be.

A viszonylag egyszerű és roppant kézenfekvő KMT-feltételek a következők:

$$k_{ij}, r_{ij} \geq 0, \text{ és}$$

$$\sum_i r_{ij} > 0 \text{ minden } j\text{-re, azaz nincs kibocsátás ráfordítás nélkül,}$$

$$\sum_j k_{ij} > 0 \text{ minden } i\text{-re, azaz minden termék termelhető.}$$

Ezek a feltevések azonban nem zárják ki, hogy a lehetséges matematikai megoldások között legyenek közgazdasági szempontból teljesen érdektelen megoldások is, nevezetesen olyanok, amelyekben csak szabad javakat termelnek. Ezért a szóba jöhető megoldások körét a

$$\sum_{ij} p_i \cdot k_{ij} \cdot x_j > 0, \quad (8)$$

pozitív értékű kibocsátás explicit kikötésével eleve korlátozták. Azt is megmutatták, hogy az adott feltevések mellett a modellnek véges számú megoldása van (legfeljebb annyi, mint a tevékenységek, illetve a termékek száma közül a kisebbik), és ezekben az egyensúlyi növekedési ütem és kamatláb nagysága továbbra is megegyezik egymással.

Egy igen lényeges mozzanat, hogy a KMT-feltevések – szemben Neumann feltevésével – megengedték, hogy a modellezett gazdaságban lehessenek a többi ágazattól *függetlenül* is működtethető *ágazatcsoportok (dekomponálható, önálló részekre bontható gazdaság)*, és ennek következtében, eltérő növekedési, illetve kamattényezővel rendelkező alternatív egyensúlyi állapotok is létezhetnek. Ezzel tehát a modell megoldásának a növekedési,

illetve kamattényező tekintetében vett unicitása nincs garantálva. Kiterjedt irodalma keletkezett a dekomponálható (önálló részekre bontható) gazdaságok jellegzetességeivel foglalkozó elemzéseknek, amelyek eleddig inkább matematikai, mint közgazdasági szempontból bizonyultak érdekesnek. Nehezen lehet igazolni a maximális növekedési ütemtől eltérő alternatív megoldások közgazdasági relevanciáját. Talán nem is véletlen, hogy Neumann ezeket eleve kizárta. Gale [1960] ugyanakkor megmutatta, hogy a KMT-feltevés mellett viszonylag egyszerű előzetes (*indekompozabilitási*) feltevésekkel szavatolni lehet, hogy a modell egyensúlyi megoldásai a növekedési, illetve kamattényező tekintetében továbbra is egyértelműen meghatározottak legyenek.

6) Neumann dolgozata nagyon keveset fed fel a modell lehetséges közgazdaságtani háttere és rokonsága tekintetében. Modellje szűkszavú közgazdasági értelmezése közben Neumann kissé talányosan csak annyit jegyzett meg: „Nyilvánvaló, hogy milyen fajta elméleti modellnek felelnek meg a fenti feltevések.” (162. o.) Hogy ez mennyire nem nyilvánvaló, azt csak a későbbi értelmezési viták igazolják.

A kortársak és a barátok visszaemlékezései is nagyon keveset fednek fel a modell lehetséges közgazdaságtani inspirációjáról. Káldor (1989) úgy emlékezett, hogy Neumann egy budapesti találkozásuk alkalmából rövid, összefoglaló jellegű közgazdasági munkák iránt érdeklődött nála, és ő Wicksell, Cassel és Böhm-Bawerk (a kor elismert osztrák és svéd neoklasszikus közgazdászai) műveit ajánlotta figyelmébe. Ugyancsak Káldor visszaemlékezése szerint, Neumann igen kritikusan nyilatkozott a neoklasszikusok módszertanáról, mert az „túl nagy hangsúlyt helyez a helyettesíthetőségre és túl keveset azokra az erőkre, amelyek kölcsönös feltételeket szabnak a növekedés számára” (uo. 162. o.). Arrow (1989) ugyanakkor határozottan azt állítja, hogy „nagyon világosnak látszik, hogy Cassel művéből indult ki” (uo. 17. o.). L. Punzo (1989), a matematikai közgazdaságtan történetének egyik neves művelője viszont úgy gondolja, hogy a bécsi Menger-kör (Schlesinger és Wald) környékén kell keresni a modell közvetlen közgazdasági és módszertani rokonságát. Samuelson (1989) a neoklasszikus rokonságot hangsúlyozta, a jelen tanulmány egyik társszerzője (Zalai, 1999) megmutatta, hogy Neumann modellje sokkal közelebbi és közvetlenebb kapcsolatban van a klasszikus (benne a marxi) hagyományokkal, mint a modern neoklasszikus irányzattal.

Neumann nem foglalkozott szisztematikusan a közgazdaságtannal, ezért nem valószínű, hogy az ő matematikusi és géniuszi magasságából egyáltalán feltűnhettek volna neki azok az értelmezési nüánszok, amelyek megkülönböztetik egymástól és szembeállítják egymással a gazdasági rendszer eltérő szemléletű (például klasszikus, neoklasszikus, marxista, neoricardianus, új klasszikus stb.) irányzatait.

A bécsi (Karl) Menger-körrel tartott szoros kapcsolata, illetve németországi munkássága mindenesetre azt valószínűsíti, hogy Neumann elsősorban a német nyelvű irodalomban találhatta meg a kétszemélyes, zérusösszegű játékok egyensúlyának bizonyítása során már használt modell közgazdasági interpretációját, akár közvetlenül (olvasmányain keresztül), akár közvetve (szemináriumokon, beszélgetéseken keresztül). Neumann valószínűleg fontosabbnak is tartotta magát a módszertant, mint a közgazdasági interpretációt. (Ne feledjük el, hogy a Hilbert-i program szellemében, amelynek az 1920-as években Neumann még az egyik kiemelkedő követője volt, a minél szélesebb körű interpretációs lehetőség csak növeli egy matematikai „metamodell” értékét.) Adva volt számára egy matematikai modell, amely már bevált egy egyensúlyi probléma elemzésében, s ehhez talált egy lehetséges

közgazdasági interpretációt. Ezt jelzi a dolgozat eredeti német nyelvű címe is, amelyben az „*egy gazdasági egyenletrendszer*” mellett egyenrangúan jelenik meg „*a Brouwer-féle fixpont-tétel egy általánosítása*” utalás.

A Neumann-modell közgazdasági értelmezése körül folyó vitában az eddig felhozott érvek és bizonyítékok nem bizonyultak perdöntőknek. A Neumann-modell közgazdasági háttere kérdésének az eldöntését megnehezíti, hogy a modellben szereplő egyensúlyi összefüggések olyannyira általánosak, hogy azok szinte bármelyik elméleti főáramba beleillenek. A modell egyensúlyi összefüggései és a kiegészítő magyarázatok ugyanis nem fedik fel, vajon Neumann elképzelése szerint milyen erők, milyen mechanizmusok idézhetnék egyáltalán elő egy gazdaság egyensúlyi állapotát. A modell nyugodtan ráilleszthető lenne akár egy kisárutermelő („népi kapitalista”), akár egy klasszikus tőkés árutermelő, akár egy kollektív irányítású („piaci szocialista”) gazdaságra. Neumann modellje ugyanis egy nagyon régi és a közgazdasági gondolkodásban igen mélyen gyökerező ideának az absztrakt matematikai metaforája. Ez pedig nem más, mint a *tisztességes és értelmes módon működő áru gazdaság ideálja*.

Ez az ideál a racionális munkamegosztás és a piaci árucsere elve alapján működő árutermelő gazdaság elvont modellje. Ebben a gazdaságban az egyes gazdasági folyamatokat egymástól elkülönült egyének és/vagy kollektívák működtetik, akik áruikat a piacon cserélik el egymással, és pedig olyan intézményrendszer és mechanizmusok keretei között, amelyek biztosítják az érdekek harmóniáját és a gazdaság hatékonyságát. A termelés természetes feltételeit illetően, egy értelmes gazdaságban elvárható, hogy az egyes anyagi ágak termelése egymással összhangban, arányosan (ha nem is feltétlenül egyenletesen), és minél gyorsabb ütemben bővüljön. A tisztességes és értelmes gazdaság árainak (csereértékeinek) pedig – hosszabb táv átlagában – olyannak kell lenniük, amelyek fedezik az elhasznált termelési eszközök pótlását, az áruk előállításában részt vevők tisztas megélhetését, továbbá az utódjaik magasabb szintű életfeltételeit megteremtő felhalmozás forrását is. Neumann egyensúlyi árai pontosan megfelelnek ennek az elvnek: minden működtetett tevékenység esetén éppen biztosítják a pótlás és az arányos bővítés fedezetét.

3. A Neumann-modell és az általános gazdasági egyensúly korai modelljei

Az előző rész záró gondolatait folytatva röviden¹ áttekintjük a gazdasági egyensúly fogalmának kialakulásához vezető fontosabb fejleményeket, amelyek alapján majd összevetjük Neumann modelljét az elődök, ill. kortársak hasonló jellegű elméleti konstrukcióival. A gazdasági egyensúly fogalma alapvetően a felvilágosodás korabeli klasszikus közgazdászoktól ered, akik a hosszabb távon harmóniát, stabilitást teremtő és hatékonyságra

¹ A téma elméletörténeti vonatkozásai iránt érdeklődő olvasók figyelmébe ajánljuk mindenkéltől a következő műveket: *Ekelund and Hébert* (1997), *Ingrao and Israel* (1990), *Mátyás Antal* (1992), *Weintraub* (1985) és *Zalai* (2000, kül. 1. fejt.).

kényszerítő természeti törvények analógiáját a mindenható szabad piac („a láthatatlan kéz” – A. Smith) egyensúlyt teremtő víziójában találták meg. Műveikben sokoldalúan, a legkülönbözőbb analógiákkal (mintha ...) és a lényegre összpontosító absztrakciókkal igyekeztek megvilágítani az akkor kiteljesedő és kifejlődő piacgazdaság működésének törvényszerűségeit. A közgazdasági elméletüket azonban nem próbálták meg formális modellekbe öntve elemezni, különösen nem egyetlen átfogó modellbe. Marx maga is átvette, és sok tekintetben továbbfejlesztette az elődök egyensúlyelméletét. Arrow (1974) a Nobel-díj átvételkor tartott előadásában kifejezetten kiemelte Marxt, mint aki az elődöknél jelentősen közelebb jutott az egyensúlyelmélet modern modelljeihez.

A 19. sz. végén tevékenykedő korai neoklasszikus közgazdászok (Walras, Jevons, Marshall, C. Menger, Böhm-Bawerk, Wieser, Wicksell stb.) számára a piac kiterjedt működése már adottság volt. Többségük erős természettudományos képzést és szemléletet kapott, ezért is bátrabban nyúltak a matematikai modellekhez, mint elődeik. Megjelentek az általános gazdasági egyensúly első matematikai modelljei, az egyensúlyi állapotot jellemző egyenletrendszerek formájában. Cournot (1838), aki a részpiaci egyensúly vizsgálatára elsőként vette igénybe a korabeli fizika matematikai apparátusát, maga is tisztában volt azzal, hogy az egyes piacok egyensúlya csak része az általános egyensúlynak. Ennek ellenére nem tartotta érdemesnek és időszerűnek egy általános egyensúlyt definiáló matematikai rendszer felírását, mert, mint írta, „...az egész rendszer figyelembevétele... meghaladná a matematikai elemzés és a gyakorlati számítási módszerek lehetőségeit (erejét), még akkor is, ha a (modell) minden konstans paraméteréhez számszerű értéket tudnánk rendelni” (uo. 127. o.).

Tehát nem a svájci Walras „találta fel” az általános egyensúlyelméletet, még csak annak matematikai modelljét sem, mégis joggal őt tartják a modern általános egyensúlyelmélet atyjának. Walras volt ugyanis az első, aki elszakadt az empiriától, és a tiszta tudomány szellemében a közgazdaságtant is absztrakt, deduktív elméletként kezelte. Alapvető művének különböző (1874, 1877) kiadásaiban, Walras több változatban is megfogalmazta az általános egyensúly matematikai modelljét. A tiszta cseregazdaság modelljével kezdte, és csak később kapcsolta be a termelést az egyensúly elemzésébe, és még később a tőkejavakat. Csak a felesleges bonyodalmakat kívánta elkerülni akkor is, amikor konstans fajlagos ráfordítási együtthatókat feltételezve írta fel és elemezte az egyensúly matematikai modelljeit. Az általános egyensúly korai modelljeinek bemutatását mégsem Walras munkájával kezdjük, hanem az ún. Walras–Cassel-moddellel, amely hatását tekintve a későbbiekben több szempontból is jelentősebbnek bizonyult.

A Walras–Cassel-modell és annak Schlesinger–Wald-féle változata

A ma neoklasszikus közgazdaságtannak nevezett áramlat kialakulásában fontos szerepet játszott a 19. század végi osztrák közgazdasági iskola. A gazdasági javakat, Carl Menger (1871) nyomán, a végtermékek és a termelési tényezők csoportjába sorolták. A végtermékek egyensúlyi árait (p_i), feltevésük szerint, a fogyasztók értékítélete (a határhasznok arányai) határozzák meg, a termelési tényezők egyensúlyi árai (w_k) pedig elvben levezethetők (visszaszámíthatók) a fenti árak és a termelési tényezők végtermékekre vonatkoztatott rá-

fordítási arányainak (c_{ki}) az ismeretében. Ez az ún. *teljes beszámítás* (Zurechnung, imputation) *elve*, amely tehát egy

$$w_1 \cdot c_{1i} + w_2 \cdot c_{2i} + \dots + w_m \cdot c_{mi} = p_i, \quad i = 1, 2, \dots, n \quad (9)$$

egyenletrendszerként írható fel, ahol n a végtermékek, m a termelési tényezők száma.

Egy bécsi tudományos kör tagjai (akiknek a szemináriumait látogatta egy magyar származású műkedvelő közgazdász, *Schlesinger Károly* is) felismerték, hogy a fenti ármeghatározási elv matematikai szempontból egy egyáltalán nem triviális problémát rejt magában. Mi biztosítja ugyanis, hogy a felírt egyenletrendszer *reguláris* lesz ($n = m$), vagyis a termelési tényezők száma megegyezik a végtermékekével, és ha még pontosan annyi is lenne, mi garantálja, hogy a kapott egyenletrendszernek van nem negatív megoldása?

Ennek a problémának az elemzése keltette fel a bécsiek érdeklődését egy svéd közgazdász, *Cassel* (1918) általános egyensúlyelméleti modellje iránt. Mivel ez a modell bizonyos tekintetben Walras modelljének egy leegyszerűsített változataként értelmezhető (bár közvetlen kapcsolat nem mutatható ki a szerzők között), az irodalomban gyakran Walras–Cassel-modellként utalnak rá. Cassel az árbeszámítási (9) képletét kiegészítette a termelési tényezők *keresleti-kínálati egyenlőségével*:

$$c_{k1} \cdot y_1 + c_{k2} \cdot y_2 + \dots + c_{kn} \cdot y_n = s_k, \quad k = 1, 2, \dots, m \quad (10)$$

ahol $y_i = y_i(p_1, p_2, \dots, p_n)$ az i -edik termék keresleti függvénye, s_k a k -adik termelési tényező rögzített mennyiségű kínálata. (A keresleti függvények kapcsán feltűnő, hogy nem jelenik meg benne a jövedelem, amely a modellben csak a termelési tényezők értékesítéséből származhat. Ennek oka az, hogy Cassel a *jövedelmet* tekintette ármércének, és ebből következően annak szintjét rögzítettnek.)

A Walras–Cassel-modell feltételeit tehát a (9) és a (10) egyenletek, valamint a keresleti függvények alkották. A fenti modell feltételei egy olyan egyenletrendszer formáját öltötték, amelyben *a változók és az egyenletek száma* ($n + m$) *megegyezik* egymással (*reguláris* egyenletrendszer), s ez lehetővé tette, hogy Cassel éljen a korábban divatos *egyenletszámlálás* módszerével. Mi több, Cassel modellje – közvetlen helyettesítések és kiiktatások révén – egy

$$d_k(w_1, w_2, \dots, w_m) = s_k, \quad k = 1, 2, \dots, m$$

alakú egyenletrendszerre redukálható, amely már csak a szűkös termelési tényezők *keresletét* (d_k) és *kínálatát* (s_k) tartalmazza. A redukált egyenletrendszer is reguláris, s erre alapozva Cassel igazoltnak tekintette modellje logikai konzisztenciáját.

A későbbiek szempontjából is hangsúlyoznunk kell, hogy Cassel azt *feltételezte*, hogy modellje egy hosszabb időszak *ex post* megfigyelésén nyugszik, és ezért eleve csak a ténylegesen *szűkös*, tehát azokat a termelési tényezőket tartalmazza, amelyek ára pozitív. Az egyenletrendszer becsült paramétereinek tehát olyanoknak kell lenniük, amelyek mellett annak létezik pozitív megoldása, éspedig a termelési tényezők megfigyelt átlagos (egyensúlyi) árai, amelyek körül az értékük ingadozik. Azt a kérdést azonban már nem vizsgálta meg, hogy mi lesz akkor, ha megváltozik a termelési tényezők kínálata. Vajon lesz-e még az egyenletrendszernek pozitív vagy egyáltalán bármilyen megoldása?

A bécsi kör tagjai úgy gondolták, hogy „*Cassel okos ötlete*” (*Punzo*, 1989), az árakat definiáló egyenletrendszer *reguláris* tétele, szolgáltatja az általuk vizsgált (9) árbeszámítási probléma megoldását. Igen ám, de Schlesinger, az osztrák iskola megközelítésének megfe-

lelően (s részben azért, mert nem tudták megfelelően értelmezni a Cassel által alkalmazott keresleti függvényt), ún. *inverz* keresleti rendszert alkalmazott, ahol az árak a függő és a keresletek a független változók: $p_i = p_i(y_1, y_2, \dots, y_n)$. Ebből kifolyólag a termékek árai helyett azok végső fogyasztása lett modelljük változója. Ennek a látszólag ártalmatlan, formai módosításnak lényeges tartalmi és módszertani következményei lettek. A Walras–Cassel-modell Schlesinger-féle alakja ugyanis már nem volt redukálható a Cassel által követett módon. A végtermékek árösszefüggése (behelyettesítve az inverz keresleti függvényeket) és a termelési tényezők kereslet-kínálati összefüggése lett a modell *redukált formája*:

$$w_1 \cdot c_{1i} + w_2 \cdot c_{2i} + \dots + w_m \cdot c_{mi} = p_i(y_1, y_2, \dots, y_n), \quad i = 1, 2, \dots, n, \quad (9/a)$$

$$c_{k1} \cdot y_1 + c_{k2} \cdot y_2 + \dots + c_{kn} \cdot y_n = s_k, \quad k = 1, 2, \dots, m. \quad (10)$$

A módosított modell ugyan szintén egy reguláris egyenletrendszer formáját öltötte, de az mégsem oldotta meg a bécsiek dilemmáját. Tekintsük ugyanis az *erőforrás-elosztási feltételeket*, azaz a (10) összefüggést kielégítő *lehetséges* (y_1, y_2, \dots, y_n) *termelési* struktúrákat. Helyettesítsük be ezeket rendre a (9/a) egyenletrendszerben a keresleti függvényekbe. A termelési tényezők árait meghatározó (9/a) egyenletrendszer továbbra is alul- vagy túldeterminált lehet, ugyanúgy, mint az eredeti árbeszámítási egyenletrendszer, hacsak a termékek és a termelési tényezők száma meg nem egyezik egymással.

„Cassel okos ötlete” tehát nem segített a bécsieknek az árbeszámítási probléma megoldásában. *Schlesinger* (1935) végül szintén a *komplementaritási elvet* alkalmazta annak érdekében, hogy az egyensúlyt definiáló egyenletrendszer irreguláris formájából adódó problémát megoldja. Nevezetesen, az egyenletek helyett egyenlőtlenségeket vezetett be:

$$c_{k1} \cdot y_1 + c_{k2} \cdot y_2 + \dots + c_{kn} \cdot y_n \leq s_k, \quad k = 1, 2, \dots, m, \quad (10/a)$$

azzal a kiegészítéssel, hogy az olyan termelési tényezők ára, amelyeknek a korlátja az egyenlőtlenség-rendszer megoldásában nem merül ki teljesen, szükségképpen nulla lesz. A (9/a) és a (10/a) összefüggésekkel adott statikus modell egyensúlyának létezését az akkor Bécsben élő *Wald Ábrahám* (1935) bizonyította be. Wald tehát Neumann-nal egy időben, és tőle minden jel szerint függetlenül, szintén az elsők között adott egzakt bizonyítást az általános egyensúly létezésére.

Vegyük észre, hogy Schlesinger és Wald a komplementaritási feltevést – Neumann megoldásától eltérően – aszimmetrikusan alkalmazták. Az egyensúlyi árakat jellemző (9/a) feltételek továbbra is egyenlőségek maradtak, tehát eleve feltették, hogy minden tevékenységet alkalmazni fognak az egyensúlyban. Ennek matematikai elégséges feltételeként Wald feltette, hogy a keresleti függvény értéke (az adott termék ára) végtelenbe tart, valahányszor az adott termék mennyisége nullához tart (minden termékre szükség van a végső fogyasztásban). Vegyük észre azt is, hogy Schlesinger és Wald a fenti megoldással Cassel *ex post* jellegű modelljét egy *ex ante* szemléletű modellt alakította át. Nem lehet ugyanis előre megmondani, mely termelési tényezők lesznek szűkösek. Mindez a modell paramétereitől függ. Ebben a tekintetben a Neumann-modell és a Cassel-modell Schlesinger–Wald-féle változata igen közel áll egymáshoz.

Cassel stacionárius növekedési modellje

Cassel a statikus modell ismertetése mellett felvázolta egy többidőszakos, arányosan bővülő gazdaság modelljét is. Cassel azonban Neumanntól teljesen eltérő módon értelmezte az arányosan bővülő, egyensúlyi gazdaság feltételeit. Cassel feltevése szerint, a termékeket közvetlenül az elsődleges erőforrásokból állítják elő, ezért a termelés szintjét és annak időbeli növekedését kizárólag csak az elsődleges erőforrások mennyisége korlátozza. Ez utóbbiak növekedési ütemét Cassel *külső* (gazdaságon kívüli) *adottságként* kezeli, és felteszi, hogy egyes komponenseik, az egyik időszakról a másikra, azonos ütemben (a tényezővel) bővülnek.

Változatlan technológiát és árarányokat, és ebből következően változatlan végső fogyasztási szerkezetet feltételezve, a termelés és a fogyasztás, illetve a termelőfelhasználás minden komponense ugyanabban a külsőleg meghatározott ütemben növekedhet. Az egyenletes növekedés következtében a termelési tényezők kereslet-kínalati egyenlőségének mindkét oldalát ugyanazzal a (pozitív) növekedési tényezővel kell beszorozni, ezért – a statikusról a stacioner modellre való áttérés során – az egyensúly (10) feltétele változatlan marad.

A termelés és a termékek felhasználása között Cassel is egyéves *késleltetést* feltételez, és ennek következtében a költségek felmerülése és a jövedelem realizálása között is egyéves eltérés jelentkezik. Emiatt a jövedelmeknek a költségeken felül kamatot is tartalmazniuk kell. Az árakat meghatározó (9) összefüggésben a költségeket tehát be kell szorozni a kamattényezővel. (Ez lesz végső soron az egyetlen változás a statikus modell összefüggéseihez képest.) A jövedelmek és kiadások második időszakra feltételezett egyenlősége miatt viszont α *kamattényezőnek* meg kell egyeznie az *egzogen* α *növekedési tényezővel*.

Az arányos egyensúlyi növekedés matematikai feltételei tehát pusztán az áregyenlet konstans α tényezőjével térnek el a statikus modell feltételeitől:

$$\alpha \cdot (w_1 \cdot c_{i1} + w_2 \cdot c_{i2} + \dots + w_m \cdot c_{im}) = p_i, \quad i = 1, 2, \dots, n,$$

$$c_{k1} \cdot y_1 + c_{k2} \cdot y_2 + \dots + c_{kn} \cdot y_n = s_k, \quad k = 1, 2, \dots, m,$$

ahol $y_i = y_i(p_1, p_2, \dots, p_n)$, mint korábban.

A tőkejavak Walras általános egyensúlyi modelljében

Cassel és Neumann modellje, mint láthatjuk, inkább *egymás komplementere*, semmint ugyanazon modell alternatív változatai. Később majd megmutatjuk, hogy Cassel modelljét át lehet alakítani úgy, hogy a szűkös termelési tényezők között elkülönítetten megjelenjenek a felhalmozott tőkejavak, mint Neumann modelljében. Mielőtt azonban erre kitérnénk, érdemes felidézni Walrast, aki ezt egyik modelljében maga is megtette már. Nézzük meg, hogyan illesztette be Walras a tőkejavakat általános egyensúlyelméleti modelljébe.

Walras feltette, hogy a gazdasági javak három nagy csoportba sorolhatók: az adott időszakban termelt *végtermékek* (köztük tőkejavak), a korábbi időszakokban termelt, azonos

jellegű felhalmozott tőkejavak és az elsődleges (nem termelhető) erőforrások csoportjába. Legyen y_i az i -edik termék személyes végső fogyasztása, z_i ugyanennek felhalmozásra fordított mennyisége. Jelölje k_i az i -edik termék tőkeként felhalmozott készletét (a tőkék kínálatát), s_k pedig a k -adik egyéb termelési tényező kínálatát. Jelöljük b_{ij} -vel a fajlagos tőkelekötési igényeket, c_{kj} -vel az egyéb termelési tényezők ráfordítási együtthatóit. Továbbá, legyen p_i az i -edik termék ára, w_k a k -adik egyéb termelési tényező ára (bérleti díja), q_i az i -edik termék mint tőkejóság fajlagos bérleti díja (költsége).

Csak az utóbbi változó értelmezése szorul némi magyarázatra. Feltevésünk szerint a termékek ára független attól, hogy elfogyasztják vagy tőkeként felhalmozzák őket, az áruk mindkét esetben ugyanakkora (p_i). Ugyanakkor a tőkejavak költségét az amortizáció és nettó megtérülés összegére lehet felbontani (Walras emellett figyelembe veszi még a tőkejavak biztosítási költségeit is, amelyektől itt eltekintünk). A tőkejavak költségét tehát a

$$q_i = (r_i + \pi_i) \cdot p_i, \quad i = 1, 2, \dots, n \quad (11)$$

egyenletekkel határozhatjuk meg, ahol π_i az i -edik tőkejóság nettó megtérülési rátája, r_i pedig az (egységnyi értékére vonatkoztatott) amortizációs rátája.

Fogalmazzuk most meg az általános egyensúly feltételeit az így kapott modellben.

A) a termékek és termelési tényezők keresletének és kínálatának egyensúlyi feltételei:

$$y_i(p_1, p_2, \dots, p_n; w_1, w_2, \dots, w_m) + z_i = x_i, \quad i = 1, 2, \dots, n, \quad (12)$$

$$c_{k1} \cdot x_1 + c_{k2} \cdot x_2 + \dots + c_{kn} \cdot x_n = s_k(p_1, p_2, \dots, p_n; w_1, w_2, \dots, w_m), \quad k = 1, 2, \dots, m; \quad (13)$$

B) egy- vagy többidőszakos beruházási készletetést feltételezve a tőkejavak keresletének és (konstans) kínálatának egyensúlyi feltétele:

$$b_{i1} \cdot x_1 + b_{i2} \cdot x_2 + \dots + b_{in} \cdot x_n = k_i, \quad i = 1, 2, \dots, n; \quad (14)$$

C) az egyensúlyi árrendszert jellemző összefüggés:

$$q_1 \cdot b_{i1} + q_2 \cdot b_{i2} + \dots + q_n \cdot b_{in} + w_1 \cdot c_{i1} + w_2 \cdot c_{i2} + \dots + w_m \cdot c_{im} = p_i, \quad i = 1, 2, \dots, n, \quad (15)$$

ahol $y_i(\cdot)$ az i -edik termék fogyasztói keresleti függvénye, és $s_k(\cdot)$ a k -adik termelési tényező kínálati függvénye. (A termékek fogyasztói keresleti függvényeiben a termelési tényezők árai a keletkező jövedelmek magyarázó változói.)

Az eddig felírt, négy csoportba sorolt (12)–(15) egyenletrendszerben a változók (x_i, z_i, p_i, q_i, w_k) száma $4n + m$, az egyenletek száma pedig $3n + m$. Az egyenletrendszer tehát még nem meghatározott (aluldeterminált), és szabadsági foka n . Ugyanakkor még nincs olyan összefüggés, amelyik meghatározná a beruházási keresletet, illetve előírná a különböző tőkejavak nettó megtérülési rátáinak egyenlőségét, amit hosszú távú egyensúly esetén elvárunk. A fennmaradó n szabadsági fok megszüntetésére tehát két lehetőség kínálkozik: vagy a beruházási keresletet, vagy a tőkemegtérülési ráták azonosságát adjuk meg további n darab (egymást potenciálisan kizáró) feltétel bevezetésével.

Walras is jól tudta, hogy az egyes tőkejavak megtérülési rátái csak akkor lesznek egyformák, ha a beruházási döntések olyanok, amelyek nyomán a felhalmozott tőkejavak szerkezete az egyensúly követelményeinek megfelelően alakul. Ha előírja az egységes tőkemegtérülési ráták feltételét, akkor a beruházások (z_i) nagyságát nem a kereslet határozza meg, hanem a többi változóra tett kikötések által, mintegy visszamaradó (reziduális) módon kap értéket a modellben. Ha viszont bevezet egy beruházási keresleti függvényt, akkor semmi

sem garantálja azt, hogy tőke megtérülési ráták kiegyenlítődnének (az egyes időszakokban a felhalmozott tőkejavak szerkezete megfelel az egyensúly következő időszaki követelményének). Walras a statikus (időtlen) modellje lezárásának dilemmáját a tőke megtérülési ráták egyenlőségének, a *hosszú távú egyensúly* feltételének bevezetésével hidalta át.² A tőke költséget meghatározó (11) egyenlet alapján ez a megoldás az alábbi

$$q_i = (r_i + \pi_i) \cdot p_i, \quad i = 1, 2, \dots, n, \quad (16)$$

n darab egyenlet egyensúlyi feltételként való elfogadását jelenti.

Az általános egyensúlyi profitráta (π) bevezetésével azonban egy újabb változó került a modellbe, ami miatt a modellnek továbbra is maradt még egy szabadságfoka. Ha azonban feltesszük, hogy a keresleti és a kínálati függvények csak az árak arányaitól függenek, a szintjüktől nem (az árakban nulladfokon homogén függvények), akkor a fennmaradó szabadságfokot az árszint megkötésével megszüntethetjük. Ezt tette Walras.

Térjünk még röviden vissza a beruházások kérdésére. A beruházások teljes volumene nyilván nem alakulhat tetszés szerint, egyensúlyban kell lennie a megtakarításokkal. A megtakarítások és a beruházások globális egyensúlyának feltételét azonban nem kell külön előírni, mivel – mint azt könnyen ellenőrizhetjük – a modell egyenletei biztosítják ennek teljesülését. Walras elképzelése szerint tehát az összberuházás szintje igazodik a megtakarításokhoz. Úgy gondolta, hogy rövid távon lehet egyensúlytalanság a tőkejavak piacán, és az eltérő tőke megtérülési rátákat eredményezhet. Hosszú távon azonban az ágazati *beruházások igazodási folyamata* (a nagyobb megtérülési rátát eredményező javakból viszonylag többet halmoznak fel, és fordítva) a megtérülési ráták kiegyenlítéséhez vezet. Statikus modelljében voltaképpen ezt a két feltevést ötvözte implicit módon.

A fentiekből is látható, hogy a tőke kezelése tekintetében Walras még nem volt annyira neoklasszikus, mint az osztrák iskola képviselői vagy akár Cassel. A vázolt modelljében a tőkék maguk is termékek, a tőkék árai maguk is a modell változói, és a tőke megtérülési ráták kiegyenlítődnése, ugyanúgy az egyensúly feltétele, mint Neumann modelljében. A klasszikus egyensúlyi profitráta tehát szükségképpen megjelenik az egyensúlyi árakat meghatározó összefüggésekben, ha figyelembe vesszük (itt a beruházásokon keresztül) a termelés körkörös voltát és hosszú távú egyensúlyt feltételezünk.

A termelés körkörösége és Leontief általános egyensúlyi modellje

A *termelés körkörösége* azonban nemcsak a beruházásokon, de a folyó termelő felhasználáson keresztül is megnyilvánul. A vázolt általános egyensúlyi modellek ebből a szempontból merőben eltérnek Neumann modelljétől. Walras, Cassel és nyomdokaikon Schlesinger és Wald eltekintettek az egyes termékek termelő (közbenső) felhasználásától, amely az összes keresletük tetemes részét teszi ki a modern gazdaságokban. Ábrázolásuk-

² Amiért azután Keynes meg is bírálta később Walrast, egyszersmind rámutatva arra, hogy a beruházások nem megfelelő alakulása nemcsak eltérő profitrátákhoz vezet általában, de elégtelen szintű effektív keresletet és az erőforrások részleges kihasználását is eredményezheti.

ban ugyanis a termékek tisztán végtermékeként tűnnek fel, így anyagköltségek sem jelennek meg a termékek árának meghatározásában. Mintha azt feltételezték volna, hogy a termelők között ugyanolyan problémamentesen megy végbe a termékek elosztása, mint egy jól szervezett nagyvállalat keretében.

Ezt a hiányosságot egyszerűen korrigálhatjuk. Tartsuk meg az előzőekben tárgyalt Walras-modell jelöléseit, és vezessük be az a_{ij} ráfordítási együtthatót az i -edik termékből a j -edik termék előállításához szükséges fajlagos ráfordítás (közbenső felhasználás) jelölésére.

Ennek nyomán a termékek és termelési tényezők keresletének és kínálatának egyensúlyi összefüggései a

$$a_{i1} \cdot X_1 + a_{i2} \cdot X_2 + \dots + a_{in} \cdot X_n + y_i(p_1, p_2, \dots, p_n) + z_i = X_i, \quad i = 1, 2, \dots, n, \quad (12/a)$$

az egyensúlyi árrendszert jellemző összefüggések pedig

$$P_1 \cdot a_{1i} + P_2 \cdot a_{2i} + \dots + P_n \cdot a_{ni} + q_1 \cdot b_{1i} + q_2 \cdot b_{2i} + \dots + q_n \cdot b_{ni} + w_1 \cdot c_{1i} + w_2 \cdot c_{2i} + \dots + w_m \cdot c_{mi} = P_i, \quad i = 1, 2, \dots, n \quad (15/a)$$

formát öltik. A többi feltétel változatlan marad.

A termékek termelő felhasználásának a figyelembevételé azonban megváltoztatja a termelés (x_i) és a termékár (p_i) korábbi jelentésének tartalmát, ezért is vezettünk be új jelöléseket (X_i és P_i). A termékek *kibocsátása* (X_i) most már nemcsak a végső (nettó) kibocsátást, hanem a *teljes* (bruttó) *termelést* jelenti. Hasonlóképpen, a *termék ára* (P_i) nem egyszerűen a tőke és az elsődleges erőforrások költsége (jövedelme), hanem tartalmazza a *felhasznált anyagok költségét is*. (Mai terminológiával élve: a hozzáadott érték a munka, a tőke és az egyéb külső erőforrások költsége, az ár pedig az anyagköltség és a hozzáadott érték összege.)

A (12/a), (13), (14), (15/a), (16) egyenletrendszer pedig, tartalmát tekintve, nem más, mint a Nobel-díjas *Leontief* (1928, 1941), nagyjából Neumann Jánossal egy időben kifejlesztett, *input-output modelljének* a feltételrendszere. *Leontief* kifejezetten gyakorlati alkalmazások céljára dolgozta ki a módszerét, ezért *Walras*nál egyszerűbb feltevésekkel élt. Így például, *Neumannhoz hasonlóan*, a végső felhasználás meghatározásánál nem használt keresleti függvényeket, és ő is eltekintett a külső erőforrások (munka, tőke és természeti erőforrások) potenciális korlátos voltától. Nála ezért nem jelentek meg a (13) és a (14) egyenletek. A végső kereslet nagyságát és a külső erőforrások árát pedig, ennek ellensúlyozásaképpen, egzogén változókként kezelte. Emiatt elmaradtak a (16) egyenletek is. Mindezek következtében *Leontief* modellje az alábbi két lineáris egyenletrendszerre egyszerűsödik:

A) a termékek keresletének és kínálatának (naturális) egyensúlyi feltételei:

$$a_{i1} \cdot X_1 + a_{i2} \cdot X_2 + \dots + a_{in} \cdot X_n + y_i + z_i = X_i, \quad i = 1, 2, \dots, n, \quad (12/b)$$

B) a termékek árainak (értékbeli) egyensúlyi feltételei:

$$P_1 \cdot a_{1i} + P_2 \cdot a_{2i} + \dots + P_n \cdot a_{ni} + h_i = P_i, \quad i = 1, 2, \dots, n, \quad (15/b)$$

ahol h_i a külső erőforrások fajlagos költsége, a fajlagos *hozzáadott érték* (a p_i változó *Walras* modelljében). Mint láthatjuk, a kapott input-output modellben meg is szűnik minden kapcsolat a naturális és az értékbeli változók, illetve egyensúlyi feltételek között.

Maga a *végző kibocsátás* – definíció szerint – a teljes termelés és a termelő felhasználás különbsége, és egyensúly esetén ez megegyezik a *végző felhasználással*:

$$x_i = X_i - (a_{i1} \cdot X_1 + a_{i2} \cdot X_2 + \dots + a_{in} \cdot X_n) = y_i + z_i, \quad i = 1, 2, \dots, n.$$

Normális körülmények között – ha az a_{ij} ráfordítási együtthatókkal jellemzett Leontief-típusú termelő rendszer produktív (bővebben lásd Zalai, 2000) – a fenti egyenletrendszer átrendezhető, és kifejezhetjük belőle az X_i változókat az x_i (azaz az y_i és z_i) változók függvényében:

$$X_i = s_{i1} \cdot x_1 + s_{i2} \cdot x_2 + \dots + s_{in} \cdot x_n, \quad i = 1, 2, \dots, n, \quad (18)$$

ahol tehát $x_i = y_i + z_i$ és az s_{ij} együtthatók, a *teljes ráfordítási együtthatók*, amelyek a végző kibocsátások egy-egy egységére vonatkoztatva adják meg a különböző termékek teljes keresletét.

Ugyanígy megkaphatjuk a termékek árait a fajlagos hozzáadott értékek és a teljes ráfordítási együtthatók ismeretében:

$$P_i = p_1 \cdot s_{1i} + p_2 \cdot s_{2i} + \dots + p_n \cdot s_{ni}, \quad i = 1, 2, \dots, n, \quad (19)$$

ahol $p_i = h_i$ a fajlagos hozzáadott érték.

Ezek az alapegyenletek jellemzik Leontief statikus input-output modelljét, amelynek szintén ismertek a dinamikus kiterjesztései. A Leontief-modellek elemzése elsősorban a nem negatív négyzetes mátrixok nevezetes matematikai (mindenekelőtt a Perron–Frobenius-féle sajátérték-) tételeken nyugszik. Általában igaz, hogy miközben számos formai és tartalmi azonosságot fedezhetünk fel Neumann és Leontief (dinamikus) modellje között, jelentősek köztük a közgazdasági szemléletbeli és a módszertani különbségek is (ezekről bővebben szintén lásd Zalai, 2000).

Befejezésképpen térjünk vissza az alfejezet elejére, ahol azzal kezdtük, hogy olybá tűnik, mintha Walras, Cassel és a többiek nem vették volna észre, hogy a *termelés körkörösége* nem csak a beruházásokon, de a folyó termelő felhasználáson keresztül is hat az egyensúlyi viszonyokra. A termékek ugyanis tisztán végtermékeknek tűnnek fel modelljeikben, a termékárak pedig látszólag csak a külső termelési tényezők költségét tartalmazzák. Minden bizonnyal igazságosabbak vagyunk velük szemben, ha feltesszük, maguk is tisztában voltak azzal, hogy ez az ábrázolási mód elvonatkoztat a termékek termelő (közbenső) felhasználásától. Minden bizonnyal csak azért tekintettek el tőle, mert látták, hogy az általános egyensúly nettó kibocsátások és árak, illetve közbenső felhasználás figyelembevételével felírt feltételeit a teljes ráfordítási együtthatók segítségével egyszerű helyettesítések révén redukálni lehet egy olyan feltételrendszerre, amelyben már csak a nettó kibocsátások és árak szerepelnek. Például, a (12/a), (13), (14), (15/a), (16) egyenletrendszerrel definiált feltételeket a Walras-modell (12)–(16) feltételrendszerére a (18) és (19) egyenletek által jelzett helyettesítések révén.

Összevetés Neumann modelljével

A bemutatott modellek és Neumann János modellje közötti hasonlóságok és különbségek részletes elemzése helyett (bővebben lásd Zalai, 1999) ehelyütt csak kiemeljük azokat a különbségeket, amelyek alapján egyértelműen Neumann modellje bizonyul érdekesebbnek, mivel számos későbbi fontos eredmény zseniális megsejtését rejtette magában:

- Neumann modelljében a termelési szerkezet és az árak *matematikai dualitása* teljes, s ezzel előrevetítette a lineáris programozás dualitási tételeit;
- a Neumann-modellben jelenik meg először a technológia leírásaként a *tevékenység-elemzési modell*, amely lehetővé teszi az *ikertermelés* és a *technológiai választék* figyelembevételét;
- Neumann elsőként ismerte fel és ábrázolta formálisan azt is, hogy a *hatékony* termelési *tevékenységek* kiválasztása és az *egyensúlyi* (hatékonysági) *árrendszerek* meghatározása egymást kölcsönösen feltételező probléma;
- Neumann a szűkös termelési tényezőket a termelésből származtatja (*körkörös termelési kapcsolatok*), ami megvilágítja a tőke és a profit természetét;
- Neumannnál a *növekedési* és a *kamattényező endogén*, az *árutermelő alrendszer hatásfoka* által meghatározott (azáltal, hogy milyen *sajátfelhasználási hányad* mellett képes a társadalom a termékeket újratermelni), Casselnél ezek *egzogén* módon, a külső erőforrások növekedési üteme által adottak;
- Cassel modelljében a *növekedési* és a *kamattényező* nem válik el egymástól, míg a Neumann-modellben megjelenő kamatláb megfelel a klasszikus profitráta fogalmának, és a növekedési ütem ettől kisebb is lehet.
- Neumann az egyensúly létezését egy *fixponttétel* alapján bizonyította, ami modelljében még egy túl erős eszköznek bizonyult, de az általános egyensúly létezésének bizonyításaiban elengedhetetlen eszköznek bizonyult, Wald ezzel szemben még teljes indukción nyugvó hagyományos módszereket vett igénybe ugyanerre a célra;
- Neumann mutatott rá elsőként a *gazdasági és a játékelméleti egyensúly* között fennálló szoros kapcsolatra.

Neumann egy fontos figyelmeztetése

Walras és Cassel korai modelljei, illetve a Schlesinger–Wald és a Neumann-modell szemlélete között van egy nagyon fontos különbség. Nevezetesen az, hogy míg Walras és Cassel számára a piaci egyensúly létezése „empirikus evidencia”, modelljeik kiinduló hipotézise volt, addig Wald és Neumann számára az egyensúly létezésének matematikai lehetősége már bizonyítandó volt. Walras és Cassel az egyensúly feltételeinek matematikai konzisztenciáját még csak az egyenletszámolás módszerével igazolták.

Neumann és Wald korára, az axiomatikus szemlélet térhódítása folytán, egy jelentős modellezés-filozófiai szemléletváltás következett be: szakítás a klasszikus fizikára jellemző közvetlen tapasztalaton nyugvó (*ex post*) modellezési szemlélettel. Az *ex post* szemléletű modellekben a változók és paraméterek mind (megfigyelhető, mérhető) *tapasztalati* kategóriák, az összefüggéseik pedig a változók között *törvényszerűen* fennálló (vélt) kapcsolatokat fejeznek ki, általában csak a modell gyakorlati alkalmazása során derülhet fény arra, hogy az adott, számszerűsített specifikáció képes-e, és milyen pontossággal képes reprodukálni (a modell megoldásaként előállítani) a megfigyelt állapotot.

Az *ex ante* szemléletű absztrakt modellekben ezzel szemben, a változók és összefüggések már *a priori* kategóriák, amelyek eredendően csak egy „matematikai struktúrát” definiálnak. Az elemzés a modellben értelmezhető kitüntetett, meghatározott tulajdonságok-

nak eleget tevő állapotok (például, az egyensúlyi) létezésének logikai lehetőségére irányul. A modell előfeltevései (axiómái) arra szolgálnak, hogy elégséges feltevéseket nyújtsanak ezek megvalósulásához. A valós jelenségekre való utalás, a modellek empirikus tesztelhetősége, gyakorlati előrejelzésekre való alkalmasságuk ellenőrzése pedig gyakran el is marad.

Az axiomaticus szemlélet térhódítása jelentősen átalakította a természettudományokat és a matematikát, és hatása fokozatosan kiterjedt a közgazdaságtan területére is, ahol egybeesett a matematikai közgazdaságtan mint relatíve önállósult tudományág kialakulásával. Miközben ezek a változások felgyorsították a tudományos fejlődést, komoly veszélyt is magukkal hoztak, amelyre idősebb korában maga Neumann János is figyelmeztetett:

„a legjobb matematikai ihletés a tapasztalatból ered, és hogy aligha lehet hinni a matematikai szigor abszolút, változatlan, minden emberi tapasztalattól elkülönült fogalmának létezésében... Ha egy matematikai diszciplína messzire távolodik el tapasztalati forrásaitól... ez súlyos veszélyt rejt magában. Egyre inkább tiszta esztétizálássá válik, egyre tisztább *l'art pour l'art*-á. ... Tapasztalati forrásától nagy távolságban vagy sok absztrakt „behatás” után a matematikai tárgyat a degenerálódás fenyegeti... Kezdetben a stílus rendszerint klasszikus; amikor a barokká-válás jelei mutatkoznak, a vészjelzés adott.” (1965, 21. és 27. o.)

A társadalomtudományok területén, ahol az empirikus kísérletezés és tesztelés lehetősége, eddig legalább is, meglehetősen korlátozottnak bizonyult, a fenti veszély hatványozottan fennáll. Ezért is van az, hogy a társadalomtudományok területén meglehetősen végletes álláspontokkal találkozhatunk az axiomaticus módszer, s a tőle szinte elválaszthatatlan matematikai nyelvezet és módszerek hasznossága megítélésében. A *kvantitatív közgazdaságtan* kialakulásának kezdetén, a kutatások fő célkitűzése az *empirikus kutatások gyakorlatának, módszertanának* a kifejlesztése volt. Eleinte kifejezett idegenkedés volt tapasztalható az elvont közgazdasági és matematikai elméletekkel szemben. Előre nem látott, köztük véletlen események azonban teljesen más irányba tereltek, elsősorban az amerikai, közgazdaságtani kutatások későbbi alakulását (bővebben erről lásd *Zalai*, 1999).

A leglényegesebb változás az volt, hogy fokozatosan átalakult a vezető kutatóknak a közgazdaságtanról alkotott felfogása: elsősorban az axiomaticus módszert formailag tökélyre fejlesztő *Bourbaki*-iskolának a matematikai módszertanról, illetve magáról a matematikáról alkotott képének kritikátlan elfogadása miatt (lásd *Weintraub és Mirowski*, 1994). Az 1940-es évek végén, illetve az 50-es évek elején a kutatások az alkalmazott *reál-tudomány* felől erőteljesen eltolódnak az absztrakt, *tiszta tudomány* irányába. Az általános egyensúlyelméleti modellek érvényességének („tudományos igazságának”) igazolását sem a gyakorlati alkalmazás, hanem inkább a matematikai, logikai konzisztencia bizonyításának irányában keresték.³

Számos munkából ismert, köztük *Kornai János* (1971) *Anti-equilibrium*-ából, hogy ennek az egyébként igen látványos fejlődésnek komoly ára volt. Nevezetesen az, hogy a logikai

³ Az általános egyensúlyelmélet modern (*ex ante*) modelljének a megfogalmazása, az egyensúly létezésének és hatékonyságának matematikai bizonyítása elsősorban *Arrow és Debreu* (1954), illetve *McKenzie* (1954) nevéhez fűződik.

konzisztencia, a matematikai elegancia és az esztétikum követelménye az empirikus relevancia fölé kerekedett. Emiatt sokan szembefordultak a matematikai közgazdaságtannal, és elutasítóan viszonyultak az általános egyensúlyelmélet és annak matematikai modelljeivel szemben. Nem lehet véletlen, hogy sem Neumann, sem Wald nem folytatta tovább az absztrakt gazdasági egyensúlyelmélet kutatását, pedig mindketten viszonylag korán az USA-ba mentek, és szoros kapcsolatba kerültek a „kvantitatív közgazdaságtani bölcsője” körül bábáskodó kutatókkal. Lehet, hogy Neumann felismerte, hogy ezek a kutatások a „barokkosodás és degenerálódás” felé tendáltak, elveszítették a „tapasztalati, empirikus forrásukat”? A tudomány bizonyára csak nyert azzal, hogy így döntött.

Az igazsághoz azonban feltétlenül hozzátartozik az is, hogy az 1950-es években a számszerűsített általános egyensúlyelméleti (CGE) modellek gyakorlati alkalmazásának a lehetősége még igencsak távolinak tűnt, továbbfejlődésük talán ezért is követett más irányt. A szükséges statisztikai adatok, matematikai algoritmusok és számítástechnikai eszközök nem álltak még rendelkezésre a gyakorlati alkalmazások számára. Az 1970-es évek második feléig kellett várni arra, hogy beinduljon és egyre szélesebb körben elterjedjen a CGE modellek gyakorlati alkalmazása (bővebben lásd például *Zalai, 1998*). Az általános gazdasági egyensúly absztrakt modelljeinek ismerete jelentősen megkönnyítette a felmerülő módszertani kérdések kezelését, de szerepük nem volt döntő a gyakorlati alkalmazások szempontjából.

A téma lezárásaképpen érdemes idézni a holland Mark Blaug ezzel kapcsolatos álláspontját:

„Egy »elméletet« nem lehet elutasítani pusztán azért, mert még nem ellenőrizhető... feltéve, hogy fontos problémákra hívja fel a figyelmet... Vitathatatlan, hogy a közgazdászok gyakran becsapják magukat azzal, hogy – mint Leontief nevezi – »implicit teoretizálással« foglalkoznak és tautológiákat közölnek, közgazdasági ismeretekhez való lényegi hozzájárulásnak álcázva. De az ilyen gyakorlat orvoslásának a célok tisztázására, s nem radikális és minden bizonnyal korai műtétre van szükség.” (*Blaug, 1962, 606. o.*)

4. Neumann hozzájárulása a játékelmélet további fejlődéséhez

Az n -személyes, nem kooperatív játékok Nash-féle egyensúlya

Neumann János mátrixjátékokra bevezetett egyensúlyfogalma inspirálta *John Nash*-t 1950-ben arra, hogy kiterjessze azt az n -személyes, nem kooperatív játékokra. Ha adott egy n -személyes nem kooperatív játék normál formában

$$G = \{S_1, \dots, S_n; f_1, \dots, f_n\},$$

akkor az $s = (s_1, \dots, s_n)$ stratégiaprofilit (Nash) egyensúlypontnak nevezzük, ha fennállnak

$$f_i(s) \geq f_i(s_1, \dots, s_{i-1}, t, s_{i+1}, \dots, s_n)$$

egyenlőtlenségek minden $t \in S_i$ és minden i esetén.

A Nash-féle egyensúlypont egyfajta stabilitást fejez ki: egyik játékosnak sem érdeke egyoldalúan eltérni az s_i egyensúlyi stratégiájától és a t stratégiát alkalmazni, ha a többiek nem térnek el a saját stratégiájuktól. Más szóval: ha egy játékos valahonnan tudja, hogy a többiek mit akarnak játszani, akkor neki is az egyensúlyi stratégiáját kell alkalmaznia. Könnyű látni, hogy ez a definíció a mátrixjátékok esetében éppen a minimax egyenlőséggel ekvivalens.

A fenti definíció csak akkor életképes, ha a játékok széles osztályára lehet bizonyítani az egyensúly létezését. John Nash az n -személyes véges játékok kevert bővítését Neumann János nyomán a következőképpen definiálta: stratégiahalmazok az eredeti véges stratégiahalmazokon értelmezett valószínűség-eloszlások ugyanúgy, mint a mátrixjátékok kevert bővítésénél, kifizetőfüggvények pedig a várható kifizetések, amelyeket a választott valószínűség-eloszlásokból nyert súlyokkal számolunk. Szintén Neumann János nyomdokait követi annak a bizonyítása is, hogy ennek a játéknak mindig létezik legalább egy Nash-egyensúlyi pontja. A felhasznált eszköz itt is egy fixponttétel, a Brouwer-féle fixponttétel.

A Brouwer-féle fixponttétel egydimenziós változata roppant egyszerű és szemléletes, több dimenzióban a bizonyítása már igen bonyolult. Mi most az egydimenziós esetet fogalmazzuk csak meg. Tekintsük a számegegyenesen a $[0, 1]$ zárt intervallumot és ennek egy önmagára való folytonos leképezését, vagyis egy folytonos $f(x)$ függvényt, amelynek az értelmezési tartománya és az értékkészlete is a $[0, 1]$ intervallum. A Brouwer-féle fixponttétel azt állítja, hogy ekkor mindig van olyan $x \in [0, 1]$ szám, hogy $x = f(x)$. Nash tételének későbbi általánosításainál a Brouwer-féle fixponttétel Kakutani-féle általánosítása is szerepet játszik. Ha szintén maradunk az egydimenziós esetről, akkor most a $[0, 1]$ intervallum minden egyes x pontjához a $[0, 1]$ egy $I(x)$ zárt részintervallumát rendeljük, és még egy bizonyos fajta folytonosságot is megkövetelünk ettől a hozzárendeléstől. Kakutani tétele azt állítja, hogy mindig van olyan $x \in [0, 1]$ szám, hogy $x \in I(x)$, vagyis, hogy x benne van a saját képében. Nem nehéz megmutatni, hogy Nash egzisztenciátétele mátrixjátékok esetében Neumann János minimax tételét adja, és így teljes joggal tekinthető annak általánosításának.

A Nash-egyensúlypont a modern közgazdaságtan egyik legfontosabb fogalmává vált, amit az is jelez, hogy John Nash 1994-ben közgazdasági Nobel-díjat kapott. Nash nagy tisztelettel nézett fel Neumann Jánosra, és neki is megmutatta felfedezését. Neumann János zsenialitására jellemző az alábbi részlet John Nash életrajzi regényéből: „Nash a záróvizsga után pár nappal felkereste Neumann. ... Neumann egy hatalmas asztal mögött ült, és drága öltönyében inkább egy sikeres bankigazgatónak látszott, mintsem akadémikusnak... Intett Nash-nek, hogy üljön le. Természetesen tudta, hogy ki ő, de látszólag zavarba hozta látogatása. Figyelmesen hallgatta, miközben fejét enyhén oldalra billentette, és ujjaival halkan dobolt. Nash hozzákezdett, hogy ismertesse az általa kigondolt bizonyítást a játék egyensúlypontjának igazolására több mint két játékos esetén. Am néhány csapongó mondat után Neumann félbeszakította, előreugrott Nash érvelésének még meg nem fogalmazott konklúziójára, és röviden annyit mondott: De hisz ez triviális! Ez csak egy fixponttétel.” (Sylvia Nasar, 2002)

A nem teljes információs játékok elmélete

Neumann János munkája inspirálólag hatott egy másik magyar származású tudós, *Harsányi János* munkásságára is, amelyért 1994-ben Nobel-díjat kapott. A magas elismerést a *nem teljes információs játékok* elméletének kidolgozásáért kapta. Ezek olyan játékok, amelyekben a játékosok nem rendelkeznek teljes információval a stratégiálmazokról és/vagy a kifizetőfüggvényekről. Maradjunk most csak annál az esetenél, amikor minden játékos ismeri a saját kifizetőfüggvényét, de a többi játékos kifizetőfüggvényeinek bizonyos paramétereikhez csak egy szubjektív valószínűség-eloszlást tud hozzárendelni. Például az oligopólium-modellben a játékosok pontosan ismerik a saját költségfüggvényeiket, de a többiekéről csak becsléseik vannak.

Harsányi János bevezette a *játékos típus* fogalmát. Az egyszerűség kedvéért feltesszük, hogy minden játékos típusa véges számú. Az oligopólium-modellben, tegyük fel, hogy minden játékos csak két típusú lehet: alacsony költséggel vagy magas költséggel dolgozó vállalat. Az i -edik játékos tudja a saját t_i típusát, de a többiekéről csak annyit tud, hogy mi a valószínűsége annak, hogy a $(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n)$ típusprofil valósuljon meg. Harsányi János alapvető feltételezése, amit *Harsányi-doktrínának* is neveznek, hogy a típusprofiloknak van egy objektív eloszlása, és az egyes játékosok szubjektív vélekedései a többi játékos alkotta típusprofilról, ebből az eloszlásból számított feltételes eloszlások, ahol a feltétel az adott játékos típusa.

Célszerű a helyzetet úgy elképzelni, hogy először a „Természet” egy minden játékos által ismert objektív valószínűség-eloszlás szerint kisorsol egy típusprofil. Minden játékos ismeri a saját típusát, de a többiekéről csak „valószínűségi ismeretei” vannak, amelyeket az objektív eloszlásból számol. Stratégiáját ennek az információnak a birtokában választja meg, és a várható kifizetését igyekszik maximalizálni. Harsányi János egy új, most már teljes információs játékot definiál, amelyben az egyes játékosok stratégiái azok a függvények, amelyekben a független változó a játékos saját típusa, a függő változó pedig egy stratégia-választás az eredeti stratégiálmazból. A kifizetések az objektív valószínűség-eloszlás segítségével számított várható kifizetések. Ennek a játéknak a Nash-egyensúlypontjait tekintjük az eredeti nem teljes információs játék egyensúlypontjának.

Megpróbáljuk az egész gondolatmenetet világosabbá tenni a tizenegyesrúgó játék némi módosításával. Egyrészt egyszerűsítünk egy kicsit, és csak azt engedjük meg, hogy a Rúgó jobbra vagy balra rúgja a büntetőt, és a Kapus is csak vagy jobbra vagy balra készülhet. Más részről bonyolítunk azzal, hogy a Rúgó játékos lehet jobb- vagy ballábás, a Kapus pedig jobb- vagy balkezes. Ezek lesznek a játékosok típusai. A jobblábás játékos jobban tud jobbra rúgni, mint balra és a jobbkezes kapus jobban tud jobbra vetődni, mint balra. Mindenki tudja magáról, hogy milyen lábás vagy kezes, de nem tudja bizonyossággal a másiktól, hogy a jobb oldala vagy a bal oldala az erősebb. Annyit tud mindenki, hogy a jobbkezesek és jobblábások (objektív) valószínűség-eloszlása a következő:

	Jobbkezes	Balkezes
Jobblábás	0,63	0,07
Ballábás	0,27	0,03

A Harsányi-doktrína ebben az esetben azt jelenti, hogy a jobblábás Rúgó szubjektív valószínűségei, amelyeket a Kapus típusához rendel: $0,63/(0,63+0,07) = 0,9$ a valószínűsége,

hogy a Kapus jobbkezes és $0,07/(0,63+0,07) = 0,1$ a valószínűsége, hogy a Kapus balkezes. A 0,9 és 0,1 tulajdonképpen annak a valószínűségei, hogy a Kapus jobb-, illetve balkezes, feltéve, hogy a Rúgó jobblábás. (Itt a Rúgó és a Kapus típusa függetlenek, de más példákban ez nincs feltétlenül így.)

Négyféle Rúgó–Kapus páros van, a sikeres büntetőik számát (10-ből) az alábbi táblázatok mutatják:

		Kapus				Kapus			
		J	B			J	B		
Rúgó	6	9					Rúgó	7	9
	8	5						7	4
		Kapus				Kapus			
		J	B			J	B		
Rúgó	4	7					Rúgó	5	8
	9	7						9	6

Ehhez a nem teljes információjú játékhoz az alábbi teljes információs mátrixjátékot rendeljük hozzá, amelyben mindkét játékosnak 4-4 tiszta stratégiája van.

A Rúgó játékos tiszta stratégiái:

y^{JJ} : ha a Rúgó jobblábás, akkor jobbra rúgja a büntetőt, ha ballábás, akkor is jobbra rúgja;

y^{JB} : ha a Rúgó jobblábás, akkor jobbra rúgja a büntetőt, ha ballábás, akkor balra rúgja;

y^{BJ} : ha a Rúgó jobblábás, akkor balra rúgja a büntetőt, ha ballábás, akkor jobbra rúgja;

y^{BB} : ha a Rúgó jobblábás, akkor balra rúgja a büntetőt, ha ballábás, akkor is balra rúgja;

A Kapus tiszta stratégiái:

z^{JJ} : ha a Kapus jobbkezes, akkor jobbra vetődik, ha balkezes, akkor is jobbra vetődik;

z^{JB} : ha a Kapus jobbkezes, akkor jobbra vetődik, ha balkezes, akkor balra vetődik;

z^{BJ} : ha a Kapus jobbkezes, akkor balra vetődik, ha balkezes, akkor jobbra vetődik;

z^{BB} : ha a Kapus jobbkezes, akkor balra vetődik, ha balkezes, akkor is balra vetődik.

A kifizetómátrix, amelyben a sikeres büntetőeknek az egyes Rúgó–Kapus párosok valószínűségeivel súlyozott átlaga szerepel, a következő:

	z^{JJ}	z^{JB}	z^{BJ}	z^{BB}
y^{JJ}	5,50	5,73	8,10	8,43
y^{JB}	6,97	7,02	8,32	8,37
y^{BJ}	6,76	6,64	5,68	5,56
y^{BB}	8,23	7,93	5,80	5,50

Ennek a mátrixjátéknak az egyensúlyi stratégiái: $y^{JB} = 0,64$, $y^{BB} = 0,36$, $z^{JB} = 0,76$, $z^{BB} = 0,24$, minden egyéb tiszta stratégia valószínűsége 0. A Rúgó játékos tehát, ha jobblábás (Ő tudja ezt!), az esetek kb. kétharmad részében jobbra lövi a büntetőt, egyharmad részében balra. Ha ballábás, akkor mindig balra. A Kapus ha jobbkezes, akkor az esetek kb. háromnegyed részében jobbra vetődik, egynegyed részében balra, ha pedig balkezes, akkor mindig balra.

Felmerülhet a kérdés: Ha mindenki tudja, hogy ő maga milyen, miért kell foglalkozni azzal is, mintha másmilyen lenne? Azért, mert ő ugyan tudja, hogy például ballábas, de a Kapus nem tudja bizonyosan, és ennek megfelelően választ stratégiát, és a Rúgónak ezt is figyelembe kell vennie.

A kooperatív játékok elmélete

Eddig olyan esetekkel foglalkoztunk, amikor a játék szabályai nem engedték meg azt, hogy a játékosok összehangolják cselekedeteiket annak érdekében, hogy közösen nagyobb haszonhoz jussanak. Ha ezt megengedjük, akkor a *kooperatív játékokhoz* jutunk el, amelynek szinte a teljes elméleti alapvetése Neumann Jánostól és Oskar Morgensterntől származik. Az ő gondolatmenetüket fogjuk követni. Olyan helyzetekkel foglalkozunk, amikor a kooperatív cselekvés a legésszerűbb, és a konfliktus abból adódik, hogy a közösen megszerzett haszonból a játékosok minél többet szeretnének maguknak megszerezni. A *közösen megszerezhető haszon* fogalma megkívánja, hogy a hasznosságokat össze lehessen adni és szétosztani. Ezt fel is tesszük a továbbiakban.

Legegyszerűbb, ha a hasznosságra mint pénzre gondolunk, amit szabadon lehet mozgatni az egyes játékosok között. Továbbra is adottnak vesszük a játékosok N véges halmazát. Ezek bármely S részhalmazát *koalíciónak* hívjuk, megengedve, hogy az üres halmaz egy olyan koalíció, amelynek egyetlen tagja sincs. A koalíciók halmazán értelmezünk egy ν függvényt, amelyet *karaktisztikus függvénynek* nevezünk. A $\nu(S)$ értéket úgy tekintjük, mint azt a maximális hasznot, amelyet az S koalíció el tud érni akkor, ha teljes mértékben csak a saját erejére támaszkodik, és nem igényli az S -en kívüli játékosok közreműködését. Az üres koalícióhoz mindig a 0 értéket rendeljük.

Példaként gondoljunk egy lakóközösségre (például egy falu és a hozzá tartozó tanyák), amely kábeltelevíziós hálózatot szeretne létesíteni. Itt az egyes házak, illetve lakások a játékosok. Legyen S egy koalíció, és jelölje $c(S)$ azt a költséget, amely akkor merül fel, ha csak az S koalíció tagjait akarjuk kábeltelevízióval kiszolgálni. $c(N)$ tehát annak a költsége, hogy mindenki megkapja ezt a szolgáltatást. Ennek a költségkímélő volta nyilvánvaló, de az egyes játékosok számára egyáltalán nem közömbös, hogy hogyan osztják fel a résztvevők a költségeket egymás között. Az egyes játékosok annak az alternatív lehetőségét is mérlegelik, hogy mi történne, ha egy megfelelő koalíció tagjaként kevesebbet kellene fizetniük, mint akkor, ha mindenki koalíciót alkot.

Ebből a rövid kis elemzésből is látszik, hogy milyen összetett probléma annak a meghatározása, hogy milyen költségosztási mechanizmust válasszon a közösség, illetve melyek azok a mechanizmusok, amelyek bizonyos értelemben stabilnak és elfogadhatóknak tekinthetők. A költség- és a hasznosztási probléma lényegében ugyanaz, mivel a $\nu(S) = -c(S)$ definícióval a költségosztást vissza tudjuk vezetni a hozamsztosztásra. Egy kooperatív játékot tehát a $\{G = N, \nu\}$ formában, a játékosok halmazával és a karakterisztikus függvénnyel adunk meg.

Igazából csak azokkal a játékokkal érdemes foglalkozni, amelyekre fennáll, hogy

$$\sum_{i \in N} \nu(\{i\}) < \nu(N),$$

ahol $\nu\{i\}$ az i -edik játékos alkotta egyszemélyes koalíció. Ezeket a játékokat *lényegesnek* nevezzük. Lényeges játékokban mindenki együtt, összefogva többet tud elérni, mintha mindenki csak egyedül cselekedne. Noha ez nem mindig szükséges, Neumann János feltette, és mintegy a kooperatív játékok tulajdonságaként kezelte a *szuperadditivitást*. A $G = \{N, \nu\}$ kooperatív játékot szuperadditívnek nevezzük, ha

$$\nu(S) + \nu(T) \leq \nu(S \cup T),$$

ahol $S \cup T$ a két koalíció egyesítése (uniója), fennáll minden olyan S és T koalícióra, amelyeknek nincsenek közös tagjaik. Ez a követelmény azt fejezi ki, hogy együtt nem járhatunk rosszabbul, mint külön-külön. Ez persze nem mindig van így, de a szuperadditív játékok osztálya még mindig olyan bő, hogy ennek kikötését nem tekinthetjük lényeges korlátozásnak.

Neumann János nyomán nevezzük *szétosztásnak* (imputation) az $\mathbf{x} = (x_1, \dots, x_n)$ számok együttesét, ha *egyénileg racionálisak*, vagyis $x_i \geq \nu\{i\}$ minden i -re és ha összegük $\nu(N)$. Az egyéni racionalitás azt követeli meg, hogy mindenki kapjon legalább annyit, mint amennyit egyedül is el tudna érni.

A közgazdaságtan egyik fontos fogalmához jutunk el, ha nemcsak az egyéni racionalitást követeljük meg, hanem az ún. *koalíciós racionalitást* is, vagyis, hogy

$$\sum_{i \in S} x_i \geq \nu(S)$$

fennáll minden S koalícióra. Azoknak a szétosztásoknak a halmazát, amelyek koalíciósan racionálisak, a *játék magjának* nevezzük. A mag egyfajta stabilitást fejez ki: ha egy olyan szétosztást javasol valaki, amely eleme a magnak, akkor nincs olyan koalíció, amelynek érdemes lenne elhagyni az N nagykoalíciót, hiszen egyedül nem tudnak elérni többet, mint amennyit a nagykoalícióban kapnak.

Neumann János a magnak egy alternatív jellemzését adta meg, amely szuperadditív játékoknál ekvivalens a fenti definícióval. Ehhez szükséges a *dominancia* fogalmának a bevezetése, ami egyébként is központi szerepet játszik Neumann Jánosnál. Ha adva van két szétosztás, $\mathbf{x} = (x_1, \dots, x_n)$ és $\mathbf{y} = (y_1, \dots, y_n)$, és egy S koalíció, akkor azt mondjuk, hogy \mathbf{x} dominálja \mathbf{y} -t az S koalíción keresztül, ha $x_i > y_i$ fennáll minden $i \in S$ -re és $\sum_{i \in S} x_i \leq \nu(S)$. Általában, \mathbf{x} szétosztás akkor dominálja az \mathbf{y} szétosztást (röviden \mathbf{x} *dom* \mathbf{y}), ha van olyan S koalíció, hogy \mathbf{x} dominálja \mathbf{y} -t az S koalíción keresztül.

A dominanciát úgy kell elképzelnünk, hogy ha éppen az a javaslat, hogy a $\nu(N)$ közös hasznat az \mathbf{y} szerint osszák szét a játékosok, akkor \mathbf{x} *dom* \mathbf{y} esetén van olyan S koalíció, amelynek egyszerre van meg a motivációja (minden tagja többet fog kapni az \mathbf{x} szétosztásban), és az ereje (ennyit maguk is produkálni tudnak) ahhoz, hogy az \mathbf{y} javaslatot visszautasítsa. Meg lehet mutatni, hogy szuperadditív játékok esetében a *mag éppen a nem dominált szétosztások halmaza* és így stabil abban az értelemben is, hogy a magban lévő bármely szétosztás esetén sincs olyan koalíció, amelynek egyszerre lenne meg a motivációja és az ereje is ahhoz, hogy elutasítsa az adott szétosztást.

A maggal kapcsolatban az egyik *alapvető probléma* az, hogy nagyon *gyakran üres*, és ekkor természetesen semmilyen előrejelzésre nem alkalmas. Neumann János mutatta meg először azt az egyszerű tényt, hogy a *konstans összegű játékok* esetében a mag mindig üres. (Egy $G = \{N, \nu\}$ játékot konstans összegűnek nevezünk, ha $\nu(S) + \nu(N \setminus S) = \nu(N)$ minden S koalíció esetén, ahol $N \setminus S$ jelöli az S -be nem tartozó játékosok koalícióját.) Minthogy a konstans összegű játékok az összes játékok között olyan nagy helyet foglalnak el, Neumann

János és Oskar Morgenstern úgy érezték, hogy egy másik, szintén a dominancián alapuló megoldástípus megfelelőbb lesz, mint a mag. Így jutottak el ahhoz a megoldáshoz, amelyet többnyire *stabil halmaznak*, vagy *Neumann–Morgenstern-megoldásnak* neveznek.

Mindkét elnevezés jogos, hiszen ez a megoldás is a stabilitás egy sajátos formáját állítja a középpontba. Az összes szétosztások egy V halmazát stabil halmaznak nevezzük, ha egyszerre rendelkezik a belső és külső stabilitás tulajdonságával:

- *belső stabilitás*: bárhogyan is választunk két szétosztást V -ből, egyik sem dominálja a másikat,
- *külső stabilitás*: bármely olyan szétosztás esetén, amelyik nincs benne V -ben, van olyan V -beli szétosztás, amely azt dominálja.

Egy V stabil halmazt Neumann János és Oskar Morgenstern úgy értelmeznek, mint egy viselkedési szabályrendszert („standard of behavior”), amelyen belül nyugodtan lehet különböző szabályokat (szétosztásokat) választani, de gyökeres változásokra van szükség, ha egy teljesen más stabil halmazra akarnánk áttérni. Egy adott játékhoz nagyon sok stabil halmaz tartozhat, ami az előbbi értelmezés szerint egyáltalán nem zavaró. Sokkal nagyobb problémának tartotta maga Neumann János, hogy vajon létezik-e minden játékhoz legalább egy stabil halmaz. Viszonylag könnyen (azért ez sem egyszerű!) lehet megmutatni, hogy minden 3 személyes játéknak van stabil halmaza. Ezt maga Neumann János bizonyította be először. 1974-ben Bondareva és társai bizonyították be ugyanezt 4 személyes játékokra. Időközben, 1969-ben, Lucas mutatott példát egy olyan 10 személyes játékra, amelynek nincs stabil halmaza. Mind a mai napig nem tudjuk, hogy mi a helyzet az 5–9 személyes játékoknál. Neumann János eredeti sejtését (és reményét), hogy minden konstans összegű játéknak van legalább egy stabil halmaza, még sem bizonyítani, sem cáfolni nem tudták, mivel az ellenpéldák egyike sem konstansösszegű. Ráadásul, az ellenpéldák mind olyanok, hogy ha a karakterisztikus függvény értékeket egy kicsit megváltoztatjuk bennük, akkor olyan játékok keletkeznek, amelyeknek már van stabil halmazuk. Így még az is lehet, hogy a stabil halmazokkal rendelkező játékok „mindenütt sűrűn” helyezkednek el, vagyis stabil halmazzal nem rendelkező játékokhoz tetszőlegesen közel van olyan, amelynek van stabil halmaza. Ez a kérdés sincs még megválaszolva.

Ez a kis izelítő a megoldatlan problémákból jelzi a stabil halmaz matematikai bonyolultságát. A konkrét esetekre való alkalmazás is általában nehéz, egyedi módszerekre és ötletekre van szükség egyes konkrét játékok esetében arra, hogy a stabil halmazokat (vagy legalábbis egyet közülük) meg tudjuk határozni.

Nézzünk egy egyszerű példát arra, amikor a stabil halmaz viszonylag egyszerű szerkezetű. Egy $G = \{N, \nu\}$ játékot *egyszerűnek* nevezünk, ha lényeges, szuperadditív és $\nu(S) = 0$ vagy 1, minden S koalíció esetén. Azokat a koalíciókat, amelyekre $\nu(S) = 1$ *nyerő*, míg azokat, amelyekre $\nu(S) = 0$ *vesztőknek* nevezük. Ha egy demokratikus parlamentben S a pártok egy koalíciója, és egy törvényjavaslat megszavazásáról van szó, akkor S nyerő, ha összes szavazataival a törvényjavaslat átmegy, egyébként vesztes. Egy nyerő koalíciót *minimálisnak* nevezünk, ha egyetlen részkoalíciója sem nyerő.

Tekintsünk egy S minimális nyerő koalíciót, és legyen V_S mindazoknak az $\mathbf{x} = (x_1, \dots, x_n)$ szétosztásoknak a halmaza, amelyekben $x_i = 0$, ha $i \in S$. Ekkor nem nehéz megmutatni, hogy V_S egy stabil halmaz. Más szóval, az S minimális nyerő koalíció tagjai bárhogyan eloszthatják egymás között az egységnyi „hasznot”, ez az elosztás rendelkezni fog a külső és belső stabilitás tulajdonságával.

A Neumann–Morgenstern-megoldás arra vonatkozóan nem mond semmit, hogy a sok stabil halmaz közül melyik fog megvalósulni, ezt modellen kívüli tényezők (például az egyes játékosok jártassága az alkudozásban és kompromisszum keresésében, szerencse stb.) fogják végső soron meghatározni.

Noha sok más koncepció létezik a kooperatív játékok megoldására, nem olyan domináns módon, mint keletkezésekor, de a Neumann–Morgenstern-megoldás továbbra is jelentős szerepet tölt be sok konfliktushelyzet elemzésében.

Felhasznált irodalom

Arrow, K. and Debreu, G. [1954]: Existence of an Equilibrium for a Competitive Economy. *Econometrica* 22, pp. 265–90.

Arrow, K. [1989]: Von Neumann and the Existence Theorem for General Equilibrium. In: Dore, Chakravarty and Goodwin [1989], pp. 15–28.

Blaug, M. [1962]: *Economic Theory in Retrospective*. Homewood, Ill.: Irwin.

Bondareva, O. N., Kulakovskaya, T. E. és Naumova, N. I. [1979]: Négyszemélyes játékok megoldása (oroszul). *Leningrad Universitet Vestnik* 6, 104–105.

Bródy András [1986]: A fizikai gazdaságtanról. Neumann egyensúlyi modelljének fél-százados évfordulójára. *Sigma* 19, 1–2, pp. 41–48.

Bródy András, Martinás Katalin és Sajó Konstantin [1986]: Gazdasági és termodinamikai mérés. *Közgazdasági Szemle*, pp. 19–27.

Brouwer, L. E. J. [1912]: Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen*, 71, 97–115.

Cassel, G. [1918]: *Theoretische Sozialökonomie*. Leipzig: Deichert (angolul: *The Theory of Social Economy*. New York: Harcourt Brace, 1932)

Champernowne, D. G. [1945]: A note on J. von Neumann's article. *Review of Economic Studies* 13, 1, pp. 10–18.

Cournot, A. A. [1838]: *Recherches sur les principes mathématiques de la théorie des richesses*. Hachette, Paris (angolul: *Researches Into the Mathematical Principles of the Theory of Wealth*. New York: Macmillan, 1929)

Dore, M., Chakravarty, S. and Goodwin, R. (szerk.) [1989]: *John von Neumann and Modern Economics*. Oxford: Clarendon Press

Ekelund, R. B. and Hébert, R. F. [1997]: *A History of Economic Theory and Methods*. 4th ed., Toronto–London: McGraw Hill

Forgó, F., Szép, J. and Szidarovszky, F. [1999]: *Introduction to the Theory of Games: Concepts, Methods, Applications*. Kluwer Academic Publishers, Dordrecht

Gale, D. [1960]: *The Theory of Linear Economic Models*. McGraw–Hill, New York

Harsanyi, J. C. [1967]: Games with incomplete information played by „Bayesian” players I–III. *Management Science* 14, 159–182, 320–334, 486–502.

Hegedűs M. és Zalai E. [1978]: *Fixpont és egyensúly a gazdasági modellekben*. Közgazdasági és Jogi Könyvkiadó, Budapest

Hicks, J. R. [1939, 1978]: *Érték és tőke*. Közgazdasági és Jogi Könyvkiadó, Budapest (*Value and Capital*. Oxford University Press.)

Hicks, J. R. [1984]: The formation of an Economist. In Hicks, J. R., *The Economics of John Hicks*. Oxford: Basil Blackwell, pp. 281–90.

Ingrao, B. and Israel, G. [1990]: *The Invisible Hand: Economic Equilibrium in the History of Economic Science*. Cambridge: MIT Press

Kakutani, S. [1941]: A generalization of Brouwer's fixed point theorem, *Duke Journal of Mathematics* 8, 457–459.

Káldor Miklós [1989]: Foreword. John von Neumann: A personal recollection. In: *Dore, Chakravarty and Goodwin* [1989], pp. vii–xi.

Kemeny, J. G. and Morgenstern, O. and Thompson, G. L. [1956]: A Generalization of von Neumann's Model of an Expanding Economy. *Econometrica* 24, pp. 115–35.

Koopmans, T. C. (szerk.) [1951]: *Activity Analysis of Production and Allocation*. New York: John Wiley and Sons

Koopmans, T. C. [1974]: *Contribution to General Discussion on Past and Future of the von Neumann Model*. In: [1974], pp. 3–4.

Kornai János [1971]: *Anti-equilibrium*. Közgazdasági és Jogi Könyvkiadó, Budapest

Lange, O. [1964]: *Politikai gazdaságtan. I.* Közgazdasági és Jogi Könyvkiadó, Budapest

Leontief, W. [1928]: *Die Wirtschaft als Kreislauf*. Archiv für Sozialwissenschaft und Sozialpolitik, 60, pp. 577–623.

Leontief, W. [1941]: *The Structure of the American Economy*. Cambridge, Mass.: Harvard University Press

Łoś, J. and M. W. Łoś (szerk.) [1974]: *Mathematical Models in Economics*. Amsterdam and New York: North-Holland

Lucas, W. F. [1968]: A game with no solution. *Bull. Amer. Math. Soc.* 74, 237–239.

Marshall, A. [1890]: *Principles of Economics*. (Reprint, 1977, London: Macmillan)

Mátyás Antal [1992]: *A korai közgazdaságtan története*. Aula Kiadó

McKenzie, L. [1954]: On Equilibrium in Graham's Model of World Trade and Other Competitive Systems. *Econometrica* 22, pp. 147–61.

Morgenstern, O. and Thompson, G. L. [1976]: *Mathematical Theory of Expanding and Contracting Economies*. Lexington: Lexington Books

Morgenstern, O. [1976]: Collaborating with von Neumann. *Journal of Economic Literature*, Sept., 14(3), pp. 805–16.

Morishima, M. [1964]: *Equilibrium, Stability and Growth*. Clarendon Press, Oxford

Nasar, S. [2002]: *Egy csodálatos elme: A Nobel-díjas matematikus géniusz, John Nash élete*. GABO Kiadó, Budapest

Nash, J. [1950]: Equilibrium Points in n-Person Games. *Proceedings of the National Academy of Sciences*, USA, 36, pp. 48–49.

Neumann János [1928]: Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100, pp. 295–320. (Magyarul: A társasjátékok elméletéhez. In: Neumann [1965], pp. 121–156.)

Neumann, J. von and Morgenstern, O. [1944]: *Theory of Games and Economic Behavior*. Princeton University Press, Princeton

Neumann János [1937, 1945, 1965]: Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes. *Ergebnisse eines mathematischen Kolloquiums*, 8, pp. 73–83. (Angolul: A Model of General Economic Equilibrium. *Review of Economic Studies* 13, pp. 1–9., magyarul: Az általános gazdasági egyensúly egy modellje. In: Neumann [1965], pp. 160–176.)

Neumann János [1947, 1965]: The Mathematician. In: Robert B. Heywood (ed.), *The Works of Mind*. Chicago: University of Chicago Press, 180–196. o. (Magyarul: A matematikus. In: Neumann [1965], pp. 11–27.)

Neumann János [1956, 1965]: Looking ahead. (Magyarul: A legújabb tudományos fejlődés hatása a gazdaságra és a közgazdaságtanra. In: Neumann [1965], pp. 100–102.)

Neumann János [1965]: *Válogatott előadások és tanulmányok*. (Ford. Augusztinovics M.) Közgazdasági és Jogi Könyvkiadó, Budapest

Pigou, A. C. [1925]: *Memorials of Alfred Marshall*. London. Macmillan

Punzo, L. F. [1989]: Von Neumann and Karl Menger's Mathematical Colloquium. In: *Dore, Chakravarty and Goodwin* [1989], pp. 29–65.

Punzo, L. F. [1991]: The School of Mathematical Formalism and the Viennese Circle for Mathematical Economists. *Journal of the History of Economic Thought*, 13, pp. 1–18.

Samuelson, P. A. [1947]: *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press

Samuelson, P. A. [1989]: A Revisionist View of von Neumann's Growth Model. In: *Dore, Chakravarty and Goodwin* [1989], pp. 100–122.

Schlesinger, K. [1935]: Über die Produktionsgleichungen der ökonomischen Wertlehre. *Ergebnisse eines mathematischen Kolloquiums*, 6, pp. 10–11.

Sraffa, P. [1960, 1975]: *Áruk termelése áruk révén*. Közgazdasági és Jogi Könyvkiadó, Budapest. (*Production of Commodities by Means of Commodities. Prelude to a Critique of Economic Theory*. Cambridge: Cambridge University Press.)

Szép Jenő és Forgó Ferenc [1974]: *Bevezetés a játékelméletbe*. Közgazdasági és Jogi Könyvkiadó, Budapest

Wald Ábrahám [1935]: Über die eindeutige positive Lösbarkeit der neuen Produktionsgleichungen. *Ergebnisse eines mathematischen Kolloquiums*, 6, pp. 12–18.

Wald Ábrahám [1936]: Über die Produktionsgleichungen der ökonomischen Wertlehre (II. Mitteilung). *Ergebnisse eines mathematischen Kolloquiums*, 7, pp. 1–6.

Walras, L. [1874, 1877]: *Elements d'Economie Politique Pure*. (*Elements of Pure Economics*. London: Allen & Unwin, 1954)

Weintraub, E. R. [1983]: On the Existence of a Competitive Equilibrium: 1930–1954. *Journal of Economic Literature*, 21, pp. 1–39.

Weintraub, E. R. [1985]: *General Equilibrium Analysis: Studies in Appraisal*. Cambridge University Press

Weintraub, E. R. [1991]: *Stabilizing Dynamics: Constructing Economic Knowledge*. Cambridge: Cambridge University Press

Weintraub, E. R. and Mirowski, P. [1994]: The Pure and the Applied: Bourbakism Comes to Mathematical Economics. *Science in Context*, No. 2., pp. 245–272.

Zalai Ernő [1998]: Általános egyensúlyi modellek alkalmazása gazdaságpolitikai elemzésekre. *Közgazdasági Szemle*, 12, pp. 1065–81.

Zalai Ernő [1999]: A közgazdaságtan metodológiájáról és a matematikai közgazdaságtanról a Neumann-modell ürügyén. *Közgazdasági Szemle*, 13, pp. 600–629.

Zalai Ernő [2000]: *Matematikai közgazdaságtan*. KJK–Kerszöv, Budapest

Zeuthen, F. [1933]: Das Prinzip der Knappheit, technische Kombination und ökonomische Qualität. *Zeitschrift für Nationalökonomie*, 4, pp. 1–24.

Götz Gusztáv

Neumann János jelentősége a meteorológiában

Neumann János sokrétű munkásságában mindössze egy-egy villanás volt, amikor meteorológiai kérdések foglalkoztatták, ám ezek az epizódok meghatározó jelentőségűnek bizonyultak a tudományterület további fejlődésében. Ahogyan az időjárás előrejelzésének kérdését ma kezeljük, amiképp az éghajlat megváltozásának lehetőségét napjainkban elemezzük, alapjaiban különbözik a Neumann tevékenységét megelőző idők gyakorlatától.

Amikor Neumann 1945 vége felé meggyőzte a princetoni Institute for Advanced Study vezetését, hogy vállaljon erkölcsi és anyagi elkötelezettséget egy elektronikus számítógépes terv megvalósításáért, az időjárás-előrejelzések készítése már közel százéves múltra tekintett vissza. Neumann ezt az évszázados tevékenységet kívánta a számítástechnika alkalmazásával gyökeresen átalakítani, lényegesen objektívebb alapokra helyezni. Milyen nehézségekkel kellett az elődöknek megküzdeniük az elképzelés valóra váltásához? Hogy a választ világosan lássuk, menjünk vissza időben a kezdetekig.

1. Az időjárás-előrejelzések készítésének kezdetei

A várható időjárás előrejelzésére irányuló régi vágyat sürgető igénnyé léptette elő a Fekete-tengerre 1854. november 14-én lesújtó vihar, amely elsüllyesztette a krími háborúban részt vevő, Szevasztopol ostromához Balaklavánál felsorakozott angol–francia flotta 38 hajóját. A katasztrófáról értesülve, III. Napóleon császár a párizsi csillagvizsgáló igazgatóját bízta meg, hogy derítse ki, vajon előre jelezhető lehetett volna-e a szélvihar érkezése. A neves csillagász, Urbain Leverrier akkor már bámulatosan pontos prognózisáról volt híres: néhány évvel korábban kiszámította, milyen pályán kell keringenie egy bizonyos tömegű, ismeretlen bolygónak, hogy gravitációs hatásával az Uránusz mozgásának észlelt zavarai megmagyarázhatóvá váljanak – és Johann Galle 1846-ban, az előre jelzett helyen valóban megtalálta a keresett, Neptunusz névre keresztelt égitestet. A császári rendelet nyomán Leverrier levelet intézett Európa csillagászaival és meteorológusaihoz, kérve őket, hogy küldjenek el neki minden olyan adatot, amely a november 12-e és 16-a közötti időjárásra

vonatkozóan a birtokukban van. A kapott 250 válasz alapján megszerkesztette az első áttekintő időjárás térképeket, és megállapította, hogy a vihar, mielőtt megérkezett a Fekete-tengerre, nyugat felől, pontosan követhetően átvonult a Földközi-tenger medencéje fölött. Ezért 1855 februárjában javaslatot dolgozott ki egy meteorológiai megfigyelőhálózat létrehozására, és az adatok távirón történő összegyűjtésére.

Leverrier javaslatát a francia kormány kedvezően fogadta. 1856-ra kiépült egy 24 állomás észleléseire támaszkodó nemzeti időjelző szolgálat, amelyhez később más európai államok is csatlakoztak. A célt a hajózás biztonságosabbá tétele jelentette: a kikötőkben horgonyzó hajókat árbcra felvont gömbök és kúpok meghatározott elrendeződése figyelmeztette a várható időjárás veszélyekre. Robert FitzRoy brit admirálist annyira elbűvölte a viharjelzés feladata, hogy 1861-ben elkezdte a napi sajtóban nyilvánosságra hozni időjárás-előrejelzéseit. Ténykedése azonban nem aratott osztatlan sikert; azt sokan bomlasztónak, a sarlatánok és csillagjósok tevékenységéhez túlságosan közelinek találták. Leverrier maga levelet írt neki, kérve, hogy ne diszkreditálja viharjelző szolgálatát. Nem tudjuk, ez a vélemény hozzájárult-e ahhoz, hogy az admirális 1865 áprilisában, búskomor rohamban önkéntesül vetett véget életének. Az viszont tény, hogy halála után a londoni Royal Society a brit kormánynak mind a viharjelzések, mind az időjárás-előrejelzések beszüntetését javasolta. A tudós társaság javaslatának megvalósítását csak a tengerészek és a nagyközönség tiltakozása tudta megakadályozni.

A társadalmi elváráshoz a kontinensen is alkalmazkodni kellett: az 1855 és 1880 között sorra szerveződő nemzeti meteorológiai szolgálatok közül nem egyet kifejezetten az időjárás-előrejelzések deklarált feladatával bíztak meg. Hazánkban Trefort Ágoston vallás- és közoktatási miniszter 1888 júniusában kelt leiratában hatalmazta fel a Meteorológiai Intézet igazgatóját az időjárás szolgálat megindítására, és az előrejelzések az 1894-ben létrehozott Telefontáviró napi programjába is bekerültek. Az évtizedek során, a tudomány és a technika fejlődésével párhuzamosan, természetesen szépen gazdagodott az előrejelzések módszertana, egyre több megfigyelési adat egyre gyorsabban állt rendelkezésre. De az alapvető eljárás lényegében mit sem változott: megmaradt azon a szubjektív tapasztalaton alapuló szinten, ami az első éveket is jellemezte. A gyakori tévedések miatt a közvélemény a meteorológust nem ismerte el egy egzakt tudomány művelőjeként.

2. Az időjárás objektív előrejelzésének első kudarca

Követhető lett volna-e mindjárt a kezdetektől szubjektív eljárások helyett egy lényegesen objektívebb időjárás-előrejelzési technika alkalmazása? *Elvileg* igen, hiszen ennek módját Vilhelm Bjerknes norvég fizikus már 1904 januárjában felvázolta. Ma is érvényes elgondolását érdemes szó szerint idéznünk: „Ha valóban úgy van, ahogyan azt minden természettudományos alapon gondolkodó ember véli, miszerint a légkör későbbi állapotai *törvényszerűen* a megelőző állapotokból fejlődnek ki, akkor fel kell ismernünk, hogy a meteorológiai prognosztika problémája ésszerű megoldásának szükséges és elégséges feltételei a következők: (1) Elegendő pontossággal ismernünk kell a légkörnek valamely idő-

pontbeli állapotát. (2) Elegendő pontossággal ismernünk kell azokat a törvényeket, amelyeknek engedelmessé az egyik légköri állapot a másikból kifejlődik.”

A Bjerknies által említett törvények – a Newton-féle impulzustétel, a tömeg megmaradásának elve és a termodinamika első főtétele –, továbbá az állapotjelzők közötti összefüggést kifejező állapotegyenlet a 19. század közepén már ismertek voltak. E törvények együttese alkotja a légköri folyamatok hidro-termodinamikai egyenletrendszerét, amely a levegő áramlási sebességének, sűrűségének és hőmérsékletének megváltozását matematikailag leírja. A kezdeti feltételek ismeretében tehát elvileg nincs semmi akadály a légkör állapota bármely jövőbeli időpontra történő kiszámításának. De az egyenletek mindegyike (a mozgásegyenletek, a folytonossági egyenlet és a termodinamikai energiaegyenlet is), matematikai alakját tekintve, nemlineáris parciális differenciálegyenlet – az ilyen típusú egyenleteknek pedig nem tudjuk megadni a folytonos analitikus megoldását. És ezzel kerülünk szembe az objektív időjárás-előrejelzés első *gyakorlati* nehézségével: közelítő eljárás alkalmazására kényszerülünk. Ennek során az állapotjelzők térbeli differenciálhányadosait egy rácshálózat diszkrét pontjaira interpolált értékek véges különbségeivel (távolságnövekményekkel) helyettesítjük, az idő szerinti differenciálhányadosokat pedig szintén véges időnövekményekkel (időlépcsőkkel) kell kiváltanunk. Ez a közelítő eljárás a numerikus integrálás, amely annál pontosabb, minél kisebb (rövidebb) véges különbségekkel közelítjük a differenciálhányadosokat, ami viszont hatalmas számtani műveletsor elvégzését követeli meg.

Az egyetlen, kivételes egyén, akit nem riasztott el a feladat vázolt munkaigénye, az angol Lewis Richardson volt. Már javában dúlt az első világháború, amikor hozzákezdett egy jól dokumentált időpont, 1910. május 20-a reggel 7 óras európai időjárás helyzetének hatórás távra szóló számszerű előrejelzéséhez. Mivel a protestáns kvéker felekezet szigorú erkölcsi szellemében nevelkedett és meggyőződéses pacifista volt, elutasította a fegyveres harcban való közvetlen részvételt. Vállalta viszont, hogy a francia hadseregnek kölcsönzött egyik mentőautó vezetőjeként sebesülteket szállítson a frontvonal mögé. Számításait – az egyenletek numerikus integrálásához szükséges aritmetikai műveletek ezreit – szabad óráiban, igen mostoha körülmények között végezte („Irodám egy halom széna volt az egyik hideg pihenőszálláson” – írta később visszaemlékezéseiben). 1917 áprilisában, a champagne-i ütközet során, sikerült kéziratát a frontvonal mögé juttatnia, ahol azonban nyoma veszett, és csak hónapok múltán leltek rá véletlenül Belgiumban, egy szénkúpac alatt.

A munkáról szóló részletes beszámolót 236 oldalas könyv formájában, 1922-ben jelentette meg a cambridge-i egyetem kiadója – annak ellenére, hogy az előrejelzés tökéletes kudarcnak bizonyult. „A legvadabb becslés sem állhatott volna távolabb a realitástól!” – jegyezte meg főnöke, Sir Napier Shaw, a brit meteorológiai hivatal vezetője, amikor meglátta, hogy a számítások a Neckar völgyének térségére hat óra leforgása alatt 145 hPa-os légnyomás emelkedést eredményeztek. Ezt a reális értéknél két nagyságrenddel nagyobb (borzalmas vihart képviselő) változást a kiindulási mezőkben jelen levő nagyfrekvenciás gravitációs hullám-összetevők idézték elő. Továbbá – amint azt Richardson maga is világosan felismerte – a megfigyelések sűrűsége és pontossága egyszerűen elégtelen volt ahhoz, hogy megbízható számításokat lehessen elvégezni. Sok-sok év telt el, mire a szakemberek kiderítették, hogy a nehézséget miként lehet megkerülni.

Hibátlan lett volna Richardson előrejelzése, ha kiindulási adatai pontosak és a gravitációs hullámoktól mentesek lettek volna? Ma már tudjuk, hogy nem, de ennek okára csak

néhány esztendő követően, 1928-ban jött rá a neves német matematikus, a később az Egyesült Államokba kivándorolt Richard Courant, és két göttingeni munkatársa, Friedrichs és Lewy. Ennek az újabb gyakorlati nehézségnek a lényege, hogy a légköri folyamatokat kormányzó egyenletrendszer *általános jellegű*, tehát a meteorológiailag fontos mozgásformákon kívül a légkörben terjedő hanghullámokat és a gravitációs hullámokat is leírja. De mivel a nagytérségű időjárás folyamatokat olyan általános állapot jellemzi, amelyben a három meghatározó erő, nevezetesen a gravitáció, a légnyomási gradiens és a Coriolis-erő közel egyensúlyban van (és így a légkörben – a nagy zivatarfelhők környezetétől eltekintve – nem alakulnak ki jelentős gyorsulások sem a függőleges, sem pedig a vízszintes irány mentén), ezért ezek a nagyfrekvenciás hullámok nem jutnak domináns szerephez. Hogy a számítások eredményei a reális értéktartományon belül maradjanak, a göttingeni egyetem három matematikusa szerint az alkalmazott távolságnövekmény és időlépcső hányadosa nem lehet kisebb a közegben fellépő, leggyorsabban terjedő hullámforma sebességénél. Modelljéből Richardson a hanghullámokat kiszűrte, a másodpercenként 300 méter körüli sebességgel haladó gravitációs hullámok létrejöttének lehetőségét azonban nem akadályozta meg. Ezért az általa használt 200 kilométeres távolságnövekmény mellett az időlépcsőnek nem lett volna szabad nagyobbak lennie 200 000/300 másodpercnél, tehát 10 percnél, azaz egy 24 órás előrejelzésnek 144 tízperces előrejelzés sorozatából kellett volna felépülnie. Ezzel szemben ő 3 órás időlépcsőt választott, ami – ha hosszabb távra folytatja előrejelzését – óhatatlanul azt eredményezte volna, hogy a számítási instabilitás következtében az állapotjelzők számított értékei az idő múlásával elfutnak a végtelenbe: a numerikusan szimulált folyamat „felrobban”.

A meteorológiai irodalom egyik legkülönösebb, de egyben leggazdagabb fantáziájú könyve végén Richardson hatalmas, színházteremhez hasonlító, képzeletbeli időjárás központot tár az olvasó elé, amelyben egymást váltva 64 000 ember dolgozik kézi kalkulátorokkal, hogy az előrejelzést az időjárás eseményeket megelőzve készítse el. „A ködös jövő egy napján talán lehetséges lesz a számításokkal az időjárás haladásánál gyorsabban előre jutni, olyan áron, amely kisebb, mint az emberiségnek a nyert információból származó megtakarítása. De ez álom...” – írta.

Neumann János volt az, aki segített valóra váltani ezt az álmot. Ennek azonban előbb meg kellett teremteni az elméleti meteorológiai előfeltételeit, ami viszont több mint két évtizedet igényelt. A kor élenjáró szakemberei ugyanis, Richardson kudarcának láttán, a gyakorlatban teljesen járhatatlannak ítélték meg a Bjerknés által felvázolt utat. Bernhard Haurwitz chicagói egyetemi tanár, az elméleti meteorológia elismert szakértője úgy vélekedett 1941-ben írt könyvében, hogy „a jövőbeli időjárásnak a termodinamika és a dinamika egyenleteinek direkt alkalmazására épülő előrejelzése ma nem sokat ígérő vállalkozásnak látszik”. Richardson kísérletére utalva, azt a következtetést vonta le, hogy „a várható időjárás dinamikai módszerekkel történő kiszámítása csak akkor válik majd lehetővé, amikor már határozottabban tudjuk, hogy adott feltételek esetén mely tényezőket kell figyelembe venni, és melyek azok, amelyek elhanyagolhatók”. Ez a nézet egészen addig uralkodott, amíg fel nem tűnt ismét egy kivételes egyéniség, ez alkalommal a Massachusetts Institute of Technology professzora és Neumann jó barátja, a svéd származású Carl-Gustaf Rossby, aki a stockholmi egyetemen Bjerknés tanítványa volt. Megjelent továbbá egy igen tehetséges fiatalember, a San Francisco-i születésű Jule Charney, aki később a princetoni Institute for Advanced Study-ban Neumann meteorológiai programjának vezetője lett.

3. A kudarc okainak elhárítása

Charney 1940-ben már a kaliforniai egyetem Los Angeles-i részlege (a UCLA) doktorképző tanfolyamának hallgatója volt, amikor az egyetemre érkezett két norvég meteorológus, Jakob Bjerknes (Vilhelm Bjerknes gyámfia) és Jörgen Holmboe. Charney matematikusnak készült, és az asztronautika területén szeretett volna dolgozni. A meteorológiáról nem sokat tudott mindaddig, amíg 1941-ben meg nem hallgatta Holmboe egyik, matematikusoknak tartott előadását. Mivel megtetszett neki ez a szakterület, a pasadenai California Institute of Technology (a neves Caltech) tanárához, Kármán Tódorhoz fordult tanácsért, aki melegen ajánlotta neki a „sokat ígérő” meteorológusi pályát. A 24 éves diák megfogadta az aerodinamika világhírű professzorának javaslatát, ami azért is figyelemreméltó, mert ekkorra már készen állt a matematikai témájú doktori tézisének jelentős része.

(Ehhez kapcsolódóan érdemes visszaemlékezni, hogy Kármán annak idején Neumann pályaválasztásában is közreműködött. Neumann 1921-ben érettségizett, és matematikus szeretett volna lenni, ám a szigorú bankár atya, Neumann Miksa semmiképpen sem akarta, hogy fia olyan tárgyat válasszon, amely nem kecsegtet gazdagsággal. Ezért az akkorigiban az aacheni egyetemen oktató barátjához fordult, hogy beszéljen a fiával, és győzze meg, mennyivel hasznosabb az üzleti pályán karriert befutni. Kármán valószínűleg nem az a személy volt, aki alkalmas lehetett egy ilyen feladatra, de a végén mindhárman elfogadták azt a kompromisszumos javaslatot, hogy a kémia legyen az ifjú Neumann egyetemi tanulmányainak tárgya.)

Új disszertációja összeállításához Charney munkájának alapját Jakob Bjerknes és Holmboe léggöri modellje képezte. A modell matematikai hátterét a két tudós „elszomorítóan elégtelennek” ismerte el, így Holmboe asszisztensére várt a feladat, hogy megalkossa a mérsékelt övi ciklonok keletkezésének egzakt képét. Említettük már, hogy a léggöri folyamatokat kormányzó nemlineáris egyenleteknek nem tudjuk felírni az általános megoldását. Egy hidrodinamikai egyenlet speciális analitikus megoldásának a meteorológiai irodalomban leginkább ünnepeelt sikerét Rossby vívta ki magának, amikor néhány évvel korábban (egy szellemes linearizálási közelítés alkalmazásával) kimutatta, hogy ez a megoldás, a már jól ismert hanghullámok és gravitációs hullámok mellett, leír egy több ezer kilométeres hullámhosszúságú, óránként mindössze 15 kilométeres sebességgel terjedő mozgásformát is. Így vált világossá, hogy az időjárási térképeken megjelenő ciklonok és anticiklonok fejlődése elméleti szempontból nem más, mint ezeknek a (ma *Rossby-hullámoknak* nevezett) képződményeknek az amplitúdójában és helyzetében bekövetkező változás. E felfedezésre építve, Charney zseniálisan oldotta meg a kapott feladatot. Értekezése volt az, amely végre megadta a léggör nagytérségű áramlási alakzatainak kialakulására vonatkozó első, igazán elhíhető mechanizmust, és amellyel 1946-ban elnyerte a doktori fokozatot, majd emellé a nemzeti tudományos tanács ösztöndíját is.

Ösztöndíjas útja során Charney elsőször, Rossby meghívásának eleget téve, a chicagói egyetem meteorológiai tanszékét látogatta meg, majd 1947 márciusának végén áthajózott Norvégiába, és az oslói egyetem asztrofizikai intézetében a hidrodinamikai egyenletek numerikus integrálásának kérdéseit kezdte tanulmányozni. Abból az észrevételből indult ki, hogy az alkalmazott hidrodinamika más ágaiban dolgozó kutatók sikeresen szűrték ki a teljes hullámspektrum érdektelen zajnak minősülő részét az alapegyenleteknek az általuk

vizsgált mozgások jellemzőire alapozott egyszerűsítésével. A légkörnek az időjárás alakításában elsődleges szerepet játszó mozgásaira négy, az egyenletek egyszerűsítéséhez alapot nyújtó, jellemző vonás sorolható fel: e mozgásformák közel vízszintesek, közel hőcsere-mentesek, továbbá egyik irányban sem lépnek fel számottevő gyorsulások. A függőleges irányú gyorsulások elhanyagolását a Richardson által is alkalmazott *hidrosztatikus közelítés* bevezetése (a függőleges nyomási gradiens és a gravitáció egyensúlyának feltételezése) jelenti, ami egyben biztosítja a hanghullámok kiszűrését. A vízszintes mentén fellépő gyorsulások elhanyagolása a vízszintes nyomási gradiens és a Coriolis-erő egyensúlyának előírásával (a *geosztrofikus közelítéssel*) érhető el. Ekkor az egyenletek numerikus megoldásaiban nem jelennek meg gravitációs hullámok. Ám amíg a hidrosztatikus egyensúly feltételezése igen jó közelítés (csak a zivatarfelhőkön belül nem alkalmazható), addig a geosztrofikus közelítés alkalmazásával eleve kizárjuk a mozgás *megváltozásának* lehetőségét: geosztrofikus légkörben soha nem alakulnának ki időjárásunk alakulásának elsődleges meghatározói, a ciklonok és anticiklonok. Így a gyorsulási tagok *teljes* elhanyagolásával – Charney hasonlatát idézve – „a fürdővízzel együtt a gyereket is kiöntjük”. Ezt a problémát ő úgy kerülte meg, hogy a horizontális gyorsulásokat a mozgásegyenleteknek csak bizonyos tagjainál hagyta el; ez az eljárás aztán a *kvázi-geosztrofikus közelítés* néven honosodott meg.

Charney, visszatérve Oslóból, nagyszabású munkáját már Neumann János mellett, az Institute for Advanced Study-ban fejezte be. És 1949-re a meteorológia elméletileg készen állt arra, hogy az időjárás előrejelzésének Vilhelm Bjerknes által kijelölt módszere, kellő számítástechnikai segédlet birtokában megvalósuljon.

4. Sikeres előrejelzési kísérletek az ENIAC-en

Neumann János – a projektív geometriában maradandót alkotó – Oswald Veblen meghívására, 1930-ban érkezett meg kvantumelméletet előadni a New Jersey állambeli princetoni egyetemre. Alig egy esztendővel később – 28 esztendőskorában – rendes tanári katedrát kapott, és ezzel ő lett az ország legfiatalabb kinevezett professzora. 1933-ban elfogadta ugyancsak Princetonban, abban az évben megnyitott Institute for Advanced Study hat matematikaprofesszori állásának egyikét, amelyet haláláig megtartott, és amellyel Albert Einstein közvetlen kollégájává vált. A harmincas években behatóan foglalkozott a folyadékok és gázok szuperszonikus turbulens áramlásának, valamint a lökés- és robbanási hullámok terjedésének kérdéseivel. Abban az időben e terület egyik vezető szakértőjének tartották. Még 1937-ben amerikai állampolgárságot szerezve kerülhetett kapcsolatba 1943-ban a Manhattan-tervvel, és így az első atombomba előállítására létrehozott Los Alamos-i kutatócsoport tagja lett. Az állampolgárság elnyeréséhez érdekes anekdota fűződik. Neumann és Kurt Gödel (a matematikai logika osztrák származású, kiemelkedő művelője) ügyét Oskar Morgenstern karolta fel, az a német születésű matematikus és közgazdász, akivel Neumann később, 1944-ben a játékelméletről és a közgazdasági viselkedésről szóló híres művet megalkotta. A bevándorlási hivatal felé autózva, ahol mindkettőjüknek bizonyítaniuk kellett, hogy ismerik az Alkotmányt és a nemzet történelmét, Morgenstern

megkérdezte őket, van-e valami kérdésük, amiben segíthetne. Gödel azt válaszolta, hogy kérdése nincs, viszont talált néhány logikai ellentmondást az Alkotmányban, amelyről majd szeretné hallani a bevándorlási hivatalnokok véleményét. Morgenstern azt ajánlotta, hogy inkább ne tegyen fel kérdéseket, hanem csak válaszoljon nekik!

Philadelphiában, a pennsylvaniai egyetem Eliakim Moore matematikus nevét viselő villamosmérnöki karán, 1943 májusában kezdődött el a hadsereg tüzérségi ágazatának a marylandi Aberdeenben levő ballisztikai kutatólaboratóriuma számára az első, teljesen elektronikusan működő számítógép, az ENIAC építése.



1. ábra. A látogatók és résztvevők 1950-ben az ENIAC számításokban. Balról jobbra: H. Wexler, Neumann János, M. Frankel, J. Namias, J. Freeman, R. Fjörtoft, F. Reichelderfer és J. Charney. A fotó az ENIAC előtt készült. Wexler, Freeman, Namias és Reichelderfer látogatók a Meteorológiai Intézetből. G. Platzman és J. Smagorinsky hiányoznak.

Neumann már korábban szoros kapcsolatba került a ballisztikai laboratóriummal: szakértelme révén 1937-ben a konzultánsa lett, 1940-ben pedig a tudományos tanácsadó testület tagjává nevezték ki. 1944 nyarán azért érkezett Aberdeenbe, hogy egyeztesse a Manhattan-terv számítási igényeit és az ENIAC előirányzott teljesítőképességét, továbbá, hogy átvegye a Philadelphiában megvalósítandó EDVAC-terv konzultánsi tisztét. Tanulmányában – amelyre ma mindenki annak első szavaival (*First Draft of a Report on the EDVAC...*) hivatkozik – elsőként foglalta egységes keretbe a korszerű számítógépek ismérveit, köztük a kettes számrendszer alkalmazását, a memória, a programtárolás és az utasításrendszer használatát.

A világháború befejeződését követően az ENIAC két fő konstruktőre, John Mauchly matematikus, tervezőmérnök és J. Presper Eckert villamosmérnök folytatta az EDVAC-

programot. A páros létrehozta a számítógépipar egyik első amerikai óriását, a UNIVAC céget. Neumann visszatért Princetonba, ahol – Veblen professzor hatékony közreműködésével – még 1945 vége előtt sikerült elérnie az Institute for Advanced Study igazgatójánál, Frank Aydelotte-nál, hogy vállaljon erkölcsi, valamint százezer dolláros pénzügyi elkötelezettséget egy elektronikus számítógépes terv megvalósításáért. A Radio Corporation of America (RCA) vezetősége, értesülve a tervről, hasonló összeget ajánlott fel annak támogatására. A projekt célja egy valóban korszerű, adat- és programtároló, grafikus megjelenítésre is alkalmas, párhuzamos rendszerű elektronikus digitális számítógép megszerkesztése volt. Ez a típusú munka igencsak rendhagyó tevékenységet jelentett az intézet kolostori falai között, ahol az alapító néhány gazdag üzletember szándékainak megfelelően kizárólag briliáns agyú tudósok – köztük 1955-ben bekövetkezett haláláig Albert Einstein – csendes, elvont szellemi alkotómunkája folyt. A tervezés 1946 elején indult el. A logikai és matematikai vonatkozású munkát Neumann helyettese, Herman Goldstine irányította (aki a háború alatt ballisztikai szakértőként, századosi rangban az ENIAC építésének katonai összekötője volt, és akivel Neumann 1944 nyarán véletlenül ismerkedett össze az aberdeeni vasútállomáson), a vezető mérnök pedig Julian Bigelow volt. A gép 1951-ben már dolgozott, de operatív működését ünnepélyes formában csak 1952. június 10-én jelentették be. Neumann a berendezést, amely logikai felépítésével a legtöbb későbbi számítógép prototípusává vált, intézete nyomán mindig IAS gépnek nevezte. Kollégái azonban – amint George Cressman, az amerikai nemzeti időjárás szolgálat későbbi igazgatója az egyik visszaemlékezésében írja – maguk között egyszerűen *Johnniac* néven emlegették. (Létezett később egy számítógép, amely hivatalosan is a JOHNNIAC nevet viselte; ezt 1954-ben, a Rand Corporation laboratóriumában építette meg Willis Ware, Neumann egyik tisztelője és volt munkatársa.)

Neumann-nak kezdettől fogva eltökélt szándéka volt, hogy jelentős erőfeszítéseket tesz a hidrodinamikai egyenleteken alapuló, objektív időjárás-előrejelzés ügyének előmozdításában: „Adjatok egy szinte megoldhatatlan problémát, és én gondoskodom róla!” – szölte neki tulajdonított mondat. Megkülönböztetett figyelmét a meteorológiai prognosztika azzal érdemelte ki, hogy abban – nem kis mértékben Rossby hatására – a leginkább komplex és nemlineáris problémát látta, amit bárki valaha kigondolt, és amely hosszú ideig jelent majd kihívást még a leggyorsabb számítóberendezések számára is. 1946 májusában, tehát alig néhány hónappal számítógépes tervének indulását követően elérte, hogy Aydelotte eljuttassa javaslatát a hadsereg haditengerészeti ágazatának kutatási és találmányi irodájához. Írásában azzal a gondolattal állt elő, hogy a hadügy támogathatna egy, az intézeti terv keretében megvalósítandó meteorológiai programot. Neumann javaslatának szövege a számszerű időjárás-prognosztikának a leglátványosabb jövőképe volt Richardson könyvének megjelenése óta. Első sorait idézve: „A program célja a dinamikus meteorológia elméletének megvizsgálása annak érdekében, hogy azt alkalmassá tegye a máris rendelkezésre álló, és a jövőben valószínűleg egyre inkább hozzáférhetővé váló nagy sebességű elektronikus digitális számítások automatikus elvégzésére. Várhatóan ezek a vizsgálatok arra is jelzéseket adnak majd, hogy milyen – akár laboratóriumi típusú, akár terep jellegű – további megfigyelések szükségesek az ilyen nagy sebességű számításokkal támogatott elméleti munka még hatékonyabbá tételéhez... A múltban gyakorlatilag semmi nem motiválta, hogy matematikai és analitikai szinten kidolgozzák a meteorológia elméletének azokat a részeit, amelyeknek a valóban hatékony gyakorlati alkalmazása 1000–100 000-szer gyors-

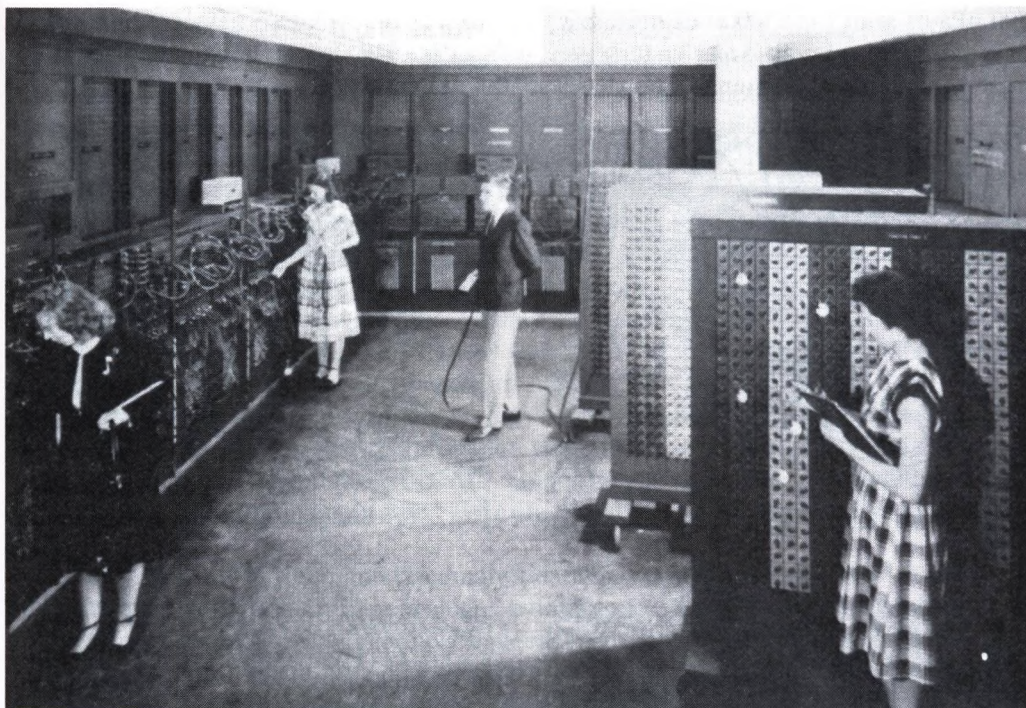
sabb számítási módszereket igényel annál, ami abban az időben lehetségesnek tűnt! Az elmélet teljes újbóli áttekintése vagy újraértékelése ezért elengedhetetlen előfeltétel.”

A haditengerészet késznek mutatkozott az öt vagy hat tehetséges fiatal szakember részvételére alapozott program évi 61 000 dolláros költségvetésének a biztosítására, így a munka 1946. július 1-jén kezdetét vehette. Neumann a következő hónapban Rossby segítségével hozzálátott olyan meteorológusok felkutatásához, akiknek a közreműködésére számíthatott. Az első lépést egy ülés megrendezése jelentette, amelyet egyszerűen „meteorológiai konferenciának” neveztek. Ezt az Institute for Advanced Study falai között, 1946. augusztus 29-én és 30-án tartott összejövetelt ma történelmi eseményként, az objektív időjárás-prognosztika első konferenciájaként tartjuk számon. A mintegy húsz résztvevő között ott volt az Egyesült Államokban dolgozó elméleti meteorológusok elitje, olyan személyiségek, mint Haurwitz és Rossby, aki Charney-t is magával hozta Chicagóból, hogy a fiatal kutatót Neumann-nal összeismertesse. A konferencián történekről ma már csak egy rövid, gépelt négyoldalas emlékeztetőből tájékozódhatunk, amely például Rossby protezsáltjának közreműködését a „Dr. Charney néhány meglehetősen elvont problémát vetett föl” mondattal összegezi.

Megalakult a princetoni meteorológiai kutatócsoport. Fő erőssége az első két évben az Egyesült Államok légierőinek egyik katonameteorológusa, Philip Thompson volt; ő írta meg később, 1961-ben a modern időjárás-előrejelzés alapjairól szóló első könyvet. Saját kérésére került a csapatba; előzőleg a UCLA-ben végzett kutatásokat, és a *New York Times Magazine* riporterének Neumann-nal és Vladimir Zworykinnal, az RCA képviselőjével készített interjújából értesült a princetoni számítógép-fejlesztési tervekről. Thompson még a Los Angelesben töltött időkben barátkozott össze Charney-val, aki 1947 tavaszán, útban Oslo felé, megállt néhány napra Princetonban. A két barát késő éjszakába nyúlóan, megszámlálhatatlan korsó sör társaságában vitatta meg, hogyan kellene a kormányzó egyenletekben a gravitációs és hanghullámokat a Rossby-hullámoktól elhatárolni. Neumann ez alkalommal találkozott másodszor Charney-val, és igen jó véleményt alakított ki magában a lelkes ifjú kutatóról. Így amikor egy évvel később Thompson princetoni küldetése a vége felé közeledett, és felmerült az utódlás kérdése, szinte magától értetődően kínálkozott a megoldás: az Oslóban dolgozó ösztöndíjas meghívása Princetonba. Így is történt. Norvégiából hazatérve, Charney 1948 júliusában érkezett meg, és a meteorológiai kutatócsoport vezetőjeként az intézetben maradt nyolc éven át mindvégig, egészen a numerikus prognosztikai program befejezéséig.

1949 nyarának végén csatlakozott a kutatócsoporthoz a norvég meteorológiai intézet kivételes matematikai adottságokkal rendelkező képviselője, Ragnar Fjörtoft. Charney és Fjörtoft, Neumann személyes közreműködésével, nyomban hozzákezdett egy reális kezdőfeltételekre alapozott számszerű időjárás-előrejelzési kísérlet megtervezéséhez. Mivel ebben az időben a Princetonban épülő IAS számítógép még távol állt a gyakorlati alkalmazhatóságtól, a numerikus integrációs számítások elvégzésére csak a Philadelphiában, 1945 decemberében egy Los Alamos-ból érkezett sürgős probléma megoldására elindított (majd hivatalosan 1946 februárjában felavatott) ENIAC kerülhetett szóba, amelyet 1947-ben átszállítottak az aberdeeni kísérleti bázisra.

A történelem két ENIAC-expedícióról emlékezik meg. Az első 1950 márciusának első vasárnapján kezdődött, amikor egy hat főből álló lelkes csapat jelent meg Aberdeenben, hogy részt vegyen egy emlékezetes próbálkozásban: Richardson álmának megvalósításában.

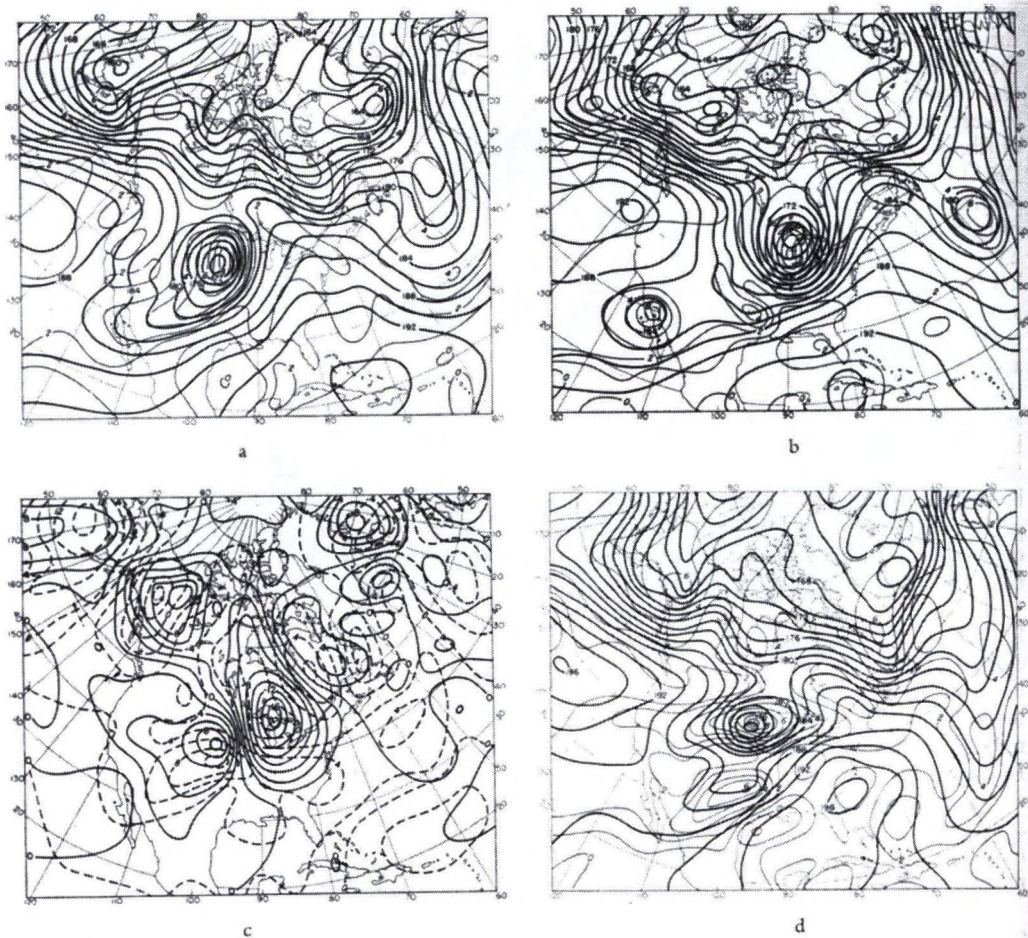


2. ábra. Elektronikus Numerikus Integrátor és Számítógép (ENIAC), 1948. július 7. Ballisztikus Kutatólaboratórium, Aberdeen Proving Ground, Aberdeen, Maryland.

A kísérlet napi huszonnégy órán át, csak rövid megszakításokkal, 33 napon keresztül folyt. A nyolcórás váltásban dolgozó csapat gerincét Charney, Fjörtoft és Neumann alkotta, akik e hosszú játszma forgatókönyvét írták, továbbá Joseph Smagorinsky és George Platzman, akik John Freeman társaságában a programozási munkát felügyelték (Rossby 1947 ősze óta már inkább a stockholmi egyetemen alapított tanszékén tartózkodott, de 14-én eljött meglátogatni őket). A programozást hatalmas vezérlőpulton, kézi dugaszolással és kapcsolókkal kellett elvégezni, és e művelet sajátosságai közé tartozott, hogy azt valamilyen okból nem volt szabad megszakítani, hanem a gép folyamatos beavatkozást igényelt. A gép egy-egy összeadást 0,2 ms alatt, szorzásokat 2-3 ms alatt végzett el. Az előrejelzésre vonatkozó számítás, amely az időlépcsőnként végrehajtandó 16 műveleti blokk negyedik tagja volt, egy 15×18 -as rácshálózat 270 pontjának mindegyikére három lyukkártya leolvasását követően történt meg. Ennek során a lyukkártyaolvasó három, jól hallható kattanó hangot adott, amit egy valamivel hosszabb idő követett, amelynek során az ENIAC elvégezte az előírt számtani műveleteket. A ciklikus munka sebessége olyan volt, hogy a teremben tartózkodók a kattanásokra könnyedén eljáráhették a feszes ritmusú gyors tánc, a dzsigg háromlépéses változatát. Egy, mára már feledésbe merült okból mindenki megkönnyebbült, amikor mindegyik időlépcsőben az említett négyes számú műveleti blokk elérézését a kártyaolvasó hangosan jelezte – és a visszaemlékezések szerint az operátorok nem egy esetben valóban élénk táncra perdültek.

Most már a hidrodinamikai egyenletek numerikus integrálásával végül megszületett az

500 hPa-os szint (az 5500 m-es magasság) áramlási mezejének első négy eredményes, 24 órás előrejelzése Észak-Amerika térségére, mégpedig 1949 első két hónapjának négy kiválasztott napján végzett mérések tényleges adataiból kiindulva.



3. ábra. Az ENIAC gépen elkészített első négy numerikus előrejelzés egyike: az 500 hPa-os szint 1949. január 5-én 0300 GMT órakor megfigyelt állapotának 24 órás prognózisa. (a) A szint magasságának és az áramlás örvényességének kiinduló eloszlása; (b) a magasság és az örvényesség 24 órával később megfigyelt eloszlása; (c) a magasság folytonos görbékkel jelölt megfigyelt és szaggatott görbékkel jelölt számított 24 órás megváltozása; (d) a magasság és az örvényesség eloszlásának a 24 órával későbbi időpontra számított képe.

A maratoni kísérlet során természetesen számtalan nehézséget kellett leküzdeni; az ENIAC hibamentes futásideje átlagosan csak egy-két óra volt, míg a javítás gyakran sok-sok órát igényelt, és lépten-nyomon felmerültek programozási nehézségek, sőt olykor kisebb balesetek is. Mindezt egy 41 oldalas, kézzel vezetett kis eseménynapló rögzíti, amelynek utolsó lapjára jegyezzük be az utókor ítéletét: *Az objektív, számszerű időjárás-prognosztika megvalósítása ezennel a realitások talajára lépett!*

A munkáról Charney, Fjörtoft és Neumann részletes beszámolót írt a Rossby által ala-

pított svéd geofizikai folyóirat, a *Tellus* 1950. évi 4. számában. Megemlítik, hogy a négy, 24 órás előrejelzés elkészítése során mintegy 100 000 standard IBM-lyukkártyát gyártottak, továbbá egymillió szorzást és osztást végeztek el (és ezek a számok megduplázódnak, ha az előzetes kísérleteket is számításba vesszük). Mivel az előrejelzési modellből a gravitációs és hanghullámokat kiszűrték, a számítás stabilitását – a Courant–Friedrichs–Lewy-féle kritériumnak Neumann által a feladathoz történt igazításával – 3 órás időlépcsőnél is sikerült biztosítani. Egy-egy 24 órás előrejelzés számítási igénye a gépen még így is 24 óra volt, tehát csak az idő múlásával sikerült „lépést tartani”. Ezt azonban az időjárás *előrejelzése* szempontjából nem tartották kiábrándítóknak; tudták, hogy az épülő IAS gép lényegesen gyorsabb számítást tesz majd lehetővé. A tanulmányhoz csatolt köszönetnyilvánítás szerint a munkát Neumann felesége, Dán Klára is segítette. Ő az ENIAC gépre történő kódolási eljáráshoz és az elkészült kód ellenőrzéséhez nyújtott útbaigazítást.

A második ENIAC-expedíciót Platzman és Norman Phillips 1951 júniusában szervezte meg. Célja egy numerikus eljárás kipróbálása volt, amire az első expedíció során már nem jutott idő. Ezzel a kísérlettel az ENIAC meteorológiai szereplése befejeződött. A gép 1955-ig működött, azután lebontották; 1962 óta a washingtoni Smithsonian Intézetben van kiállítva.

Charney, Fjörtoft és Neumann tanulmánya hatalmas érdeklődést váltott ki. 1952-ben az Institute for Advanced Study meteorológiai kutatócsoportján kívül már három olyan jelentős központ működött a világon, amely figyelmét a numerikus időjárás-prognosztika kérdéskörére összpontosította. Aztán az 1950-es évek vége felé, részben a számítógépes technika fejlődésének, részben a légköri modellek fokozatos tökéletesedésének köszönhetően, számos nemzeti meteorológiai szolgálat jutott arra az elhatározásra, hogy az objektív prognosztikai módszert bevezeti az operatív időjárás-előrejelzések készítésének folyamatába. A modellek ebben az időben a nagy térségű légköri mozgások naponkénti változékonyságának nagyjából a 65 százalékát voltak képesek helyesen leírni, tehát még távol voltak attól, hogy a gyakorlott előrejelzési szakemberek méltó versenytársaivá váljanak. A helyzet inkább az volt, hogy ezek a hagyományos eljáráson nevelkedett előrejelzők az összes addig használt segédletük mellé olyan új információt kaptak, amire ránézhetnek, és elmondhatják, hogy „nos, ez az, ami a számítógép szerint történni fog”. Azután belátásuk szerint felhasználták vagy elvetették ezt az információt.

Ezekben az 1950-es években a hagyományos *szubjektív* előrejelzési technika, valamint a Neumann-féle *dinamikus* előrejelzési technika mellett létezett egy harmadik módszer is, a *statisztikus* prognosztika. Ez utóbbi legelterjedtebb típusa a *lineáris prognosztika* volt. Az ilyen regressziós eljárással például X város holnapi hőmérséklete egy a állandó, plusz egy Y város valamely b állandóval szorzott mai hőmérséklete, plusz valamely Z hely c állandóval szorzott tegnapi légnedvessége, plusz más hasonló tagok formájában állítható elő, és az állandók optimális értékének megbecslésére régen kidolgozott, jól megalapozott matematikai módszerek álltak rendelkezésre. A statisztikus előrejelzések készítésének hívei nem hittek abban, hogy a Neumann-féle dinamikus előrejelzési technika felülmúlhatja az ő módszereiket, de természetesen csak az idő és a konkrét tapasztalat adhatott választ a kérdésre, hogy nekik van-e igazuk vagy Neumann csapatának. A helyzetet azonban kiélezte, hogy az újságok szenzációként kezelték a számítógépes prognózisok kezdeti sikereit, és olyanokat írtak, hogy a számítógép belátható időn belül megoldja akár a tetszőlegesen hosszú távra szóló előrejelzések feladatát is. Ezen a ponton Norbert Wiener, a Massachu-

setts Institute of Technology (MIT) matematikaprofesszora szükségét érezhette annak, hogy a témáról elmondja saját véleményét. 1956-ban, a kaliforniai Berkeley-ben *Nemlineáris előrejelzés és dinamika* címmel előadást tartott, és abban rámutatott, hogy a légköri folyamatok dinamikai előrejelzése az a típusú feladat, amelyhez elégtelen számú, pontatlan megfigyelés áll rendelkezésre, ezenfelül a folyamatok nemlineárisak, tehát csak közelítő módszerek jöhetnek számításba; mindebből pedig összességében a lehetséges hibák széles skálája adódik. Következtetését szó szerint idézve: „Rossz eljárás azt a típusú tudományos módszert, amely az asztronómia vagy a ballisztika precíz, sima működéséhez tartozik, olyan tudományra alkalmazni, amelyben a hibák statisztikája széles, a megfigyelések pontossága pedig csekély. A félig egzakt tudományokban, amelyekben a megfigyelések ilyen jellegűek, az eljárásnak kifejezetten statisztikusnak, és a csillagászatban használnál kevésbé dinamikusnak kell lennie.”

Ezt a megnyilatkozást többféleképpen lehetett felfogni. Például William Aspray abban a könyvében, amely Neumann János munkásságáról szól, és az MIT Press gondozásában, 1990-ben jelent meg, az esetet úgy állítja be, mintha Wiener megtámadta volna a „Neumann-féle irányvonalat”. Ráadásul a támadás mögé odavetíthette az Egyesült Államok egyik legkiválóbb oktatási intézménye, a 19. század közepén Cambridge-ben alapított Massachusetts Institute of Technology hatalmas presztízsét. Erre, a Harvard Egyetem szomszédságában, a Charles folyó partján épült intézményre, amelyet valaha egyszerűen Boston Tech-ként emlegettek, de ma már mindenki csak MIT-ként ismer, Erich Segal, az író szavait idézve, „azok, akiknek hite szerint a tudomány vallás, úgy tekintenek, mint a tudomány Vatikánjára, ahol Nobel-díjasokból áll a bíbornoki testület”.

Mások viszont úgy látták, hogy Wiener, a neves matematikus, a kibernetika egyik megalapítója (az elnevezés is tőle származik) konstruktív véleményt fejtett ki, amelyre érdemes odafigyelni. Hosszabb idő után visszatekintve, ezt még inkább így értelmezhetjük, mert a dinamikai tábor és a statisztikus tábor vitája is hozzájárult ahhoz a nagy jelentőségű, új felismeréshez, amely nemcsak az időjárás-előrejelzések módszertanának új alapokra helyezését, hanem egész természetszemléletünk bizonyos fokú átértékelését is maga után vonta.

5. Determinisztikus valószínűségi időjárás-előrejelzések

Bár az 1950-es évek közepén az MIT meteorológiai tanszékén oktató Edward Lorenz maga is statisztikus meteorológusként tevékenykedett, de nem osztotta kollégáinak a regressziós eljárások fölényét valló nézetét, amelyhez számukra a fő hivatkozási alap Norbert Wiener előadása volt. Úgy vélte, hogy Wiener állításait, amelyek „bizonyára korrektek voltak, de nem a legegyszerűbben érthető nyelven voltak leírva”, sokan félreértelmezték. 1956 augusztusában, a Wisconsin állambeli Madisonban nagyszabású távprognosztikai értekezletet rendeztek, amelyen megjelentek a statisztikus előrejelző közösség jeles képviselői is. Lorenz a maga előadásában javasolta a Wiener-féle állításokra hivatkozó nézet ellenőrzését olyan módon, hogy kiválasztásra kerül a prognosztikai jellegű hidrodinamikai egyenle-

teknek egy határozottan *nemlineáris* rendszere. Számítógéppel megvalósítható ennek a rendszernek a hosszú távú numerikus integrációja, és ezután, a megoldásokat úgy tekintve, mint hogyha valóságos légköri állapotjelzők értékeinek megfigyelt időbeli sorozata lenne, az ismert eljárások alkalmazásával ellenőrizhetővé válnak a statisztikus prognosztika optimális lineáris előrejelzési formulái. Ha ezek a formulák valóban fölérnek bármelyik más típusú előrejelzési eljárással, akkor *tökéletesen kell működniük*, hiszen a kiválasztott nemlineáris dinamikai rendszer segítségével bárki tökéletesen „prognosztizálni” tudja az értékeket, egyszerűen úgy, hogy ugyanazokkal a kezdeti feltételekkel másodszor is lefuttatja a számítógépes programot.

Elgondolásának megvalósításához Lorenz maga kezdett hozzá. Első dolga a megfelelő prognosztikai egyenletrendszer kiválasztása volt. Ebben hivatásos meteorológusként és amatőr matematikusként járt el. Noha a kísérlet elvégzéséhez a nemlineáris rendszerek széles köréből elvileg *bármelyik* alkalmas lett volna, azt remélte, hogy vizsgálódása során bizonyos részeredmények is elérhetők, ha egy olyan bonyolultabb, *meteorológiai* értelemmel rendelkező egyenletrendszert választ, amelyik hasonlít a légkör nagy térségű viselkedését szimuláló egyenletek rendszerére. Hosszas keresgélés után egy 12 változót tartalmazó egyenletrendszer mellett kötött ki, amelynek integrálásával beigazolódott gyanúja. A standard statisztikus eljárásokat az így generált adatokra alkalmazva, azonnal kiderült, hogy a lineáris előrejelzési technikák távolról sem tökéletesek.

A meteorológiai jelentéssel *is* bíró nemlineáris egyenletek megválasztásának igazi gyümölcsei ekkor kezdtek beérni. Az egyenletek megoldásait „légköri adatokként” is értelmezni lehetett: a numerikus eljárás a légkör állapotjelzőinek értékeit (az „időjárást”) hatórás időközönként léptette előre. Az adatok elemzése nagyon hamar nyilvánvalóvá tette a modellezett időjárás alakulásának teljesen szabálytalan, látszólag véletlenszerű jellegét. Ez az irreguláris viselkedés a meglepetés erejével hatott, mivel a determinisztikus előrejelzési modell autonóm rendszert alkotott, azaz nem tartalmazott semmiféle időben változó külső kényszert; az időjárás aperiodikus változékonyságát a rendszer belső mechanizmusai szabályozták. Ezzel adott is volt az első részeredmény, amelyet később számtalan kísérlet igazolt, és amely egy nagyon fontos tényre tanított meg bennünket: a légkör alapvető kvalitatív tulajdonsága az *irreguláris szabad változékonyság*. Az időjárás és az éghajlat szabálytalan ingadozásainak nem szükségszerű feltétele a légkörre ható külső kényszerek (például a naptevékenység, a vulkántevékenység vagy az emberi tevékenység) valamelyikének szabálytalan váltakozása, de még az sem, hogy bármelyik külső kényszer megváltozzék.

De az igazi meglepetés még hátra volt. Kísérletsorozata egyik pontjánál Lorenz elhatározta, megismétli a számítás egy szakaszát, hogy részletesebben is megvizsgálhassa a történeteket. Megállította számítógépét, betáplálta a számoknak azt a sorozatát, amit a gép valamivel korábban, ennek az érdekes szakasznak az elején kiírt, és ezekkel a kezdeti feltételekkel újra elindította a most már részletesebb időbeli bontást biztosító futtatást. Saját gépe (amely akkortájt, több mint húsz évvel a személyi számítógépek piaci megjelenése előtt, igencsak ritkaság volt), egy Royal-Mc Bee LGP-30 típusú, íróasztal méretű berendezés, kellemetlen zajt okozva működött. Ezért egy kis csöndre vágyva, lesétált a hallba egy csésze kávé meginni, majd úgy egy óra múlva tért vissza szobájába. Ez idő alatt a gép nagyjából két hónap időjárását produkálta. A kinyomtatott számok azonban egyáltalában nem hasonlítottak a korábbi futtatásnál adódottakra. Lorenz azonnal az egyik elektroncső kimerülésére vagy valamilyen más géphibára gyanakodott. Mielőtt kihívta volna a szervizt,

úgy gondolta, hogy megpróbálja maga megkeresni, hol történhetett a hiba – tudván, hogy így rendszerint felgyorsítható a javítás menete.

A két adatsor összevetése nem mutatott hirtelen szakadást. Az új értékek kezdetben megegyeztek a korábbiakéval, de nem sokkal ezt követően az utolsó decimális helyértéken kezdtek egy, majd több egységgel különbözni egymástól. Később az eltérés átterjedt az eggyel, majd a kettővel balra álló helyértékekre is. Végül kiderült, hogy valójában a különbségek többé-kevésbé egyenletes tempóban, körülbelül négynaponként kétszereződtek meg, majd a második hónap során a két futtatás eredményei közötti hasonlóság teljesen eltűnt. Ez elég volt Lorenz számára, hogy tudja, valójában mi is történt: azok a számok, amelyeket a második, részletesebb futtatáshoz kezdőértékeként a számítógépbe betáplált, nem az eredeti, *pontos* számok voltak, hanem olyan *kerekített* értékek, amelyeket a gép az első futtatáskor kinyomtatott. Tehát a két futtatás eredményeinek a különbözőségéért a kezdeti feltételek megadásának kerekítési hibáit terhelte a felelősség: a kiindulási apró eltérések következetesen növekedtek, és a végén teljesen eluralták a második megoldást. Ma ezt a jelenséget a *nemlineáris dinamikai rendszerek viselkedésének a kezdeti feltételekre mutatott érzékenységeként* ismerjük.

Lorenz az eredményeit 1963-ban, *Determinisztikus aperiodikus áramlás* címmel az Egyesült Államok vezető meteorológiai tudományos folyóiratában tette közzé – érdemleges figyelemben azonban a társtudományok részéről abban az évtizedben alig részesült. A fizikusoknak és a matematikusoknak megvannak a maguk folyóiratai, éppen elég fáradságot jelent azok követése, miért is olvasgatnának meteorológusok számára írt cikkeket. Csak amikor 1970 táján más kutatók is felfigyeltek a Lorenz által tapasztalt viselkedésre, akkor derült ki, hogy mindez már évekkel korábbról „ismert tény”, és akkor vált közleménye az egyik legtöbbször idézett tanulmányá. Aztán a nemlineáris determinisztikus rendszerek önmagát nem ismétlő (aperiodikus, látszólag véletlenszerű), és a kezdeti feltételekre érzékeny függőséget mutató viselkedésének James Yorke, a marylandi egyetem matematikusa 1975-ben figyelmet felkeltő elnevezést adott: *káosz*.

A kaotikus viselkedési formát a háromszáz éves newtoni mechanika korábban nem ismerte. 1963 előtt tudtuk, hogy léteznek időben változatlan (stacionárius) állapotok, szabályosan ismétlődő (periodikus) állapotváltozások, továbbá egyidejűleg több, egymástól lineárisan független alaphétfrekvenciával jellemezhető kváziperiodikus állapotváltozások is. E három „jól viselkedő” (reguláris) változás mellett a „különös”, kaotikus viselkedés megjelenésének ténye új terminológiát vezetett be a fizika fogalmi készletébe: a *korlátozott előrejelezhetőség* fogalmát. Kiderült ugyanis, hogy nemlineáris kaotikus rendszerekben a kezdeti feltétel megadásának hibája az idővel exponenciálisan – minden hatványfüggvényénél gyorsabb ütemben növekszik. És mivel a kezdőfeltételek abszolút pontos megadása irreális absztrakció, a felfedezés megcáfolta a klasszikus mechanisztikus determinizmus Pierre Simon Laplace márki által 1814-ben megfogalmazott elvének *gyakorlati* érvényét, nevezetesen a Laplace-démonnak azt a képességét, hogy ha ismerné a világmindenség pillanatnyi állapotát, akkor „számára semmi sem volna bizonytalan, és a jövő éppúgy, mint a múlt, jelen volna szemei előtt”.

Bebizonyosodott tehát, hogy a számítógépes kísérletezés is vezethet eddigi világgépünk alapvető átfogalmazását szükségessé tevő eredményekhez, *valószínűségi szemléletet* követelve meg nemcsak a több évszázados newtoni mechanikában, hanem például a tiszta matematika legradicionálisabb ágában, a Gauss által „a matematika királynőjének” nevezett

számelméletben is. Ezért a korlátozott előrejelezhetőség felfedezését az 1980-as évek végén sok neves kutató egyenesen paradigmaváltásként értékelte. Kijelentették, a 20. század tudományát három dolog teszi majd emlékezetessé: a relativitáselmélet, a kvantummechanika és a káosz. Azt állították, hogy a relativitáselmélet végzett az abszolút tér és idő létezésének newtoni illúziójával, a kvantumelmélet megsemmisítette az ellenőrizhető mérési eljárás lehetőségének szintén newtoni álmát, a káoszelmélet pedig leszámolt a determinisztikus előrejelezhetőség laplace-i képzetével. Más tudósok sokkal visszafogottabban fogalmaztak, mondván, hogy a káosz felismerése mindössze egy több évszázados tévedést korrigált, amely nem nyújt új világképet; új törvények felfedezése nem kapcsolódik hozzá, hanem csak az ismert törvények eddig el sem képzelt bonyolultságával szembesít.

Bárhogyan értékelünk, az tény, hogy a meteorológia – amelynek egyik legfontosabb gyakorlati feladata az időjárás előrejelzése – arra kényszerült, hogy teljesen új alapokra fektesse a Neumann-féle prognosztikához történő elméleti hozzáállását. Az időjárás-előrejelzések hibáinak bizonyossága kötelező elvé tette, hogy egyetlen prognózis sem tekinthető teljesnek a prognózis megbízhatóságának egyidejű prognosztizálása nélkül. Más megfogalmazásban ez az elv azt jelenti, hogy időjárás-előrejelzéseket csak a statisztika eszköztárát felhasználó *valószínűségi* formában szabad kibocsátani – ebben az értelemben tehát tökéletesen megalapozottnak bizonyult Norbert Wiener 1956-ban megfogalmazott gyanúja!

A meteorológiai gyakorlatban előrejelezhetőségi időtávként azt az időszakot tekintjük, amelynek során a kezdeti feltételek előírásának tipikus hibái közel megháromszorozódnak, pontosabban e -szereződnek, ahol $e = 2,718$ a természetes alapú logaritmus alapszáma. A numerikus kísérletek szerint az időjárás alakításában meghatározó szerepet játszó tényezőre, a középtroposzféra (tehát körülbelül az 5 km-es magassági szint) nagytérségű áramlási mezejére vonatkozóan ez az időtáv átlagosan 25 nap körül van. Konkrét esetekben a hibák növekedésének üteme erősen függ az áramlási képtől: léteznek időjárási helyzetek, amelyeknek a jövőbeli alakulása hosszabb időre prognosztizálható másokénál. A kezdeti feltételek megadásában rejlő bizonytalanság természetesen nem az egyetlen hibaforrás, amelynek következtében az előrejelzések előbb-utóbb elveszítik érvényüket. Noha a Charney, Fjörtoft és Neumann által elvégzett korszaknyitó kísérlet óta eltelt fél évszázad során az előrejelzési modellek – részben elméleti ismereteink gyarapodásának, részben a számítástechnika fejlődésének köszönhetően – nagymértékben tökéletesedtek, még napjainkban is további javításokra szorulnak. Ezért az említett, átlagosan 25 napos időtáv az időjárás előrejelezhetőségének *elméleti* felső határát jelenti, a *gyakorlatban* ma még csak legfeljebb egy-két hétre vagyunk képesek használható előrejelzéseket kibocsátani. Mint tudjuk, készülnek ennél hosszabb távra is prognózisok, ezek azonban nem egy-egy adott időpontnak az időjárását, hanem valamilyen hosszabb időszak (néhány hét, hónap vagy évszak) átlagolt légköri állapotának a szokásostól várható eltérését adják meg.

Felmerül a kérdés: anélkül, hogy tagadnánk a légkör jövőbeli állapotainak a fizika törvényei alapján egyértelműen *determinált* voltát, miként önthetjük az időjárás-előrejelzéseket *valószínűségi* formába? A Neumann-féle prognózisok az első évtizedekben *kategorikus* előrejelzések voltak: a nemzeti előrejelző központok a legjobbnak ítélt kezdeti feltételekből kiinduló *egyetlen* prognózist készítettek el. Ezek a kezdeti feltételek a leggondosabb ellenőrzési eljárások alkalmazása ellenére is bizonytalanok. Már az 1970-es évek elején felvetődött a gondolat, hogy a kezdeti feltételeknek a hibahatáron belüli módosításaiából, tehát az ugyan-csak lehetséges kezdeti állapotokból kiindulva, egyidejűleg *több* (többször tíz) kategorikus

előrejelzés *együttesét* kellene elkészíteni. Természetesen nem tudhatjuk, hogy melyik előrejelzés bizonyul majd helyesnek (vagy legalábbis a leginkább megbízhatónak), azt azonban joggal feltételezhetjük, hogy az együttes tagjainak szórása jó indikátora az adott időjárási helyzet előrejelezhetősége mértékének. Minél nagyobb fokú az előre jelzett állapotok közötti eltérés, annál nagyobb az adott időjárási helyzet prognózisához kapcsolódó bizonytalanság.

A számítógépek teljesítőképessége az *együttes* (vagy *ensemble*) előrejelzési technikának a gyakorlati bevezetését csak 1992 végére tette lehetővé. Az angliai Readingben működő Európai Középtávú Időjárás Előrejelző Központ, amelynek hazánk is együttműködő állama, a hagyományos (és kontrollként kezelt) kategorikus előrejelzés mellett napjainkban 50, az említett módon perturbált előrejelzést is elkészít, mégpedig tíznapos időtávra és az egész Földre vonatkozóan. Az $50 + 1$ előrejelzés alapján a légköri állapotjelzők különböző várható értékeihez meghatározott valószínűségek rendelhetők, ha pedig a prognózist mégis kategorikus formában kívánjuk megfogalmazni, akkor a matematikai statisztika módszereivel igazolható, hogy az optimális előrejelzést az ensemble tagjainak átlaga (a sokaságátlag) képviseli.

6. Neumann gondolatai a természeti környezet megóvásáról

Ha valaki az 1950-es évek elején megkísérli a Neumann-féle numerikus prognosztika első modelljeinek kormányzó egyenleteit hosszabb távra, mondjuk egy hónapos időszakra integrálni, olyan áramlási alakzatokat nyert volna, amelyeket egészen biztosan nem látott még senki a természetben. Ennek egyszerű oka az, hogy ezek a modellek konzervatív jellegűek voltak: nem tartalmazták sem a napsugárzás által előidézett külső termikus kényszert, sem a légmozgások súrlódásos disszipációját. Az ilyen eljárás jó közelítésnek bizonyult az *időjárási* folyamatok egy-két napos előrejelzéséhez, viszont alkalmatlan volt arra, hogy szimulálja a nagy földi légkörzést, és megmagyarázza a légkör szokásos viselkedésének (az *éghajlati* képnek) a kialakulását.

Az általános cirkuláció modelljének prototípusát az Institute for Advanced Study-ban Charney csoportjához 1951 szeptemberében csatlakozott Norman Phillips dolgozta ki. Neumann volt az első tudós, aki az ilyen típusú modellezésnek a jelentőségét és a természeti környezet károsodásával kapcsolatos kutatásokban történő alkalmazhatóságát azonnal felismerte. Egyedi problémák – például a levegő és a víz szennyeződése, a zajártalom vagy a földi erőforrások kimerülése – már a 20. század első felében is jelentkeztek. Az a tény, hogy az emberi tevékenység a környezet általános és folyamatos degradálódását okozza, csak az 1960-as évek végére vált nyilvánvalóvá. Ugyancsak erre az időre tisztázódott, hogy a környezeti problémákat általában lehetetlen egyedileg orvosolni. A természeti környezet oszthatatlan egésznek alkot, megóvása interdiszciplináris, és egyben nemzetközi együttműködést igénylő feladat.

E gondolat korai előfutára volt Neumann 1955-ben megjelent esszéje, amely azt a kérdést elemezte, hogy túlélhető-e a technika fejlődése, és amelyben elsőnek foglalkozott a

globális klímamódosulásnak azzal a vetületével, hogy mit jelenthet ez a probléma az országok közös gondolkodása és a nemzetközi stratégiák kidolgozása szempontjából. Felvetette az „iparszerű” éghajlat-szabályozás lehetőségét. A sarkvidéki jégmezőknek például kompromisszummal történő beszórását, amivel a napsugarak erős visszaverődése által előidézett hőveszteség csökkenthető lenne. Egyidejűleg azonban azt is hangsúlyozta, hogy a földi éghajlat instabilis rendszer, amelyben kis zavarok nem kívánt folyamatokat is elindíthatnak. Ezért óvott mindenfajta mesterséges beavatkozástól mindaddig, amíg ismereteink nem elegendően mélyek az előidézett hatások pontos előre látásához.

Neumann kezdeményezésére 1955. október 26–28-án az Institute for Advanced Study ismét meteorológiai konferenciának adott otthont. Ez alkalommal Neumann a világ harminc élenjáró meteorológusát hívta meg annak a kérdésnek a megvitatására, hogy milyen szerep vár a jövőben az általános légkörzés, és ezen keresztül az éghajlati folyamatok számítógépes modellezésére. Az ülést – az intézet igazgatói állását akkor betöltő – J. Robert Oppenheimer professzor nyitotta meg, a híres atomfizikus, akinek a vezetésével a Los Alamos-i nemzeti kutatóközpontban, a Manhattan-terv keretében, az első atombombát kifejlesztették. Neumann, első előadóként, a klímaingadozások előrejelzésének kérdését taglalta, felvázolva a rövid és hosszú távra szóló meteorológiai előrejelzések módszereinek fejlesztésénél követendő stratégiát. És ezen az összejövetelen mutatta be a hallgatóságnak Norman Phillips az IAS gépen futtatott általános légkörzési modelljével nyert első eredményeket.

E konferencia előadásai – szemben a kilenc esztendővel korábban rendezett princetoni ülés anyagával – néhány évvel később könyv alakjában is megjelentek. Az ott kitűzött kutatási irányvonal nyomán beszélhetünk a mai értelemben vett éghajlat-dinamikáról és klímaprognosztikáról, tehát az összejövetel éppen úgy mérföldkövévé vált a klimatológia történetének, mint az 1946-os konferencia a számszerű időjárás-előrejelzés fejlődésének. Elméleti meteorológusok, akiknek a szemléletmódja gyökeresen eltért a hagyományos éghajlattan művelőinek eszmevilágától, hidro-termodinamikai oldalról kezdték megközelíteni a klíma megváltozásának kérdéskörét. A fejlődés ezen a téren – a számítástechnikai háttérnek és a társadalmi elvárásoknak is köszönhetően – valóban hihetetlenül rohamos volt. 1983-ban Smagorinsky már joggal állíthatta, hogy akkorra „a klimatológia a leíró geográfia tiszteletreméltó ágából a fizika tudományának kvantitatív diszciplínájává alakult át”.

Az emberi tevékenységnek a globális éghajlatra (és még inkább a helyi klímára) gyakorolt hatását e fejlődése ellenére még napjainkban is sűrű homály fedi. Tudjuk, hogy az üvegházhatást előidéző gázok (elsősorban a szén-dioxid, a metán és a dinitrogén-oxid) légköri koncentrációja az ipari forradalom kezdete óta eltelt másfél évszázad során jelentősen növekedett. Tudjuk, hogy a Föld felszínének éves középhőmérséklete a 20. század során 0,6 fokkal emelkedett. És tudjuk azt is, hogy a melegedési tendencia az elmúlt három évtized során töretlen volt. E tények tükrében beszélhetünk-e bizonyítottan egy folyamatban levő, antropogén eredetű klímaváltozásról? Ez az a kérdés, amelyre sok kutató szerint ma még nem vagyunk képesek egyértelmű választ adni. Nem tudjuk, hogy a fokozódó üvegházhatásból eredő melegedést regionális léptékben mennyire hatékonyan ellensúlyozhatja a napsugárzást elnyelő légköri aeroszol koncentrációjának az emberi tevékenység nyomán megfigyelhető növekedése. Nem tudjuk, hogy egy klímamódosulási folyamat pontosan milyen mértékben jár együtt bolygónk felhőborítottságának, illetve a felhők optikai tulajdonságainak a megváltozásával, márpedig például egy műholdas technikával alig kimutatható, igen

csekély felhőzetnövekedés is elegendő lehet, hogy a viláágűrbe több napsugárzást visszaverve, kompenzálja az üvegházgázok hatását. És nem tudjuk azt sem, hogy a megfigyelt melegedési trend nem tartozik-e még mindig a Lorenz munkásságával kapcsolatban említett szabad változékonyság keretébe. Sok még a teendő a Neumann által kezdeményezett kutatások területén, hogy a gazdasági és politikai döntéshozók elé kétségbenvonhatatlan tényeket tudjunk tárni éghajlatunk épségét biztosító intézkedések szükségességére.

7. Epilógus

1955-ben Neumann Washingtonba költözött. Eisenhower elnök az Atomenergia Bizottság tagjává nevezte ki. A princetoni intézet, amely feltételezhetően az ő személyére való tekintettel fogadta el annak idején a számítógépes projektet, távozásával szabadulni igyekezett ettől a programtól. Charney és Phillips szerettek volna meteorológus szakot alakítani a princetoni egyetem kebelén belül, de nem találtak fogadókézségre, ezért 1956 közepén mindketten az MIT meteorológiai tanszékére mentek el dolgozni. Távozásuk után, tíz év múlva – a Smagorinsky által létrehozott geofizikai folyadékdinamikai laboratórium keretében végzett kutatások révén – Princeton immár a klímadinamikai elemzések amerikai fellegrára lett.

A történettudomány nem szereti (és nem fogadja el) a „mi lett volna, ha” kezdetű kérdésvetéseket. Ezért mi se boncolgassuk, hogy mi történt volna, ha Neumann János, pályafutásának egy-egy rövid időszakára, nem lép a meteorológia színpadára. Az időjárás megbízható előrejelzése, a nagyközönség tájékoztatásán túlmenően, fontos gazdasági, élet- és vagyonbiztonsági, és nem utolsósorban honvédelmi feladat. A légköri folyamatokkal foglalkozó tudomány egész történelme folyamán szinte azonnal részesült a legújabb technikai vívmányok áldásaiból, így a modern távközlés, majd a radarok és a mesterséges holdak nyújtotta információszerzési lehetőségekből. Előbb vagy utóbb bizonyára a számítástechnika is belépett volna az említett sorba. Neumann érdeklődésének, befolyásának és szervezőképességének köszönhetjük, hogy ez a lépés *azonnal* bekövetkezett.

A légköri folyamatok objektív leírásának tudománya az alkalmazott fizikai és alkalmazott matematikai diszciplína. Művelői nemcsak sokat *kaptak*, hanem új gondolatokat voltak képesek *adni* e két társtudományágnak. Neumann ezt a lehetőséget ismerte fel, miután érdeklődése az 1940-es évek elején a hidrodinamika nemlineáris egyenleteinek kezelésére, majd a meteorológiai problémákra terelődött. Azt vallotta: „Ha az emberek nem hiszik el, hogy a matematika egyszerű, az csak azért van, mert nem ébrednek annak tudatára, hogy az élet mennyire bonyolult”. Ennek a gondolatnak a kifejtése is elősegítette, hogy a régi iskolán nevelkedett időjárás-előrejelzők a 20. század utolsó harmadában egyre inkább elfogadták és kezdték értékelni mindazt, „amit a számítógép mond”. Az ezredforduló küszöbére érkezve, a szubjektív módszerekre alapozott eljárásokat világszerte teljes mértékben felváltotta az objektív időjárás-prognosztika. Neumann Jánost – akit Kármán Tódor, Szilárd Leó, Teller Ede és Wigner Jenő mellé, a *magyar marslakókként* emlegetett, kivételes intelligenciával megáldott csapatba választott be a nemzetközi tudós társadalom – ma már a szakma kiemelkedő egyéniségeként tiszteli a meteorológusok *egész* közössége.

Kovács Győző–Szelezsán János

Gondolatok Neumann János *First Draft of a Report on the EDVAC* című, 1945 júniusában megjelent tanulmányáról

Ezt a tanulmányt Neumann János írta a második amerikai elektronikus számítógép tervezésével kapcsolatban. A világ számítástechnikai szakemberei ezt a tanulmányt – amit Neumann János szerzőtársak nélkül alkotott és tett közzé – a számítógép-tervezés első és legátfogóbb leírásának tartják, ami lehetővé tette a számítógép iránt érdeklődő tudományos szakembereknek, mérnököknek és matematikusoknak, hogy a negyvenes évek közepén a számítógépekről szóló, minden fontos információhoz hozzájussanak. Talán nem túlzás azt állítani, hogy ez a „First Draft...”, „Vázlat...”, „Jelentés” (mindhárom elnevezést használjuk a következőkben) tette lehetővé, hogy számos egyetemen és kutatóintézetben elkezdődjék a számítógépek tervezése és építése, azt is mondhatnánk, ez volt az a hólabda, ami elindította a számítógépes korszak – azóta is tartó – lavináját.

Írásunkban lényegében a „First Draft...”-ot ismertetjük, de „beágyazzuk”, ugyanis – mielőtt rátérnénk ismertetésére – röviden elmondjuk, hogy milyen fontosabb mérföldkövek voltak a számítástechnika történetében Neumann előtt (az „időszámítás előtt”) és azt is, hogyan jutott el Neumann a „First Draft...” megírásáig.

Írásunk végén – szinte „Függeléként” – a Jelentésben szereplő elveken alapuló ténylegesen megtervezett EDVAC gép utasításrendszerét, valamint az EDVAC környezetében Neumann János közreműködésével létrehozott „magas szintű nyelvet” (3. fejezet) ismertetjük.

1. Út a „First Draft...”-ig, a Neumann János előtti informatikatörténelem

Mérföldkövek a Neumannig vezető úton

Számos kutató feltételezi, hogy az idők homályába vesző őskorban, az őstársadalomban, amikor a számfogalom, a számlálás, a mennyiség fogalma már kialakulóban volt, az ősember is használt számolást megkönnyítő eszközöket, elsősorban a mindig kéznél lévő ujjaikat. Ennek a számolóeszköznek tulajdonítják, hogy az emberiség nagy részénél a tízes

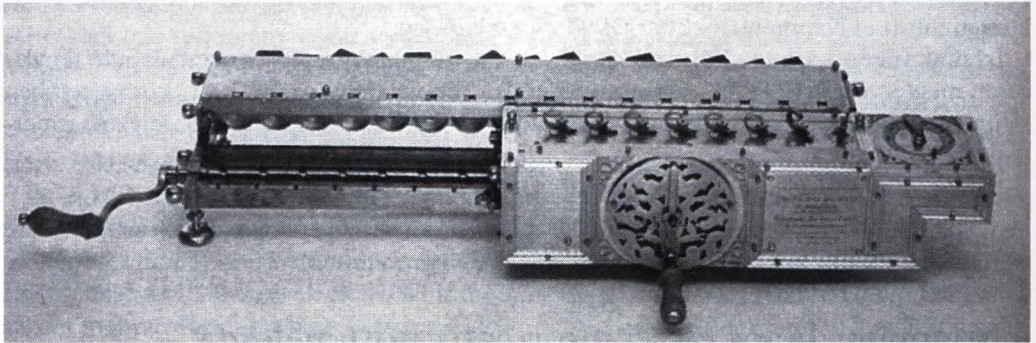
számrendszer terjedt el; ha – netán – az emberek nem kétszer öt, hanem kétszer négy ujjal születnek, akkor nagyon valószínű, hogy ma nem tízes, hanem nyolcas számrendszerben végeznénk a műveleteket.

A számolás megkönnyítésének az igénye tehát már a nagyon korai embereknél is megvolt, az ember mindig szeretett volna eszközöket találni számolási feladatainak a megoldására. Előbb kavicsokat (calculus) vett föl a földről, amiket rendbe rakott és ezekkel számolt, azután kitalálja a kavicsokkal való számolás modern változatát, a golyós számológépet, a számolótáblákat, felfedezi a matematikát, és hamarosan már meghatározza a matematika törvényszerűségeit is.

Az idő előrehaladtával az ember egyre több matematikai problémát akart megoldani, amihez már nem volt elég a két keze, a golyós számológép is kevés, az sem elég, hogy ismeri az algebra szabályait, egyre csak azon gondolkodik, hogy a számolás műveletét egy számológéppel megkönnyítve, még bonyolultabb feladatokat tudjon megoldani.

Azután felnéz az égre – már az ősember is megnézegette az égitesteket, amelyekből még isteneket is kreált magának –, azután felfedezi az égi mechanikát (a napszakok és az évszakok változását, a betlehem-i csillagot, a vést hozó üstökösöket, a Nap- és Hold-fogyatkozásokat stb.), sőt hamarosan ki is tudja számolni bizonyos égi jelenségek bekövetkezését. Ehhez viszont számolnia kell, sőt sokat kell számolnia, ami gép nélkül egyre lehetetlenebb.

Így nem csoda, hogy az első kerekes számológépeket csillagászok találják fel, Al-Kassi, a szamarkandi obszervatórium vezetője (1393–1449), valamint Wilhelm Schickard (1592–1635) Tübingenben, a csillagászat tanára és matematikus. Híres számológép-konstruktőr volt Blaise Pascal (1623–1662), aki ugyan nem volt csillagász, de a számológépét csillagászok is használták. Időben majdnem kortársa volt Gottfried Wilhelm Leibniz (1646–1716), polihistor, sok mindennel foglalkozott, megalkotta például a számológépéhez a Leibniz-kereket, amit – számológépek konstruálására – még a 20. században is használtak.



1. ábra. Godfried Wilhelm von Leibniz (1646–1716) matematikus, a mechanikus – négy alapműveletes – számológép megalkotója. A gép legfontosabb alapeleme a lépcsős Leibniz-fogaskerek volt, amit – mechanikus számoló-berendezésekben – még a 20. században is használtak. (1670)

Azután Angliában megjelent Charles Babbage (1791–1871), aki ugyancsak közeli kapcsolatban volt a csillagászzal és kora neves csillagászaival, de inkább matematikus volt, és megszállottja a számolás gépesítésének. Anglia ekkor válik tengeri nagyhatalommá, az óceánokon való hajózásban kitüntetett szerepet kap a csillagászat és az időmérés, hiszen a világtengereken csak a pontos idő és az égitestek alapján lehet tájékozódni. A tökéletes ha-

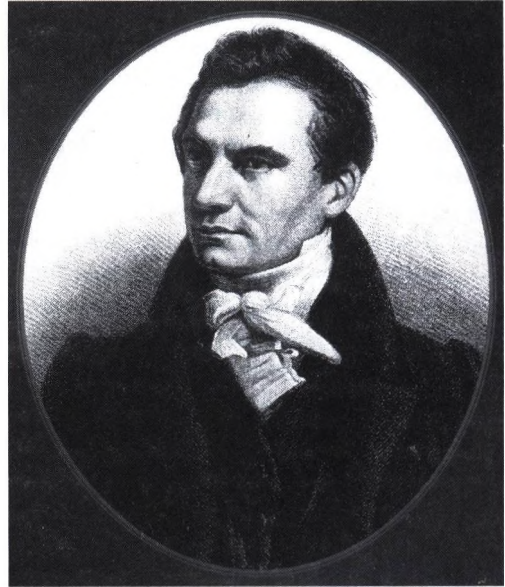
józási órát 1759-ben John Harrison órás-mester alkotja meg, amivel elnyeri a pontos tengeri óra elkészítésére kiírt díjat.

A királyi csillagászok közzé teszik a csillagászati évkönyveket, bennük azokat a táblázatokat, amiknek alapján – megfelelő mérőműszerek (szextáns, oktáns stb.), valamint a már említett pontos óra segítségével – a világ bármelyik pontján pontosan meg lehet határozni, hogy a hajó hol tartózkodik. Egy, a táblázatokat kiszámoló gép megalkotása egyre sürgetőbbé vált, ugyanis a csillagászati módszerek és a matematikájuk is állandóan fejlődik, ezért a hajósok egyre újabb táblázatokat igényelnek, ezt az egyre nagyobb számítási igényt már csak gépekkel lehet kielégíteni.

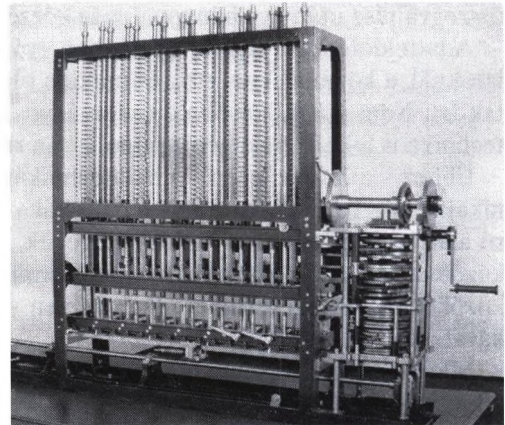
Ekkor lép a színre Charles Babbage, aki célul tűzi ki, hogy a hajózási csillagászati táblázatokat géppel számítsa ki, sőt ki is nyomtassa. A Királyi Csillagászati Társaság már 1820-ban kitünteti a „*Gépek matematikai táblázatok kiszámításánál való alkalmazásának tapasztalatai*” c. munkájáért, igaz Babbage először még egy gőzgépmeghajtású számológépre gondolt.

A brit kormány 1823-ban dönti el, hogy támogatja Babbage – csillagászati táblázatokat számoló – *differenciagépének* a megépítését, amit az alkotónak sohasem sikerült tökéletesen befejeznie. Mint megannyi más zseni – mielőtt a gépet befejezte volna –, új gondolat ébredt benne, kitalálta, hogy az általa kidolgozott elvek és a differenciagép alapján építeni tudna egy általános célú számológépet is, ami nem csak kötött formulák alapján képes számolni, hanem bármilyen formulát – aminek meg lehet írni a *programját*, és azt – *lyukkártyákon* – a gépnek meg lehet adni – ki lehet számolni. A találmányát *analitikus gépnek* nevezte.

Babbage ezzel feltalálta a programozott mechanikus számítógépet. Nagy szerencséjére csatlakozott hozzá egy fiatal matematikusnő, Ada Byron, Lady Lovelace (1815–1852), aki azonnal megértette a gép nagyszerűségét. Babbage egyik olaszországi előadásáról Luigi F. Manebrea tábornok (1809–



2. ábra. Charles Babbage (1791–1871). Matematikus és csillagász, a Királyi Csillagászati Társaság tagja, az első, programozott, mechanikus számológépek – a *Difference Engine* és az *Analytical Engine* – feltalálója.



3. ábra. Babbage második gépének, az *Analytical Engine*-nek a működő másolata, amit – az eredeti tervek alapján – születésének a 200. évfordulójára a londoni Science Museumban építettek meg. (1991)

1896) írt beszámolót, aminek az angol fordítását Lady Lovelace végezte el – az előadásnál háromszor hosszabb – jegyzetekkel. Babbage zseniális alkotását és a gép programozását a világ ezekből a Lady Lovelace által készített jegyzeteből ismerte meg.

Babbage-nak egyik gépét sem sikerült tökéletesen befejeznie, noha a már működő differenciagépét az 1862-es londoni világkiállításon is bemutatták.

Hermann Goldstine könyvéből idézet:

„Egy Pehr Georg Scheutz (1785–1873) nevű svéd nyomdász Babbage gondolatától indítva megépítette a differenciagépet, és 1854-ben Londonban – Babbage jelentős segítségével – be is mutatta. (...)

Scheutz ugyanis 1834-ben Edward fiával együtt egy, az *Edinburgh Review*-ban megjelent cikkben olvasott Babbage differenciagépéről. A cikk felkeltette érdeklődésüket, és hozzáfogtak egy lényegesen módosított, tökéletesített változat megépítéséhez. 1834-ben a gép működőképes volt. (...) 1853-ra Scheutz befejezte a gép tökéletesített változatát, ami 1855-ben a párizsi világkiállításon aranyérmert nyert.”

Babbage Scheutz gépéről – ebben az időben – így írt: „(...) A gép pontos másolatát a Messrs. Dunkin and Co. cég készítette el az angol kormány számára, és jelenleg a Sommerset House-ban, az anyakönyvi irattárvezető irodájában használják.”

A hajózási – csillagászati – számológépigény után felmerült a nagytömegű adatok feldolgozásának a szükségessége is, ennek az első lépése a népszámlálási adatoknak a korábbiaknál gyorsabb feldolgozása volt. A sokszereplős történetből Hermann Hollerith (1860–1929) nevét kell megemlítenünk, aki 1889-ben kapott a *népességi statisztikák táblázatba foglalására* szabadalmat, az első Hollerith-gépek segítségével az 1890-es amerikai népszámlálást nagy sikerrel bonyolították le. Amíg a korábbi – 1880-as és a még korábbi – népszámlálások alkalmával az adatok kézi feldolgozása általában a következő népszámlálásig tartott, addig – 1890-ben – az első gépi feldolgozásnál az első adatokat a lyukkártyák összegyűjtése után egy hónappal már közzé lehetett tenni.

Adatfeldolgozás céljából az iparban gyártott Hollerith-gépek megjelentek a nagyvállalatoknál, a közműveknél stb., és minden olyan helyen, ahol sok adatot tároltak és dolgoztak fel. Nem sokkal később a tudományos élet – éppen a csillagászat –, valamint a haditechnika is felfedezte a gépi számolásban rejlő tudományos és katonai lehetőségeket.

Ehhez – természetesen – a haditechnikának is fejlődnie kellett, például a lövészet technikája a tüzérek művészetéből a ballisztika tudományává vált. Azután – az első világháború alatt – megjelentek a hadirepülőgépek, amiket földi légelhárítással szerettek volna leküzdeni, így azután kifejlődött a haditechnikai méréstudomány, amit – hogy hatékony eszközöket és módszereket eredményezzen – egyesíteni kellett a katonai számolástechnikával.

Mielőtt elérnénk Neumann Jánoshoz, egy példával mutatjuk be ennek a fejlődésnek egy állomását.

Az ellenséges repülőgépek megsemmisítéséhez a hadseregnek egy olyan eszköze volt szüksége, ami a repülőgépet tűzzel fogadó lövegeket a repülőgép mozgásának, az időjárási viszonyoknak, a lövegek adatainak ismeretében automatikusan célra állítja. Ez a löelemképző. A II. világháború előtt ezt a feladatot a leghatékonyabban egy magyar mérnök,

Juhász István (1894–1981) oldotta meg, aki Budapesten kifejlesztette a GAMMA-Juhász löelemképzőt, amit számos ország: például Svédország, Argentína, Kína, Svájc, Hollandia, Norvégia, Finnország és Irán – majd a második világháború után Egyiptom hadseregeiben is hadrendbe állítottak. A löelemképző valójában egy bonyolult mérőrendszerrel összekötött vezérlőrendszer és az ezt irányító, ballisztikus fixmemóriával rendelkező elektromechanikus analóg számítógép volt.

Időben most értünk el a negyvenes évek első éveiben Neumann Jánoshoz és az elektronikus számolóeszközökhöz. Az amerikai hadvezetés, a negyvenes évek elején – látva a németországi eseményeket – biztosra vette, hogy nem maradhat ki az Európában dúló háborúból, ezért igyekezett a háborús részvételre felkészülni. Azt is látták, hogy a harcot az I. világháború után kifejlesztett modern haditechnikai eszközökkel – mozgékony gépesített hadsereggel, modern repülőekkel és tüzérséggel kell megvívnia, ezért elérkezettnek látta az időt, hogy a tudományt – elsősorban a matematikát és a fizikát – is „hadrendbe” állítsa.

Már a harmincas években elkezdődött az európai – főleg zsidó tudósoknak: Einstein, Wigner, Kármán, Oppenheimer, Szilárd, Teller, nem utolsósorban Neumann és mások – a náci megszállás elől Amerikába menekülése, ezek az emberek óriási mértékben megerősítették az ország tudományos és így védelmi kapacitását.

Hamarosan behívták katonai szolgálatra Hermann H. Goldstine-t, a ballisztika tanárát, matematikust, aki azt a feladatot kapta, hogy a Ballisztikai Kutató Laboratóriumban modernizálja – többnyire az első világháborúban készült – tüzérségi és bombázási táblázatokat. Goldstine felmérte a feladatot, és megállapította, hogy az akkor használt matematikai eszközökkel – Bush-analizátor, két jelfogós számítógép (Aiken és Stibitz jelfogós gépei), elektromechanikus asztali számológépekkel stb. – nem lehet a munkát rövid idő alatt elvégezni, egy új elvű, hatékony és nagyságrendekkel gyorsabb számolóeszközt kell keresni vagy kifejleszteni.

Közben elindult a Manhattan-terv, a versenyfutás a német tudománnyal, hogy ki tudja előbb megalkotni és a háború eldöntésére bevetni a minden eddiginél hatékonyabb fegyvert, az atombombát. A tudósok csatasorba álltak, egyik részük Los Alamosban atombombát fejlesztett, a másik részük pedig megalkotta a Pennsylvanai Egyetem Moore (villamosmérnöki) intézetében az első nagyon gyors számítóeszközt, az ENIAC-ot.

Ezt követően találkozott Hermann Goldstine az aberdeen-i állomás peronján Neumann Jánossal, elindult egy csodálatos együttműködés, ami – talán nem túl nagy szavak – elvezetett az első számítógépekhez és – hosszú távon – az informatikai társadalomba.

Hogyan jutott el Neumann János a First Draftig?

Neumann János „ismerkedik” a számológéppel (Moore School)

Neumann Jánost 1943-ban kinevezték tanácsadónak az atombomba előállítását előkészítő Manhattan-projekthez, egyúttal az aberdeeni Ballisztikai Kutató Laboratóriumba is. Az itt felmerülő numerikus matematikai problémák megoldásához mindinkább szüksége lett volna valamilyen mechanikus gyors számolóeszközre. Ezt többször szóvá is tette. 1944

januárjában levelet írt Warren Weavernek, az Alkalmazott Matematikai Bizottság elnökének. Azt kérdezte, hogy a kormánynak milyen eszközei vannak a számítások elvégzésére. Weaver beszélt neki Aikenről, Stibitzről és Wallace Eckertről. A Moore Schoolt nem említette meg, mert nem tartotta reális kezdeményezésnek.

Pedig Moore Schoolban John Mauchly és J. Presper Eckert vezetésével – John V. Atanasoff elveinek a felhasználásával – ekkor már elkezdték építeni az ENIAC (*Electronic Numerical Integrator and Computer*) első elektronikus számolóeszközt. Az 1923-ban létrehozott „Moore School of Electrical Engineering” a Pennsylvania Egyetem önálló elektromérnöki részlege, „kara” volt. Írásunkban Neumann Jánosról beszélünk, de az előbbi két név megemlítése nélkül ezt több okból nem tehetjük meg. Ha Neumann Jánost a számítógépek atyjának nevezzük, akkor nem volna nagy merészség azt állítani, hogy John Atanasoff, valamint Eckert és Mauchly a nagyapák.

Találkozás a pályaudvaron

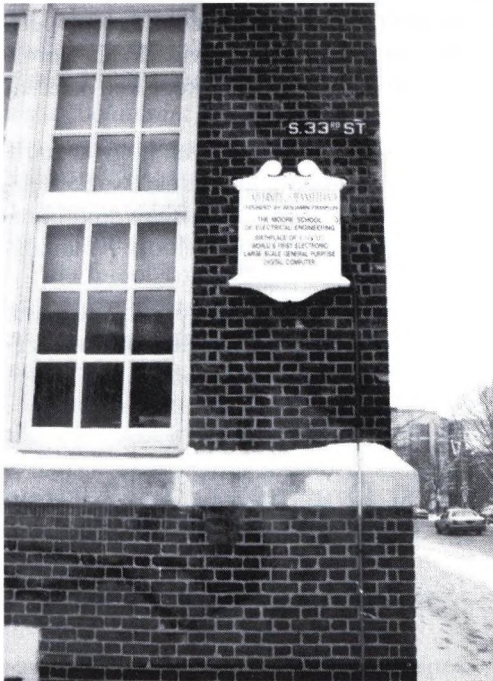
Tanácsadóként Neumann sokat utazott Los Alamos, Aberdeen és Princeton között. 1944 júniusában az aberdeeni pályaudvar peronján állt, amikor Goldstine csatlakozott hozzá. (Hermann H. Goldstine matematikus – mint láttuk – a hadsereg, a Pentagon képviselőjeként vett részt az ENIAC projektben.) Goldstine azonnal megismerte Neumannt, mert néhány évvel korábban részt vett egy konferencián, ahol az előadók egyike Neumann János volt.

„Nyilvánvaló az, hogy Neumann rám emlékezett volna, lehetetlen” – mondja Goldstine. Mindenesetre Goldstine odament Neumannhoz, bemutatkozott.

Neumann János nagyon kedvesen fogadta Goldstine bemutatkozását, és megkérdezte, hogy mi járatban van itt. Goldstine elmondta, hogy az ENIAC projektben vesz részt. Mindjárt el is mesélte Neumann-nak, hogy ez a ENIAC gép 333 szorzást végez másodpercenként. Neumann-nak a szemei felcsillantak. A beszélgetésről Goldstine azt mondja, hogy a találkozás pillanatok alatt egy doktorátusi vizsgává változott, Neumann mindenről kikérdezte az ENIAC géppel kapcsolatban. Goldstine megemlíti azt is, hogy Neumann teljesen megértette, miről van szó, és hogy mit tud az ENIAC.

Látogatás a Moore Schoolban

Néhány nappal később Neumann felhívta Goldstine-t és arra kérte, hogy mutassa be neki Moore Schoolt, ahol az ENIAC gép épült. Mauchly így emlékszik vissza arra az



4. ábra. A Pennsylvániai Egyetem villamosmérnöki kara, a Moore School, ahol az ENIAC épült. Az ENIAC emléktábla

időre. „A látogatásról valamikor augusztus első hetében hallottunk. Mondták, hogy ide jön egy kiváló matematikus, beszélni akar velünk, és könnyen meglehet, hogy nagy segítséget nyújthat nekünk, mivel ő a világ legnagyobb matematikusa. Izgatottan vártam, hogy találkozzam ezzel az emberrel és kiderüljön, hogy miben tud segíteni.”

A látogatás végül szeptember első hetében létre is jött. Goldstine bemutatta a Moore Schoolban dolgozó munkatársakat, és megmutatta az akkumulátorokat (mai kifejezéssel regisztereket). Ez egy kicsivel gyorsabban számolt, mint Neumann János. „Neumann János megnézte a kiállításunkat, és mindenről mindent elmondunk neki. Utána bementünk egy szobába, és elkezdtünk beszélgetni egy tábla előtt. Ismertettük vele a gép terveit, amit éppen építettünk. Neumann világosan és azonnal megértette, hogy ennek a gépnek milyen számítási képessége lesz, és ennek az ENIAC gépnek a Los Alamosban (ahol az atomkutatások folytak) felmerülő problémák megoldásában nagy szerepe lehet. A beszélgetés közben Neumann egészen előre ment, és elkezdett gondolkodni azon, hogyan lehetne a mi terveinkhez képest jobb számológépeket építeni” – mondta Mauchly.

Neumann, a tanácsadó

Neumann János ettől kezdve tanácsadóként bekapcsolódott az ENIAC projektbe, és több alkalommal is meglátogatta Moore Schoolt. Lelkes támogatójává vált a Moore Schoolban folyó munkának. Állandó kapcsolatban állt Mauchly-val és Eckerttel. Neumann különösen a számítógép logikai felépítésének az elemzésében volt nagy segítségükre. A technikai részek abban az időben nem keltették fel Neumann érdeklődését.

Neumann János közreműködése az ENIAC-ban nem csak szakmailag volt jelentős, hanem azzal is, hogy a fejlesztés végén bekapcsolódva, nevének presztízse miatt a tervezők további 105 000 dollárt kaptak a Pentagontól 1944 októberében.

Neumann megjelenésével a korábban szokásos informális beszélgetések helyett napi-rendekben meghatározott üléseken vitatták meg a problémákat. Neumann János nagyon aktív volt ezeken az üléseken, rengeteg javaslatot tett. Az ENIAC-kal a fő probléma a programozható memória (az adatok tárolásához szükséges szerkezet) volt. Látták, abból a célból, hogy a gép igazán jól használható legyen, szükség van arra, hogy mind az utasításokat, mind pedig az adatokat „meg tudja jegyezni” (tárolni tudja), és hogy az utasítások anélkül lehessenek megváltoztathatók, hogy egy operátornak dugaszolással, billentyűk lenyomásával kellene változtatni a programon (mint az ENIAC esetében).

Eckert és Mauchly közben megkapta az ENIAC-ra a szabadalmat. Gyakorlati szakemberek voltak, és világosan látták a gépeiknek a kereskedelmi lehetőségeit. A Moore Schoolban sokan nem hitték, hogy a számítástechnikából kereskedelmi haszon is származik. Ebbe a csoportba tartozott – érdekes módon – Neumann János is. Meg volt győződve arról, hogy ezek a munkálatok voltaképpen a háború miatt voltak fontosak, vagy esetleg a „tisza tudomány” szempontjából. A dolog szabadalmi része egyáltalán nem érdekelt.

Annak, hogy Neumann János „becsatlakozott” a Moore School Projektbe, két közvetlen hatása is volt. Az első, hogy elismertette a projektet, például a Nemzeti Védelmi Kutatási Bizottsággal, aminek eredményeképpen további támogatást kaptak a projekthez.

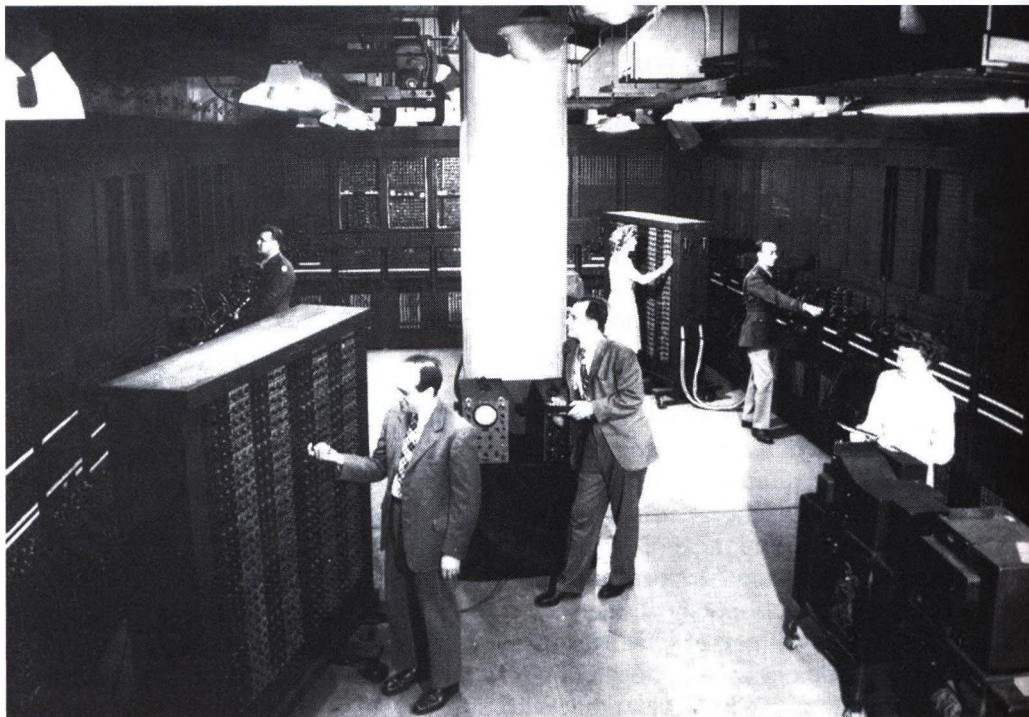
A második eredmény az volt, hogy miután Neumann János is érdekelt volt az ENIAC-programban, a Moore Schoolal szerződést kötöttek egy új, sokkal nagyobb teljesítményű számítógépnek, az EDVAC-nak a kifejlesztésére.

Neumann először mint potenciális felhasználó (user) érdeklődött a Moore Schoolban folyó munkálatok iránt, de később már mint tudományos és technikai tanácsadó a projekt egyik főszereplőjévé vált; számos konferencián vett részt, amelyeken javaslatokat fogalmazott meg az EDVAC logikai tervezésével kapcsolatban. Amikor Neumann Los Angelesben volt, állandó levelezésben állt Goldstine-nel, aki informálta őt a projekt haladásáról; Neumann Goldstine közvetítésével juttatta el javaslatait a projektben résztvevőknek.

Röviden az ENIAC gépről

Arthur W. Burks, aki maga is oszlopos tagja volt az ENIAC-ot létrehozó csoportnak, „From ENIAC to the Stored-Program Computer: Two Revolutions in Computers” c. cikkében az ENIAC-kal kapcsolatban a következőket írja: „...két fontos forradalom volt a számítógépek területén. Az első az elektroncsövek alkalmazása, amelyek segítségével gyors, megbízható, nagy teljesítményű általános célú számítógépet lehetett építeni. Ez a fejlesztés John Atanasoff-fal indult, aki egy speciális célú elektronikus számítógépet épített, és az ENIAC-ban kulminált. ... Az ENIAC forradalmi jelentőségű volt: ez volt az első elektronikus, digitális, általános rendeltetésű tudományos számítógép, amely 1000-szer gyorsabban végezte a számításokat, mint elektromechanikus versenytársai. ... A másik forradalom a tárolt programú számítógép létrehozása volt.”

A gép tervezését 1943 áprilisában kezdték, 1943 júniusában már elindult a beruházás, 1945 végén fejezték be.



5. ábra. Az ENIAC, a világ első elektronikus számolóberendezése, 1943 és 1946 között épült.

A fő cél voltaképpen lövéstani számítások elvégzése volt. A gép kb. 18 000 elektroncsövet tartalmazó áramkörökből épült fel.

Főbb jellemzői:

- Különálló, „félautonóm” egységekből állt, a vezérlés részben elosztott volt (az egyes egységekben lévő vezérlést egy „koordináló” vezérlés „kötötte össze”).
- Külső programozású volt, a programokat dugaszolással vitték be a gépbe.
- Decimális számrendszerben dolgozott.
- Lyukkártyás bevitellel (input) és kivittel (output) működött.

Fontosabb egységek:

- 10 decimális jegyet + előjelet befogadni képes 20 akkumulátor (összeadó egység)
- 1 szorzóegység
- 1 osztó/gyökvonó egység
- Inicializáló egység, szinkronizátor
- Központi programadó („Master programer” a programhurkok vezérlésére)
- 1 kártyaolvasó, 1 kártyalyukasztó

Műszaki adatai:

- Műveleti idők: összeadás 200 μ sec, szorzás 2,8 msec, osztás 2,4 msec
- 125 kártyát olvasott be, illetve 100 kártyát lyukasztott percenként
- Energiafelhasználás 174 kW, alapterület 180 m²
- Hatásfok: 70%
- Kiszolgáló személyzet 6 fő/műszak

A gép egy évig működött Philadelphiában (Los Alamos részére végzett számításokat), majd Aberdeenbe telepítették át (itt 1955-ig használták).

Az új gép: az EDVAC

A Moore School megrendelést kapott a hadügyminisztériumtól egy új számítógép kifejlesztésére. Az új gép neve EDVAC (Electronic Discrete Variable Computer = Elektronikus diszkrét változójú számítógép).

A „Jelentés”

1944 szeptemberétől Neumann János aktív résztvevője volt az EDVAC projektnek.

1945. június 30-án Neumann Jánostól egy 101 oldalas dokumentum érkezett a Moore Schoolba, Los Alamosból. A címe a következő volt: „First Draft of a Report on the EDVAC” („Az EDVAC-ról szóló je-



6. ábra. Az EDVAC, a Moore School-ban tervezett második, de már tárolt programú, soros működésű, elektronikus (elektroncsöves) számítógép. (1945)

lentes első változata”, a továbbiakban röviden „First Draft...”-nak vagy Jelentésnek hívjuk). Ebben a dokumentumban Neumann János az EDVAC gép teljes leírását adja, összefoglalva mindazon gondolatokat, amelyek Moore Schoolban jöttek létre az új géppel kapcsolatban. Az anyagban leírta a gép logikáját, a gép felépítését. Neumann a gépet gyors működésű, automatikus, digitális számítógépszernek nevezte.

Neumann János elképzelését jelentősen befolyásolta Warren McCulloch neuropszichológus és Walter Pitts matematikus munkássága, akik 1943-ban egy dolgozatot publikáltak, amelyben az emberi agy működését, illetve az agy pszichológiáját úgy írták le, mint egy, az agy által elvégzett logikai kalkulust. Neumann János az elkészített dokumentumban erre hivatkozott, felhasználta McCulloch és Pitts elképzeléseit. Vizsgálta a neuronok és az emberi idegrendszer egyéb területeinek működését, és összehasonlította az automatikus számítógéppel. Ezzel az első lépést tette meg a számítógép „antropomorfizálása” terén; ami önmagában is nagyon jelentős.

Csak Neumann neve szerepel a Jelentésen. Jogos-e?

Goldstine sokszorosította a dokumentumot, szétszította Moore Schoolban, majd elküldte a másolatokat sok más, a dolog iránt Amerikában és Nagy-Britanniában érdeklődő tudósnak. A Moore Schoolban, Eckertnél és Mauchly-nál nagy elégedetlenséget váltott ki, amikor kiderült, hogy milyen széles körben terjedt el a dokumentum. Nagyon haragudtak Goldstine-re és voltaképpen Neumannra is. Ennek első oka az volt, hogy a munka szigorúan titkos minősítéssel folyt, és azzal, hogy a „First Draft...”-ot nyilvánosan osztogatták, megszegették a biztonsági előírásokat. A második ok az volt, hogy Mauchly és Eckert szabadalomért folyamodtak az EDVAC-ügyben, és nyilvánvalóan nem volt kétségük afelől, hogy ezt meg is kapják. Attól féltek, hogy a publikáció csökkenti annak esélyét, hogy szabadalmaztathatják az EDVAC-ot.

Az elkeseredettség harmadik, talán a legfontosabb oka az volt, hogy a Neumann-féle dolgozat lényegében semmilyen hivatkozást nem tartalmazott a Moore Schoolban működő többi kutatóval kapcsolatban. Eckert és Mauchly nem tartották helyénvalónak, hogy az általuk elért eredmények Neumann nevéhez fűződnek majd, noha azok némelyike tényleg nem volt Neumannhoz köthető. Tudták, hogy a csoportos kutatásoknál az eredményt a legismertebb névhez kötik, ráadásul annak a nevéhez, aki legelőször publikálja azt. Eckertnek nem volt lehetősége publikálni a dolgokat, mert az EDVAC gép katonai titkot jelentett. Semmi sem mutat arra, hogy Neumann el akarta orozni mások munkáját, hiszen a dokumentumra ráírta, hogy „Frist Draft...”, amivel jelezni akarta, hogy a dolgozatot szét kell osztani a többi kutató között, abból a célból, hogy megtegyék a megjegyzéseiket, korrekciójukat, mielőtt a végleges dolgozat létrejön. Goldstine akciója azonban végleges helyzetet teremtett, amelyen már nem lehetett segíteni.

Neumann János a „számítógép atyja”?

Többen azt mondják, akik a számítástechnika korai történetét feldolgozták, hogy a tárolt programú számítógép ötlete voltaképpen nem Neumann Jánostól származik, mert egy ilyen elven működő gép létrehozásának lehetősége a Moore Schoolban már azelőtt felvetődött, mielőtt Neumann János a színre lépett. Eckert már 6 hónappal azelőtt írt erről egy emlékeztetőt, amikor Neumann János egyáltalán nem hallott még a Moore School-pro-

jektről. Goldstine ezzel szemben a következőképpen védi meg Neumann Jánost: „A jelentés mesteri analízisét és szintézisét adja mindazon elgondolásoknak, amelyek az EDVAC-kal kapcsolatban 1944 őszétől 1945 tavaszáig megszülettek. Nem minden gondolat származik tőle, de a jelentősek igen. Nyilvánvaló, hogy Neumann János megírva a „Jelentést”, oly módon kristályosította ki a számítógéppel kapcsolatos gondolkodásmódot, ahogy senki más ezt előtte nem tette meg.”

Néhány történetíró azt állítja, hogy a Jelentés címlapján több szerző nevére is hagytak helyet. Tényleg igaz, hogy két címlap készült. Az elsőt, amelyen csak a Jelentés címe és Neumann neve szerepelt, volt hely több szerző részére is. A második, végleges címlapon azonban semmi nem mutatott arra, hogy Neumann János neve mellé más neveknek is kellett volna kerülniük.

„Neumann János hozzájárulása az EDVAC projekthez valóságos és fontos volt, és általában is az első számítógép létrehozásában játszott szerepe nem volt kicsi, de a Jelentés publikálása ezt beárnyékolta” – mondta Goldstine.

Az a körülmény, hogy a „First Draft...”-on csak Neumann neve szerepelt, és ezzel ő lett a tárolt programú számítógépek „atyja”, Moore Schoolban, az EDVAC projekt széteséséhez vezetett. Ehhez még hozzájöttek a szabadalmi problémák is, így együtt egy sor konfrontáció keletkezett Neumann és az EDVAC tervezői teamje között, aminek következtében sokan otthagyták a Moore Schoolt, és más egyetemi intézetekbe mentek át, vagy saját céget alapítottak. Végül is ez azt eredményezte, hogy az EDVAC további tervezése és fejlesztése lényegében megállt. Egyébként Burks, aki szintén egyik résztvevője volt az EDVAC projektnek, a Neumann jelentéssel kapcsolatban a következőket írja. „Kétségtelen, hogy mások nevét is szerepeltetnie kellett volna. Véleményem szerint elsősorban P. Eckertet és J. Mauchly-t, H. Goldstine-t és engem is be kellett volna írni a jelentés szerzői közé.”

Egyes történetírók azzal kelnek Neumann János védelmére, hogy a jelentést kézírással küldte Moore Schoolba abból a célból, hogy a Hadügyminisztériummal kötött W-670-ORD-4926 számú szerződés teljesítéseként adják be, küldjék el a Pentagon azon osztályára, ahonnan a projektet megrendelték.

2. A First Draft

Az eddig elmondottakkal azt kívántuk bemutatni, hogyan került a korszak nagy (legnagyobb) matematikusa a számítógépek közelébe, hogyan jutott el a „First Draft...” (a „Jelentés”) megírásáig, ami miatt voltaképpen őt tekintik a „számítógépek atyjának”.

A „First Draft...” magyar nyelven még nem jelent meg. Mi nem tűzhetjük ki célul, hogy ezt pótoljuk, de az alábbiakban részletesen ismertetjük fő mondanivalóját. Mondatainkat több helyen akár pontos idézetnek is tekinthetjük, de nem tesszük idézőjelek közé.

Sok helyen megjegyzéseket fűzünk a Neumann-féle filozófiához.

E fejezet íróinak szerencsésük volt: a „First Draft...” keletkezésétől számított alig több mint egy évtizeddel „belecsöppentek” az első hazai számítógép építésének és programozá-

sának munkálataiba, és ettől kezdve saját élményeik alapján növekvő csodálkozással figyelhették, mi lett a Neumann-elv alapján épülő EDVAC-ból. Lenyűgözően meredek fejlődési ív, megjósolhatatlan perspektívával.

Mielőtt rátérnénk a „First Draft...” ismertetésére, elemezzük, milyen előzmények vezettek ahhoz, hogy ez az összefoglaló mű létrejöjjön.

Közvetlen előzmények. Általános áttekintés

Annak érdekében, hogy a „First Draft...” megírásával kapcsolatban az olvasó Neumann János egyedülálló teljesítményét értékelni tudja, röviden összefoglaljuk az olvasó számára érdekes előzményeket, azokat, amelyekből a „First Draft...”-ban szereplő elvek, gondolatok születtek (vagy születhettek). Azt is szeretnénk, ha az olvasó megértené, hogy a „First Draft...” megírásakor Neumann János miért dolgozott teljesen egyedül, a munkatársai nélkül a dolgozaton (bár néhány részletében felhasználta a korábbi munkájukat), illetve miért indult el az EDVAC leírásával egy addig nem járt, a formális logikával jellemzett úton, ami a mai modern számítógépek megtervezésének a technikájához vezetett. Ebből az összefoglalóból – talán – még az is megérthető lesz, hogy miért nem volt szükség a „First Draft...”-on – annak ellenére, hogy a címében „vázlatnak” készült – további munkára és kiegészítésre.

Még valami. Neumann megjegyezte, azért írta meg ezt a jelentést, hogy *„hozzájáruljon a nagysebességű számítógépekkel kapcsolatos ismeretek további fejlődéséhez, továbbá a lehető legkorábban és legkiterjedtebben tárgyalja az e tárggyal kapcsolatos tudományos és műszaki gondolásokat”*.

Neumann egyik munkatársa elmondta, hogy Neumann a számítógépet mindig is a tudomány közkincsévé kívánta tenni, azért volt ellene minden szabadalmaztatási kísérletnek, hogy a tudományos világot ne korlátozza bizonyos dolgoknak egy néhány embernek a birtokában lévő tulajdonjoga. Azt is mondták, hogy a számítógépek fejlesztésének a korai szakaszában Neumann János nem gondolt arra, hogy egyszer majd a számítógépeket gyárakban fogják tömegével előállítani, az volt az elképzelése, hogy a számítógép egy olyan eszköz lesz, amiket különböző egyetemek és kutatóintézetek a maguk képére és hasonlatosságára, a felhasználandó célra fogják egyedileg előállítani. Ebben az esetben mindenféle szabadalmaztatási eljárás valóban akadályozta volna a számítógépek gyors elterjedését.

A „First Draft...” tárgyalása előtt megvizsgáljuk az EDVAC-ot közvetlenül megelőző számolóberendezések logikai és műszaki felépítését, külön felhívva a figyelmet az azokból átvett megoldásokra és részegységekre, valamint az eltérésekre és az azonosságokra.

- Az EDVAC megtervezése előtt Amerikában szinte alig voltak mérnökök és matematikusok, akik a számolóberendezések tervezésében, a műveletek vezérlésében, egyáltalán a számítógépes elektronikus áramkörök kialakításában tapasztalatokkal rendelkeztek volna.
- Az említett elektromechanikus Hollerith (lyukkártyás) adatfeldolgozó gépek az első elektronikus számítógépek szempontjából azért voltak fontosak, mert John Mauchly és Presper Eckert, az ENIAC tervezői az ENIAC programozásának az áramköreit és

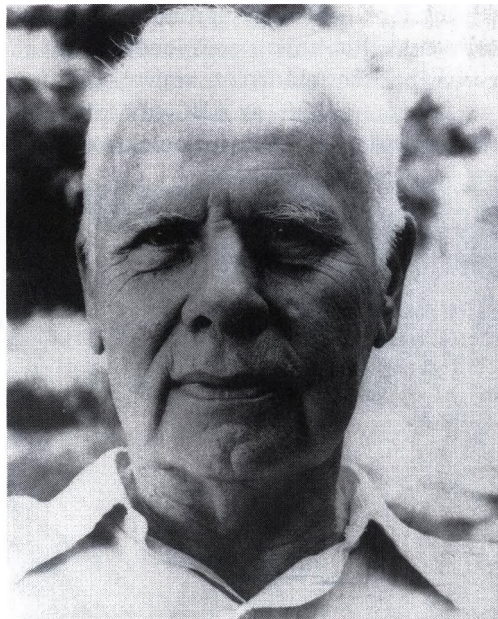
módszerét ezektől a gépektől vették át. Már a Hollerith-gépeket is felhasználta Wallace J. Eckert tudományos – csillagászati – számítások elvégzésére, de a gép konstrukciója erre csak némi változtatásokkal volt alkalmas. Wallace J. Eckert munkásságának az óriási jelentőségét abban lehetne összefoglalni, hogy az IBM figyelme az adatfeldolgozás irányából a tudományos számítások felé fordult.

- A számítások gépesítése szempontjából ugyancsak óriási jelentősége volt – a harmincas évek végén megépült – két jelfogós számítógépnek, az egyiket – a MARK I-et – Howard H. Aiken vezetésével a Harvard egyetem és az IBM közösen fejlesztette, a másikat pedig – a Model I–V-öt, a részben automatikus számológépet – a Bell Laboratórium, George R. Stibitz irányításával építették meg. Ezek a relés számológépek tulajdonképpen Charles Babbage mechanikus építésű, analitikus gépét másolták le, így – noha az ENIAC-ba szinte semmiféle megoldásuk nem került át – mégis az ENIAC közvetlen „rokonának” voltak tekinthetők.

- Valójában a jelfogós számítógép-technikának már nem volt ideje kifejlődni, mert alig két-három év múlva a számítástechnikában megjelent az elektroncső, ami több nagyságrenddel gyorsabb kapcsolóelem volt, így – az éppen megszületett jelfogós technikát – gyorsan használaton kívül helyezte. Egyébként Aiken egy 1937-es kiadatlan tanulmányában már célzó a számítógépnek az egyik legfontosabb jellemzőjére: „...*(a gép) egy számítást a matematikai műveletek természetes sorrendjében végezzen el*”. Ezt a megjegyzést, közvetve – talán belemagyarázással – úgy is lehetne értelmezni, hogy a gép a matematikai műveletek programját automatikusan hajtsa végre.

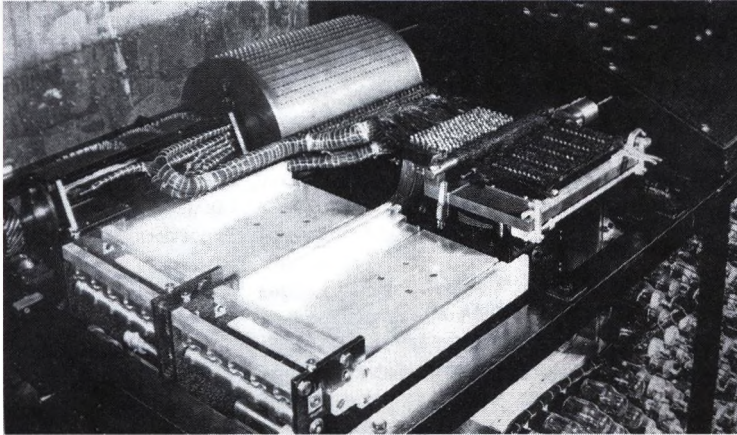
Ezért írja azt Hermann H. Goldstine a jelfogós számológépekről: *... a Harvard Számítás-technikai Laboratóriumában található gépek önmagukban semmiképpen nem voltak a jövő előfutárai. Egy újszülött tudományág korai időszakában igen sok újszerű ötlet lát napvilágot, és ölt testet valamilyen szerkezet alakjában. A tudományos alkalmazás választóvize azután az alkalmatlanokat kiszűri; ez a technológiai analógiája a természetes kiválasztás darwini elméletének, amely szerint a legéletképesebb egyed marad meg. Amikor egy új, fontos gondolat első ízben napvilágot lát, egy sor módosulásra lehet számítani, és ez mindaddig folytatódik, amíg egy új, stabil helyzet nem kristályosodik ki.*

- A számolóberendezések fejlődésében ez a módosulás – John Vincent Atanasoff fellépésével – viszonylag gyorsan bekövetkezett. 1937-ben vált világossá Atanasoff számára, hogy elektroncsövekkel – kettes számrendszerben – számolóáramköröket lehet



7. ábra. George R. Stibitz, a Bell Laboratórium elektro-mechanikus, részben automatikus számítógépének, a Model V. jelfogós gépnek a tervezője. (1937)

építeni, több nagyságrenddel gyorsabbakat, mint azt elektromechanikus kapcsolóelemekkel bármikor is ki lehetett volna fejleszteni. Atanasoff feltalálta és alkalmazta a kapacitív dobtárolót, amivel – Clifford Berry-vel közösen – egy elektronikus számológépet épített, az ABC-t (Atanasoff Berry Computer), ami – elszigeteltsége miatt – csak közvetve hatott az elektronikus számítógépek tervezésére.



8. ábra. John Vincent Atanasoff, matematikus, az elektronikus számítógép feltalálója. Az első elektroncsöves számítógépet – az ABC-t (Atanasoff Berry Computer) – 1939-ben Clifford Berry fiatal villamosmérnökkel együtt építette meg.

– Amikor Atanasoff ötletét John W. Mauchly meghallotta, azonnal felkereste az Iowa State College-ben, ahol egy hét alatt megismerte az elektroncsöves számítógépek áramköröket és kikérdezte Atanasoff-ot további elképzeléseiről. Az eredmény Presper J. Eckert villamosmérnökkel közösen született meg: az ENIAC.

Az utókor érdekes vitákat generál az

ENIAC-kal kapcsolatban. Az egyik kérdés, hogy az ENIAC, amikor megépült, egy gyors, elektronikus számológép (calculator, mint például a MARK I) volt-e, vagy pedig számítógép (computer), mint például az EDVAC, az EDSAC, illetve az IAS gép? A hazai, de a nemzetközi meghatározás szerint is, számológépnek nevezik azt a számolóberendezést, ami a matematikai és logikai műveleteket el tudja végezni, a programot egy külön tárolóban (például lyukkártyán, lyukszalagon vagy más tárolóeszközön stb.) őrzik, onnan utasításonként olvassák be és hajtják végre. Ha ezt a meghatározást elfogadjuk, akkor az ENIAC számológép és nem számítógép volt.

- Az ENIAC tervéről az a Nemzetvédelmi Kutatási Bizottság azt mondta, hogy az nem más, mint a gyors számológépprogram leírásának az elektronikus másolata. Számos szakértő – többek között Goldstine – szerint is az ENIAC valóban egy mechanikus számológépnek volt az elektronikus másolata. Felfedezhetők voltak benne az elektronikusan megépített „tízfogú számológerekek” (gyűrűs számlálók), az átvitelképzés, minden akkori mechanikus számológépelemnek az elektronikus megfelelője.
- Goldstine az ENIAC leírásában erről így ír:

Az aritmetikai egységek, az akkumulátorok működésmódjukban nem különböztek túlságosan a kor elektromechanikus, digitális gépeitől. Ezek az elektromechanikus készülékek számlálókerekeket tartalmaztak, amelyek egy elektromos jel hatására egy fokkal elfordultak. Az elektronikus gyűrűs számlálók ezekkel a kerekkel analóg módon működtek. (...) Az ENIAC-ban egy-egy számláló szorosan kapcsolt flip-flopokból állt, amelyeket úgy kapcsoltak össze, hogy a számlálóban mindig csak egy flip-flop volt 1 állapotban, az összes többi 0-ban. Ha a számláló elektromos jelet észlelt, ez a flip-flop visszaállt 0 állapotba, a következő flip-flop

pedig 1 állapotba billent át. A számlálót úgy lehetett nullázni, hogy egy előre megadott flip-flopot, amelyet első helyértéknek neveztek, 1 állapotba kellett állítani. Az ENIAC számlálói mind gyűrűs számlálók voltak, ami egyszerűen annyit jelent, hogy az első és az utolsó hely megfelelő összekapcsolása révén a számláló, ha az utolsó állapotban volt, és áramimpulzust kapott, ismét visszatért az első állapotba, azaz az utolsó hely után ismét az első hely következett. Működési elvében tehát egy gyűrűsszámláló nagyon hasonlított a számlálókerekéhez, csak sokkalta gyorsabb volt azoknál. (...)

...az ENIAC decimális gép volt, amely (tíz számrendszerbeli) 10 jegyű, előjeles számokkal tudott dolgozni, minden akkumulátor – azok az egységek, amelyek az összeadás és a kivonás műveletét végezték – 10 darab 10 fokozatú és egy 2 fokozatú gyűrűs számlálót tartalmazott, az utóbbit a szám előjelének a jelzésére. A 10 fokozatú számláló állapotai feleltek meg rendre a 0, 1, ... 9 számjegyeknek. Valamennyi akkumulátorban átviteli áramkörök kötötték össze a számlálókat úgy, hogy ha valamelyik számláló a 9. fokozatról visszaváltott a 0-ra, a következő, tőle balra elhelyezett számláló egy áramimpulzust kapott, jelezve, hogy a maradékátvitel megtörtént.

Pontosan úgy, mint a néhány korszakkal előbbi fogaskerekes – Schickard, Leibniz, Pascal számológépekben. Éppen ezért, az ENIAC-nál sokkal korszerűbb elveken épült fel a korábban megszületett ABC, Atanasoff gépe.

- Az ENIAC egy nagyon gyors, programozható számológép volt, a mechanikus számológépek analógiájára gyűrűs számlálókkal (elektronikus számkerekekkel) tízes és nem kettes számrendszerben számolt, a gépet valójában dugaszoló táblákon, Hollerith-rendszerben lehetett programozni. Az ENIAC-ban nyoma sem volt azoknak a logikai áramköröknek, amelyek a modern tárolt programú számítógépeket – így először a technikatörténetben – az EDVAC-ot is jellemezték.
- Az ENIAC vezérlése – enyhén szólva – ugyancsak nagyon bonyolult volt. Ismét Hermann Goldstine-t idézzük:

...hogyan lehetett az ENIAC-ot egy adott feladat elvégzésére utasítani. Ez igen bonyolult vállalkozás volt, és ez a bonyolultság volt az egyik oka annak, hogy a gépből nem készült több. (...) A legtöbb egység önálló programvezérléssel (program control) rendelkezett, amely meg tudta jegyezni, hogy az egységnek egy adott műveletet, például egy összeadást el kell végeznie, és jelezni is tudta, hogy ezzel végezt. Általában több programvezérlésre is szükség volt egy aritmetikai művelet elvégzéséhez. Ha például az 1. akkumulátorban tárolt számot hozzá kellett adni a 2. akkumulátorban tárolthoz, az 1. vezérlést utasítani kellett, hogy vigye át az akkumulátor tartalmát, a 2.-ét pedig, hogy fogadja be. (Zárójelben megjegyezzük, hogy a két akkumulátor között összeköttetést is kellett létesíteni, hogy a számokat valóban át lehessen vinni.)

Ezért mondta Goldstine az egyik előadásában, hogy az ENIAC – talán – legnagyobb hibája a programozás megoldatlansága volt, a programozók egy feladat programját körülbelül három hétig írták, míg a gépen a feladat három perc alatt futott le.

Ennek ellenére, az újjelentő L. J. Comrie (1893–1950) matematikus és csillagász, aki először használta a „tekerős” mechanikus számológépeket és a Hollerith adatfeldolgozó gépeket csillagászati számításokra, az ENIAC-ot látva, nagyon költőien írja le a benyomásait:

Tik ... tak ... tik ... tak ...

Három másodperc ... ezer szorzást már elvégzett a gép. Az óra tovább ketyeg ... Egy órán belül egy millióval végez.

Így talán érthető, hogy Neumann János az EDVAC tervezésénél se a műveletvégző egy-

ség, sem a vezérlések megtervezésénél az ENIAC tapasztalatait egyáltalán nem használta fel, szinte mindent maga talált ki. Annak egyelőre nincs nyoma, hogy az EDVAC-megoldások kinek jutottak még az eszébe, a munkatársainak-e vagy pedig csak Neumann Jánosnak. Alapos okunk van feltételezni, hogy a „First Draft...” logikájának a megoldásai, különösen a 6–12. fejezetben leírt áramkörök Neumann János fejében születtek. Akkor viszont miért szerepeltetett volna a dolgozatban társszerzőket?

- Az ENIAC alkotóit nem csak az ENIAC vezérlésnek a bonyolultsága, de az ezzel szoros kapcsolatban lévő programozási problémák is izgatták. Goldstine ezzel kapcsolatos korabeli megjegyzései:

1944-re tehát a Moore Intézetben működő csoport tagjai körében általánosan ismertté vált, hogy egy digitális gép számára az utasításokat numerikus formában lehet szalagon tárolni. (...)

Különösen engem (...) foglalkoztatott sokat az ENIAC programozásához használt eszköz (a dugaszolható programtábla. KGy) esetlensége és a tárolóregiszterek kis száma. Lelki szemünk előtt már a ballisztikai egyenletek megoldásánál jóval bonyolultabb számítások elvégzése lebegett.

...a Moore Intézetben sokat beszélgettünk arról, hogyan lehetne az ENIAC működési hiányosságait kijavítani. Ismeretes egy Eckerttől és Mauchlytól származó keltezetlen jelentés, amely egy késleltető művonalnak nevezett, a gép tárcapacitásának bővítésére szolgáló eszközt ír le. (...) A leírt eszköz (...) kulcsfontosságúnak bizonyult a fejlődés következő fázisa során.

Míndez arra utal, hogy a Moore Intézetben folyó elméleti gondolkodás 1944 augusztusának végére már meglehetősen előrehaladott állapotban volt. (...) Ekkor hozakodott elő Eckert azal a gondolattal, hogy a késleltető művonalat fel lehetne használni információk tárolására. (...) Különös szerencse, hogy éppen, amikor az ötlet fölmerült, lépett színre Neumann János. (...)

John Grist Brainerd, a Moore Intézet igazgatója így ír:

Lehetetlen az ENIAC tárcapacitását olyan mértékben növelni, ... ami a gyakorlati célú nemlineáris parciális differenciálegyenletek kezeléséhez szükséges volna. (...) Jelenleg két olyan elméleti elképzelést ismerünk, amelyet kiindulópontként lehetne kezelni. Az egyik ikonoszóp csövek esetleges alkalmazása, amelyre nézve dr. Neumann megbeszéléseket folytatott dr. Zworykinnal az RCA Kutató Laboratóriumából, a másik pedig a késleltető művonal felhasználása információk tárolására, amellyel kapcsolatban már vannak tapasztalataim.

Ismét Goldstine:

- *Az új gép logikai tervén való munkálkodás pontosan Neumann ízlése szerint való volt. E tevékenység során a formális logika terén folytatott korábbi munkássága döntő szerephez jutott. Mielőtt Neumann színre lépett volna, a Moore Intézetben dolgozó csoport elsősorban a valóban súlyos technológiai problémákra összpontosította a figyelmét, ő viszont a megérkezése után átvette a vezető szerepet a logikai problémák terén. Megkezdődött ily módon a csoport kettéoszlása technológusokra – Eckert és Mauchly – és logikával foglalkozókra – Neumann, Burks és jómagam. Ez természetes munkamegosztás volt, az elmentések azonban az idők múltával egyre növekedtek, és végül a csoport kettészakadásához vezettek.*
- *Így biztosra vehető, hogy a modern számoló és főleg vezérlő, elektronikus áramköröknek a leírását először Neumann János tette közzé a szóban forgó, EDVAC-dolgozatban, a First Draft írásába – megkockáztatható – valószínűleg azért sem vonta be-*

le a társakat: Mauchly-t, Eckert-et, Burks-öt, de még Goldstine-t sem, mert attól félhetett, hogy Mauchly és Eckert az ENIAC irányába vinnék el az EDVAC terveit (ez persze csak spekuláció), még az is feltételezhető, hogy a logikusan felépített gondolatait megzavarták volna a többiek részéről – írás közben – elhangzó hozzászólások.

- Neumann az egyetlen logikus és lehetséges megoldást választotta, az EDVAC terveit, a „First Draft...”-ot miután megbeszélte a munkatársakkal – Los Alamos-ban, az íróasztala mellett, a saját logikájának megfelelően írta meg. Egy ilyen tanulmányt közösen nem is lehetett volna elkészíteni. Azzal az utólagos véleménnyel nem nagyon lehet vitázni, hogy a dolgozatot Neumann – feltételezhetően – ideiglenes leírásnak szánta, a címében foglalt „Draft” minősítés – talán – éppen ezt mutatja.
- Meg kell jegyezni, hogy az EDVAC-ot megelőző számolóberendezések mindegyike az adatokat sorosan tárolta és a műveleteket sorosan hajtotta végre. Egy korabeli leírásban olvastuk, ennek az volt az oka, hogy a számológépek konstruktőrei a drága aktív elemek nagy részét (előbb jelfogók és kapcsológépek, majd pedig elektroncsövek, sőt tranzisztorok is) „kispórolták” a számológépekből. Úgy okoskodtak, ha egy bináris számológép harmincbites számokkal számol, akkor egy 30 bites műveletvégző áramkört kell megépíteni, amiben ezeket a hosszú számokat egy időben lehet összeadni. A harminc bites műveletvégző áramköröket – a korabeli konstruktőrök szerint – csak sok alkatrészrel lehetett volna megépíteni, így pl. a sok elektroncső miatt romlott volna az áramkörök biztonsága (sok alkatrész, sok hibaforrás), ezen kívül az áramkör sokba is került volna, nem beszélve arról, hogy hosszabban és bonyolultabban lehetett volna az áramköröket kialakítani. Ezért feltehető – amikor Neumann János az EDVAC-leírást készítette – a korabeli elvek szerint, már csak szokásból is, soros számítógép megépítésére gondolhatott.
- Neumann azért is a soros számítógép mellett tette le a voksot, mert akkor fedezték fel és vették alkalmazásba a soros számítógépeknek a már említett egyik legfontosabb elemét, a késleltető művonalat, amivel a legkönnyebben soros számítógépeket lehetett építeni. Ennek a fontos eszköznek a leírásához ismét Hermann Goldstine dolgozatát vesszük segítségül:
- *...a bennünket érdeklő típus az ún. ultrahangos késleltető művonal. (...) Ezek az ultrahangos eszközök úgy működnek, hogy a késleltetni kívánt elektromos jelet ultrahangjellé alakítják át; ezt valamilyen folyadékon keresztül vezetik, majd ismét visszaalakítják elektromos jellé. A késés abból a tényből fakad, hogy a jelek a folyadékon sokkal lassabban haladnak át, mint az elektromosság a vezetéken. Higanyban például egy ilyen jel sebessége 1450 m/sec, míg az elektromos jel sebessége a vezetéken a fény $3 \cdot 10^8$ m/sec sebességéhez áll közel. Ha tehát megfelelően választjuk meg a folyadékot tartalmazó edény hosszát, akkor előre meghatározott nagyságú késleltetés érhető el.*

A késleltető művonal, pl. egy higannyal töltött cső, aminek a két végét egy-egy piezoelektromos kvarckristállyal zárták le. A kristály, ha elektromos feszültséget kap, megváltoztatja az alakját, illetve – mechanikus hatásra – a kristályban elektromos feszültség ébred. Így módon a kristályra vezetett impulzusok hatására a kristály rezegni kezd, ez a folyadékban hanghullámokat kelt, amiket a kimeneti kristály – akár másodpercenként több milliószor – ismét elektromos impulzusokká alakít vissza. Egy 1,45 méter hosszú, higannyal töltött csőben 1 milliszekundumos késleltetést lehetett elérni.



9. ábra. A soros számítógépek memóriája, a késleltető művonal. A képen Isaac Auerbach kezében látható a UNIVAC-ban használt (1946/47) higanyos művonal.

rolási kapacitása 20 szó volt, egy-egy bináris számjegy tárolására egy kettős triódával (két félelektroncsővel) épített flip-flopot alkalmaztak. Az EDVAC esetében egy késleltető művonal 1000 bitet tudott tárolni, a művonalon cirkuláltatott bitek energia-vesztésének a pótlására elég volt tíznél kevesebb elektroncső, így a művonalas tárolóval a memóriához szükséges elektroncsövek számát – az ENIAC flip-flopos tárolójához viszonyítva – körülbelül a századrészére lehetett csökkenteni. Ezért azután Neumann a tervezésnél az ENIAC 20 szós tárkapacitásával szemben az EDVAC tárkapacitását minimálisan 2000 szóra választotta.

Egy érdekesség, amit éppen itt érdemes megjegyezni. A számítógépek építésénél a legkényesebb elem – az ENIAC-nál és az EDVAC-nál, de még a mai modern PC-nél is – a memória volt. A késleltető művonalak után megjelentek a tárolócsövek (Neumann „igazi” számítógépében, az IAS gépben is ezzel épült a memória), majd a ferritmemóriák, az utóbbiaknak az volt a legnagyobb baja, hogy a memórialapokat kézzel kellett fűzni, a technológia ezért nem volt automatizálható és miniatürizálható sem.

1969-ben az Intel alapítói – Gordon Moore és Bob Noyce – bejelentették az első félvezető alapú, integrált áramköri memóriát, a 3101-es, 64 bites RAM-ot, amiben 1 bit tárolása körülbelül 100-szor többbe került, mint a korabeli, széles körben használt ferritmemóriában. Az Intel piacpolitikáját a szakma értetlenül nézte, hogyan lehet egy ennyire drága termékkel egy kiforrott memóriatechnikával szemben versenyezni? Ez az Intel memória egyébként a valamikori ENIAC memóriára hasonlított, ugyanis egy-egy bitet egy-egy triódapár helyett egy-egy tranzistorpárból kialakított flip-flop tárolta, amit a szilícium egykristályban tudtak „kifaragni”. Az új és nagyon drága terméket az Intel mégis a piacon tartotta, sőt vásárolták is, ha nem is nagy tömegben, mert a ferritmemóriával szemben a félvezető RAM picit volt, kevés energiát fogyasztott és benne volt az automatizált tömeggyártás lehetősége. Nem kellett hosszú ideig várni, az Intel megalkotta az integrált áramköri

Ha a kimenő és a bemenő oldalt összekötjük, akkor a csőbe vezetett hangminta periodikusan ismétlődik, amíg a veszteségek miatt a jel a csőből el nem tűnik. Ha az energiavesztés pótlására a visszavezetett jeleket felerősítik, akkor máris készen áll a késleltető művonal mint tároló, amiben a bevezetett információt – amíg az erősítés megvan – tárolni lehet.

Ha a jel, amit a bemenő oldalra vezettek, 0,5 mikroszekundum hosszú volt, akkor egy ilyen csőben akár ezer bináris számjegyet is tárolni lehetett.

- A késleltető művonalat az ENIAC-ban alkalmazott tárolóval összehasonlítva azonnal megérthető, hogy Neumann miért fogadta el azonnal az új találmányt, amit Eckert mutatott be az EDVAC tervezőinek, többek között Neumann-nak is. Az ENIAC tá-

tömeggyártási technológiát és a picinyke Intel memóriachip, a ferritmemóriát néhány perc alatt a padlóra küldte. Egy iparág eltűnt, amit felváltott egy másik – az integrált áramkörti technika – ami a megjelenése óta uralja a piacot.

- Az EDVAC egyik legnagyobb különbsége – az ENIAC-hoz viszonyítva – az volt, hogy kettes számrendszerben számolt, így Neumann eleve elhagyta a gyűrűs számlálót, azaz az elektronikus fogaskereket. Ezzel a változtatással az egész gépnek az aritmetikai egysége, amit a tanulmány CA-nak nevez, megváltozott. A korabeli irodalomból kitűnik, hogy Neumann valószínűleg nem ismerte John Vincent Atanasoff munkáját, csak később találkoztak, a találkozással kapcsolatban Atanasoff azt mondta: *Neumann volt az első szakember, akivel a számítógépekről értelmesen lehetett beszélgetni, és akitől sokat tanultam.* Atanasoff 5-6 évvel az EDVAC előtt leírta, sőt John Mauchly-nak el is mondta, hogy elektronikus számolóeszközökbe, amik kétállapotú elemekből – akkor elektronsövekből – épülnek fel kizárólag a kettes számrendszerű aritmetikai műveleteket lehet beépíteni.
- Neumann úgy gondolta, hogy az EDVAC-ban elég 8 decimális számjegyből (27 bitből) álló számokkal műveleteket végezni, *ez elegendő* – mondta – *számos probléma megoldására.* Neumann tehát a megoldandó feladatokból vezette vissza a megoldásra kerülő feladatban az alapadatok nagyságát. Ennek ellenére – valószínűleg a konstruktőrök szerették a kerek számokat – több korabeli számítógép is 30 bites (+előjel) szavakkal számolt. 30 bites szavak esetében a bináris eredményt előbb nyolcas számrendszerre fordították le, majd utána a nyolcas számrendszerű eredményből számították ki – tízes rendszerben – a végeredményt. Később ez a módszer lényegesen egyszerűsödött.
- Neumann a Jelentésben (7–8. fejezet) részletesen sorba veszi a műveletvégző egységeket, az összeadó (kivonó) egységet, az előjelekkel való műveleteket, a bináris (tizedes) pont meghatározását, a szorzást (külön foglalkozik a szorzás pontosságával, hiszen, ha két 30 bites számot összeszorunk, akkor egy 60 bites számot kapunk eredményül) és az osztást. Egyébként voltak olyan korabeli számítógépek, amelyek a szorzás, különben elvesző részét, egy külön regiszterben tárolták, ezt a regisztert az osztás pontosságának növelésére is fel tudták használni. Neumann – hasonlóképpen a szorzáshoz – az osztás elvégzésének is megadta a pontos leírását.
- A Jelentésben külön fejezet foglalkozik a bináris (a köznyelvben tizedes) ponttal. Ez azért is lényeges volt, mert a gép pozitív és negatív, egynél kisebb abszolút értékű, fixpontos számokkal számolt, amiket azután számos első generációs számítógép – éppen a „First Draft...”-ból átvett. Az első hazai számítógép, az M-3 is 30+1 bites szavakkal dolgozott, fixpontos gép volt, az adattartománya pedig a fenti tartományba esett. Ez a korlát részben megkönnyítette, részben megnehezítette a bináris vessző, illetve az eredmény konverziója után a tizedes vessző helyének a meghatározását (9.0 fejezet).
- Neumann az egyéb aritmetikai műveletekkel is foglalkozott, mint például a négyzetgyökvonással, amit a legtöbb korai számítógépben egy szubrutinnal intéztek el, az utasításrendszer négyzetgyökvonást nem tartalmazott. Neumann erre is ad megoldást. Egyébként felteszi a kérdést, hogy milyen műveletek szükségesek a számítógépben? Az alpműveletek ugyanis az összeadás és a kivonás, a többi művelet – a szorzás, az osztás és a négyzetgyökvonás – már visszavezethetők az első két műveletre. De

– mondja Neumann – miután a műveletek gátja a viszonylag kicsi memória volt, más útja is van a magasabb műveletek elvégzésének, például a beépített logaritmusos és aritmetikai – mondjuk – szorzótábla. Számos korai számítógéppel lehet találkozni – ilyen volt például Kozma László egyik antwerpeni gépe is, amelyikbe egyszerezgábla volt beépítve, egy másikba pedig egy Wheatstone-hidas (analog) áramkör, amivel – jó pontossággal – meg lehetett becsülni a magasabb műveletek végeredményét (10.0 fejezet).

- A Vázlat – már volt szó róla – nagyon sokat foglalkozik a memóriával, hiszen Neumann az EDVAC számítógéppel az ENIAC-nak több nagy problémáját is meg akarta oldani, ezek közül az egyik a memória elégtelensége volt. A másik a programozás lassúsága, ami ugyancsak a memória problémájára volt visszavezethető.

Az EDVAC leírás kétféle memóriát említ, a már ismertetett késleltető művonalat, illetve az *ikonoszkópot* (*speciális, impulzusokat tároló elektroncső*), ami a későbbi Neumann–Goldstine IAS-fejlesztésekben főszerepet kapott. A késleltető művonal ugyanis meglehetősen bonyolult módon, főleg soros számítógépekben volt használható, míg az ikonoszkóp inkább – az akkor még nem létező, sőt elvetett – párhuzamosan működő számítógépekben, ehhez az architektúrához a korabeli memóriák közül éppen az ikonoszkóp-memória volt az ideális. Az nyilván a technikatörténetbe való belemagyarázás, hogy Neumann még az EDVAC-riportot írta, de már az IAS számítógépre gondolt, inkább az a valószínű, hogy a tervezéshez szükséges információk összeszedése alkalmával felmerült az ikonoszkóp-memória, amit az EDVAC nem használt. Abban a pillanatban azonban, amikor az IAS gép tervezése elkezdődött, Neumann-nak az ikonoszkóp az eszébe jutott, és a sokkal modernebb – párhuzamos működésű – IAS gépet ezzel a memóriával tervezte és építették meg.

- A „First Draft...” tanulmánynak talán a legnagyobb értéke, hogy Neumann ebben írja le pontosan a később „*Neumann elv*”-nek nevezett *tárolt program* elvet. A „Draft” szinte alig észrevehetően fogalmaz (14.0 fejezet): „*Az utasítások az M-ből* (memória) *kerülnek a CC-be* (central control, központi vezérlőegység), *ugyanonnan, ahol a numerikus adatokat is tárolják.*” Egyébként csak ennyi és nem több a tárolt program elve.
- Kiegészítésként azt kell még tudni, hogy minden Neumann előtti amerikai gép külön tárolta az adatokat, láttuk az ENIAC-nál flip-flopokban, a jelfogós gépekben jelfogókkal, amikből olyan sebességgel lehetett az adatokat megkapni, mint amilyen sebességgel a műveleteket végrehajtották. Az utasítások tára minden gépben – még Kozma László MESz-1-ében is – külön volt. Az ENIAC-nál dugaszoló táblán, más gépekben lyukkártyán, lyukszalagon tárolták az utasításokat, azaz a programot, mert az utasításokat általában lassabban hajtották végre, mint a műveleteket.

Neumann János, amikor az ENIAC nagyon nehézkes programozásán töprengett, jött rá arra, hogy a két tárat, a programtárat és az adattárat egyesíteni kell, annál is inkább, mivel az utasításokat is számokkal kódolták, tehát egy utasítás megjelenésében pontosan olyan volt, mint egy adat és fordítva.

Hermann Goldstine egy hosszú beszélgetés alkalmával elmondta, hogy az elv kitalálása után – amire egyértelműen azt mondta: *Neumann János fejéből pattant ki az isteni szikra* – egymást túllícitálva kezdtek el hangosan gondolkodni. A következőket sorolták fel:

- Ha a közös tárban együtt van az adat és a program, akkor az adatokat is lehet programnak nézni, és fordítva.

- Ha a programot lehet adatnak nézni, akkor azon műveleteket is lehet végrehajtani.
- Ha a programon műveleteket lehet végrehajtani, akkor egy programmal egy másik programból egy harmadikat lehet létrehozni.
- Ha ez így van, akkor egy számítógép a saját programját módosíthatja – ezt már nem mondták ki: a számítógép önmagát is fejlesztheti (evolúció).
- A tárolt program elve mind az EDVAC-nak, mind az utána következő számítógépeknek – azóta is – a „kőbe vésett” törvénye, amin a konstruktőrök – bár számtalanszor megkísérelték – eddig sohasem tudtak túllépni.
- Neumann – ne tévesszük össze a két elnevezést, ezt az üzemmódot ugyanis néha *sorosnak* mondják, de nem az – a tanulmányban állást foglalt az utasításoknak az egymás utáni (soros vagy inkább sorban való) végrehajtása mellett. Az ENIAC ebben is más volt, ott megengedett volt a több utasításnak az időben egyidejű végrehajtása is.
- Még valamit az ENIAC-ról, ismét egy Hermann Goldstine-idézet:

...Neumann vitába szállt velem: vajon át lehetne-e alakítani az ENIAC-ot egy tárolt programú géppé. Azt javasolta, hogy az ENIAC két függvénytáblája közül az egyiket arra használjuk, hogy ebben tároljuk a program leírásához szükséges utasításokat, és az ENIAC-ba egyszer s mindenkorra huzalozzuk bele ezeket az utasításokat. Még néhány más trükk segítségével fel lehetett használni az ENIAC különböző egységeit az utasítások végrehajtásának a vezérlésére.

Az ENIAC átalakításához több hónapig tartó fárasztó programozási munkára volt szükség. Minthogy Neumann nagyon türelmetlen volt, hogy mikor lehet az ENIAC-ot az új módon használni Los Alamosban, ezért 1947. június 7-én kinevezte a feleségemet a Laboratórium konzultánsává. Ebben az időben Neumann már így írt: „...nagyon elkötelezett vagyok Adelenek a leveleiért. Nick és én már az ő új kódjával dolgozunk, és eddig kiválóan találtuk. (...)” Adele Goldstine, az általa kidolgozott rendszert továbbadta Richard Clippingernek, aki akkoriban a Ballisztikai Kutató Laboratórium Számítástechnikai Laboratóriumának a vezetője volt. (...) 1948. szeptember 16-án az új rendszer lefutott az ENIAC-on.” Ez a Goldstine megjegyzés annyit jelent, hogy az ENIAC is, 1947-ben – Neumann és nem Mauchly, valamint Eckert – javaslatára átalakult kvázi tárolt programú géppé, amit – a gép életének második felében – már így használtak.

- Még egy érdekes adalék a tárolt program elvének a kitalálásához.

Egy 1994-es látogatás alkalmával Konrad Zuse egy levelet küldött *Hermann Goldstine*-nek, akivel *sohasem volt alkalma találkozni*. A nagyon barátságos hangú levélben Zuse arra kérte Goldstine-t, hogy a publikációiban említse már meg: a tárolt program elvét nem Neumann János, hanem ő maga, Zuse találta ki. 1935 óta, amióta elhatározta, hogy számítógép-fejlesztő lesz, nem is épített más gépet, csak olyanokat, amikben a program és az adatár egybe volt építve. A választ Zuse nem tudta megvárni, mert – sajnos – hamarosan eltávozott az élők sorából.

Valóban Konrad Zuse már nagyon korán alkalmazta a tárolt program elvét, anélkül, hogy azt nagydobra verte volna. Nem is gondolt rá – mondta –, hogy az egy ilyen jelentős elvnek számít. A háború következtében ugyanis nem nagyon hallott az amerikai számítógép-fejlesztésekről. Konrad Zuse egyik tragédiája éppen a háború és a náci uralma volt, akik a német tudomány Németországban élő nagy alakjait elzárták a tudományos világtól. Zuse másik tragédiája az volt, hogy mint németet – noha nem volt náci – a háború után se a keleti, se a nyugati országok tudományos emberei nem nagyon fogadták be köreikbe.

Éppen ezért a tárolt program elvét helyes volna a továbbiakban Neumann–Zuse elvnek nevezni.

Goldstine az EDVAC megoldásainak a szerzőségéről összefoglalóan így ír:

„Vannak bizonyos részek, amelyek szerzősége nyilvánvalóan köthető egyik vagy másik emberhez... arról, hogy egy probléma megoldására föl lehetne használni az akusztikus tartályt, Pres Eckerttől hallottunk először. Más elgondolások esetében a helyzet bonyolultabb volt. Olyan bonyolult, hogy az, aki a gondolatot elsőként felvetette, saját magát beszélte le róla, kétszer vagy akár háromszor is megváltoztatva a véleményét. Sokszor nem az tette a javaslatot, akiben a gondolat elsőként felmerült. Az ilyen esetekben gyakorlatilag lehetetlen a szerzőséget megállapítani.”

- Egy másik idézet Paul Armertől, a számítógép-tudomány egyik vezető alakjától, amit a Datamation 1962. 8. számában tett közzé, amikor az alábbi „Burks–Goldstine–von Neumann: Preliminary Discussions of the Logical Design of an Electronic Computing Instrument” (1946. június 28.) című – az IAS számítógépről szóló – dolgozat megjelent. Ehhez írt Armer bevezetőt:

Ki találta fel tulajdonképpen a tárolt program elvét? Ez talán nem is különösebben érdekes másnak, csak a munkálatokban közvetlenül részt vevőknek, akik számára fontos lehet, hogy kit illet az elismerés. Az elv megszületett, és a későbbi generációk biztosan minden bizonnyal az emberi haladás egyik mérföldkövének fogják tekinteni... A legtöbb joggal az itt idézett dolgozat szerzői, valamint a Pennsylvaniai Egyetemen a J. Presper és John Mauchly vezetésével működött munkacsoport tagjai követelik maguknak az elsőséget. Nem kétséges, hogy a találmányban másoknak is van részük, köztük nem utolsósorban Babbage-nek...

Az itt idézett dolgozat mindazonáltal a számítástechnika meghatározó jelentőségű dokumentuma. Nem csak a tárolt programú számítógépek tervezésének elveit adja meg, de számos rendkívül bonyolult problémát is előre lát, és zseniális javaslatokat tesz megoldásukra. A dolgozatban leírt gépet (amely IAS gép, Princetoni gép és Neumann gépneveken közismert) meg is építették, azután lemásolták (bár sohasem pontosan), majd a másolatokról is másolatokat készítettek...

E dolgozat megszületésekor az automatikus számolás elve szilárd alapokon állt (már a Harvardon megépített MARK I óta), és annak a hatalmas lépésnek a megtételére is sor került (az ENIAC megalkotásával), amelyet az elektronika alkalmazása jelentett. Annak az ugrásnak a horderejét, amelyet a tudományágnak akkori helyzetében a dolgozat részletei jelentettek, ma már nehéz tárgyilagosan felmérni...

- Neumann a „First Draft...”-ban szinte egyetlen szót sem szentel a programozásnak, sokkal inkább az EDVAC-nak – elsősorban – a logikai felépítésével foglalkozik. 1945-ben Neumann írt egy rendezőprogramot, de még ekkor sem foglalkoztak a programozásnak és a kódolásnak az alapos elemzésével. Még talán azt a megjegyzést is meg lehet kockáztatni, hogy a programozás – az összámtógépeknél – még nem volt tudomány, csak a hardver fejlesztése számított annak. Mára ez a vélemény alaposan megfordult.
- A programozás tudományának az első alapos elemzésére a „Preliminary Discussion...”-ban, az IAS géppel kapcsolatban került először sor. Ennek első „termése” a *folyamatábrászerű programábrázolási mód*, azt is szokták mondani, hogy magas szintű programozási nyelv volt, amit ugyanebben az évben Goldstine – Neumann közreműködésével – dolgozott ki.

Ettől kezdve Goldstine és Neumann számos, a programozással foglalkozó tanulmányt írt, ezek közül az egyik legfontosabb a „*Goldstine–von Neumann: Planning and Coding Problems for an Electronic Computing Instrument*” volt. Erről röviden a későbbiekben szólnunk.

A First Draft részletes ismertetése

Írásunk ezen része azoknak az olvasóknak készült, akik valamivel részletesebben is meg szeretnének ismerkedni a dolgozattal. A Jelentés 15 fejezetből áll. A fejezetek számozását, és a fejezeteken belüli pontok jelölését is az eredeti szerint pontosan követjük.

A Jelentés első hat fejezetét részletesen, szinte szó szerinti fordításban ismertetjük. Ezt azért tesszük, mert ebben a részben Neumann János voltaképpen az alapelveket írja le, egy „előzetes összefoglalót” ad.

A 7–15. fejezeteket kevésbé részletesen ismertetjük, ennek az oka, hogy ezekben a fejezetekben Neumann János „áramkörü mélységben” írja le javaslatait, írásunk előírt terjedelme viszont nem teszi lehetővé, hogy ezt mi is megtegyük.

Megjegyzéseinket dőlt betűvel szedjük.

1.0 Definíciók

- 1.1. A Jelentés: nagyon nagy sebességű, automatikus, digitális számítási rendszerekkel (computing system) és ezek logikai vezérlésével foglalkozik.
- 1.2. Egy automatikus számítási rendszer olyan berendezés, amely utasítások végrehajtására képes abból a célból, hogy számításokat végezzen jelentős bonyolultságú problémák megoldására, például nemlineáris parciális differenciálegyenletek numerikus megoldása céljából.

Megjegyzés: *ebben a pontban látszik, hogy Neumann a számítógépet teljességében matematikai gépnek tekintette, arra, hogy mire lesznek képesek ezek a gépek a harmadik évezred elejére, talán soha nem is gondolt.*

Az utasításokat részletesen meg kell adni. Az utasításoknak tartalmazniuk kell minden numerikus információt, amely a probléma megoldásához szükséges, mégpedig a kezdő és peremfeltételeket, a fix paraméterek értékeit, sőt azokat a függvénytáblákat is, amelyek a probléma leírásában szerepelnek. Az utasításokat kártyán, lyukszalagon, mágnesezett acélszalagon vagy huzalon, filmen stb. kell megadni. A feladatok megoldására szolgáló eljárások megadására kódokat (utasításokat) kell használni, amelyekkel le lehet írni a megoldandó probléma logikai és algebrai lépéseit. A gépnek adott utasításoknak alkalmasaknak kell lenniük arra, hogy – további emberi beavatkozás nélkül – teljességgel végrehajthatók legyenek. A műveletek végrehajtása után a gépnek fel kell tudnia jegyezni az eredményeket a fent említett formában. Az eredmények numerikus adatok.

Megjegyzés: *Ezekből a mondatokból még jobban látszik, hogy Neumann a számítógépet például differenciálegyenletek megoldására szánta, hiszen neki akkor erre volt szüksége (az atombombával és a ballisztikai rakétákkal kapcsolatban is).*

- 1.3. Nyilvánvaló, hogy a berendezésnek általában jóval több numerikus adatot kell produkálnia, mint amit végeredményként várunk. Ez azt jelenti, hogy a numerikus output csak egy része a ténylegesen képződő számoknak.
- 1.4. Természetesen feltesszük, hogy a gép automatikus működése közben hibátlanul végzi el a számításokat, hibátlanul hajtja végre az utasításokat. Az is nyilvánvaló, hogy egy berendezés hibátlan működése csak bizonyos valószínűséggel várható. Egy komplikált gép esetén és ráadásul egy hosszú műveletsorozat elvégzése közben nem reális feltételezni, hogy a hiba keletkezésének valószínűsége elhanyagolható. Abból a célból, hogy a hibákat felismerjük, és korrekciókat tudjunk végrehajtani, mindenképpen szükséges az emberi beavatkozás.

Bizonyos mértékig ezek a hibás működések elkerülhetők, a gép felismerheti a leggyakrabban előforduló hibákat (automatikusan) megjelölve, kiírva ezek létrejöttét és helyét egy kívülről látható jelsorozat segítségével, és megáll, vagy esetleg maga hajtja végre, mégpedig automatikusan, a korrekciókat és folytatja a működését.

Megjegyzés: Ez a probléma évek múlva is foglalkoztatja Neumann Jánost. Elméletileg vizsgálja azt, hogy lehet-e olyan automatát szerkeszteni, amelyik a saját hibáit „felismeri” és kijavítja. Egy érdekes matematikai tételt is kimond és bizonyít.

2.0 A rendszer főbb alkotóelemei

- 2.1. A gép működésének elemzése után kézenfekvőnek látszik, hogy a gépnek az alábbi fő alkotóelemekből kell állnia.
- 2.2. Mivel a létrehozandó berendezés elsősorban számológép, ezért az alapl műveletek: összeadás (+), kivonás (-), szorzás (\times), osztás (\div) műveletének elvégzésére kell képesnek lennie, és erre speciális alegységekkel kell rendelkeznie.

Felvetődhet, hogy vajon szükséges-e mind a négy alapl művelet, illetve a négy alapl művelethez tartozó berendezés. Az is meggondolandó, hogy az olyan műveleteket, mint a $\sqrt{\quad}$, $\sqrt[3]{\quad}$, sgn , $|\quad|$, vagy \log_{10} , \log_2 , \ln , \sin stb. függvényeket és inverzeiket nem kell-e beépíteni. Lehet, hogy éppen szűkíteni kell a műveletek körét, elhagyva az osztást (\div) és akár a szorzást (\times) is. Felvetődhet, hogy sokkal rugalmasabb szerkezet szükséges, például olyan, amely függvénytáblák létrehozására képes, vagy éppen szukcesszív approximációs módszert valósít meg.

Megjegyzés: A szukcesszív approximációval (a fokozatos közelítés módszerével) sok matematikai feladatot lehet közelítőleg (tehát nem pontosan) megoldani, ezért kézenfekvő volt ennek a módszernek a gépbe való beépítése.

Mindenesetre leszögezhető, hogy a szerkezetnek rendelkeznie kell egy központi aritmetikai egységgel és ez alkotja a számológép első specifikus főelemét, amit CA-val (Central Arithmetic) jelölünk.

- 2.3. Szükség van a számológépben egy logikai vezérlőegységre, amely gondoskodik az egymás után következő műveletek végrehajtásáról. A létrehozandó berendezésnek eléggé rugalmasnak, vagyis (amennyire csak lehetséges) általános rendeltetésűnek kell lennie. Különbséget kell tenni a speciális utasítások (amelyekkel egy adott partikuláris problémát lehet megoldani) és az általános vezérlés között, amely végrehajtja az előbbi műveleteket. A vezérlőegységnek az utóbbi a feladata. A vezérlőegység

tehát a második fontos eleme a számológépnek, jelöljük röviden CC-vel (Central Control, központi vezérlő).

2.4. A harmadik alapegység a memória.

- a) Az osztási, illetve szorzási műveletek végrehajtása közben a részeredmények tárolásához mindenképpen szükség van memóriára. Kisebb mértékben ugyan, de az összeadás és a kivonás is igényel memóriát (az átvitelt több pozíción keresztül kell mozgatnunk), és különösen szükséges a memória a négyzetgyök vagy a köbgyök kiszámításánál.
- b) Az utasítások is, amelyek a feladat megoldásához szükségesek, szintén egy memóriában tárolandók.
- c) Számos (matematikai) probléma megoldásánál bizonyos speciális függvények fontos szerepet játszanak. Ezeket rendszerint táblázat alakjában adjuk meg, máskor pedig analitikus kifejezésekkel. Egyszerűbb és gyorsabb egy fix táblázatból kikeresni az értékeket, mint mindig újból kiszámítani, ahányszor csak szükség van rájuk. Rendszerint elegendő olyan táblázatokat használni, amelyekben 100–200 pontban adjuk meg a függvényértéket, és interpolációt célszerű használni.

Megjegyzés: Az interpolációs módszer lényege: adott pontokban ismert függvényértékekből egy speciális formulával számítjuk ki (közelítőleg) a függvénynek a közbülső pontokban az értékeit.

Lineáris vagy akár kvadratikus interpoláció sok esetben nem elegendő, célszerű inkább köbös vagy negyedrendű interpolációt, akár magasabbat használni. A 2.2. pontban említett függvényeket hasonló módon kezelhetjük, tehát a függvények értékeit interpolációval számítjuk ki. Sőt a reciprokot (tehát az $\frac{1}{x}$ függvényt) is érdemes lehet ily módon kezelni, így az osztás műveletét szorzásra tudjuk visszavezetni.

Megjegyzés: Mai szemmel nézve nagyon érdekes, hogy Neumann János ekkor még úgy gondolta, hogy a függvényeket táblázat alakjában kell bevinni, és nem direkt módon – bármely pontban – közelítő formulákkal kiszámítani. Ma a legegyszerűbb zseb-számológép is az összes elemi függvény (\sqrt{x} , $\sin x$, $\cos x$, e^x stb.) értékét egy adott pontban programmal számítja ki (ráadásul elég nagy pontossággal). Természetesen Neumann János jól ismerte azokat a módszereket (formulákat), amelyekkel az elemi függvényeknél „komplikáltabb” függvények értékei jó közelítéssel kiszámíthatók, de úgy látszik, nem gondolt arra, hogy az aritmetikai egység elég gyorsan boldogul a formulák (képletek) kiszámításával.

- d) Differenciálegyenletek megoldásánál a kezdeti, illetve a peremfeltételek nagy numerikus anyagot jelenthetnek, ezeket is meg kell jegyezni, tárolni kell, tehát ehhez is szükséges a memóriaegység.
- e)–f) Ezekben a pontokban további differenciálegyenleteket elemez abból a szempontból, hogy megoldásuk folyamán milyen adatokat kell a memóriában megjegyezni.
- g) Külön elemzi a szukcesszív approximációknál megjegyzendő adatokat; ezek számára is szükséges a memóriaegység, hiszen az approximációs lépések közbülső eredményeit ideiglenesen ugyan, de meg kell őrizni.

h) Bizonyos statisztikai kísérletekben előforduló rendezési problémákhoz is szükséges memória abból a célból, hogy rögzítsük azon adatokat, amelyekre a rendezést végezzük. Neumann János megemlíti, hogy ezek a nagyon gyors szerkezetek egy érdekes lehetőséget teremtenek majd a statisztikai feladatok megoldásánál.

- 2.5. Összefoglalva a harmadik egységgel kapcsolatos mondanivalónkat: a berendezéshez mindenképpen kívánatos egy jelentős nagyságú memória. A memória különböző részein természetükben különböző műveleteket kell elvégezni, és jelentősen különbözhetnek rendeltetésükben, mégis a teljes memóriát mint egy egységet kell tekintenünk, biztosítani kell a különböző részek közötti felcserélhetőséget.

Megjegyzés: Neumann itt voltaképpen arról ír, hogy a memóriában nem kell külön helyet kijelölni az adatoknak, utasításoknak, az induló adatoknak (pl. kezdőfeltétel) és az eredményeknek (részeredményeknek) stb., ezek egységesen tárolhatók.

A berendezésnek tehát egységes memóriát kell tartalmaznia, ami a harmadik specifikus része a számológépeknek; jelöljük ezt az egységet M-mel.

- 2.6. Ezek a specifikus (vagyis a rendszerre jellemző) egységek CA, CC (együtt C) és M voltaképpen megfeleltethetők az emberi szervezetben létező, az ember idegrendszerét alkotó asszociatív neuronoknak. Meg kell még vizsgálnunk, mi felel meg az érző (szenzor, afferens) és mi a motoros (mozgató, efferens) idegrostoknak. Voltaképpen az input, ill. az output felel meg a neuronrendszerben lévő szenzor-, illetve motoros elemeknek.

A numerikus (és más) információknak a C és M egységek közötti átvitelét olyan mechanizmusoknak kell megoldaniuk, amelyek ezen egységek részét képezik. Szükséges az is, hogy az eredeti, külső információt is be lehessen vinni a számítógépbe, és a végső információkat, tehát az eredményeket ki lehessen vinni a számítógépből a külvilág számára. A külső információknak, vagyis a gépbe beviendő adatoknak a létrehozása direkt módon, emberi tevékenység útján történik, gépeléssel, lyukasztással, fotografikus fényimpulzusokkal, amit valamilyen billentyűkkel hozunk létre, vagy valamilyen fémszalag mágnesezése révén. Ezeket statikusan tároljuk.

A számítógépnek tehát rendelkeznie kell azon képességgel, hogy kezelni tudja a bevitelt és a kivitelt (az inputot és az outputot), mégpedig úgy, hogy bizonyos speciális médiumok olvasására, illetve ezekre való írásra legyen képes. Ezeket a médiumokat külső médiumoknak fogjuk hívni és R-rel jelöljük.

- 2.7. A berendezésnek rendelkeznie kell olyan egységekkel is, amelyek képesek a numerikus vagy más információkat átvinni R-ből C-be és M-be. Ezek az egységek a gép inputját képezik. Ezt tekintjük az számológép negyedik specifikus alegységének, és I-vel jelöljük. Jó megoldásnak az látszik, hogy az információkat R-ből (I-n keresztül) M-be visszük, és sohasem direkt módon a C-be.
- 2.8. A számítógépnek rendelkeznie kell egy olyan egységgel is, amely ki tudja vinni (a lényegében csak numerikus) információt a C-ből és az M-ből R-be. Ezek az egységek a gép outputját jelentik, és ez az ötödik alegysége a gépnek, amit O-val jelölünk. Úgy gondoljuk, hogy itt is az átvitelt M-ből O-n keresztül valósítjuk meg, közvetlenül R-be és soha nem C-ből.
- 2.9. Az output információt, amely R-re megy ki, úgy tekintjük, mint a szerkezet által produkált végeredményt, ami a megoldott probléma eredménye. Ezt jól meg kell különböztetni a közbülső eredményektől, amelyek végig M-ben maradnak (amikről a 2.4. pontban beszéltünk).

Ezen a ponton egy fontos kérdés merül fel. A gépben keletkező eredmények természetü miatt mindenképpen szükséges, hogy a memóriának, az M-nek legyen olyan képessége is, amit az R át tud venni az M funkcióiból, mivel amellet, hogy az R többé-kevésbé emberileg érzékelhető információt tartalmaz, memóriatulajdonsággal is rendelkezik, és ebben a minőségében kiegészítheti M-et.

3.0. A tárgyalás menete

Ebben a fejezetben Neumann János az előbbi definíciókat felhasználva elmondja, hogy milyen módon fejtí ki a továbbiakban az öt alapegységről szóló részletes javaslatát. A fejezetben belül a 3.3.-ban még egyszer visszatér a hibák felfedezésére, lokalizálására és a bizonyos feltételek mellett a lehetséges korrekciókra is; megemlíti, hogy ezt mindenképpen tárgyalni kellene. Azt is mondja, hogy nem tudja ezt a nagyon fontos tárgykört tárgyalni, bár megkísérli, legalábbis érinteni, ahol ez fontosnak látszik majd.

4.0. Elemek, szinkronizáció, neuronanalógia

4.1. Minden digitális számológépezet (gép) diszkrét (egymástól elkülönülő) egyensúlyi állapotokat felvevő, relé típusú elemeket tartalmaz.

Egy ilyen elem két- vagy többállapotú lehet. Ezek lehetnek teljesen egyensúlyi (stabil) állapotok, amelyekben az elem külső hatás nélkül is megmarad az adott állapotban, miközben valamilyen megfelelő külső hatásra egyik egyensúlyi állapotból képes átmenni egy másik (egyensúlyi) állapotba.

Kétállapotú elem esetén egyik állapotban marad, ha nincs külső hatás, míg a másikba kerül, ha adott külső hatás éri.

Az eddigi számolóberendezések különféle mechanikus vagy elektronikus szerkezetű elemeket használtak: fogaskerekeket, amelyek 10 vagy több állapotot reprezentálnak úgy, hogy akár mechanikusan, akár elektromos impulzus útján más fogaskerekeket mozgatnak egy fokkal arrébb (új állapotba), vagy telefonreléket, amelyek elektromágneses hatással nyitnak vagy zárnak egy elektromos áramkört. Lehetséges az első kettő kombinációja is. Végül természetesen szóba jöhetnek elektroncsövek is, amelyek szintén alkalmasak adott állapotok reprezentálására.

Minden ilyen szerkezet autonóm módon ütemezi magát, a benne lévő elemek szukcesszív ütemezése révén. Lehet az ütemezést egy fix órával is végezni, amely előre definiált periódusokban ad ütemjeleket. Az ütemező lehet pl. egy forgó tengely, vagy egy kristályvezérlésű elektromos oszcillátor. Az *elem* fogalmát a fenti értelemben használjuk, és a berendezést *szinkron* vagy *aszinkron* típusúnak hívjuk attól függően, hogy az ütemezés órával vagy autonóm módon történik.

4.2. Megemlítjük, hogy a magasabb rendű élőlények neuronjai az előbbi értelemben elemeknek tekinthetők. „Minden vagy semmi” tulajdonsággal rendelkeznek, azaz két állapotuk van, nyugalmi és ingerületi. A neuronok teljesítik mindazon paramétereket, amit az előbb az úgynevezett elemekről elmondtunk. Egy ingerületi állapotban lévő neuron standard hatásokat fogad be, tengelyfonalakon (axonokon) keresztül. Ilyen vonal azonban két különböző módon kapcsolható a következő neuronhoz. Egyfelől egy ingerületi szinapszissal úgy, hogy a hatás ingerületbe hozza a neuront. Másfelől

egy úgynevezett inhibitor szinapszissal, amelynek olyan tulajdonsága van, hogy teljes mértékben megvédi az ingerülettől, bármelyik más neuronból is jöjjön az. A neuronnak határozott reakcióideje van (a hatás befogadása és a kimeneti hatás közötti idő), ez az úgynevezett szinaptikus késés.

Követve W. S. McCulloch és W. Pitts dolgozatát, amely 1943-ban jelent meg („*A logical calculus of the ideas immanent in nervous activity*”, Bull. Math. Biophysics Vol 5., (1943)), és ha ugyanúgy, mint ők, elhanyagoljuk a neuron működésének bonyolultabb aspektusait, az ingerküszöböt, az időleges összegezést, a gátlást stb., akkor meg lehet mutatni, hogy az egyszerűsített neuron működése teljességgel utánozható, imitálható egy telefonrelével vagy egy elektroncsővel. Az idegrendszer a szinaptikus késleltetést tekintve valószínűleg aszinkron működésű.

- 4.3. Világos, hogy a nagyon gyors számítási eszközökbe vákuumos elektroncsöveket kell beépíteni. Az elektroncsövekkel nagyon jó, mikroszekundum nagyságrendű reakcióidő (szinaptikus késleltetés) érhető el. Ez olyan teljesítmény, amelyet semmilyen más eszközzel nem lehet megközelíteni. Nyilvánvaló, hogy a tisztán mechanikus eszközök ilyen szempontból nem jönnek számításba, és a telefonrelék reakcióideje is 10 millisekundum nagyságrendű, vagy talán ennél is nagyobb. Érdekes megjegyezni, hogy az emberi neuron szinaptikus késleltetése 10^{-3} , azaz 1 millisekundum nagyságrendű.

A további vizsgálatainkban az elektroncsöveket tekintjük a szóban forgó számolóberendezések elemeinek. Megkíséreljük a csövek minden paraméterét elemezni, olyan elektroncsöveket felhasználva, amelyek a kereskedelemben kaphatók. (Ne használjunk nagyon bonyolult elektroncsöveket, és olyanokat sem, amelyek alapvetően új funkciókkal rendelkeznek.) Végül megemlítjük, hogy a szinkronberendezések lényeges előnyökkel járnak.

5.0. Az aritmetikai műveletek fontosabb elvei, tulajdonságai

- 5.1. Vizsgáljuk meg a központi aritmetikai egységet, a CA-t.

Az elektroncső – mint egy áramszelep vagy *kapu* – egy „*minden vagy semmi*” típusú szerkezet (vagy legalábbis közelítőleg az), ugyanis a rácsra bocsátott feszültségtől függően vagy átengedi, vagy nem engedi át az áramot. Két triódából vagy egy pentódából ún. *trigger áramkörök* hozhatók létre, amelyeknek két egyensúlyi állapotuk van.

Mivel a számítógépekben az elektroncsöves áramköröknek kell kezelniük a számjegyekből álló számokat, célszerű olyan aritmetikát használni, amelyben a kezelendő számjegyeknek két értéke van. Ez azt jelenti, hogy a bináris számrendszert kell alkalmazni.

Megjegyzés: Láttuk, hogy a korábbi számológépek egy része tízes számrendszeren alapult; például az elektroncsöves ENIAC is (aminek az alapeleme a „számkeréknek” megfelelő, tízes, „elektronikus gyűrűs számolókerék” volt), míg a másik része kettes számrendszerben működött (például John Atanasoff elektroncsöves, ABC számítógépe). A tízes rendszerű elektronikus számológépek építése nyilván azzal függött össze, hogy a mechanikus számológépekben főleg fogaskerekeket (esetleg fogasléceket) használtak.

Neumann János – ahogyan korábban Atanasoff is – észrevette, hogy az elektroncső nagyon alkalmas két stabil állapot megvalósítására, így szinte teljesen természetes, hogy

a kettes számrendszeren alapuló aritmetikát kell használni. Neumann John Atanasoff gépet és munkáját csak később ismerte meg, igen jó véleménnyel volt róla.

- 5.2. A bináris számrendszer alkalmazása jelentősen egyszerűsítheti a szorzási és osztási műveletet is. A bináris aritmetikának egyszerűbb logikai struktúrája van, mint minden másnak (például a tízes számrendszerbelinek).

Megemlítjük, hogy azon numerikus anyag, amit az ember direkt módon kezelni tud, természetesen tízes számrendszerben van felírva. Ez azt jelenti, hogy az R-ben, tehát az inputban és outputban szereplő számok decimálisak. De mindenképpen előnyös a tisztán bináris aritmetikát alkalmazni a CA-ban, és ugyanilyeneket a központi vezérlőegységben, a CC-ben is. Ebből viszont az következik, hogy az M-nek, tehát a memóriának, bináris számokat kell tartalmaznia.

Az előbbieket azt is jelentik, hogy szükségünk van decimális-bináris, illetve bináris-decimális konverzióra, amelyeknek az I-ben (az input részben), illetve az O-ban (az output részben) kell szerepelniük. A konverziókhoz sok aritmetikai művelet szükséges, ezért ezeket a számításokat a leggazdaságosabb a CA-ban elvégezni, és nyilván szerepet kap az I-vel és az O-val kapcsolatban lévő vezérlőegység, a CC is. A konverzióban szereplő aritmetikai műveleteknek teljesen binárisaknak kell lenniük.

- 5.3. Ezen a ponton egy másik elvi kérdés is felvetődik. Minden olyan létező számológépben, ahol az elem nem elektroncső, az elem reakcióideje elegendően hosszú ahhoz, hogy bizonyos „*teleszkópolást*” lehessen a lépések tekintetében megtenni, mind az összeadási, mind a kivonási, de még inkább a szorzási, illetve osztási műveleteknél.

Elemezzük a dolgot a bináris szorzásnál. Az ésszerű pontosság sok differenciálegyenlet megoldásánál mintegy 8 decimális számjegy, azaz a kerekítési hiba 10^{-8} alatt van. Ez 2^{-27} -nek felel meg bináris számrendszerben, azaz 27 bináris számjeggyel kell ábrázolnunk a számokat. Így a szorzásnál 27 jegyű a szorzandó, 27 jegyű a szorzó, és a szorzásnál a feladat a 0 és az 1 számjegyek szorzását jelenti, majd utána pozicionálásuk és kombinálásuk szükséges. Ez összesen lényegében 27^2 , azaz 729 lépést jelent, a bináris szorzatoknak az összegyűjtése, valamint kombinálásuk az időt megkettőzheti. Így mintegy 1000–1500 (elemi) lépés szükséges két 27 jegyű bináris szám összeszorozásánál.

Nyilvánvaló, hogy decimális számrendszerben jóval kevesebb, 8^2 , azaz 64 lépés szükséges a szorzás elvégzéséhez, ami szintén duplázandó, tehát körülbelül 100 a lépések száma. Azonban ez a kis szám azzal jár, hogy egy szorzótáblát kell alkalmaznunk, vagy más módon komplikálódna a berendezés. Ezen az áron viszont beépíthetünk a gépbe több direkt bináris egységet. A továbbiakban a decimális műveletekkel éppen ezért nem foglalkozunk.

- 5.4. Az 1000–1500 lépésből álló szorzási művelet elvégzése minden nem elektroncsöves berendezésben elfogadhatatlanul lassú. Minden ilyen szerkezetben – kivéve a legújabb speciális reléekkel működőkben – a reakcióidő több mint 10 milliszekundum. Ezekben a gépekben 10–15 másodpercig tartana a két 8 decimális jegyből álló szám szorzása, míg ez 10 másodperc a gyors asztali kalkulátorokban és 6 másodperc a standard IBM-szorzókon.

Hogy elkerüljük ezt a hosszú időt, a műveletek (lépések) *teleszkopizálására* van szükség, azaz annyi műveletet (lépést) kell elvégezni egyidejűleg, vagyis szimultán módon, amennyit csak lehet. Ez nem egyszerű, de például az összeadásnál és a kivo-

násnál a megfelelő (az azonos pozícióban álló) számjegyek lépésenként egyszerre kezelhetők, és az átviteli számjegy is egyszerre kezelhető a soron következő pozícióban lévő két számjeggyel. A szorzásnál a részszorzatok szimultán módon képezhetők és pozícionálhatók, hiszen a bináris számrendszerben egy ilyen részszorzat vagy maga a szorzandó (mert 1-gyel szoroztunk), vagy 0. A részösszegek összeadásánál az említett módszer szerint járunk el. A szorzásnál az összeadást gyorsíthatjuk a részösszegek külön, páronkénti összeadásával.

Természetesen, ha a műveletek időtartamát két összeadás egyszerre történő elvégzésével akarjuk felezni, kétszer annyi összeadó szükséges.

A nem elektroncsövekkel működő gépeknél a fenti módszer alkalmazása teljesen indokolt a műveleti idők csökkentése céljából, hiszen itt ez nagyon fontos; és ebben nagy mérnöki tapasztalat halmozódott fel.

Egy valóban univerzális automatikus számolóberendezés, amely a fenti elvek alapján működik, több mint 10 000 elemből áll.

- 5.5. Úgy tűnik, hogy egy elektroncsöves berendezésnél a fentiekkel ellentétes eljárás sokkal ígéretesebb.

Mint a 4.3. pontban már mondtuk, egy nem túl bonyolult elektroncsőnél a reakcióidő egy mikroszekundumnál rövidebb lehet. Ilyen idő mellett egy nem manipulált szorzás elvégzéséhez szükséges időtartam elfogadható: 1000–1500 lépés, azaz 1–1,5 milliszekundum, ez jóval kevesebb, mint bármilyen nem elektroncsöves szerkezet esetén; így nem biztosítható az embernek az input-output végeken történő szinkron (ugyanolyan ütemű) módon való beavatkozási lehetősége (ütemkülönbség miatt).

(Neumann János itt a részletek tekintetében utal egy {} pontra, amit a First Draftban végül is nem dolgozott ki.)

Ami az összeadást, illetve kivonást illeti, ezek jóval gyorsabban végrehajthatók.

Két 27 bináris számjegyű összeadás esetén legfeljebb kétszer 27, azaz mintegy 30–50 lépés szükséges. Ez 0,03–0,05 milliszekundumot tesz ki.

Az osztási művelet időigénye kb. annyi, mint a szorzásé, és a gyökvonásé sem lényegesen több.

- 5.6. Ezen aritmetikai műveletek gyorsítása nem látszik szükségesnek; legalábbis addig nem, amíg teljességgel meg nem ismerjük ezeket a nagyon nagy sebességű szerkezeteket. Kérdéses, hogy jó megoldás-e a teleszkopikus módszerrel történő műveletgyorsítás az ehhez szükséges elemek számának többszörözése árán. A sokkal komplikáltabb (azaz a több elemből álló) elektroncsöves berendezés sokkal szélesebb tűrést igényelhet az ütemek tekintetében, azaz hosszabb reakcióidő lehet szükséges. A pontos kvantitatív hatásokat nehéz megbecsülni.

Úgy tűnik tehát, helyes az alábbi álláspont: a berendezésnek olyan egyszerűnek kell lennie, amilyen csak lehet, azaz a lehető legkevesebb elemből kell állnia. Ezt úgy tudjuk elérni, hogy nem végzünk el a gépben két műveletet szimultán módon, ha ez jelentősen megnövelné a szükséges elemek számát. A berendezés így megbízhatóbban működik, és az elektroncsövek sokkal gyorsabban (magasabb frekvenciával) hajthatók meg.

- 5.7. Az, hogy a fenti elv alkalmazását milyen határig erőltetjük, függ a mindenkori rendelkezésre álló elektroncsövek fizikai karakterisztikáitól. Előfordulhat, hogy az optimum nem a fenti elv 100%-os alkalmazásánál van, hanem valamilyen kompromisszu-

mos megoldás szükséges. Ez mindig függ majd az elektroncső technika pillanatnyi állásától.

Megjegyezzük, hogy a nagy sebességű digitális számolóberendezésekkel kapcsolatos elképzelés, az erről való gondolkodás mostanáig az előbbiekkal ellenkező irányba mutatott; azaz eddig a szükséges elemek többszöröse árán is a teleszkopikus eljárásokkal való műveletgyorsítást részesített előnyben. Mindenesetre tanulságos volna – annyira, amennyire csak lehet – az ellentétes álláspontot is elemezni.

Megjegyzés: Az 5.3.–5.7. pontokat eléggé részletesen ismertettük. Ennek több oka is van. Az egyik az, hogy ezekből látszik, hogy Neumann János mennyire konkrétan, részletesen elemezte a problémákat. Nagyon sok dolgozatában elvégezte a numerikus számításokat. Az, hogy ez mennyire jellemző Neumann Jánosra, mutatja az alábbi szöveg is. (1947-ben írt dolgozatában szerepel.):

„Megbeszéléseket folytattam az IBM munkatársaival New Yorkban (590 Madison Avenue) a SSEC (Self-Sequencing Electronic Computer = Önműködő Elektronikus Számítógép) ügyében.

Úgy tűnik, hogy ez a gép ebben az esetben néhány hétre bérelhető. Az ára előreláthatólag 300–400 \$ egy ténylegesen felhasznált órára. Az árat illetően az alábbi mondható:

A gép két 14 jegyű decimális számot 20 msec alatt szoroz össze. Ezzel párhuzamosan 20 msec szükséges egy gépi utasítás előhozásához. Úgy gondolom, hogy egy szorzási művelet adminisztrálásához 3-4 utasítás szükséges. Így tehát célszerűnek látszik egy szorzásra 70 msec-ot vagy ellenőrzéssel együtt 140 msec-ot venni. Ez 7 szorzást jelent másodpercenként, azaz 25 000-et óránként. Óránként 350 \$ mellett 1,4 centbe kerül egy szorzás.

Egy ember esetén 10 decimális jegyű szorzás („Friden” vagy „Marchant” asztali számológéppel) 10 másodpercet vesz igénybe. A szorzást kísérő más műveletekre egy 4-szeres, az ellenőrzésre egy 2-szeres és az emberi gyenge hatékonyság miatt egy 2-szeres faktort célszerű venni. Ez 160 másodpercet, vagyis 3 percet tesz ki szorzásonként, azaz 20 szorzás esik egy órára.

Nagyon jó számítócsoportoktól szerzett ismereteim alapján ezek a számok nem pesszimisztikus becslésből adódnak. Ez 800 szorzást jelent egy 40 órás héten. 50 \$-t véve egy számolóember heti béreként és 2-szeres faktort az általános költségek miatt, 12,5 centet kapunk egy szorzásra.

Ez azt jelenti, hogy egy SSEC ezen árak mellett $12,5/1,4 = 9$ -szer olcsóbb, mint a számoló-ember.”

A továbbiakban Neumann János kiszámítja, hogy hány embert pótol a számítógép. Figyelembe véve, hogy a számítógép heti 40 órás működése alatt 30–50 százalékos produktív időt jelent, a számítógép 500 embert pótol.

Ezután („Más számítógépek” címmel) a kifejlesztés alatt álló újabb számítógépek teljesítményét és egyéb paramétereit is elemzi.

Úgy látja, hogy az új gépek sebessége 5–50-szer nagyobb lesz, mint a SSEC-gépé. Ezek naponta 16–24 órát üzemelnek, és 5, esetleg 7 napot egy héten. Produktivitásuk több mint 50%-os lesz. Ezekből a számokból azt következteti ki, hogy az új gépek 10 000–100 000 emberrel lesznek ekvivalensek. Neumann János előrejelzése szerint ilyen gépek 1–3 év múlva jelennek meg, digitálisak lesznek, 2000–4000 elektroncső-

vet tartalmaznak. (A SSEC 12 000 elektroncsövet és 20 000 elektronmechanikus relet tartalmazott, közli Neumann János ugyanitt.)

Elemzi az alkatrészek meghibásodási valószínűségét is. Nyolcóránkénti egy hibás művelet esetén műveletenként 10^{-12} hiba-valószínűség adódik. Megjegyzi, hogy a telefonrelék, 10^{-8} – 10^{-9} hiba-valószínűséggel működnek, az elektroncsövek ennél rosszabbul. Úgy látja, hogy a 10^{-12} -es számot nem könnyű elérni.

A numerikus megoldási módszer kidolgozása és a stabilitás elemzése után Neumann János pontosan meghatározza a műveletszámokat és megvizsgálja, hogy a különböző számú osztópontok esetén mennyi gépidő szükséges a SSEC-gépen.

„Tehát... 170 000 szorzás tartozik egy feladathoz, azaz 210 emberhét vagy 7 óra a SSEC-gépen; 40%-os hatékonyságot feltételezve a gépen ez 18 óra, azaz $2,2 \times 8$ óras nap.”

Vagy másutt:

„Egy „nagy” feladat, amelyben az x osztópontok száma 25, a t osztópontoké 120, azaz $m = 25$, $n = 120$, tehát 3000 csomópont esetén

$$3000 \times 16 \text{ sec} = 48\,000 \text{ sec}, 13,3 \text{ órát igényel}”$$

Megjegyzés. Ezt a megjegyzésünket bátortalanul tesszük meg. Több mint 50 évvel a „First Draft...” megjelenése után, a jelenlegi hardverek ismeretében könnyű csodálkozni azon, hogy Neumann János „azon filozofálgat”, vajon érdemes-e elektroncsöves berendezésekben az elemi műveleteket (tehát a bitekkel végzett műveleteket) párhuzamosan végezni (teleszkopizálni), hiszen az input (output) oldalon álló ember úgyszemint tudja ilyen gyorsan táplálni a gépet. Igaz persze, hogy ha a műveletidők felezése miatt 5 m^3 helyett 10 m^3 -es gépet kellett volna építeni, az jelentős különbség. Ma a chippek világában kétszer annyi chip beépítése lényegében nem jelent problémát. Természetesen enyhén szólva eléggé „el-nagyoltan” elemeztük a dolgot, de a lényeget akartuk kiemelni.

6.0. Az E-elemek

6.1. Az 5.0. pontban megvizsgáltuk azokat az alapelveket, amelyek szerint a CA-t működtetni kell. Vizsgáljuk meg most részletesebben technikai oldalról is.

Nyilvánvaló, hogy elemzésünket az alapelemek működésére kell alapoznunk.

Az ideális eljárás az elemek vizsgálatára: úgy tekinteni azokat, amilyenek valójában ti. elektroncsövek. Mivel nagyon sok alternatív lehetőség van az aritmetikai eljárások, logikai vezérlések stb. elrendezésére, ezért hasonlóan sok lehetőség van a különböző típusú, méretű elektroncsövek és más áramköri elemek használatára (például a gyakorlati teljesítmények alapján). Ez azonban nagyon megnehezíti a dolgunkat.

Éppen ezért vizsgálatunkat egy hipotetikus elemre alapozva végezzük, amely úgy működik, mint egy elektroncső – pl. mint egy trióda a hozzá megfelelően választott RLC áramkörrel – de önálló (izolált) egységként elemezhető anélkül, hogy rádiófrekvenciás elektromágnességi vizsgálatot is el kellene végeznünk.

6.2. Az emberi idegsejt, amit az előbbieken már tárgyaltunk, megfelel erre a célra. A továbbiakban ez képezi vizsgálataink alapját. Pontosán definiálnunk kell azonban ezen elem jellemzőit.

A vizsgálandó elemet E-elemnek fogjuk hívni, és egy O körrel jelöljük. Az O elem ingerületi és inhibitor hatásokat vesz fel, és egy hozzá kapcsolódó tengelyen saját ha-

tásokat bocsát ki; jelöljük ezt a következőképpen: $O-$. A tengely el is ágazhat, ezt így jelöljük: $O-<$, $O-<-$. Az emisszió a tengelyen egy szinaptikus késleltetéssel halad tovább; tegyük fel, hogy ez minden E-elemre ugyanakkora; jelöljük τ -val. A késleltetést a tengelyre rajzolt nyíllal jelöljük, így: $O\rightarrow-$, $O\rightarrow-<$. Ez a nyíl egyúttal a tengely indulópontját és irányát is jelzi.

- 6.3. Ebben a pontban a szerző elemzi az emberi idegrendszerben a szinaptikus késleltetés és a tengelyen a hatás terjedési sebessége közti különbséget, és összeveti ezt az elektroncsöves áramkörökkel. Megállapítja, hogy az emberi idegrendszerben a τ szinaptikus késleltetésnek nagy a szórása; az E-elemknél viszont egységes, azonos τ késleltetés feltételezését javasolja abból a célból, hogy *szinkronizálni* lehessen a berendezés különféle egységeinek működését.

Ebből a célból egy központi órát javasol (legjobbnek egy elektromos oszcillátort gondol), amely minden τ periódusban egy τ' hosszúságú $\left(\frac{1}{5}\tau \leq \tau' \leq \frac{1}{2}\tau\right)$ impulzust bocsát ki.

Megjegyzés. A szerző két helyen is „ígéri” ($\{\}$ hivatkozással), hogy később ezt részletezi, de ezt nem teszi meg.

- 6.4. Mindegyik E-elem egy előtte lévő elemtől vagy egy $-O\rightarrow-$ ingerületi szinapszison $\rightarrow O\rightarrow-$, vagy egy inhibitor szinapszison keresztül veszi fel a hatást. Mint 4.2.-ben mondtuk, 1, 2, ill. 3 küszöbértékű E-elemekkel foglalkozunk, azaz olyanokkal, amelyek 1, 2, ill. 3 szimultán bejövő hatásra kerülnek ingerületi állapotba.

Használni fogunk dupla, azaz 2τ szinaptikus késleltetésű E-elemeket (jelöljük így: $-O\rightarrow\rightarrow-$) és keverteteket is (jelöljük így: $-O\rightarrow-<\rightarrow-$). Így nagyobb flexibilitással csatolhatunk össze egyszerűbb struktúrákat, és ezeket elektroncsöves áramkörökkel meg tudjuk valósítani.

- 6.5. Ha megvizsgálunk néhány tipikus E-elemet, arra a következtetésre juthatunk, hogy a legtöbb ilyen elem 1-2 elektroncsövel megvalósítható.

7.0. A + és \times műveletek áramkörei

Megjegyzés. Ezt és a későbbi fejezeteket is az előbbiekhöz képest lényegesen kisebb részletességgel ismertetjük.

- 7.1. Főbb mondanivalók:

A valós számokat bináris számjegyekkel reprezentáljuk (30 bináris jegy elegendő).

A bináris számjegyeket a CA adott pontján és időpontban valamilyen hatás jelenléte (1) vagy hiánya (0) jelenti.

- 7.2. A különféle E-elemek ábrázolásához különféle *blokkjeleket* használ a szerző ($\supset-$ az input, \rightarrow az output).

- 7.3. A szerző felrajzolja és elemzi az összeadó áramkört.

- 7.4. A szorzási áramkör lényegesen különbözik az összeadásétól. (A szorzáshoz a részlet-eredményeket meg kell jegyezni.)

- 7.5.–7.6. Az E-elemek memóriaként is használhatók.

- 7.7.–7.8. A szerző részletesen leírja a szorzási művelet áramkörét.

8.0. A – és az \div műveletek áramkörei

- 8.1. A kivonás műveletéhez szükség van az előjelre. Legyen ez a bal első bináris jegy (0 a –, 1 a + előjel).
- 8.2. Ebben a pontban a szerző megépíti (felrajzolja) a kivonás műveletének áramkörét.

9.0. A bináris pont

- 9.1. A bináris pont definiálása a szorzás és az osztás műveleténél feltétlenül szükséges.
- 9.2. Mivel két 30 jegyű szám szorzata 60 jegyű lesz, de az eredmény tárolására 30 jegy áll rendelkezésre, ezért 30 jegy elvész. Megoldódik a probléma, ha a bináris pontot az előjel után helyezzük el, vagyis ha a gépben csak -1 és 1 közé eső számok ábrázolását tesszük lehetővé.
- 9.3. A fentiek azt jelentik, hogy a számítási feladatot úgy kell transzformálni (2^{-i} -vel való szorzással), hogy a beírt számok és a műveletek eredményei -1 és $+1$ közé essenek. **Megjegyzés.** *Illusztrációképpen arra, hogy Neumann milyen részletességgel „táralja” mondanivalóját, pontosan lefordítunk néhány mondatot.*
- „a) Az összeadást és a kivonást nem lehet elvégezni, ha az eredmény nem -1 és $+1$ közé esik (hanem -2 és $+2$ közé).
- b) Nem lehet az osztást elvégezni, ha az osztó kisebb (abszolút értékben), mint az osztandó.
- Ha ezeket a szabályokat megszegjük, az összeadó, a kivonó, az osztó egység ugyan eredményt ad, de ez nem lesz az összeg, a különbség vagy a hányados.”
- 9.4. Ebben a pontban a szorzás és az osztás kerekítésével foglalkozik a szerző. (Bevezet egy „kerekítési szelep”-et.)

10.0. A gyökvonás áramköre

Más műveletek

- 10.1.–10.2. A szerző megmutatja, hogy a gyökvonás áramköre nem tér el lényegesen az osztás áramkörétől.
- 10.3. Miután megszerkesztettük az összeadás, kivonás, szorzás, osztás, gyökvonás műveletének áramkörét, vajon hogyan lehet ezeket integrálni a CA-ban?

Mielőtt erre válaszolnánk, vizsgáljuk meg, feltétlenül fontos-e minden fenti műveletet beépíteni a CA-ba.

Felvetődhet, hogy a logaritmustábla alkalmazásával a szorzás, osztás, gyökvonás műveletét vissza lehet vezetni összeadásra és kivonásra, de egy ilyen táblának legalább 10^9 értéket kellene tartalmaznia. A következtetés: a szorzás legyen az alapl műveletek között.

Az \div és a $\sqrt{\quad}$ műveletek közelítő formulák alkalmazásával visszavezethetők az összeadás, kivonás, szorzás műveletére. De ha a szorzás hosszú ideig tart, akkor ezek a műveletek is hosszú ideig tartanak.

Következtetés: a +, –, \times , \div , $\sqrt{\quad}$ műveleteket be kell építeni a CA-ba.

10.4. Milyen más műveleteket célszerű még a $+$, $-$, \times , \div , $\sqrt{\quad}$ műveleteken kívül beépíteni a CA-ba?

A köbgyököt nyilván nem érdemes. Szóba jöhetnek még a logaritmusfüggvény, a trigonometrikus függvény és inverzeik. Ezeket hatványsoraikkal lehetne kiszámítani, de ehhez a gépben meg kellene építeni a megfelelő logikai sémát. Ehelyett célszerűbb a függvénytáblázatok alkalmazása.

11.0. A CA felépítése

A műveletek teljes listája

11.1.–11.2. A CA általában két valós számmal végez műveletet. Az operandusokat az M memóriából kapja, ezek a CA két inputján I_{ca} -n és J_{ca} -n kerülnek be a CA-ba és az eredmény a CA outputján (O_{ca}) keresztül megy át az M memóriába. A memória különböző részei közötti kapcsolat megteremtése felesleges, ez a CA-n keresztül valósulhat meg.

A szerző részletesen leírja a memória és a CA közötti adatforgalmat.

11.3. Ebben a pontban a szerző felveti, hogy szükség van a következő műveletekre is:

- egy szám előjelének kiértékelése,
- két szám közötti reláció megállapítása,
- két lehetséges akció közül az egyik kiválasztására (azaz a számítási folyamatban az elágaztatásra),
- a számoknak kettes számrendszerből 10-esbe és fordítva való konvertálására.

A fentiekre további 5 műveletet javasol. A Neumann-féle gép 10 műveletet tud elvégezni, „utasításrendszere” (Neumann nem így hívja) 10 utasításból áll: $+$, $-$, \times , \div , $\sqrt{\quad}$, i , j , s , bd , db , ahol i a I_{ca} -ból az O_{ca} -ba, j a J_{ca} -ból az O_{ca} -ba való közvetlen átvitel, s a szám előjelének kiértékelését, bd a binárisból decimálisba, db a decimálisból binárisba való konvertálás műveletét jelenti.

12.0. A memória kapacitása. Általános elvek

Megjegyzés. Neumann – az előző fejezetekből is látszik (7.4., 7.5., 7.6., 7.7., 8.3., 10.2., valamint a CA leírásában) – a tanulmányban nagyon sokszor visszatért a memória működésére, ugyanis éppen az EDVAC volt az első amerikai számítógép, amelyik a tárolt programelvének megfelelően egy memóriában tárolta mind az utasításokat, mind pedig a programokat. Ezért nem volt mindegy, hogy Neumann milyen memóriát választ ki, és mekkorára tervezi meg a memória tárolási kapacitását. Az sem volt közömbös, hogy mennyi idő alatt lehet és kell – az adott memóriában – elérni a memóriában tárolt adatokat.

12.1. Ebben a fejezetben írja le Neumann, hogy az EDVAC memóriájának egy soros, vagy késleltető elvű, ciklikus típusú memóriát választott, miután – az ENIAC után ez a memória látszott az EDVAC céljára – soros gépről lévén szó – a legalkalmasabbnak.

12.2. Valószínűleg Neumann előtt még senki sem gondolta végig – rendkívül érdekes módon ezt a „spekulációt” először nem egy villamosmérnök, hanem egy matematikus

tette meg –, hogy milyen körülményeknek kell egy számítóeszköz memóriájának megfelelni, valamint azt sem, hogy az akkor ismert memóriáknak milyen előnyei és hátrányai vannak.

A memória *kapacitásának* az eszköznek azt a tulajdonságát nevezte, hogy hány bináris számot tud tárolni.

Ami az elemi kapacitás egységet illeti, Neumann már a 7.1.-es fejezetben meghatározta, hogy a gépnek előjeles, 30 bites (azaz 31 bites) bináris számokkal kell a műveleteket végeznie, tehát ilyen hosszúságú bináris számot (később az egységet szóznak nevezték, Neumann idejében még nem).

Megjegyzés. Ez a 30-as szinte „kultikus” szám, ami 9 decimális számnak felel meg. Más leírások szerint is, a 30 bites „számolási egység” Neumann-tól ered, ugyanis – már korábban is volt erről szó – Neumann úgy becsülte, hogy körülbelül ez a nagyságrend már alig kezelhető mechanikus számolóeszközökkel, az ilyen pontosságú számításokhoz érdemes elektronikus gépeket használni. (Neumann „elegendőnek látta”, ha egy számítógép 27 bites számokkal végez műveleteket, ezt – valószínűleg az ugyancsak ilyen méretű utasítás mérete miatt, később kerekítette fel 30+1 bitre.)

Neumann óriási tekintélyét mutatja, hogy egy néhány évvel későbbi számítógép-konstruktőr elmondta, a gépébe – szinte gondolkodás nélkül – ugyancsak 30 bináris számjegyből való számokkal tervezte meg a műveleteket, még csak nem is gondolt arra, hogy – például 32 bites számokkal számoljon. Eszébe sem jutott, hogy 32 bites aritmetikai egységet építsen, pedig a 32 bites (2 hatványa) bináris szám később nagy könnyebbségeket jelentett volna, a gépnek – például – a szövegfeldolgozásra való felhasználás esetén.

Neumann-nak, miután az EDVAC-ot tárolt programú gépnek tervezte, számolnia kellett a tárolandó utasítással is, ami nem lehetett hosszabb, mint amilyen hosszúak a tárolt adatok voltak. Ezért az EDVAC programozására háromcímű utasításokból álló utasításrendszert hozott létre, ami később zsákutcának bizonyult. Az EDVAC kísérletek alatt Neumann megvizsgálta a programozók szokásait és úgy találta, hogy a háromcímű utasítások bonyolultsága miatt a programozók – általában – az utasításban csak egy címet használtak, így a háromcímű utasításrendszerben rejlő lehetőségeket a programozók nem voltak képesek kihasználni.

Éppen ezért – az ENIAC és az EDVAC után – Neumann harmadik gépe, az IAS gép – noha ugyancsak 39+1 bites számokkal számolt – már egycímű utasításokkal épült, a 40 bitben azonban két egycímű utasítást tároltak. 10 bit a műveleti kód és 10 bit a cím. Ezzel az IAS gép vezette be először a későbbi modern számítógépekben – széles körben használt – egycímű utasításokat.

12.3. Ezek után Neumann részletesen megvizsgálta – a 2.4.-ben leírtaknak megfelelően – (a)-tól (h)-ig a különböző típusú műveletek memóriai igényét, a megoldás időszükségletét, ebből kívánt ugyanis következtetni az alkalmazott memória valóságos nagyságára és költségére.

Ebben a fejezetben is látható, hogy Neumann bizonyos részleteket csak jelzett {}, később – valószínűleg – vissza akart ezekre a témákra térni, az oda tartozó további fejtegetések részletesebb kidolgozására. A fejezetet olvasva az embernek a szerzővel kapcsolatban az jut az eszébe, hogy nem is egy matematikus, sokkal inkább egy mérnök „spekulál” így – egy-egy feladattal kapcsolatban – a memória időigényére. Ami egyébként igaz,

a legenda szerint Neumann mint vegyész mérnök szokott rá az illetén való mérnöki spekulációra és számításokra. Persze – más fejezeteket (ahol például a formális logikával írja le a különféle műveletek végrehajtását – 7.0., 8.0., 9.0., 10.0.) olvasva, azonnal előbukkan Neumann matematikusi zsenialitása, szemben a kétségkívül erős mérnöki kvalitásaival.

- 12.4. A 12.3. fejezetben felsorolt memóriaigény-számításokat Neumann ebben a fejezetben összegezi, a különféle feladatok megvizsgálása után a memória kapacitását – némi túlbecsléssel – 8000 és 2000 közötti minor ciklusban határozza meg. *Miután 1 minor ciklus kettő az ötödiken egységnyi (unit), így a szükséges memóriakapacitás: 2^{18} , illetve 2^{16} egységre jön ki. Neumann azt is megállapítja, hogy a számítógépekben az „üveg nyaka” (bottleneck) a memória. Ha a mai PC-re és az egyre bonyolultabb operációs rendszerekre, valamint alkalmazói programokra gondolunk, ez ma sincs másként. **Megjegyzés.** Ebben az időben a problémát részben a memória ára okozta, ugyanis a kutatóintézetekben és a gyárakban szinte nem volt memóriaválaszték. Igaz, Atanasoff már évekkel korábban feltalálta a kapacitív dobót, de rajta kívül senki sem használta. A mágneses dob akkor még egyáltalán nem létezett, a Williams-csövet és a ferritmemóriát sem találták még fel. Egyedül a higanyos és a nikkelléltető művonal jöhetett – memóriaelemként – számításba, talán az EDVAC-kal kapcsolatban ezért is gondolkodott Neumann soros számítógép tervezésében.*

Neumann – a számítások után – kimondja, hogy a számítógépben legalább negyedmillió egységet tartalmazó memóriára volna szükség.

- 12.5. Ennek a pontnak a címe: „Hogyan lehet egy negyedmillió egység kapacitású memóriát építeni?”

Megjegyzés. *Neumannból ismét előbukik a mérnök. Nem csak felméri az igényt, de elgondolkozik azon is, hogy egy ilyen memóriát meg is kell építeni.*

Neumann elgondolkozik és számol. Egy késleltető művonalon 30 bites bináris számokat lehetett tárolni, ami annyit jelentett, hogy negyedmillió számhoz 8000 művonalra lett volna szükség, amit – abban az időben – nem tartott megvalósíthatónak, ezért legfeljebb kb. 65 000 tárolt bináris számra gondolhatott, amit már „csak” 2000 művonalat jelentett volna, ugyanis – abban az időben – ez is hatalmas mérnöki munkát jelentett, ami nagyon nehezen lett volna kivitelezhető.

Megjegyzés. *Neumann nem adott megoldást, nem is adhatott, mert nem volt villamosmérnök. A fennmaradt nagyszámú levélből tudható, hogy számos – memória-áramkörökre vonatkozó – kérdést tárgyalt meg Mihály öccsével és még többet Bay Zoltánnal, az utóbbtól leginkább a legújabb elektronikus felfedezésekkel kapcsolatban kért tanácsot: hogyan lehet ezeket az újabb eszközöket memória céljára felhasználni.*

Ebben a fejezetben végül is meghatározza annak a memóriának a paramétereit (kapacitás, ciklusidő stb.), amivel a legtöbb feladatot el lehetett látni, miközben azal nem nagyon törődött, hogy a szükséges memóriát hogyan fogják előállítani. Ez – végül is – a villamosmérnökök gondja maradt.

- 12.6.–12.7. Neumann folytatja a spekulációt, sőt néha bemerészkedik a villamosmérnök konstruktőrök területére is. Tanácsokat ad: hogyan lehet megfelelő mennyiségű elektroncsővel, csökkentett késleltetési idővel, rövidebb szinkronizáló impulzusokkal bizonyos vezérlési, valamint tárolási feladatokat ellátni, és így a kívánt kapacitású memóriát előállítani.

Ezt követően Neumann ismét tovább számol, szinte „hangosan” gondolkodik. Látható módon előtte alaposan tanulmányozta az áramkörök technikáját, az ember arra gondol, hogy Neumann mindezt elsősorban azért tette, mert meg akarta teremteni a számítógép-tervezés módszertanát és nem is azért, hogy egy új elvű, működő számítógépet hozzon létre.

Megjegyzés. *Ismerve a később történeteket, az IAS gép tervezését és építését, Neumann valószínűleg itt gondolta végig, hogy a jövőben nem EDVAC-szerű, soros számítógépet kell tervezni és építeni, mert a soros gép és a szinkronizált soros memória **zsáktutca**, ami sehová sem vezet. Valószínűleg ezért hagyta abba a „First Draft...” kidolgozása után azonnal, mindazoknak az elveknek – néhány kivétellel – az alkalmazását, amiket a „First Draft...” írása során kidolgozott és tért át az akkor még ismeretlen és sehol sem alkalmazott párhuzamos gépi architektúrára. Az újabb neumann-i spekuláció azt eredményezte, hogy az IAS gép végül kevesebb elektroncsővel épült meg (így – elektroncsöves gép révén – az üzemeltetés során kevesebb alkatrész tudott elromlani), mint az EDVAC, és körülbelül 30–50-szer gyorsabban hajtotta végre a műveleteket. Ennek az eredménynek a nyomán a fejlesztők szinte azonnal leálltak a soros számítógépek tervezésével, és a kutatóintézetek, valamint a cégek áttértek a párhuzamos számítógépek építésére. Még Eckert és Mauchly is.*

- 12.8. Neumann leírja, hogy a memóriaszámítások mind a késleltető művonalas memóriára vonatkoztak, nem zárja ki azonban, hogy később ne találnának ki más, alkalmas memóriákat.

Neumann elsőként az *ikonoszkópra* gondolt, aminek az ernyőjén 200×500 , azaz 200 000 pontot lehetett tárolni. Mindehhez egy nagyon bonyolult, speciális elektroncsőre és abban egy elektronsugarra volt szükség, amivel a bináris információt az ernyőre – világító pontok formájában – fel lehetett írni, és onnan ki lehetett olvasni. Ez az „ötlet” azért is érdekes volt, mert addig az ikonoszkópot képfelbontásra és képátvitelre használták, eredetileg nem számítógép-memóriának tervezték. Talán Neumann látta meg benne először, hogy a képpontok bináris adatok tárolására is alkalmasak.

A fejezeten belül, az (a)-tól (d)-ig pontokban, Neumann a memória számos problémájával foglalkozik, amikre megpróbál részben becsléseket, részben pedig megoldásokat adni. Ezek közül néhányat érdemes felsorolni.

Neumann kiszámolta, hogy például 200 000 pont tárolásához az elektronsugarat milyen pontossággal kell vezérelni, hogy a tárolást biztonságosan meg lehessen oldani. Foglalkozott az információ beírásával és kiolvasásával, a memória sebességéből adódó kérdésekkel, a tárolt információ megcímzésével, de még néhány – az ikonoszkópban megoldandó – technikai problémával is, mint például a memória tárolási biztonsága.

Megjegyzés. *Az ikonoszkópot – szemben a késleltető művonalas memóriával – nem igen volt célszerű soros számítógép memóriájaként alkalmazni, ami például abból is látszik, ahogyan Neumann a kiolvasott információ késleltetésével foglalkozott. Erre azért volt szükség, hogy mind a műveletvégző (CA), mind pedig a vezérlő (CC) egység képes legyen a feladatát végrehajtani. Ezekbe az egységekbe ugyanis az információ sorosan, tehát a bitek egymás után léptek be.*

Az ikonoszkóp – technikailag – tipikusan párhuzamos memória volt, amint később Neumann az IAS számítógépben párhuzamos memóriaként használt fel. Ismét érdemes arra gondolni, nagyon valószínűsíthető, hogy Neumann az ikonoszkóp-memória tanulmányozása során már előre, egy párhuzamos számítógépen gondolkodott (az előregondolás egyébként is tulajdonsága volt), tehát az EDVAC-ot tervezte, miközben az IAS gépen is gondolkodott. Ez persze csak spekuláció.

Talán ezért írta Neumann a fejezet végére, hogy ezek után „ésszerűbbnek látszik a késleltető művonalas memória analízisének a folytatása”, de azért nem feledkezett meg az ikonoszkópról.

13.0. A memória (M) szervezése

13.1.–13.5. Ebben a fejezetben a 12.6. és a 12.7. fejezetekben leírtak alapján a memória belső szervezésével foglalkozik, meghatározza azokat a memória működéséhez szükséges logikai elemeket, amiket a rendszer kialakításához meg kell építeni. Műszaki részletekkel nem törődik, inkább az elemeknek, pontosabban a szükségesnek tartott részegységeknek a funkcióit írja le.

Ez a fejezet a következő 14.0. fejezet fordítottjának is felfogható, ugyanis a memória (M) oldaláról mutatja be, hogy a számítógép többi részével (pl. C vagy CC) az M hogyan fog kommunikálni, illetve hogyan kell az M – mai szóval architektúráját, akkor szervezést mondtak – kialakítani, hogy a számítógép más részei által előírt logikai követelményeknek meg tudjon felelni.

A fejezet pontosan leírja nemcsak az egész memória, hanem a részegységek szervezését, valamint lehetséges kapcsolati rendszerét is, amibe – az eredeti leírásban – csak azoknak érdemes mélyen belemenni, akik az EDVAC memóriájának a működését részletesen és mélyen meg akarják ismerni.

Megjegyzés. *A technikatörténettel foglalkozó kutatók a „First Draft...” tanulmányozása közben értik meg igazán, hogy az EDVAC terveit, amit Neumann szinte bit mélységig kidolgozott, miért lehetett a némileg korszerűsített, de az EDVAC szervezésével szinte teljesen azonos EDSAC és BINAC számítógépeknél felhasználni. Valószínű – a leírásban sűrűn előforduló, {} jellel kiemelt részek is ezt bizonyítják – Neumann az EDVAC leírásában csak a szerinte leglényegesebb kérdések megoldására koncentrált, így számos részprobléma, amire a fenti jel utal, megoldatlan marad.*

Talán nem tiszteltelenség Neumann dolgozatáról feltételezni, hogy a „First Draft...”-ban hibák és ellentmondások is előfordulhattak, aminek a valószínűsége – ismerve Neumann precizitását – nagyon kicsi, de nem kizárható. Ezeket a problémákat az EDSAC tervezésénél Wilkes, míg a BINAC tervezésekor Eckert és Mauchly is föltételezhetően felfedezték és kijavították, amire számos példa volt, például később, az IAS „klónok” tervezése során is.

Nagyon valószínű, amit nemcsak az EDSAC, hanem a BINAC példája is bizonyított, hogy a „First Draft...” nyomán minden számítógép-tervezőnek sokkal könnyebb volt a dolga, Neumann ugyanis elvégezte a munka nagyobbik – logikai – részét, mint azoknak, akik korábban – Neumann munkája előtt – igyekeztek számítógépeket alkotni.

14.0. A CC és az M

Megjegyzés. Ebben a fejezetben a Central Control (CC, központi vezérlőegység) és a memória (M) kapcsolatát elemzi Neumann János, pontosabban azt, hogy a memóriában lévő utasítások hogyan „vezérlik” a központi vezérlőegységet.

- 14.1. Ebben a pontban a CC elemzésében mélyedünk el. Ehhez az utasításrendszert kell áttekintenünk, mivel a CC feladata ezen utasítások fogadása, értelmezése és más egységekkel való végrehajtatása.

Feladatunk tehát a CC-t vezérlő utasításlista létrehozása, azaz a berendezésben használt kód leírása, és ezek matematikai és logikai jelentésének definiálása.

Vizsgáljuk meg a CC-t és főleg kapcsolatát az M-mel.

Az utasítások az M-ből kerülnek a CC-be, ugyanonnan, ahol a numerikus adatokat is tároljuk. Az M tartalma ún. minor-ciklusokból áll, minden minor-ciklusnak egy olyan jelet is kell tartalmaznia, amely azt jelöli, hogy számról vagy utasításról van-e szó.

Az utasítások 4 csoportba oszthatók:

- csoport: 10 speciális műveletet hajtanak végre (+, -, ×, ÷, három átviteli művelet, két konverziós művelet (decimálisból binárisba, binárisból decimálisba)).
 - csoport: a számokat a memória egyik helyéből a másikba viszik át.
 - csoport: ide tartoznak azok az utasítások, amelyek a CC-nek az M-mel való kapcsolatát változtatják meg: egyik memóriarekeszről egy másikra adják át a vezérlést, hogy onnan vegye a CC a következő végrehajtható utasítást.
 - csoport: az input és az output berendezések vezérlését végzik.
- 14.2.–14.5. Ezekben a pontokban pontosan megadja Neumann János, hogy az egyes utasítások hogyan kerüljenek át a memóriából a CC-be és hogyan végezzék a vezérlést. A következőket külön kiemeljük:

A memóriában egymás melletti rekeszekben található utasítások végrehajtását egyszerűen lehet vezérelni. Lenniük kell azonban olyan utasításoknak is, amelyek kivételes esetként arra utasítják a CC-t, hogy egy tetszőleges M-beli pontra (rekeszre) tegyék át a CC kapcsolatát.

Megjegyzés. Ezekben a pontokban azt írja le a szerző, hogy a minor-ciklusokkal hogyan valósulhat meg a vezérlésátadás.

15.0. A kód

- 15.1. A memóriának az előbbi fejezetben történő specifikációja lehetővé teszi, hogy definiáljuk a kódokat, amelyek a CC logikai vezérlését (és ezen keresztül az egész berendezését) valósítják meg.

Az M memóriaegységekből áll, amelyeknek mindegyike egy hatás jelenlétével vagy hiányával jellemezhető. Ezek az egységek az 1, illetve 0 bináris számjegyeket reprezentálják. Ezeket az egységeket 32 egységből álló minor-ciklusokba csoportosíthatjuk; ezek a csoportok szerepelnek majd az alábbi bevezetendő kódokban.

Megjegyzés. Egy minor-ciklus voltaképpen egy rekeszt jelent, azaz a memóriában egy olyan területet, amelynek bináris jegyekből álló tartalma egy egységet (számot, utasítást) reprezentál. Neumann János a Jelentésben még nem használja a rekesz elnevezést. Mint

korábban említettük, Neumann János művonalas memóriára alapozta tervét; ebben például 32 jegyből álló bitsorozat valóban „ciklusban” keringett a művonalban, így tárolta a memória, tehát egy minor-ciklus egy rekesznek felelt meg.

Az egy-egy csoportban szereplő bináris számjegyek, egy $i_0, i_1, i_2, \dots, i_{31}$ sorozatot alkotnak, amelyet $I = (i_0, i_1, i_2, \dots, i_{31}) = (i_\nu)$ alakban is írhatunk.

A minor-ciklusokat két osztályba sorolhatjuk: a számok, illetve utasítások osztályába. Ezt a két kategóriát az első bittel, tehát i_0 -val különböztetjük meg; jelölje $i_0 = 0$ azt, hogy számról, $i_0 = 1$ pedig azt, hogy utasításról van szó.

- 15.2. Számoknál a 31. egység a bináris számjegyeket, illetve az előjelet reprezentálja.

Az aritmetikai műveletek természetéből kifolyólag (tekintettel az átvitelre) a bináris számjegyeket jobbról balra töltjük fel, azaz $\xi = i_{31}, i_{30}, i_{29}, \dots, i_1$ alakban állnak elő a számok. Az utolsó betöltött jegy i_{31} az előjel, mégpedig $i_{31} = 0$ a +, $i_{31} = 1$ a – előjelet jelenti. A bináris pontot az előjel után helyezzük el, vagyis az ábrázolandó ξ számot mod(2)-vel a $(-1, 1)$ intervallumba transzformáljuk, azaz:

$$\xi = i_{31}, i_{30}, i_{29}, \dots, i_1 = \sum_{\nu=1}^{31} i_\nu 2^{\nu-31} \pmod{2}, \quad -1 \leq \xi < 1.$$

- 15.3. Az utasítások 31 bite, az utasítás hatását adja meg.

Az alábbi utasításcsoportokat definiáljuk.

α) Ide tartoznak a CC azon utasításai, amelyek hatására a CA elvégzi a 14.1. a) csoportban szereplő 10 utasítást. Jelöljük ezeket sorban a 0, 1, 2, ..., 9 számokkal, illetve az ezeknek megfelelő 4 jegyű bináris számokkal.

Megjegyzés: Ebben a pontban Neumann János részletesen elemzi, hogy a CA különböző egységeiben hogyan kell kezelni az operandusokat a műveletek elvégzése után, mikor kell törölni az adott egységet, mikor nem, illetve hogyan lehet megoldani, hogy az előző művelet eredményét a következő műveletben fel lehessen használni (például az xy szorzat eredményét hogyan lehet hozzáadni az előtte elvégzett művelet z eredményéhez, azaz képezni a $z + xy$ összeget).

Annak jelölésére, hogy egy művelet után az O_{ca} -t (az aritmetikai egység eredményregisztere) törölni kell-e vagy nem, a c -vel jelölt bitet kell használni ($c = 0$ törlést, $c = 1$ a törlés elmaradását jelenti).

β) Ebbe a csoportba azok az utasítások tartoznak, amelyek hatásaként a számok M-ből CA-ba mennek át.

γ) Olyan utasításokra is szükség van, amelyek ún. direkt betöltést végeznek, vagyis az utasításszámláló által megcímezett rekeszt követő rekesz tartalmát töltik az I_{ca} -ba.

δ) Szükségesek olyan vezérlőutasítások is, amelyek hatására a számok CA-ból M-be kerülnek.

ϵ) Az ebbe a csoportba tartozó vezérlőutasítások az utasításszámláló által megcímezett rekeszt követő címre adják át a vezérlést.

θ) Az utasítás hatására a számok CC-ből CA-ba kerülnek át (pontosabban, az O_{ca} -ból az I_{ca} -ba).

ζ) Ezek az utasítások a CC és az M kapcsolatát az M egy másik helyén található minor-ciklusára (rekeszre) irányítják át (vezérlésátadás).

η) Ebbe a csoportba az input, output utasítások tartoznak.

(I) Típus	(II) Jelentés	(III) Rövid jel	(IV) Kód- szimbólum																																				
			Minor-ciklus $I = (i_\nu) =$ $(i_0 i_1 i_2 \dots i_{31})$																																				
Szám vagy utasítás (γ)	$A \bar{\xi} = i_{31} i_{30} \dots i_1 = \sum_{\nu=1}^{31} i_\nu 2^{\nu-31} \pmod{2};$ <p>$-1 \leq \xi < 1$ számot tárolja. Az i_{31} bit az előjel, 0 a + jelet, 1 a – mínuszjelet jelöli. Ha a CC ehhez a minor-ciklushoz fordul, akkor az utasításként működik, amely ξ-t I_{ca}-ba viszi át. Nem ez történik azonban, ha ez a minor-ciklus közvetlenül egy $w \rightarrow A$, vagy $wh \rightarrow A$ után következik.</p>	$N\xi$	$i_0 = 0$																																				
Utasítás (α)+(δ)	<p>Ezek olyan utasítások, amelyek CA-ban hajtódnak végre. w a 11.4.-ben szereplő művelet valamelyike. Az alábbi táblában a w decimális oszlopban a műveletek decimális, a w.bináris oszlopban bináris alakban, a w oszlopban szimbólumokkal szerepelnek</p> <table border="1" data-bbox="294 924 861 1153"> <thead> <tr> <th>w. decimális</th> <th>w. bináris</th> <th>w w</th> <th>w. decimális</th> <th>w. bináris</th> <th>w</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0000</td> <td>+</td> <td>5</td> <td>0101</td> <td>i</td> </tr> <tr> <td>1</td> <td>0001</td> <td>–</td> <td>6</td> <td>0110</td> <td>j</td> </tr> <tr> <td>2</td> <td>0010</td> <td>\times</td> <td>7</td> <td>0111</td> <td>s</td> </tr> <tr> <td>3</td> <td>0011</td> <td>\div</td> <td>8</td> <td>1000</td> <td>db</td> </tr> <tr> <td>4</td> <td>0100</td> <td>$\sqrt{\quad}$</td> <td>9</td> <td>1001</td> <td>bd</td> </tr> </tbody> </table> <p>A III. oszlopban szereplő jelek jelentései: – h azt jelenti, hogy az eredmény megőrződik az O_{ca}-ban – $\rightarrow \mu\rho$: az eredmény a μ major-ciklus ρ minor-ciklusába kerül átvitelre – $\rightarrow f$: közvetlenül az ε utasítás végrehajtása után az eredmény a végrehajtás alatt álló utasítást követő minor-ciklusba kerül – $\rightarrow A$: az eredmény I_{ca}-ba kerül. – ahol nincsen \rightarrow jel, ott nem történik tárolás (az eredmény megőrzésétől eltekintve)</p>	w. decimális	w. bináris	w w	w. decimális	w. bináris	w	0	0000	+	5	0101	i	1	0001	–	6	0110	j	2	0010	\times	7	0111	s	3	0011	\div	8	1000	db	4	0100	$\sqrt{\quad}$	9	1001	bd	$w \rightarrow \mu\rho$ vagy $wh \rightarrow \mu\rho$	$i_0 = 1$
w. decimális		w. bináris	w w	w. decimális	w. bináris	w																																	
0		0000	+	5	0101	i																																	
1		0001	–	6	0110	j																																	
2	0010	\times	7	0111	s																																		
3	0011	\div	8	1000	db																																		
4	0100	$\sqrt{\quad}$	9	1001	bd																																		
Utasítás (α)+(ε)		$w \rightarrow f$																																					
Utasítás (α)+(θ)		vagy $wh \rightarrow f$																																					
Utasítás (α)		$w \rightarrow A$																																					
Utasítás (β)		vagy $wh \rightarrow A$																																					
Utasítás (ζ)		wh																																					
Utasítás (β)	A μ major-ciklus ρ minor-ciklusában lévő szám I_{ca} -ba kerül.	$A \leftarrow \mu\rho$																																					
Utasítás (ζ)	Az utasítás CC-t μ major-ciklus ρ minor-ciklusával köti össze.	$C \leftarrow \mu\rho$																																					

15.4. Ebben a pontban a szerző megvizsgálja, hogy a 31 bitből álló utasításoknak milyen legyen a szerkezete.

Megállapítja, hogy kerülni kell több funkció beépítését az utasításba.

15.5. Vannak azonban olyan teendők, amelyeket célszerű összevonni egy utasításba; pl. az (α) típusú utasításoknál a művelettel együtt meghatározható, hogy a művelet eredménye hova kerüljön. Neumann pontos elemzést végez arra nézve, hogy az egyes utasítások által végzendő teendőket hány számjeggyel (bittel) lehet elérni. Azt hozza ki, hogy 32 bináris jeggyel mintegy 50%-os átlagos „hatékonyság” érhető el.

15.6. Az előbbi pontokban végzett elemzések alapján Neumann az 1. táblázatot állítja elő.

Utószó a First Drafthoz

Ezzel végére is értünk a First Draft ismertetésének. Reméljük, sikerült a fontosabb gondolatokat kiemelni, érzékeltetni a munka „első vázlat” jellegét és különösen Neumann gondolkodásmódját, elemzési módszerét, a problémák „tálalásának” módját.

3. Egy utasításrendszer, egy „magas szintű” nyelv

Az EDVAC gép utasításrendszere

Az előbbieken (a First Draftban) láttuk, hogy Neumann János miképpen képzelte el a gép utasításrendszerét. Az EDVAC gép tervezett utasításrendszere lényegében követte a Neumann által leírt elveket (a sémát).

Az EDVAC gépben a tervezett utasítások szerkezete az alábbi volt:

műveleti kód	cím1	cím2	cím3	vezérlőbitek
5 bit	6 bit	6 bit	6 bit	

A gép tehát 3 című volt.

A gép tervében az alábbi 15 utasítás szerepelt (**a**, **b**, **g** memóriacímek, **a**, **b**, **g** pedig konstans operandusok).

- 1) **add a b g**: Az **a** és **b** tartalmát összeadja, és az eredményt **g**-be helyezi.
- 2) **sub a b g**: A **b** tartalmából kivonja **a** tartalmát, és az eredményt **g**-be helyezi.
- 3) **mul a b g**: Az **a** és **b** tartalmát összeszorozza, és az eredményt **g**-be helyezi.
- 4) **neg a b g**: Ugyanaz, mint a **mul** utasítás, de **g**-be az eredmény -1 -szeresét teszi.
- 5) **c a b g**: Ha az **a**-ban tárolt szám nagyobb, a **b**-ben tároltnál, akkor a következő utasítást a **g** címről veszi.
- 6) **x a b g**: Ha $a > b$, akkor a következő utasítást a **g** címről veszi.
- 7) **t – – g**: A **g** memóriacímen tárolt utasításra adja át a vezérlést.

- 8) **p a b g**: Az **a**-ban tárolt számot **g** pozícióval balra tolja, és az eredményt **b**-be teszi.
- 9) **q a b g**: Az **a**-ban tárolt számot **g** pozícióval jobbra tolja, és az eredményt **b**-be teszi.
- 10) **i a b g**: A soron következő (eggyel nagyobb című) memóriarekeszben lévő **a b g** utasításban az **a, b, g** címekeket az **a, b, g** értékekkel megnöveli.
- 11) **e a b g**: Kiemel bizonyos biteket az **a** című rekesz tartalmából, és ezeket elhelyezi a megfelelő pozíciókban a **b** című rekeszben. A szóban forgó biteket a **g** tartalma határozza meg.
- 12) **fn – b g**: **g** számú szót beolvas **b**-be, és az **n** szalagot tovább mozgatja.
- 13) **bn – b g**: Ugyanaz, mint előbb, de a szalagot visszafelé mozgatja.
- 14) **fn a – g**: Kiír **g** szót az **a** rekeszből a szalagra, és a szalagot előre mozgatja.
- 15) **bn a – g**: Ugyanaz, mint előbb, de a szalagot visszafelé mozgatja.

Ez az utasításrendszer a First Draftban szereplő elvek alapján készült ugyan, de egyes történetírók szerint az EDVAC gép nem pontosan eszerint működött.

Neumann és az első „magas szintű” nyelv

Már az EDVAC tervezése közben felmerült, hogy szükség volna olyan nyelvre, amivel a matematikai eljárásokat a gép számára, a gépi kódnál jobb (kényelmesebb) módon lehet leírni.

Az első ilyen kísérletet nem Amerikában tették meg, hanem Európában. Németországban Konrad Zuse 1945-ben kidolgozott egy Plankalkül nevű nyelvet, ami lényegében a Hilbert-féle ítéletkalkuluson alapult. A kéziratot csak 1972-ben publikálták, bár rövid ismertetőt 1948-ban és 1959-ben Zuse már közölt.

Az óceán túlsó partján is foglalkoztak a gépi kódnál magasabb szintű programozási nyelvek kidolgozásával. Goldstine kezdte el elemezni, hogyan lehetne olyan módon leírni az eljárásokat – akkor még csak a matematikai (numerikus) eljárásokra gondoltak –, hogy azt a „kódoló”, aki a gép nyelvére, gépi kódra írja át az eljárást, könnyen meg tudja ezt tenni, és az ember számára is jól áttekinthető legyen. A feladat megoldásához csatlakozott Adele Goldstine (Goldstine első felesége), Neumann János és Arthur Burks is.

A kidolgozott „nyelv” filozófiája egészen más volt, mint a Zuse által javasolté. Az előbbi nagy súlyt helyezett a típusdeklarációkra, míg Neumannék ezzel egyáltalán nem törődtek.

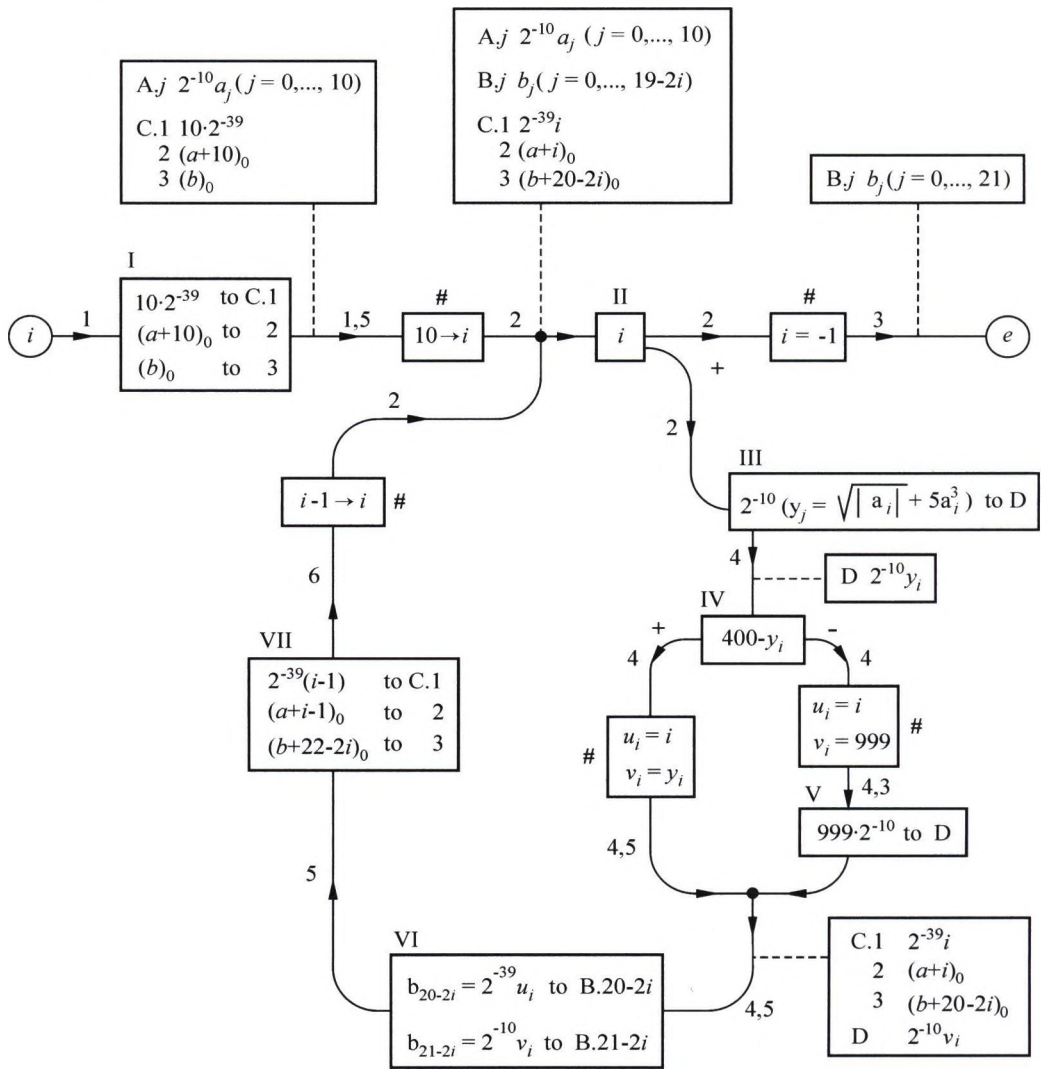
A Neumannék által kidolgozott „nyelv” voltaképpen a blokkdiagramok őse volt (ők „flow-diagramnak”, később „flowchartnak” nevezték).

A projekt 1946-ban indult, és 1947-re lettek kész vele. Folyóiratokban ugyan nem publikálták (kéziratban „Planning and Coding Problems for an Electronic Computing Instrument” címet viselte), de gyorsan elterjedt, és a később létrejött programnyelvek mellett is hosszú évtizedeken keresztül a programterveket blokkdiagramokkal ábrázolták.

A Goldstine–Neumann-diagram téglalapokból (bokszközből, dobozokból) és az ezeket összekötő nyilakból állt. (Lásd a 10. ábrát.)

A dobozok négyfélék lehetnek.

- a) Római számokkal jelölt *operációs dobozok*, amelyekben a memóriában végzendő átviteleket jelölték, azaz a memóriarekeszek (akkor még nem így nevezték) változásait



10. ábra. Az első magas szintű (grafikus) nyelv, amit Neumann János, Adele és Hermann Goldstine dolgoztak ki: a programok leírására és gyors áttekintésére használt folyamatábrára. (1947)

adták meg. Voltaképpen az elágazás nélküli (értékadásokat jelentő) programrészek jelölésére szolgáltak az ilyen típusú dobozok.

- b) *Alternatív dobozok* (szintén római számokkal jelölték), amelyeknek + illetve – jellel jelölt kimenetük van. Ezek a dobozok vezérlésátadást jelentettek: a dobozban szereplő mennyiség előjelétől függően jelölték ki az eljárás további irányát, tehát logikai feltételek szerinti elágazást adtak meg vele.
- c) *Helyettesítő dobozok*, amelyeket a # jellel jelöltek és ezekben a „→” jelet használták. Az ilyen dobozokban nem a gép által elvégzendő műveleteket jelölték. Például az „ $i - 1 \rightarrow i$ ” tartalmú doboz azt jelenti, hogy azt i (index) $i - 1$ -re változik. Ma azt mon-

danánk, hogy i ciklusváltozó értéke 1-gyel csökken. (Neumannék még nem vezették be a mai értelemben vett ciklus fogalmát.)

- d) Állítást tartalmazó dobozok (állításdobozok), amelyeket szintén # szimbólummal jelöltek, amelyekben a külső jelölések és a vezérlés kurrens állapotát adták meg. Az ábrán három állításdobozt látunk. Az egyikben az „ $i = -1$ ” állítás szerepel, a másik kettőben az, hogy az u_i, ν_i („változók”) kimenetek, ezen a ponton értékeket vesznek fel. (Mai szemmel ezt jogosabban hívhatnák „beállítási” dobozoknak, mert az i, u_i, ν_i változók értékeit állítják be. Azt is mondhatnánk, hogy ezek kommentárok.)

A helyettesítő és az állításdobozok nem jelentenek a számítógép által elvégzendő műveleteket.

Az ábrán szereplő dobozokban sok helyen látunk 2^{-n} -nel való szorzásokat. Ez azzal függ össze, hogy – mint a First Draftban is láttuk – az akkori gépek [az EDVAC is] fixpontosak voltak, mégpedig a bináris pont [vessző] úgy volt elhelyezve a memóriában binárisan ábrázolt x számban, hogy $-1 < x < 1$ teljesüljön, vagyis a gépben csak egynél abszolút értékben kisebb számok voltak ábrázolhatók, azaz a szám legmagasabb helyiértéke 2^{-1} , a következő 2^{-2} i -t, az utolsó 2^{-n} volt. (n a szám hosszát, azaz a bitek számát jelenti, EDVAC-nál ez 40 volt.) Például a memóriában ábrázolt 101101 szám ($n = 6$) az

$$1 \cdot 2^{-1} + 0 \cdot 2^{-2} + 1 \cdot 2^{-3} + 1 \cdot 2^{-4} + 1 \cdot 2^{-5} + 1 \cdot 2^{-6}$$

menyiséget, vagyis a

$$0.101101$$

számot jelenti.

A fixpontos gép komoly gondot jelentett a programozónak. A megoldandó feladatot úgy kellett transzformálni, hogy minden szám – az input adatok, a közbülső részeredmények, a végeredmény – abszolút értékben egynél kisebb legyen. Például az $\frac{a}{b}$ hányados kiszámításánál előbb meg kellett vizsgálni (a programban), hogy $a < b$ teljesül-e; ha nem, akkor a -t el kellett osztani (annak idején úgy mondtuk, hogy normálni kell) egy olyan c számmal, hogy $\frac{a}{c} < b$ teljesüljön. Természetesen a c -vel való osztás következményeit végig kellett vinni a teljes „képleten”.

Az ábrán látható táblázatban szerepel még egy érdekes jelölés, például $(a+i)_0, (b)_0$ stb. Általánosságban x_0 azt jelentette, hogy x egy memóriacím (egy egész szám), mégpedig úgy, hogy a gépben x_0 a $2^{-15}x + 2^{-39}x$ alakban van jelen, tehát egy olyan bináris szót jelöl, amelyben x kétszer szerepel: a 9–20 és a 29–40 pozíciókon (balról számítva). Az ilyen x_0 típusú számokat utasítások módosítására használták; egy 40 bites szó mindkét felét lehetett módosítani vele (pl. hozzá lehetett adni ezt a számot egy utasításhoz, miáltal az utasításban szereplő címek módosultak).

Megjegyzés. Az EDVAC-kal kapcsolatos fenti paraméterek (a szám hossza, a cím két részre bontása) nem a First Draftban leírt EDVAC, hanem feltehetőleg inkább az EDVAC után épült IAS gép adatai.

Amikor egy folyamatdiagram elkészült, a programozó elkezdhette a „dobozok” gépi kódra történő leképezését (ezt „statikus kódolásnak” nevezték).

A folyamatábra csak egy eszköz volt, ami megkönnyítette a program írását, de nagy

szükség volt a programozó problémamegoldó képességére, a gépi kódú utasítások ismeretére. Nem gondoljuk, hogy az Olvasónak érdemes volna dobozról dobozra menve elemezni az ábrán látható folyamatdiagramot, ha mégis, úgy segítünk azzal, ha megmondjuk, hogy az ennek alapján elkészítendő program a 10 elemű $(a_1, a_2, \dots, a_{10})$ tömb elemeire kiszámítja az

$$y_i = \sqrt{|a_i|} + 5a_i^3$$

kifejezés értékét, és amennyiben $400 - y_i > 0$, akkor rögzíti (például kinyomatás céljából) az (i, y_i) , ellenkező esetben pedig az $(i, 999)$ értékpárt.

A Goldstine–Neumann folyamatdiagrammal kapcsolatban megemlítjük, hogy ebben a szubrutin fogalma nem szerepel (tehát a szubrutin jelölésére nem tesznek javaslatot).

4. Epilógus

A „*First Draft...*” – Neumann János számítástechnikai alkotásai közül az első – amivel nem csak Amerikában, de mindenütt a világon elismerést szerzett mind a Pennsylvania Egyetemnek, mind az ott dolgozó kutatóknak és természetesen saját magának is. Feltételezhető, hogy ennek a tanulmánynak és persze későbbi számítástechnikai alkotásainak köszönhetően lett Neumann János világhírű matematikusból a „*számítógépek atyja*”.

A számítástechnika történetében viszonylag kevés olyan alkotás található – ilyen a „*First Draft...*” is –, ami a megjelenése után több mint 50 évvel pontosan úgy érvényes, mint ahogyan a megjelenésekor volt. A „*Vázlat*” megítélésében az sem lehet közömbös, hogy a bírálók legfeljebb azért szóltak hozzá a tanulmányhoz, mert kimaradtak a névsorból, azért senki sem, hogy ahhoz valamit hozzátegyen, netán abból elvegyen. Annak ellenére, hogy a címben benne van, hogy „*draft*”, ez a vázlat a maga valóságában tökéletes, egy költő azt mondhatná: „Olyan, mint egy csiszolt, tiszta gyémánt. Nincs benne hiba.”

Korábban már említettük, hogy az EDVAC szerzősége körüli viták, a szabadalmaztatással kapcsolatos torzalkodások oda vezettek, hogy Presper Eckert és John Mauchly szakítottak nemcsak Neumann Jánossal és Hermann Goldstine-nel, hanem a Pennsylvania Egyetemmel is.

Az ENIAC-ot a kormány – hivatalosan – 1946. június 30-án vette át a Pennsylvania Egyetemtől (PENN). A háborús évek alatt az Egyetem az ENIAC fejlesztésében kifáradt, ezért a gépet szíves örömet visszaadták a hadseregnek (egy púppal kevesebb – gondolták az egyetemen). Az ENIAC-ot 1946. november 9-én kikapcsolták és leszerelték, majd Aberdeenbe, a Ballisztikai Kutató Laboratóriumba szállították. A gép újraélesztése 9 hónapig tartott, 1947. július 23-án állították ismét üzembe.

Az Egyetem nem volt korrekt sem Mauchlyval, sem Eckerttel, sem Goldstine-nel, sem Neumann-nal szemben, ugyanis az ENIAC után számítógép-tervezéssel és – egy ideig – magával a számítógép-tudománnyal sem foglalkoztak. Az ENIAC alkotói azt is nagyon sérelmezték, hogy – a fenti négy kiváló szakembernek – katedrát sem ajánlottak fel az egyetemen. Az addig – a számítógép-tudományban – vezető pozícióban lévő egyetem számítás-



11. ábra. A princetoni IAS (Institute for Advanced Study) központi épülete.

1996-ban az Egyetem óriási ünnepséget tartott az ENIAC születésének az 50. évfordulóján. Az ENIAC építői közül, akik még éltek, ott voltak. A díszvendég Hermann Goldstine volt. Többen nehezményezték, hogy a megemlékezésen szinte egyedül Goldstine beszélt Neumann-nak, az ENIAC tökéletesítésében játszott szerepéről, Atanasoff érdemeiről, az egyetemi szónokok, de a kormány képviselői erről nem szóltak, kizárólag Eckert és Mauchly kétségtelen érdemeit emlegették.

Eckert és Mauchly még azt sem várták meg, hogy az ENIAC számológépből kvázi tárolt



12. ábra. Az ENIAC születésének az 50. évfordulója Philadelphiában. A képen Hermann Goldstine és Neumann János titkárnője. (1996)

technikai fejlesztései leépültek, csak évekkel később kezdték el újra a számítástechnikai konferenciákat szervezni, amivel megpróbálták a régi hírust visszaszerezni. A számítógép-tudományban a vezető szerepet a princetoni IAS, a UNIVAC cég és számos más egyetem, nem utolsósorban az IBM vette át, amelyek fejlesztették és továbbfejlesztették Neumann János és társainak munkáját.

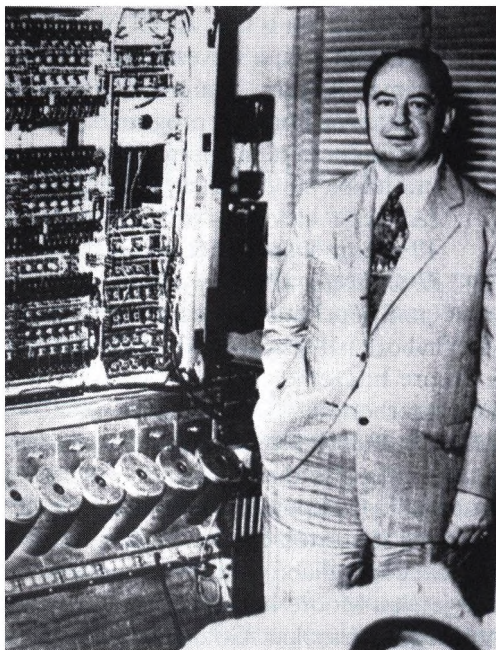
programú számítógéppé alakuljon át, úgy hagyták el az egyetemet, és megalakították Philadelphiában a saját közös vállalatukat, az Electronic Control Co.-t. Ezután hozzáfogtak az EDVAC terveit követő, ugyancsak higanysós készletelésű művonalakkal dolgozó, BINAC számítógép tervezéséhez és gyártásához. Egyébként Amerikában ez volt az első számítógép, ami az ENIAC

után épült. Ezt követően megalapították az „Eckert–Mauchly Computer Corporation”-t, megtervezték és megépítették a UNIVAC-nak nevezett gépet, amit 1951-re fejeztek be. Ezzel hamarosan elkezdődött az amerikai számítógépipar szédületes fejlődése.

Neumann sem akarta a számítógépek fejlesztését befejezni, mert már akkor ott motoszkálhatott benne az IAS gép terve.

Az IAS gépben Neumann és Goldstine – minden korábbi géphez képest – forradalmi változásokat vezettek be. Miután ez az írás az EDVAC-ról és nem az IAS gépről szól, ezeket a változásokat – Goldstine könyvéből vett idézetekkel – csak nagyon röviden foglaljuk össze:

- Az IAS gépet Burks–Goldstine–von Neumann: „Preliminary Discussion of the Logical Design of an Electronic Computing Instrument” című alapvető munkája írta le.
- A tanulmányban a szerzők, a számítástechnika tudományának az egyik legfontosabb döntését hozták meg: az IAS-ben *nem* az EDVAC-nál alkalmazott soros tárolási módot alkalmazták, hanem a *párhuzamos* szervezési módot, „*aminek az eredménye egy ún. párhuzamos működésű gép (parallel machine) lett, szemben az EDVAC-kal, amelynek a szervezése soros volt... (...)* E két rendszer közötti fő különbség az összekapcsolás elvégzésének a módjában rejlik; a párhuzamos működésű gépben a megfelelő számjegyek párhozait egyidejűleg adjuk össze, míg egy soros szervezésű gépben e számpárok összekapcsolása egymás után történik...”
- Eleinte az IAS gép tervezőiben még élt az alkatrészeket „feleslegesen pazarló” műszaki megoldástól való félelem, amit Neumann oszlatott el, aki kiszámolta, hogy a párhuzamos működésű gép – a látszat ellenére, miután a vezérlőegysége egyszerűbb, mint a soros gépeké – kevesebb alkatrészből felépíthető, mint a soros számítógépek. A körülbelül egyforma kapacitású párhuzamos IAS gép összesen 2000, míg a soros EDVAC 3000 csövet tartalmazott. A Neumann-gép – a párhuzamos működés miatt – 600 mikroszekundum alatt végzett el egy szorzást, míg az EDVAC 3,0 milliszekundum alatt. Az IAS gép az ikonoszkóp memóriájába 25 mikroszekundum alatt helyezett el egy szót, az EDVAC pedig – a soros memóriába – 200 mikroszekundum alatt.
- A két gép között teljesítménykülönbség volt, ezért – az IAS megszületése után – soros számítógépeket gyakorlatilag alig építettek. Ettől kezdve az IAS gép volt az etalon, világszerte és főleg Amerika-szerte ezt a gépet másolták. Az IAS gép közelebbi rokonai (mai kifejezéssel klónjai): a JOHNNIAC, az ORDVAC, az ILLIAC, de még a UNIVAC 1100-as és az IBM 700-as és 7000-es szériái, ezenkívül még más – párhuzamos



13. ábra. Az elkészült IAS gép. Előtte az alkotó, Neumann János. (1952)

mos, külföldi – gépek is voltak, például a már említett első hazai elektronikus számítógép az M-3-as is.

- A gép struktúráját illetően úgy határoztak, hogy a gépben ábrázolt szám 39 bitből és egy előjelbitből fog állni.
- Az IAS gép az utasításrendszerekben is forradalmi változást hozott, ez a „forradalom” még ma is tart. Az EDVAC háromcímű gép volt, az oka egy korábban elterjedt hiedelem volt, a tervezők azt mondták, minél több című egy gép, annál gyorsabban számol. Azután – amint ezt a tanulmány magyarázatában már említettük – Neumann rájött arra, hogy egy programban a két- vagy háromcímű utasítások második vagy harmadik címét nem használják, egy utasításba legtöbbször csak egy címet írnak, így az utasítások címeinek egy része az esetek többségében kihasználatlan marad. Ezért az IAS gépet eleve – először a számítástechnika történetében – egycíműre tervezték, egy 40 bites szóban így két utasítást (10 bit a művelet, 10 bit pedig a cím kódolására – lásd: *H. Goldstine: A számítógép Pascaltól Neumannig*) tudtak elhelyezni. Azóta is csak egycímű gépeket építenek.

A háború befejezése után – 1946 őszén – brit látogatók nagy létszámú csoportja érkezett a Moore Intézetbe, közöttük Maurice Wilkes is, aki kézbe kapta a „First Draft...”-ot, és megismerhette az ENIAC-kal és az EDVAC-kal kapcsolatos tervezési elgondolásokat – közöttük a tárolt program elvét is. Ezután visszatért az angliai Cambridge-be, ahol megépítette a világ első tárolt programú számítógépét, az EDSAC-ot (Electronic Delay Storage Automatic Computer), ami előbb készült el, mint az IAS, Neumann János számítógépe.

Jöttek látogatók Svédországból, aminek nyomán megépült a BESK, a dánok Svédországból és Angliából vették át a tapasztalatokat, így épült meg a DASK (Danish BESK). A szovjetek a Moore Intézetből postán kérték el a számítógépek építésének megkezdéséhez szükséges leírásokat, Goldstine – a már említett munkájában is leírta – hogy meg is kapták. Valószínűleg ennek nyomán épültek meg az első szovjet számítógépek, a MESzM és a BESzM, majd később egész sor különféle architektúrájú és teljesítményű számítógép az ország minden részén.

Neumann János alapvető munkája, akaratának megfelelően, elterjedt az egész világon.

Legendi Tamás

Neumann mint a sejtautomata megalkotója, elméleti és gyakorlati követői

Hogyan jutott el Neumann a sejtautomata fogalomig?

Előljáróban néhány gondolatot szeretnék kifejtetni az alkotó emberről, a génuszról vagy zseniről, hogy bizonyos mértékig keretek közé helyezzük el Neumann-nak a sejtautomata fogalom megalkotásával létrehozott tudományos teljesítményét, az eredmény, az érték szempontjából.

Természetesen mind ezt, mind Neumann mint ember, mind polihisztorkénti bemutatását csak vázlatosan adom elő, mivel ennek a könyvnek más helyein is szóba kerül, vagy éppen a részletes tárgyalás meghaladná a fő téma kereteit.

Az alkotó ember általában újat hoz létre, művészeti, tudományos, műszaki vagy bármely más területen. Sok esetben analitikusan elemzi a világot, majd összerakja másképp az elemeket, van, hogy egyetlen gondolatszikkral teremt újat, sokszor hosszú évtizedek munkájából áll össze az újdonság, nem ritka, hogy a véletlen is segít, de rendszerint azoknak, akik azt felkészülten tudják fogadni (mint Flemingnek a penicillin felfedezésekor). Sokszor jólismert tények, dolgok összefüggésére másként tekintve születnek a találmányok.

Néha a fentiek szerint különbséget tesznek felfedezők és feltalálók között. A körülmények is sokszor szerepet játszanak, még olyan is, mint a lázas állapot, vagy az álom, erről Sellye professzor írt érdekesen *Az álmotól a felfedezésig* című könyvében.

Nos, a különböző alkotások között minőségi különbségek vannak, ezek megítélése időben, társadalmi változásokat követően változhat, a tudományon belül sincs egzaktul definiálva, mégis másként ítéljük meg Newton és Einstein teljesítményét, mint sok más tudósét.

Mindenesetre az eredmény(ek)nek elsősorban a minősége, sokszor a váratlansága, a világgépet megváltoztató jellege emeli géniusszá megalkotóját. Az olyan tudóst, aki egész tudománynak vagy tudományágnak, vagy akár egy fontos részterületnek az alapjait veti meg, a legtöbb esetben zseninek tekintenek, jobb esetben már saját korában, mint Röntgent vagy Pasteurt, vagy később, mikor felismerik eredménye szükségességét – Bool algebrai eredményei is később, a számunkra fontos számítógépek matematikai alapjának a részévé váltak.

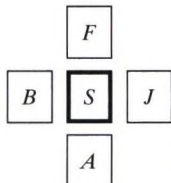
Nos, Neumann teljes tevékenysége során legalább öt területen alkotott maradandóan újat, és ennek alapján mai (túl)specializálódott világunkban modern polihisztorként is tekinthetjük, a sejtautomata fogalom megalkotásával az automataelméleten belül egy új világot, egy rész tudományágot vagy fontosnak és elterjedőnek minősíthető részterületet teremtett meg (és ennek alapján – is – tarthatjuk géniusznak).

Természetesen nem a semmiből azért („de a semmi helyén”) – az általános automataelmélet a matematika részeként már létezett, a számítógép létrehozásában Neumann társával együtt maga is részt vett és a „teremtés” során, mint látni fogjuk, fontolóra vette gyakorlati jellegű modell megalkotását is az önreprodukción. Ugyanis az önreprodukciónak kérdése sokat foglalkoztatta Neumann, erről a területről indult el a sejtautomata fogalom kialakítása felé. Anyagi jellegű modelleken gondolkozott például, folyadékban lebegő robotalkatrészekből robotok robotokat hogyan állítanának elő, de aztán sorra elvetette ezeket az ötleteket (amelyek általánosításait is gondolat kísérletekben tárgyalja Lem Summa technologiae-ja). Egyes források szerint befolyásolta döntésében Ulam, lengyel származású matematikussal folytatott konzultáció is.

Mindenesetre leszögezhetjük, hogy Neumann matematikai modellt alkotott, egy új matematikai részterületet hozott létre, amely többirányú matematikai vizsgálódás tárgyává lett, de éppen ebben az írásban elég részletesen fogjuk fejtegetni, hogy a matematikai sejtautomata-modell milyen viszonyban áll az anyagi valósággal, milyen értelemben modellezi azt, és ettől nem teljesen függetlenül miért alkalmas nagysebességű számítógépek elvi alapját is képezni.

Az önreprodukciónak modelljéhez talán a sok sejtből álló, de eredetileg azonos, egyforma (mai elnevezéssel) összejtéből kialakítható sokfunkciós szervezet képe lebeghetett előtte. A matematika egyszerűsítő-általánosító szűrőjeként adódhatott a(z elméleti modellben akár végtelen) sok egyforma (viszonylag kis) véges automata rendszere, kétdimenziós elrendezésben, talán az agy neuronhálózatának példája sugallta azt az egyszerűsítést, hogy helyiek, lokálisak legyenek (csak) a kapcsolatok ezek között az elemek/sejtek/automaták között. Az így kialakuló modell nagyon elegáns is, ez bizonyosan szempont volt Neumann számára. Na jó, de ez a modell egyelőre üres, még akkor is, ha hozzávesszük, hogy alapvető működésének szabályai egyszerűen adódnak: minden sejt/véges automata bemenetén a saját és szomszédai állapotértékei jelennek meg, ezek hatására a sejtekből álló sejtter minden egyes eleme egyszerre (párhuzamosan) veszi fel új értékét. Egy ilyen lépést szokás billenési lépésnek is nevezni, szemléletesen egy állapotmátrix-sorozatot képez a billenések hatása.

A sok kis automata ebből a szempontból egyetlen nagyobb automataként viselkedik, matematikai szakkifejezéssel és függvényként történő felfogásban a lokális állapotfüggvény, amely minden sejtben azonos, indukál egy globális állapotfüggvényt (amely azt adja meg, hogy egy állapotmátrix egy billenési lépés során milyen állapotmátrixba megy át). Már ez is rengeteg későbbi matematikus feladatot eredményezett – milyen a viszony a lokális és a globális állapotátmenet-függvény között.

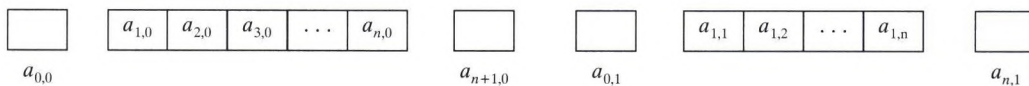


$S_{ij} = f(A, B, F, J)$ az új érték nem függ a saját állapottól
vagy $S_{ij} = f(S, A, B, F, J)$ tipikusabb: az új érték függ a saját állapottól is
Pontosabban a lépésszám (n) megadásával:

$$S_{n+1} = f(A_n, B_n, F_n, J_n), \text{ illetve}$$

$$S_{n+1} = f(S_n, A_n, B_n, F_n, J_n).$$

1. ábra. Sejtátmenetfüggvény

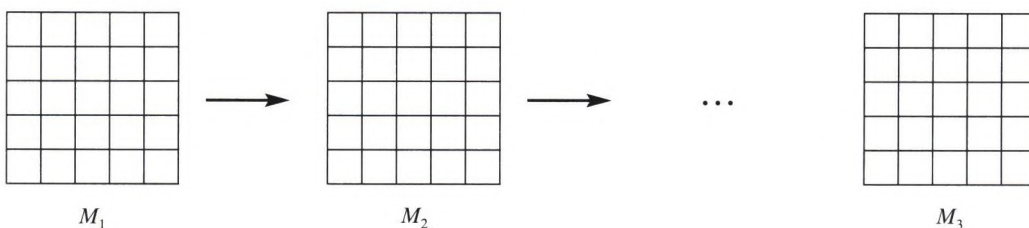


a_{ij} $i = \text{hely (sorszám)}$ $j = \text{idő (lépésszám)}$

$$a_{ij+1} = f(a_{ij}, a_{i-1j}, a_{i+1j})$$

A széleken 1-1 dummy cell-t (üres vagy álsejtet) tételeznek fel a homogén leírás kedvéért, ezek állapota vagy 0, vagy a külvilág állítja be (sejtprocesszorban itt érkezik a feldolgozandó input).

2. ábra. 1 dimenziós (véges) sejtér működése



Az f lokális átmenetfüggvény (lásd 1. ábra) generál (indukál) egy Φ globális (állapot) átmenetfüggvényt úgy, hogy minden billenési (átmenetfüggvény-végrehajtási) lépésben minden egyes sejtben egyidejűleg végrehajtódik az f lokális (állapot) átmenetfüggvény.

$M_1, M_2, \dots, M_j, \dots$ a sejtér (sejtmező) állapotmátrixai.

3. ábra. 2 dimenziós véges sejtér: mátrixsorozat mint globális átmenetfüggvény

Vegyük észre, hogy ez a konstrukció nagyon szabályos, rendezett, egyszóval homogén jellegű – rögzített a szomszédság adott szabályos geometria mellett, azonos az átmenetfüggvény minden sejtben.

De még mindig nem tud semmit, amíg konkrét állapotfüggvényekkel nem dolgozunk.

Neumann, aki a műszaki életben konstruktorként végzett alkotó munkát, a matematikában is szerette a konstruktív bizonyításokat. A sejtautomata felépítése pedig egyenesen ezt kínálja fő megoldásként!

Neumann két nagy eredeti, konkrét bizonyítást alkotott meg, egy lokális állapotfüggvény definiálásával és két „szerkezet”, két kiindulási részállapotmátrix megadásával.

Az egyik „szerkezet”, „konfiguráció” képes volt önreprodukcióra – természetesen matematikai értelemben (egészen tág értelemben), ha bármely automata leírását megkapta (a külvilágból, illetve behelyezve a sejtérbe), akkor generálta azt a konfigurációt, amely ugyanúgy „viselkedett”, mint a leírás szerinti automata – ha a saját leírását kapta meg, vagy eleve „benne volt”, akkor „saját maga további példányát állította elő” a sejtérben.

A másik szerkezettel/konfigurációval azt bizonyította Neumann, hogy a sejtautomata-rendszer mint automata a fenti működési szabályokkal univerzális, azaz mindent ki tud számítani, abban az értelemben, ahogy az általános kiszámíthatóságot a Turing-géppel reprezentálják (mert arról bizonyították, hogy ekvivalens azzal), mert ez a második konfi-

guráció Turing-gépet modellezett a sejtautomatában (műszaki területen szimulációról, emulációról beszélünk hasonló értelemben, hétköznapi nyelven szólva: ugyanúgy viselkedik, ugyanazt tudja csinálni).

Neumann konkrét 29 állapotú átmenetfüggvénye nagyon bonyolult volt (mások számára), a konstruktív bizonyítások is nehezek, és nem is voltak teljesen viszonylag korai halála miatt.

Neumannról mint emberről sok mindent írtak már, itt csak annyit szeretnék megemlíteni, hogy nevét legszívesebben John von Neumann formában használta, büszke volt családja bárói címére és nagyra értékelte befogadó/fogadott hazáját is, nevében is kifejezve ezeket a kötődéseit. Matematikus és konstruktőr volt, de ne gondoljanak szobatudósra, az alkalmazások, amelyek számára alapot teremtett, vagy amelyeken maga is dolgozott, már utalnak erre, de kellemes társasági ember is volt, anekdotáival, sziporkázó beszédstílusával sok társaságot hódított meg. Értékes képessége volt – a matematikusokra közhelyszerűen nem jellemző – fejszámolói tudása.

A sejtautomata mint az anyagi világ folyamatainak modellje

Neumann megalkotta a sejtautomata fogalmát önreprodukciós modell létrehozásához, egyben konstruktív matematikai bizonyítással univerzalitására is rámutatott.

Azonnal adódtak matematikai vizsgálódási területek a matematikának ebben az új fejezetében – utaltam a lokális és globális átmenetfüggvények közötti összefüggések vizsgálatára, amelyeket ugyan nem fogok részletezni (később a sejtautomaták meglepő globális tulajdonságairól viszont egy alfejezet fog teljesebb képet adni), de a most következő sorok alapvetően ennek az összefüggésnek az anyagi világban való megjelenését fogják tárgyalni.

Ez is arra mutat rá, hogy hol a valóság absztrahálásából (általánosításból, attól való elvonatkoztatásból), hol a matematika egyes ágaihoz tartozó „üresnek tűnő” önmagukban zárt rendszerek létrehozásuk után „új életet kezdenek élni” sokszor, az elmélet szintjén is bővülnek, vagy felhasználhatók az anyagi világ modellezésére, alkalmazások kezelésére, feladatok megoldására.

Mit is jelent a sejtautomata fogalom az anyagi világban? Mi felel meg ennek a fogalomnak?

Elég könnyű rájönni – azok a folyamatok, ahol a helyi, lokális történések szabályai határozzák meg a teljes folyamat lezajlását. Akár szinte minden folyamatot felfoghatunk ilyennek, az anyag különböző mozgásformáitól kezdve a társadalmi folyamatokig. Ez nem is olyan meglepő, hiszen a véges automata fogalmáról is elmondható hasonló állítás, a sejtautomata pedig ezekből áll, nagyon szabályos szerkezetben egy nagyobb automataként is értelmezhetően. Persze az ilyen általános kijelentések még nem feltétlenül tesznek lehetővé felhasználhatóságot is. Tehát keressünk speciális területeket – igen szerencsésnek mondható például a folyadékok áramlása vagy a gázokra vonatkozó törvényszerűségek leírása.

Ezeknél valóban a szabályosan leírható helyi, lokális folyamatok generálják, indukálják a teljes mozgásfolyamatot. Szokványos leírásuk differencia-, majd differenciálegyenletek formájában történt, képlet jellegű matematikai megoldással. Majd a numerikus matemati-

kában megjelent a grid (némi egyszerűsítéssel háló, négyzetrács) fogalma, és az ezen történő számítási eljárásokkal történő megoldás. Vegyük azonban észre, hogy a grid szinte sejtautomata, legalábbis elrendezését tekintve – ha pedig azt nézzük, hogy minden pontjában azonos algoritmust kell végrehajtani, akkor már tényleg nyugodtan állíthatjuk, hogy a sejtautomata természetes modellje ezeknek a feladatoknak, a bevezető mondatra visszapillantva egyáltalán nem meglepő módon. Természetesen ez sem rengeti meg a világot még gyakorlati szempontból, mert ha a sejtautomata modellel akarnánk megoldani ezeket a feladatokat, ugyancsak lassan menne a dolog.

De máris van egy indokunk sejtautomata felépítésű számítógép készítésére és használatára, mert a teljesítmény drámaian megváltozik, ha a gridet hardverben megépített, önmagukban is nagy teljesítményű és nagyszámú (több ezer – többmillió) sejt/nanoprocesszor hajtja végre.

Egy nagyszámú sejtet tartalmazó sejttérből mint fő műveletvégző elemből és egyéb komponensekből álló rendszert a továbbiakban **sejtprocesszornak** nevezem.

Kezdhettem volna persze egy jóval egyszerűbb feladat, a biliárdozás sejtes megoldásával is, de annak mind fontossága, mind sorosabb jellegű végrehajtása miatt – szemléletesége ellenére – kevésbé hatékony a megoldása, kevésbé fajsúlyos a bemutatása.

Térjünk át a valódi világ egy másik fontos szegmensére, a látásra és ennek a digitális számítógépekhez kapcsolódó feladataira, amit digitális számítógépeken képfeldolgozásnak neveznek. Mi ennek a fő jellemzője? Különösen a kép előfeldolgozási feladatokon, de sok képtranszformációs feladaton és egyéb képfeldolgozási műveleteknél is jellemző a nagyszámú képponton vagy egyéb képleíró objektumon (például vektorokon) végrehajtandó nagyszámú azonos művelet. Ismét ott tartunk, hogy a feladat anyagi világbeli természete miatt sejtautomatával természetes módon modellezhető egy feladat. Ez valószínűsíti nagysebességű sejtprocesszorral történő megoldhatóságát (később látni fogjuk, hogy ez nem feltétlenül közvetlenül egy képnek a beültetésével történik, de mindenképpen a feladat természetéhez igazodó módon és sokkal hatékonyabban, mint egy hagyományos, akár szuperszámítógépen, vagy – a majd a hardverről szóró részben tárgyalt általános vektorprocesszorokkal, mivel a sejtprocesszor speciális vektorprocesszornak is tekinthető).

1. Neumann elméleti követői

Neumann zseniális, de elég bonyolult konstrukcióját, befejezetlen kézírata alapján *Burks* rendezte sajtó alá (szinte újjáírva, részben feltalálva egyes részeket), és ez az alapvető munka sokak érdeklődését keltette fel, hiszen egy új, további felfedezésekre csábító világot tárt fel.

A sokfelé elágazó elméleti munkák közül itt négy irányt emelek ki:

- Neumann viszonylag bonyolult konstrukciójának egyszerűsítése;
- különböző paraméterekkel (szomszédok száma, állapotok száma stb.) rendelkező sejtautomaták viszonyának, különösen egyenértékűségüknek (ekvivalenciájuknak) a vizsgálata;

- olyan érdekes feladatok megoldása, amelyek látszólag csak központi vezérléssel hajthatók végre, de lokális, párhuzamos megoldásuk is feltalálható/kidolgozható;
- matematikai értelemben vett egyszerű önreprodukció igen egyszerű (kis állapotszámú/szomszédságszámú, egyszerű átmenetfüggvényű) sejtautomatákkal.

Neumann konstrukciójának egyszerűsítése

E. F. Codd nevéhez fűződik a viszonylag bonyolult neumann konstrukció egyszerűsítése, érthetőbbé és átláthatóbbá tétele.

Codd mindössze nyolc állapottal és négy szomszédsejttel és (emiatt is) „olvashatóbb”/érthetőbb átmenetfüggvénnyel oldotta meg mind a Turing-gép modellezését („szimulációját”), mind az önreprodukciós feladatot. Hallatlan tömörségű monográfiájának ez még így is csak a kisebb részét alkotja.

Különböző sejtautomaták viszonyának vizsgálata

Codd volt az (is), aki megvetette – konkrét vizsgálatokkal – ennek a sejtautomata-kutatói ágának az alapjait monográfiája másik részében, kiemelem a különböző sejtautomaták ekvivalenciájának vizsgálatait.

Érdekes eredmény, hogy a négyállapotú nyolcszomszédos sejtautomaták ekvivalensek a codd sejtautomatával, tehát az állapotszám és a szomszédságszám bizonyos mértékig „pótolhatja” egymást. Ezek a vizsgálatok feleslegessé tehetik és teszik is ugyanazoknak a feladatoknak a vizsgálatát különböző sejtautomatákban, egyben lehetővé teszik az egyes feladatoknak abban a sejtautomatában történő bemutatását, ahol az a legegészségesebb, legáttekinthetőbb.

Banks kétállapotú sejtautomaták ekvivalenciáját mutatta be nagyobb szomszédságszámú sejtautomatákkal, igazolva, hogy még a legkisebb állapotszámú sejtautomaták is univerzálisak (persze modellezési erő/számítási teljesítmény sincs ingyen, több lépéssel (idővel), több sejttel kell „fizetni” – amikor már hardver realizálásban gondolkodunk, akkor a klasszikus *trade-off* elv alapján választunk).

Sejtautomata megoldások központi vezérlésű jellegű feladatokra

Akár szűkebb értelemben vett elméleti matematizálásról is beszélhetnénk ezekkel a feladatokkal kapcsolatban, de mégis jelentősek – mutatják a sejtautomata szervezés/„konstrukció” erejét, lehetőségeit, túlmenően azon a lehetőségen, hogy a sejtautomata a világ jelentős számú jelenségének lehet a modellje (mindez azért fontos, mert mind a jelenségek természetének feltárását, mind számítógépes szimulációjuk, az adott esetben különösen *sejtprocesszor*on történő nagysebességű szimulációjukat teszi lehetővé).

Egy feladatosztályt emelek ki a sok közül – a szinkronizálási problémákat.

A sejtautomata-elméletben megszokott végtelen számú sejt szerepel itt is, azonban nagy (de véges) számú sejt esetén is fogós problémáról van szó. Tehát a csak szomszédakkal „érintkező”/kommunikáló sejteknek hogyan legyen „órájuk”, hogyan működjenek együtt – egy rendszerben ehhez valamilyen szinkronizálásra van szükség. Egy ilyen leegyszerűsített, de nem könnyen megoldható feladat a sortűz (firing squad) probléma.

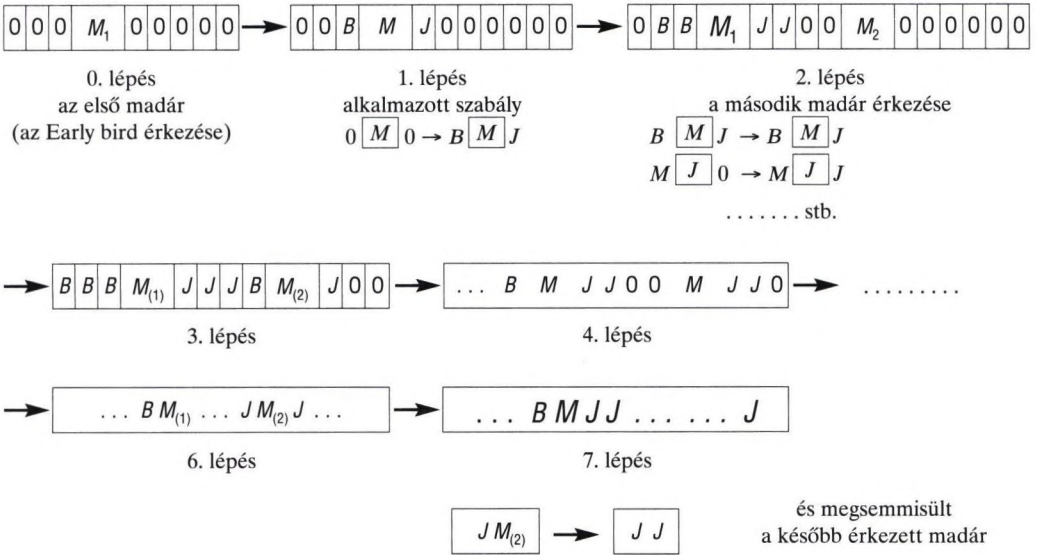
Maradjunk egydimenziós sejtautomatánál, legyen a lineárisan elhelyezkedő sejtek mindegyikének két (a jobb és a bal oldali) szomszédja. A sortűznek az felel meg, ha egy üres vagy akár tetszőleges kezdeti állapotból valahányadik lépésre minden sejt egy kijelölt különleges T(űz!) állapotba megy át – természetesen –, úgy, hogy előzőleg egyikük sem veszi fel ezt az állapotot.

Nos, véges sejtterben működne egy olyan megoldás, hogy a sejtekben lenne egy számláló (ez sejtautomata – rész-átmenetfüggvény segítségével megoldható), mondjuk a jobbszélő sejt kezdeti állapotába bekódoznánk az összes sejt számát, az átmenetfüggvény további részét pedig úgy konstruálnánk meg, hogy minden sejt eggyel kisebb számot adjon át a szomszédjának, a számlálók csökkenő módon „lefelé” számoljanak (countdown számláló), és amikor a 0-t elérik, akkor menjenek át T állapotba.

Végtelen térben ez már így nem megy, ott különleges „interferáló” hullámokat kell indítani, és azok kölcsönhatása vezet a szinkronizáláshoz. Akit érdekel, próbálja meg megoldani a feladatot, vagy utánanézni a teljes megoldásnak a szakirodalomban.

Egy másik ilyen feladat az Early Bird-probléma, ahol egy kitüntetett („madár”) állapotot vesz fel időnként (a külvilágból – az ebben az esetben szintén egydimenziós sejtautomata peremén megjelenő érték hatására) egy-egy sejt. Maga a feladat arról szól, hogy a sejtautomata tudja megkülönböztetni a korábban érkezett madara(ka)t és a későbben érkezett(ek)et „ölje meg”, azaz egyetlen „madár” állapotú, „a legrégebben/legkorábban érkezett madarat” szimbolizáló sejt legyen csak egy idő után a sejtautomatában. Nos, születtek erre is több ezer állapotú megoldások a sortűz problémához hasonló hullámok generálásával, de sikerült a gyakorlati célú rendező, összeadó, szorzó stb. *sejtalgoritmusok* készítése során kiérlelt optimalizálási technikák, vagy egyszerűen csak tapasztalat alapján az alábbi ötletet kidolgoznom: mindössze 5 (öt) állapotú sejtautomatára van szükség (!): „madár” állapot természetesen, „jobb toll”, „bal toll”, „semleges” és már ezekkel el lehet magyarázni a megoldás lényegét: a madár minden lépésben „kibocsát” egy jobb szárnyat és egy bal szárnyat – így minden lépésben „egyre hosszabb szárnyai” lesznek, tulajdonképpen ezzel lesz „erősebb” a később érkezőknél; ha közben érkezik egy újabb madár, az is hasonlóképpen jár el, és amikor találkozik – egymás mellé ér – egy bal szárny és egy jobb szárny, akkor annihilálják, megsemmisítik egymást, ha egy jobb szárny vagy bal szárny madár mellé érkezik, akkor annihilálja, semleges állapottá alakítja át – így a hosszabb szárnyú madár „győz”. Ez természetesen csak a lényege, a váza a megoldásnak, a precíz bizonyításhoz a részletek kidolgozása és teljes indukció jellegű gondolatmenet is szükséges, hiszen például az első és a második madár közé is érkezhet madár stb. Szükséges még természetesen egy ötödik, „semleges”, „tétkitöltő” állapot is.

M_1, M_2, \dots -nek egyaránt az M állapot felel meg, csak a szemléletesség kedvéért jelöltük így.



- 0 - semleges állapot
- M - madár
- J - jobb hullám
- B - bal hullám

A szabályoknak csak egy (kis) része szerepel itt, nem klasszikus formában (is).

4. ábra. Az Early bird probléma megoldásának a lényege a működés egy részének bemutatásával

Kiváló matematikus barátom, Katona Endre dolgozta ki a precíz bizonyítást. Engem még izgatott a probléma folytatása, általánosítása, és 2 dimenzióra is adtam egy ötletnél több, de szintén vázaltszerű megoldást, ami az előzőhöz hasonlóan került teljes kidolgozásra. Akinek van kedve, gondolkodjon el a 2, esetleg a több dimenzióra történő megoldás lényegén!

A történethez még hozzátartozik, hogy egy német sejtautomata-kutató részletes, esztétváltató jellegű bizonyítást dolgozott ki arra, hogy négy állapottal már biztosan nem lehet megoldani a problémát. A matematikusok számára így kerek, lezárható a vizsgálódás – legalábbis addig, amíg egy újabb matematikus elő nem áll egy kapcsolódó újabb problémával – **Vollmar** professzor tréfás, de sok igazságot magában rejtő matematikusdefiníciója szerint:

A matematikusok elsősorban problémacsináló emberek, de előbb-utóbb valamelyikük meg is oldja a problémát, mert problémamegoldók is.

Egyszerű önreprodukció sejtautomatákkal

Természetesen ebben az esetben is matematikai értelemben vett önreprodukcióról lesz szó, az egyszerűséget az jelenti, hogy itt nem is teljes sejtstruktúrák vagy véges automaták beültetéséről, szimulációjáról lesz szó, hanem „csak” alakzatok (állapotmátrixok) többszörözéséről, a legegyszerűbb esetben pedig arról, hogy egyetlen (azonos) állapotú sejtek maradnak csak a sejtautomatában („kihalás” vagy pedig egy vagy több más állapot is fennmarad egyáltalán).

Két példát fogok bemutatni: *Conway* híres *Game of Life* (Életjáték) átmenetfüggvényét és a *Lindenmayer*-függvényét.

Az életjáték esetében tulajdonképpen függvénycsaládról kellene beszélni, mert a lényegét néhány egyszerű szabály képezi, amelyek alapján sokféle konkrét átmenetfüggvényt konstruálhatók.

Az alap gondolat az evolúció egyes kiragadott elemeinek/alapszabályainak megvalósítása sejtautomatával, nevezetesen az életben maradásnak a populáció területi sűrűségével való összefüggésének a leképezésével. Ha túl sűrű a populáció („sokan vannak”), akkor nincs elegendő élelem a populáció tagjainak: sok vagy akár az összes tagnak el kell pusztulnia (végül), ha viszont nagyon ritka a populáció („kevesen vannak”), akkor nem tud szaporodni a populáció (itt nyilván az ivaros szaporodás a modell), és kihal. A konkrét állapotfüggvények esetén ezek a nagyon egyszerűnek, szinte primitívnek látszó alapszabályok – akár kétállapotú sejtekből álló sejtautomaták esetén is elég bonyolult „viselkedést”, állapotmátrix/alakzat sorozatokat eredményeznek.

A populáció(k) tagjainak adott állapot, az üres térnek egy kitüntetett állapot felel meg értelemszerűen.

Már az első feltétel esetén is észrevehető, hogy a túl sűrű populáció „ritkítását” előíró szabály nem szükségszerűen vezet kihaláshoz, kisebb sűrűség mellett stabilizálódhat, vagy újra növekedésnek indulhat a populáció. Valóban, a gyorsan fellángoló, majd kisebb-nagyobb ingadozásokkal mindmáig tartó életjáték-átmenetfüggvény konstruálási láz (még díjakat is kifizettek, például végtelen növekedést eredményező átmenetfüggvény-kezdőállapotmátrix párra) már a legelején azt bizonyította, hogy egyes függvények kihalást eredményeznek, mások „befagynak”, „helyben járnak” vagy úgyis megfogalmazták, hogy „egy lépésben oszcillálnak”. Vannak több lépésben oszcilláló, „ciklizáló” esetek is, a végtelen növekedésre az egyik híressé vált példa az „ágyú”, amely néhány lépésenként „kilőtt” egy „golyót”, és ezt a végtelenségig folytatta.

Az olvasó figyelmébe ajánlom, hogy könnyű és nagyon szórakoztató ilyen függvényeket programozni, életjátékot szimulálni. Aki erre valamiért nem adja rá a fejét, annak a sokféle szimulátor közül az egyik modern, PC-ken futó kifejezetten életjáték-szimulátort, *Mirek Celebration*-ját (megtalálható a Neten, letölthető) ajánlom kipróbálásra (egyszerűen – is – használatba vehető, kész függvényekkel, de gazdag eszköztárat kínál annak, aki hajlandó beletanulni egy kicsit és izgalmasabb szimulációkat futtatni, vagy éppen kifejezett játékos csinálni belőle egy vagy több játékos számára.

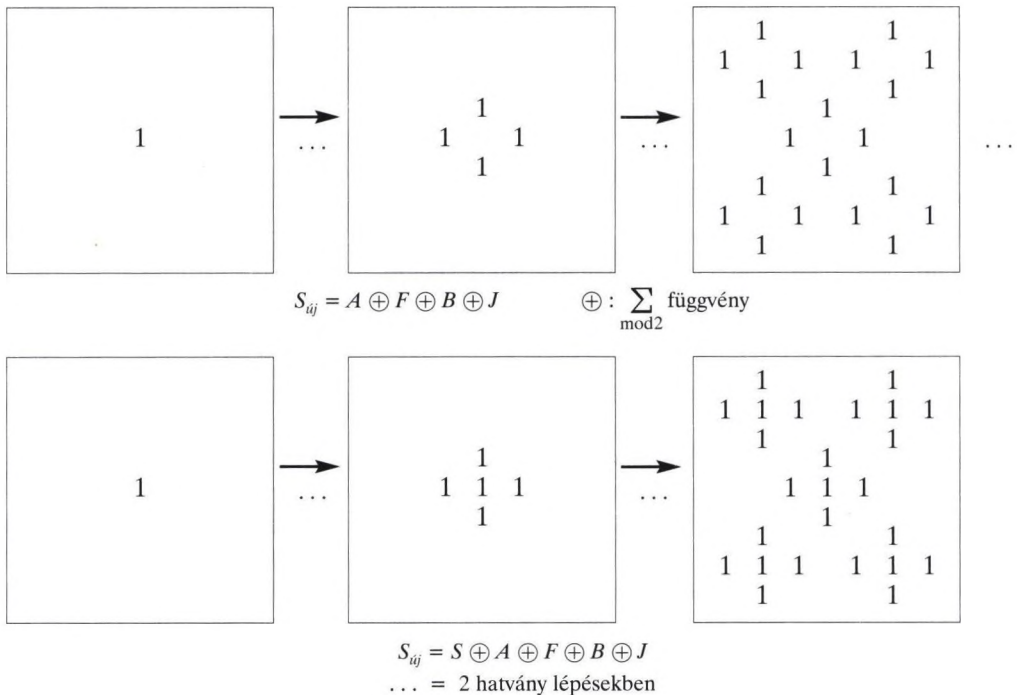
A *Lindenmayer*-függvény első ránézésre egy igen egyszerű, jól ismert alapfüggvény: a szumma moduló 2 függvény (értelemszerűen kétállapotú sejtekből álló sejtautomatában).

De mit tud művelni ez az egyszerű függvény egy sejtautomatában, ahol sok vagy végtelen sok „példányban” „fut” – az már elég meglepőnek bizonyul.

Némi gondolkozás után adódnak általánosítások és variációk a hagyományos sejtautomaták körében is, nagyobb állapotszámra kiterjeszteni szorzattérként (ekkor az egyes sejtek állapotbitjei által alkotott bitmátrixokra vonatkozik az alapfüggvény), vagy szumma modulo N függvény alkalmazása N állapotú sejtekből álló sejtautomatában.

Visszatérve az alapváltozatra – legyen az állapotfüggvény a négy szomszéd szumma moduló 2 érték(e), tehát csak a négy szomszédtól függ az új érték. „Rajzoljunk be” a 0 állapotú sejtekből álló sejtautomatába egy kis, 4 darab 1 állapotú sejtből álló rombuszt.

4 (billenési) lépés múlva 4 darab rombuszt fogunk látni! Egyes 2 hatvány lépésekben pedig további hasonló arányú „szaporodást” fogunk látni (azért nem minden 2 hatvány lépésben, mert a rombuszok között „nincs mindig elég hely” – ez távoli analógia csak az életjárték-függvénnyel).



5. ábra. Egyszerű önreprodukció a Lindermayer-függvénnyel (2 állapotú sejtek)

Végtelen sejtautomatában ez végtelen „szaporodáshoz” vezet, véges sejtautomata esetében pedig az állapotmátrix méretétől is függ, mi történik – 2 hatvány méretű négyzet úgy fog ciklizálni, hogy visszatér a kezdőállapothoz! Egyéb méretek mellett szabálytalanabb viselkedést mutat a sejtautomata, a „szélekhez érkező” „utódokból” „mutációk”, „torzszülöttek” alakul(hat)nak ki.

Érdekes eredményekhez vezet, ha a sejt mátrixban a szomszédságot tőrúszként definiáljuk, azaz a felső sejt sor egyik szomszédait az alsó sejt sor – és viszont – alkotja, és ugyanez

történik a bal oldali és a jobb oldali sejtoszloppal. Lényegében rövidebb ciklusok alakulnak ki.

A sejtautomata változatosságának kimerítésétől azonban messze vagyunk, az egyszerű alap ellenére! A fentiek alapján el is lehet „játszani” egy darabig a felsorolt változatokkal, illetve más és más kezdőállapot alakzat, szomszédság választásával (pl. a saját állapot is számítson bele a függvényértékbe! Mi fog történni?).

De igazi játékot is lehet játszani sokféleképpen, pl. 2 játékos különböző alakzatokat helyez („rajzol”) be a térbe, és ezek „harcolni” fognak egymással, közös „utódaik”, mutációk keletkeznek.

Különböző konkrét játékszabályok alkothatóak – pl. befejezés eldöntése lépésszám, vagy egyik fél alakzatának teljes eltűnése, vagy a két alakzat közötti arány jelentős eltolódása esetére; megadott időpontokban „tartalék alakzatok” behelyezhetősége – egyfajta sejtautomata go-szerű játék jöhet létre.

Még mindig nincs vége, sőt most jön a java talán. Nem lebecsülve a matematikai konstruktív vizsgálódást, a játékot (amely az egyik legkomolyabb dolog a világon, egyben máris egy *alkalmazás*) éles, gyakorlati célú alkalmazás is kihozható ebből az egyszerű sejtautomataalapból.

Később kerül részletesebben bevezetésre a térben-időben változó átmenetfüggvényű sejtautomata fogalma – gyakorlati okokból (mivel a homogén sejtautomaták programozhatósága igen nehéz, hatékonyan pedig szinte lehetetlen). Ez a fogalom, illetve általánosítás segít itt is, már egydimenziós, de különösen kétdimenziós, a szumma moduló 2 függvényhez hasonló egyszerű további függvények alkalmazásával (térben inhomogén sejtautomata esetén az egyes sejtekhez 1-1 meghatározott állapotfüggvény tartozik; időben inhomogén esetben pedig ez a hozzárendelés időben változik) kódolás-dekódolás, tehát nagysebességű titkosítás végezhető, egyéb sejtautomata tulajdonságok – pl. a fentebb említett ciklizálás – segítségével veszteségmentes tömörítés is végrehajtható!

2. Neumann sejtautomatájának gyakorlati követői Sejtautomata számítógép, emulátor, sejtprocesszor létrehozása és alkalmazása

Az első hazai kezdeményezés: a Codd-ICRA sejt és sejtautomata

Fáy Gyula elévülhetetlen (hazai) érdeme, hogy a sejtautomata neumann-i fogalmát megismertette a hazai elméleti és gyakorlati szakemberekkel szemináriumok és kis sorozatban terjesztett kéziratok formájában, majd Codd munkásságát is bemutatta, és ennek alapján műszaki konstrukciót is javasolt, felismerve, hogy a sejtautomata alkalmas nagysebességű, nagyfokú párhuzamos működésű programozható számítógép konstrukciójának alapjául.

Ehhez kisebb-nagyobb módosításokkal átvette a Codd-sejtet és a homogén felépítésű sejtter elvét, valamint a neumanni, illetve a codd sejt „programozásának” alapvető fogáit.

Ezek azonban alapvetően konstruktív matematikai bizonyítások célját szolgálták, a hatékonyság kérdése igazából fel sem merült (később több elméleti vizsgálat foglalkozott a már említett ekvivalencia kérdéseken kívül egyes feladatok megoldásának lépésszám („idő”) igényével különböző sejtautomaták esetén).

A sejtautomata-átmenetfüggvény jelentős részét lefoglalta a kezdeti „programbetöltés”, azaz olyan kezdeti állapotmátrix betöltése a sejtter széléről, amely azután feldolgozza a szintén a szélről érkező adatokat. Ezen később „dummy cell”-ek, egyszerűsített szélsejtek alkalmazásával javítottak. Láthatjuk, hogy megjelent, ha csak minimális mértékben is az inhomogenitás fogalma és a későbbiekben is általánosan használt elv, az adatok szélről történő betöltése és a sejtteren keresztül áthaladva történő feldolgozása.

Ezt az elvet a kisebb méretű sejtprocesszoroknál szinte kötelező alkalmazni, nagyobb méretű sejttereken már célszerű meghaladni és biztosítani a sejtter belsejébe történő adatküldést is, ami elméleti szempontból inhomogén statikus vagy inhomogén változó szomszédságot is jelenthet.

Műszaki szempontból a – matematikai oldalról nézve ideális – Codd-sejt elviselhetetlenül nagy, mérete meghaladja egy 4 bites mikroprocesszor méretét.

Programozása az egyforma, azonos állapotfüggvényű sejtek, a sejtter totálisan homogén volta miatt rendkívül gazdaságtalan, egyszerű műveletek elvégzéséhez is igen sok ilyen nagyméretű sejt szükséges.

Rendkívül ötletdús megoldások születtek, a későbbi sejtprogramozás csírái is, de talán elég lesz egyetlen példát említeni – a rendezést (szortolást): a rendezendő anyag minden egyes bitjének feldolgozásához egy 40×50 Codd-ICRA sejtből álló konfiguráció (állapotmátrix) volt szükséges (a felhasználandó hardver mennyisége 1 kisméretű sejtre csökkenthető a későbbi térben-időben inhomogén, kisméretű, optimalizált átmenetfüggvényű sejteket tartalmazó sejtprocesszorokban és a végrehajtási idő is jelentősen csökken, így több nagyságrenddel jobb eredmény érhető el).

Összefoglalva azt mondhatjuk, hogy a hazai bevezetés, a műszaki felhasználás ötletében úttörő munka volt a Codd-ICRA projekt, azonban tanulságos abból a szempontból, hogy egy matematikai konstrukciót nem triviális hatékony műszaki megoldássá alakítani át, bármilyen remek nyersanyag is.

Toffoli sejtautomata emulátorrendszere

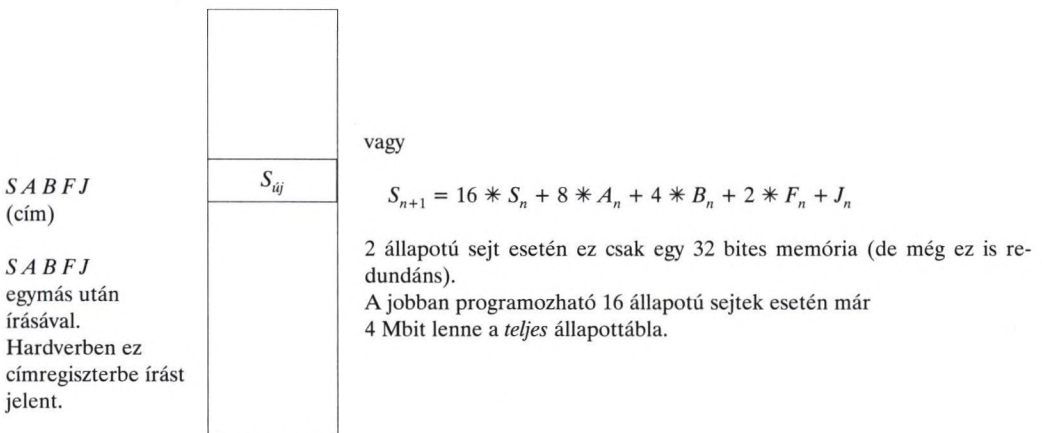
Tommaso Toffoli a római IAC (Istituto per le Applicazioni del Calcolo – a számítógépek alkalmazásával foglalkozó) intézetnél fizikusként kezdett érdeklődni a sejtautomaták iránt, mint ahogy már fejtegettem előzőleg is a lehetőséget, fizikai jelenségek modellezéséhez. Csakhogy a sejtautomaták szimulációja igen lassú, a fizikai jelenségek meg elég nagyok voltak. Így aztán Toffoli programozható célhardver, emuláció mellett döntött.

Mielőtt ennek a bemutatására rátérnék, megemlítem, hogy vizsgálatai során több matematikai, sejtautomata-elméleti eredményre is jutott, amelyek közül legalább egyet érde-

mes közelebről is szemügyre vennünk. A fizikai folyamatok megfordíthatósága igen fontos kérdés a fizikán belül, de filozófiai és számítástechnikai vetületei is vannak, ez utóbbiak például programok teszteléséhez, „belövéséhez” is fontosak. Ez az első ránézésre nem túl bonyolultnak látszó feladat utasításszinten gyakorlatilag megoldhatatlan a belső programozású, szekvenciális gépeken belül, legalábbis utasításszinten, különösen a felülről adatok miatt, de a vezérlés jellege miatt is (egy interpreter jellegű megoldás pedig igen nagy idővesztést okozna). Az igényeket elérhető mértékű ráfordítással kielégítő megoldás azért létezik, backtrack („visszatekerés”) formában, ez lényegében időnkénti teljes program és adatállapot-mentést jelent.

Visszatérve a fizikai gondolkodásmódra, a folyamatok megfordíthatósága az idő folyamásának megfordításával lehet egyenértékű adott peremfeltételek mellett, így valóban kardiális, központi kérdés. A sejtautomata-rendszereken belül többféle lehetőség is kínálkozik, és Toffoli ezek közül egy univerzálisat ragadott meg és bizonyított, nevezetesen azt, hogy minden sejtautomatához található/konstruálható egy eggyel nagyobb dimenziószámú sejtautomata, amelyhez van inverz átmenetfüggvény is, amelynek végrehajtásával a sejtautomata mintegy „visszafelé megy/működik”, azaz a lokális átmenetfüggvény által indukált globális függvény inverz globális függvénye szerint az állapotmátrix-sorozat fordított sorrendben változik.

Az univerzális számítógépen egyetlen sejt „billenésének”, egyetlen állapotátmenetének számítása is több logikai és/vagy aritmetikai, illetve címszámítási művelet végrehajtását igényli, ami többlépésnyi gépi ciklusidőt vesz igénybe. Ennek hardvergyorsítása is főleg ezen eszközök közül való választással történhet. Mivel az aritmetikai műveletek nagyobb hardvert és több végrehajtási időt igényelnek, maradnak a logikai műveletek, vagy az átmenetfüggvény kimenő értékének memóriában történő elhelyezése úgy, hogy a szomszédos sejtek (esetenként ezek közé beleértve a saját sejt) állapotértékét együttesen címzésre használjuk fel, így különböző szomszédságoknak más-más cím és azon a címen lévő állapotérték olvasható ki.



6. ábra. Toffoli sejtautomata-emulátorrendszere (állapot) átmenetfüggvénytábla (memória) 2 dimenziós sejtérhez

Toffoli, aki ekkor már az MIT-en dolgozott, ez utóbbi megoldást választotta (mivel tet-szőleges állapotátmenet-függvény programozható megoldása univerzális kapuáramkörökkel lehetséges ugyan, de alkatrészszámban és a vezérlés viszonylagos bonyolultsága miatt egy sejt esetén ez nem versenyképes megoldás – nagy sejtmező esetén alternatívát jelent, ahogy az SP1 sejtprocesszornál konkrétan kidolgozott megoldás is mutatja majd a későbbi tárgyalás során).

Toffoli memóriabázisú emulátorában egy memória egy teljes átmenetfüggvényt tartalmazott, így a szomszédság kimenete – némi egyszerűsítéssel, mert azért szinkron áramkör-ről van szó – egyszerűen odahuzalozható volt a memória bemeneteire, és a kimeneten jelent meg az előre betöltött függvényérték. Természetesen gondoskodni kellett a kezdeti betöltésről, a sejtállapotmátrix elemeinek oda- és visszajuttatásáról és a működés szinkronitásáról, mindez viszonylag kevés hardver/szoftver ráfordítás mellett – a CAM-4, majd a CAM-6, illetve a CAM-8 egy PC-kártya formájában lett realizálva és legalább két nagyságrendnyi gyorsítást eredményezett a végrehajtás során.

Képzeld el a gyorsítást, ha sok sejt lenne hardverben realizálva!

Vagy akár az egész sejtmező, vagy legalábbis egy jelentős része!

Több tízezer vagy több millió sejt dolgozna egyidejűleg egy feladaton!

Ez azonban szinte megoldhatatlan a Toffoli-sejt ehhez a célhoz képest nagy mérete miatt.

A Toffoli-sejt elemzése nagyméretű sejtmezők megvalósítása szempontjából

Amellett, ha figyelmesebben megvizsgáljuk a konstrukciót, későbbi munkánk szempontjából szerencsés dolgot figyelhetünk meg – ez a megoldás már matematikai értelemben is gazdaságtalan.

A teljes átmenetfüggvény tárolása felesleges, mert redundáns! Négyszer-öttször annyi bemenő (cím)bittel rendelkezik, mint ahány bitből áll a kimenete. Ha faként fogjuk fel az átmenetfüggvényt, amelynek levelein az új állapotértékek vannak, akkor gráfelméleti oldalról is adódik a redundancia, a méretének csökkenthetősége.

TRANSCCELL függvényleíró és optimalizáló szoftverünkkel logikai műveletekkel megadva átmenetfüggvényeket optimalizáltunk, és a végső szakaszban optimalizált memóriatartalommal alakítottuk. 1-2 nagyságrenddel (!) kisebb memóri igény adódik, természetesen némi címszámítási veszteség árán, amit azonban (fix) célhardver végez el, és ennek mérete és végrehajtási ideje elhanyagolható a nyereséghez képest.

Mivel későbbi sejtprogramozási tapasztalataink azt is megmutatták, hogy a teljes szomszédság, illetve valamennyi bitjére nincs szükség a bemeneti oldalon, a gyakorlatban is versenyképes alternatívát dolgoztunk ki memóriabázisú sejtekből álló sejtmezők tervezésére és építésére. A konkrét technológiai peremfeltételek és a felhasználás igényei szerinti sejtmező méret döntheti el, hogy memóriabázisú (mint az M1 és az X1 sejtprocesszor) vagy központosított átmenetfüggvényű mikroprogramozott jellegű (mint az SPP chip és az SP1 sejtprocesszor) konstrukciót válasszunk.

3. Hungarian cellprocessor project (magyar sejtprocessor projekt)

Ezen a néven vonult be a nemzetközi szakirodalomba, itthon sokáig szegedi sejtprocessor projekt néven volt ismert, mert a Kalmár László akadémikus, majd Gécseg Ferenc professzor vezetésével működő szegedi JATE Számítástudományi Tanszékén, illetve a mellette működő MTA Automataelméleti Kutató Csoportnál folyt a kutatás, majd ennek eredményei alapján a Mikrosejt Gmk-nál, utóbb a CELLWARE Kft.-nél folytatódott a fejlesztés és a kissorozatú gyártás. A munkák sokoldalú anyagi támogatást kaptak, egyebek mellett az OTKA és az OMFb forrásaiból pályázati alapon, a Mikroelektronikai Kormánybiztostól és az ipari minisztertől kiemelt fejlesztési alapokból.

Ezek a kutatás-fejlesztési munkák több mint 200 emberévnyi ráfordítást igényeltek, amelyek a szerteágazó munkák volumene miatt igen hatékony végrehajtást igényeltek. Különösen eleinte, kevés forrás birtokában, nagyobb mértékben hallgatói munkákra is támaszkodtunk (többtucat diplomamunka is készült a JATE-n, az ELTE-n, a GAMF-on, a BME-n), de ez később is folytatódott kisebb mértékben és eredményesen megvédett kisdoktori munkákkal egészült ki a JATE-n és a BME-n (már az intézmények felsorolása is mutatja a téma komplex jellegét).

Milyen felismerések képezték e kutatás-fejlesztési munkák alapját?

A kiindulópont természetesen a neumann-i sejtautomata-fogalom volt. Bemutattuk, hogyan egyszerűsítette ezt Codd, és hogyan próbált a Codd-ICRA project sejtautomata elvű számítógépet készíteni. Elemeztük, hogy igen kis hatékonyságú ez a megoldás, bár modellértékű a sejtautomata alapú hardverfejlesztés szükségességének felismeréséhez.

A szekvenciális (soros) működésű számítógépek korlátainak és a korai párhuzamos működésű hardverek, majd a vektorprocesszorok és a nagypárhuzamosságú sokprocesszoros rendszerek megismerése során kimutattuk az azonos processzorokból álló, homogén felépítésű rendszerek előnyeit és hátrányait, valamint a – viszonylag – nagyméretű processzoroknak a hardver elemi szintű alkatrészeinek szintjén történő nagyfokú kihasználatlanságát.

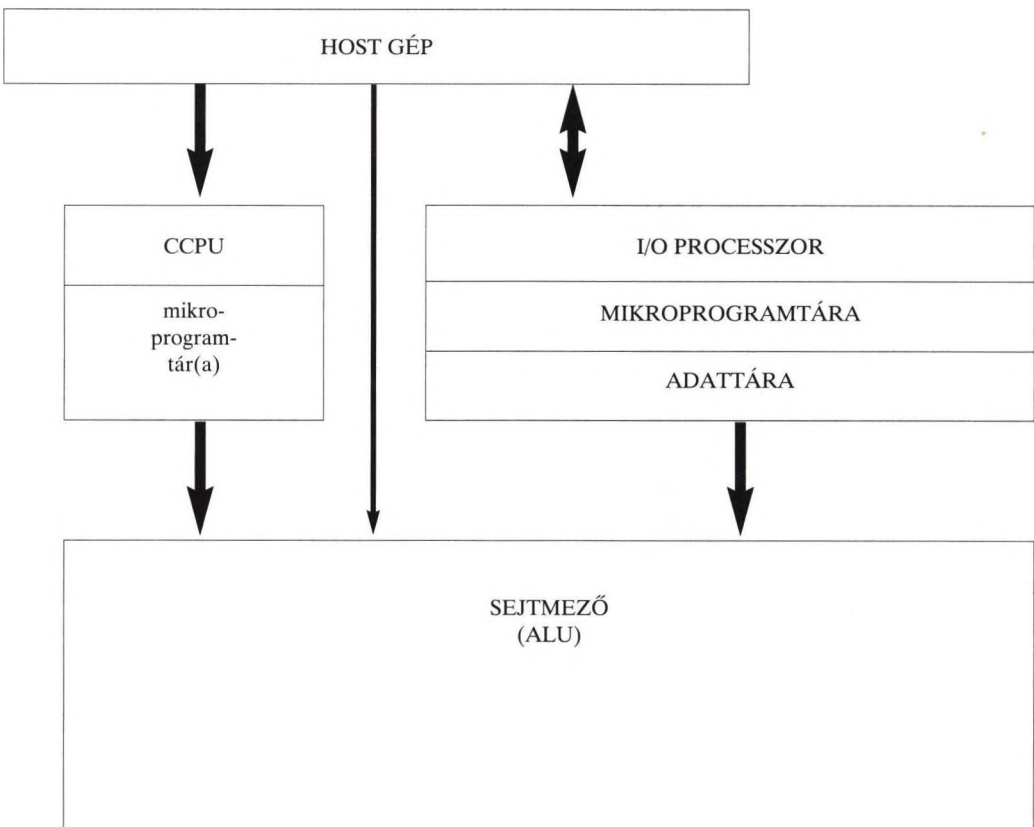
Toffoli sejtautomata emulátorai már tükrözik ezeket a felismeréseket, és viszonylag szűk feladatosztályon alkalmazható igen gyors, viszonylag egyszerű, memóriabázisú sejtet és e köré felépülő sejtautomata emulátorcsaládot épített és sejtprogramozott. Ez óriási lépés volt a sejtprocesszorok létrehozása felé.

Első közelítésben két jelentős hátrány azonban azonnal felismerhető: ez még csak emulátor, és a sejt méretei miatt nagy sejtprocessor aligha építhető belőle gazdaságosan; a második, még lényegesebb felismerés, ami egyben az előző probléma megoldása felé vezető utat is mutatja, pontosabban kikényszeríti ennek meg- vagy feltalálását az az, hogy Tof-

foli az igen redundáns **teljes** állapot-átmenetfüggvényt tárolja sejtjében. A kutatási-kísérleti szakaszban létrehoztuk a TRANSCELL átmenetfüggvény-optimalizáló programot, amely ezt konkrét állapot átmenetfüggvényeken szemléletesen bizonyította (amúgy igen plauzibilis volt a viszonylag sok bemenetű és kisszámú kimenetű függvény mint függvény, illetve a függvénynek megfelelő fa, illetve memóriatartalom redundanciája).

A tömörített (optimalizált) átmenetfüggvény kiértékelése többlethardvert igényelhet, ennek minimumra szorítása rendkívül fontos gyakorlatban is használható sejtprocesszor(ok) létrehozásához. Segítséget nyújt ehhez a munkák során kialakult sejtprogramozási gyakorlat is, amely megmutatta, hogy nincs szükség valamennyi szomszéd sejt összes állapotbitjére adott operátorok (például sejt mikroprogramokkal megvalósított aritmetikai műveletek, rendezés – szortolás –, képfeldolgozási műveletek) esetén, így az egyes sejtek bemenetére kevesebb bit kerül, kisebb lehet az átmenetfüggvény, és így az egyes sejtek is.

Ha nem emulátort akarunk építeni, hanem nagypárhuzamosságú sejtprocesszort, akkor valamennyi sejtet hardverben meg kell valósítani, ami drámaian fontossá teszi az egyedi sejt méret leszorítását a lehetséges minimumra úgy, hogy az egyes feladatok megoldásához – sejtprogramozásához – azért viszonylag ne túl sok sejt legyen szükséges.



7. ábra. Sejtprocesszor egyszerűsített sematikus rajza és működése

4. A megépített sejtprocesszorok

A sejtprocesszormodell

Az első elkészült rendszer sejtprocesszormodell volt, tekintve a rendelkezésre álló kisintegráltságú alkatrészeket (SSI) és az ebből adódó szekrényméretet, az aránylag egyszerű sejtprogramozást támogató szoftvert, a központosított átmenetfüggvénytárból adódó sebességveszteséget.

Egyébként azonban rendelkezett a későbbi sejtprocesszorok lényeges komponenseivel – központi vezérlő egységgel, input-output egységgel, 16×16 darab kétállapotú sejttel, össze volt kapcsolva egy anyagéppel (host), ahonnan a feladatokat és azok adatait kapta, és ahová az eredményt visszaadta.

A méretkorlátok miatt központosított függvénytárból kapták a kvázisejtek sorosan-párhuzamosan az átmenetfüggvényeket, így térben-időben inhomogén átmeneti függvények végrehajtása volt lehetséges (a részbeni párhuzamosságot azt biztosította, hogy az adott időpontban ugyanazt az átmenetfüggvényt végrehajtó sejtek ténylegesen párhuzamosan működtek).

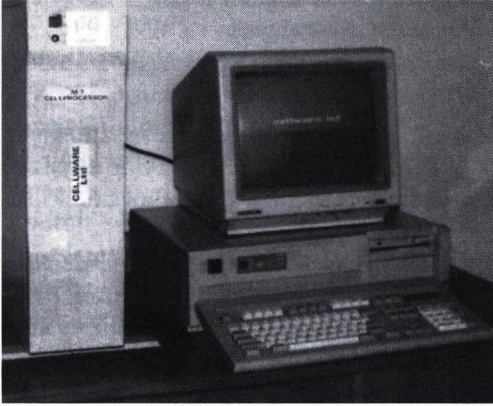
Kisszámú sejtprogrammal rendelkezett, ezek azonban szemléletesek voltak, aritmetikai alapműveleteken kívül, képfeldolgozó operátorokkal és jellegzetes sejtes átmenetfüggvényekkel, mint az előzőekben bemutatott Lindenmayer- (szumma moduló 2) függvénnyel és az életjátékfüggvénnyel. Így a sejtprocesszormodell az architektúra kipróbálására, bemutatási célra, sejtprogramozás megtanulására, gyakorlására, optimalizálásra igen alkalmas volt, körülbelül egy nagyságrendnyi gyorsítást is elért az anyagéphez képest. A rendszerarchitektúrát és a kétállapotú sejtet jómagam specifikáltam, a rendszer kivitelezésében részt vevő kollégák közül kiemelem a hardvertervezésben Tóth József, a szoftvermunkákban, tesztelésben Ghymes Balázs munkáját.

Az M1 sejtprocesszor

Az első, már tényleg sejtprocesszornak nevezhető rendszer közepes integráltságú (MSI) alkatrészekből készült, 16 állapotú (4 bites) sejtekből állt, amelyekből 256 darabot tartalmazott 16×16 elrendezésben.

Ezek a sejtek memóriabázisúak voltak, minden sejt a kezdeti betöltéskor megkapta az átmenetfüggvényeit, így a működés lényegében valódi párhuzamossággal történt (hardveroptimalizálási okok miatt a szomszédos sejtek bitjei nem feltétlenül egyszerre jutottak el a végrehajtó sejtbe, így tulajdonképpen egy billenési lépésen belül egy sorosan, időben változó függvénysorozat hajtódott végre). A rendszer központi vezérlőegységet, input-output processzort is tartalmazott, anyagépe egy korabeli IBM PC volt. Az egész rendszer már elfért egy nagy PC-toronyban (4 nagy nyomtatott áramkörü lapon a sejtmező és egy ötödiken a többi funkcionális egység), a tervezésnél már ilyen, viszonylag kis sejtmező mellett is lát-

szott az ismétlődő áramkörü részegységek (lapok és lapon belüli sok ismétlődő rész) hatalmas előnye az inhomogén rendszerekkel szemben.



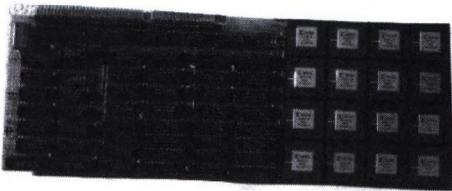
8. ábra. M1 sejtprocesszor

rutinkönyvtárként (ma: dll-ként) jelent meg. Tehát természetesen nem az egész alkalmazói program gyorsult fel ennyire, csak a sejtprocesszorral végrehajtható részek, vektor-, mátrixprocesszási feladatok végrehajtása, képfeldolgozási feladatok esetén ezek részaránya a végrehajtási időt (tehát nem a programsorokbeli részarányt) tekintve magas, így a teljes alkalmazás is jelentősen gyorsul.

A 16 állapotú sejtet és a rendszerarchitektúrát a sejtprogramozókkal és a hardveresekkel folytatott konzultációk, a tőlük kapott igények és lehetőségek figyelembevételével jómagam specifikáltam. A hardvertervezés Zsótér Antal eredményes munkája volt, amihez járult az építés és tesztelés során végzett értékes teljesítménye is. A sejtprogramozás pedig főleg a CELLAS sejt-szimulációs rendszerünkön már kipróbált sejtalgoritmusok átültetése volt Katona Endre (és) vezetésével különösen Palágyi Kálmán ért el kitűnő eredményeket a többi sejtprogramozó áldozatos munkája mellett.

Az X1 sejtprocesszor

Fő célunk, az igen nagyszámú, többtízezres, többmilliósejt sejtprocesszorok megvalósítása felé vezető úton saját tervezésű, ennek alapján legyártott nagyintegráltságú áramkör elkészítése szükséges, amelynek fajlagos tervezési ideje és így költsége is viszonylag alacsony, lehetőségeinket azonban még ennek ára is meghaladta (bár a később bemutatásra kerülő SP1 sejtprocesszorhoz egy szerény, egyszerű kivitelű saját chipet sikerült az X1 sejtprocesszort követően megtervezni és külföldön meg is építeni).



9. ábra. X1 sejtprocesszor

Az alkatrészpiacon ekkoriban jelentek meg az első PGA (programozható gate array, kaputömb) áramkörök. A hardver viszonylagos merevségét a szoftver különösen nagyfokú rugalmasságával ötvöző PGA és FPGA (field programmable gate arrays – ezek többször is átprogramozható PGA-k) áramkörök kísérleti berendezések gyors elkészítését, kissorozatú gyártás gyors beindítását teszik lehetővé.

Funkciójuk révén homogén felépítésűek, lényegében mátrixelrendezésben „őskapukat” (vö. őssejt) tartalmaznak, amelyekből programozás (betöltés) útján tetszőleges kapuáramkör alakítható ki, az összekötéseket – közvetlen vezetékek helyett – egy hatalmas kapcsolómező – amely természetesen szintén programozható és a program betölthető – biztosítja. A működési sebesség megközelíti a szokásos hardversebességet, a kihasználtság természetesen igen alacsony tranzisztorszint felől nézve: 1-2 nagyságrenddel több kapu lehetne képezhető a programozhatóság miatt felduzzasztott hardverstruktúra miatt.

Mivel a megvalósításnál az egyes „őskapuk” programozását egy kis memória biztosítja, egy sejtprocesszorra nagyon emlékeztető struktúra áll előtünk! És ezek az áramkörök sorozatban készülnek, beszerezhetők. Persze egy kis szerencse is kellett, hogy a Xilinx cég legelső FPGA-ja pont olyan legyen, hogy körülbelül tízkapunyi kapuáramkör megvalósítására elegendő helyen egy egész sejtet ki tudjunk alakítani. Szándékosan használtam ezt a kifejezést, mert természetesen nem a szabványos módon a sejt áramkörét vittük be, hanem az „őssejt” memóriáját használtuk fel átmenetfüggvény-tároló memóriának!

Így végül 64 darab kétállapotú sejtet tudtunk kialakítani egy darab Xilinx PGA-n belül, ami egy 1024 sejtes közepes méretű PC-kártya létrehozását tette lehetővé. Ennek programozása korlátozottabb és nehezekebb, mint az M1 sejtprocesszoré, kevesebb feladat valamivel lassabban oldható meg vele, de mérete és ára sokkal kisebb, illetve alacsonyabb természetesen. Tehát nem véletlen, hogy a témaválasztás (nagyhomogenitású, sok kis processzorból álló rendszer) a megvalósítást ilyen módon is támogatta, mondhatnók a szerencse a jó irányba indulókat támogatja.

Az X1 sejtprocesszor megépítésénél már fel tudtuk használni összes előző tapasztalatainkat, a legnehezebb feladat volt az 1024 darabos kétállapotú sejtmezőbe „cipőkanállal” beszorítani éles feladatok sejtprogramjait, így itt Katona Endre, Köles Péter és Palágyi Kálmán kiemelkedő teljesítményét szeretném hangsúlyozni a hardver és a rendszerszoftver tervezőinek és építőinek eredményeit nem lebecsülve.

Hangsúlyozom, hogy egy sejtprocesszor komplex rendszer a tranzisztor- vagy kapuszinttől kezdődően az alkalmazói könyvtárakig, vagy éppen egy-egy teljes alkalmazás megoldásával bezárulóan.

Az SP1 sejtprocesszor

Az integráltsági fok növekedésével előtérbe kerülne a térben-időben inhomogén átmenetfüggvényű sejtprocesszorok, különösen a központosított átmenetfüggvénytárral rendelkező típusaik, mivel így lehet a legtöbb sejtet egy chipre integrálni, kétségtelenül végrehajtási idővesztés árán, ugyanis az átmenetfüggvény-végrehajtás így soros-párhuzamossá válik, mint ezt részleteztük már előzőleg.

Az SP1 sejtprocesszor létrehozását az tette lehetővé, hogy bár pénzügyi lehetőségeink

továbbra is szerények maradtak (a téma igényeihez képest), de angliai együttműködő partnerünk segítségével már módunk volt egy egyszerű, kb. 9000 kaput tartalmazó gate-array (kaputömb) áramkör megtervezésére és kísérleti gyártására a világ egyik legnagyobb custom design (testreszabott) chiptervező és -gyártó cégénél.



10. ábra. SP1 sejtprocesszor

A sejtprocesszor modellnél megismert egyszerű kétállapotú sejtet valósítottuk meg, amelyből ezzel az eljárással 64 darabot tudunk elhelyezni egy chipen (lapkán). Emellett a megoldás mellett szólt a sejt kis mérete, valamint az is, hogy igen alaposan ismertük és teszteltük már a sejtprocesszor modellben – ugyanis a külföldön végzendő tervezésre és kipróbálásra fordítható pénz és így az időtartam is igen korlátozott volt, arról nem is beszélve, hogy elsőre hibátlanoknak kellett lennie az áramkörünknek, mert javítási fordulóra a fenti okok miatt nem lett volna lehetőség.

Mindamellettt fontos új áramköri részegységet kellett megterveznünk és beépítenünk a lapkára a növekedő rendszerméret miatt. Az előző sejtprocesszorokban a viszonylag kis sejtmezőméret miatt a sejtek egyszerűen össze voltak huzalozva, így jött létre az összefüggő sejtmező. Előzőleg csak a sejtmező szélén voltak széláramkörök, amik az anyagéppel történő adatforgalmat valósították meg. Most azonban a sejtmező már 8×8 -as elrendezésű, 64 sejttes darabkákból áll, ezek szélén olyan sok csatlakozásra van szükség, ami közvetlen összeköttetésekkel az adott technológia korlátai (a lapka maximális lábszáma) miatt nem volt megvalósítható. Így az egyes lapkákra olyan széláramköröket kellett felhelyezni, amelyek a szomszédos lapkákkal soros-párhuzamos végrehajtással (idővesztés árán) emulálták a közvetlen összeköttetést. Jellemzően ez az eljárás beleillett a rendszer térben-időben inhomogén sejtprogramozásához szükséges végrehajtási mechanizmusba (mikroprogramozott vezérlés), bár célszerű volt külön $I \times O$ vezérlőprocesszorra bízni a speciális mikroparancsok miatt, amik természetesen biztosították a sejtmező szélén lévő lapkákon található szélsejteknek az anyagéppel való kapcsolatát is. Ez a megoldás két fontos lehetőséget is biztosított – az egyik Tóth József javaslata alapján további kis kiegészítéssel „ugrató áramkörök” hozzáépítésével lehetővé tette adatok két billenési lépés (átmenetfüggvény-végrehajtás) között nagy sebességgel távolra küldését is (ez felfogható korlátozottan – sejtmezőcsoportok közötti – változtatható szomszédságú sejtmező létrehozásának is), ami a sejtprogramozás egyik legnagyobb korlátját, a teljes lokalitást oldja fel. Igazán nagy jelentősége ennél a kis sejtmeződarabokat tartalmazó rendszernél még nincs, de gondolva létrehozható nagyobb rendszerekre, áttörést jelent a sejtprogramozás hatékonyságának javításában. (A megoldás funkcionális értelemben, távolról hasonlít a mai P2P adattovábbításhoz az interneten.)

A másik lehetőséget jómagam építettem rá ennek az ötletnek a „tetejére” – mivel nem igazán tetszett, hogy további idővesztést okoz a teljes sejtmező-összeköttetés létrehozása (ugye ez annyiban sorosan történik, hogy biztosítjuk az összeköttetést – eljuttatjuk minden lapkához a szomszéd sejtek értékeit, majd végrehajtjuk az átmenetfüggvényt, és ez a kétlépéses ciklus ismétlődik), megkíséreltem átlapolni időben ezt a két folyamatot. Több

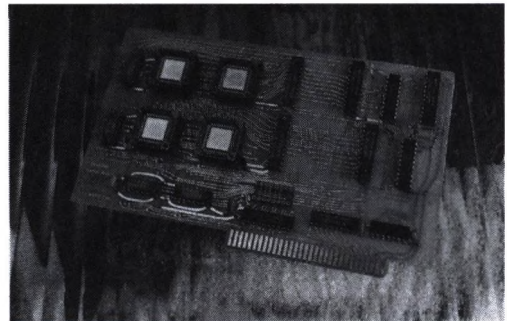
lehetőség közül végül az vált be, hogy a sejtmezőt kétrétegűvé tettem, így az egyik rétegben input-output műveletek, a másik rétegben átmenetfüggvény-végrehajtás zajlik egyidejűleg, majd a két rétegben a két folyamat felcserélődik. Ez természetesen kétszer annyi ideig tart, de nem kell a sejteket, illetve a sejtmezőt megkettőzni, hanem csak az egyes sejtek állapot- és szomszédállapot-tárolóit, így az egész rendszer hatékonysága kb. 40%-kal javul. A mai PC-kben utasításvégrehajtási szinten a multi-thread konstrukciókban jelennek meg hasonló gondolatok.

Így jött létre a $(2 \times)8 \times 8$ sejtet tartalmazó SPP chip, amelyből egyetlen nagyméretű kártyára 8×8 darabot felhelyezve $(2 \times)16K$ -s sejtmezőt építettünk, központi vezérlővel és input-output processzorral együtt alkotva teljes rendszert, az SP1 sejtprocesszort, amelyet egy IBM PC-hez mint hostgéphez (anyaggéphez) kapcsolva helyeztünk üzembe.

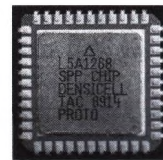
Kisméretű oktatókártya SPP chipekből

Mivel igény mutatkozott a nem teljesen egyszerű sejtprogramozás, illetve mikroprogramozás oktatására, a viszonylag lassú szimulátorok mellett célszerű volt hardveroktató – gyakorló – sejtprocesszort is tervezni és építeni. A különböző adódó lehetőségek közül kompromisszumként négy darab SPP chipből álló sejtmezőt tartalmazó, minimális kiegészítő hardverrel ellátott kisméretű, olcsó kártya mellett döntöttünk.

Fontos szempont volt, hogy a sejtmező ne legyen nagyon kicsi, a szimulátoroknál legyen lényegesen gyorsabb, de nem feltétlenül kell csúcsebbségen futnia. Így a hardver a PC órajelét használva, a PC-n a sejtprocesszornak a végrehajtási idejében nem nagy időigényű részeit szimulálva olyan eszközhöz jutunk, amely a „nagy” sejtprocesszor lényeges funkcióit a PC-nél bő 2 nagyságrenddel gyorsabban hajtja végre, azonos vezérlőparancsok hatására, illetve a sejtprogramok ugyanúgy hívhatók meg a PC-s főprogramból, mint az SP1-ben.



11. ábra. SPP oktatókártya



12. ábra. SPP chip

Sejtprocesszor-alkalmazások

A sejtprocesszorok alkalmazásai a hardver nagyfokú párhuzamos működésén és így nagy számítási teljesítményén alapulnak, korlátot jelent a vektorprocesszorszerű működés az igen kisméretű „nanoprocesszorokon”, sejteken. Az ellentmondást részben feloldja a

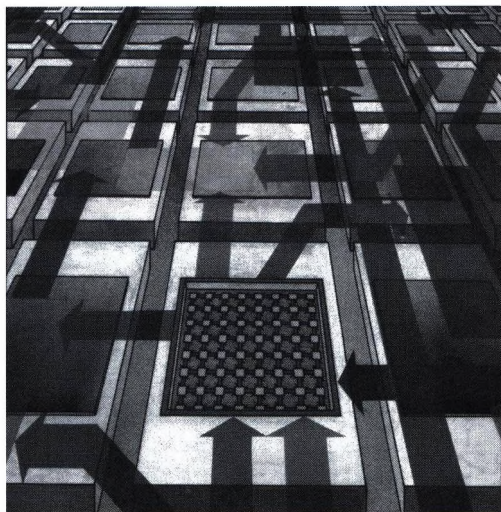
hardver térben-időben-szomszédságban inhomogén volta és a fentiekkel összhangban kialakított sejtprogramozási metodológia.

Rendkívül természetesen adódnak a vektorokon, mátrixokon, homogén felépítésű adattömbökön végezhető műveletek, a sejtter méretétől függően „átfolyatva” a sejtteren, vagy igazán nagy sejtmező esetén az adatokat a sejtterbe betöltve kerülnek feldolgozásra ezek az adatok. Ezek az általánosan használható műveletek alkalmazások széles osztályait nyitják meg.

Ki szeretném emelni a képfeldolgozást, mivel ennek ma szerteágazó területei és felhasználásai vannak az orvosi diagnosztikától (pl. a CT vagy computertomográf, amely lényegében egy háromdimenziós röntgenberendezés, kétdimenziós metszeteket készít, és ezekből kell kiszámolni a háromdimenziós képet) az arcfelismerésig, általánosabban a képfelismerésig, amelyek robotok működését, termelési folyamatokat támogathatnak.

Sejtprocesszorral leginkább képeken végezhető transzformációk végrehajtása jön szóba, vagy kicsit általánosabban kép-előfeldolgozási feladatok, amelyek a transzformációkon kívül adatkivonást – a kép jellemző tulajdonságainak megállapítását, képrészletek kiemelését stb. – végeznek.

Sejtprocesszor projektünk során végeztünk feldolgozásokat a Paksi Atomerőmű zajdiagnosztikai feladataihoz (itt az erőmű belsejében lévő zajforrásokot kellett elemezni, olyan helyekről szerezve információt, ahová a sugárzás miatt ember nem léphet be, de a különböző frekvenciatartományokból – nem csak a hallható hangtartományból – érkező jelsorozatok fontos információt hordoznak az erőmű biztonságos működéséről. A folyamatok gyorsak, így ahhoz, hogy be tudjunk avatkozni idejében,



13. ábra. Sejtprocesszor architektúra

gyors kiértékelésre van szükség, ezért merült fel sejtprocesszor használata.

Egyik fő alkalmazásunk képfeldolgozás alapú térképészeti felhasználás volt a Térképészeti Intézet részére. Itt úgynevezett szintvonalas térképekből (amelyeken az azonos magasságú pontokat görbék kötik össze, ezek a pontok többnyire mérési adatokból származnak, és ez a térképfajta a térképész számára sokatmondó ugyan, azonban a térképfelhasználók számára a domborzati térkép az érthetőbb, szemléletesebb) kellett háromdimenziós domborzati térképet számítani ki.

Nagyon nagy egyszerűsítéssel rengeteg interpolációt kellett számítani, a feladat ennél sokkal összetettebb, a sejtprocesszorral való számolás esetén pedig külön megoldandó feladat volt, hogy a síkságon és a hegyvidéken ugyanaz az algoritmus helyesen működjön. Az ötletek és a kivitelezés Katona Endre és munkatársai érdeme, jómagam a munka egy holtpontján egy fokozatos közelítés ötlettel/javaslattal tettem egy keveset hozzá ehhez az eredményhez (amit leleményesen nagyítás-kicsinyítés támogatására is kihasználtak a kollégák). Az elért gyorsítás a megrendelt 4–8-szoros helyett körülbelül százszorosra sikerült, ami reálissá tette Magyarország teljes nagyfelbontású térképének a tervezettnél évekkel hama-

rabb történő előállítását. Ehhez azért még arra is szükség volt, hogy mivel az együttműködés olyan jól sikerült, hogy a munka előfeldolgozás részében is kaptunk feladatot (a papíron lévő szintvonalas térkép digitalizálásában), a nyers digitalizált kép hibajavítását. A hatalmas adatmerek miatt adódott itt is sejtprocesszálásra alkalmas részfeladat, ha nem is annyira központi rész, mint a domborzatszámításnál (legalábbis a futási/végrehajtási időt tekintve). Ez a feladat ugyanis nem volt teljes egészében algoritmizálható, hanem interaktív módon, emberi közreműködéssel történt. Itt az segít sokat, ha az embernek nem kell sokat várni (és ebben is segített a sejtprocesszor), valamint, ha a teljes algoritmizálás/automatikus végrehajtás nem is megy, azért nagy segítség, ha a számítási rendszer alternatívákat kínál fel, és ezekből már „csak” választani kell az interaktívan dolgozó térképésznek (és itt mind a sebesség, mind néhány új sejtprogram is javította a teljesítményt).

Végül egy tréfás, de mutatós „alkalmazást” említek meg – valódi időben – real-time – készülő egyszerű karikatúrák készítését. A kamerába néző arcon néhány egyszerű geometriai műveletet elvégezve igen hatásos képekhez lehet eljutni, kiállításokon kiváló nézőcsalogató volt (egy nemzetközi innovációs kiállításon közönségdíjat is kapott, ahol maga a sejtprocesszor a zsűri osztott I. díját kapta). Budapesttől Szombathelyig, Berlinton Hannoverig és Párizstól Edmontonig mindenütt örömet szerzett az ötéves gyermekektől a 86(!) éves néniig, sokan nézték meg magukat szívesen a különböző pózokban és kértek/vettek a papírra nyomtatott változatokból (amelyekben a mai scannerekben, nyomtatókban alkalmazott interpolációs jellegű megoldással finomítottuk a képet). A legmonumentálisabb – de rövid – bemutatónk a Vörösmarty téren volt egy óriástáblán (általában monitoron, néha kivetítőn mutattuk be a karikatúrákat).

Nemzetközi együttműködések, közös eredmények

A sejtprocesszorprojekt eredményeit publikáltuk (több mint 100 cikkben), részt vettünk a párhuzamos számítási rendszerekkel foglalkozó főbb európai konferenciákon, programbizottsági tagként is közreműködtem azokon, illetve alapító és társelnökként a PARCELLA berlini konferenciasorozaton, amelynek kiadványai az Akademie Verlag-nál, a North-Holland-nál és a Springer Verlag-nál jelentek meg, így igazi kelet-nyugati, nemzetközi sejtautomata/sejtprocesszor témakörű konferenciát sikerült létrehozni (1982–1996). (Itthon a Neumann Társaságon belül pedig megalapítottam a Párhuzamos Számítási Rendszerek Szakcsoportot, ahol előadások és kiadványsorozat foglalkozott a témakörrel.)

Több éven át működtünk együtt MTA együttműködési szerződés keretében az NDK Tudományos Akadémiája ZKI intézetével. A PARCELLA szervezésében is oroszlánszerpet vállaltak (az egyesítésig), a társelnök és társszerkesztő dr. Gottfried Wolf volt.

Több mint tíz éven át három közös projektünk volt az MTA és a DFG (Deutsche Forschungsgemeinschaft) egyezménye alapján a Branschweigi Műgyetemmel, ahol Roland Vollmar professzor irányításával folytak sejtautomata-elméleti és sejtstimulációs vizsgálatok. Munkáink jól kiegészítették egymást, illetve a témák megvitatása mindig adott új ötleteket is. A közös munkáról, illetve a kapcsolódó saját munkákról több kötetet adtunk ki, amelyek – részben másutt is megjelent – cikkeket, illetve munkabeszámolókat tartalmaztak.

Irodalom

von Neumann, J.: Theory of automata: construction, reproduction, homogeneity. Part II of "Self-Reproducing Automata" edited by A. W. Burks. University of Illinois Press, Urbana, Illinois, 1966

Leben und Werk von John von Neumann (editors: T. Legendi, T. Szentiványi): Bibliographisches Institut Mannheim/Wien/Zürich 1979

E. F. Codd: Cellular Automata. ACM Monograph Series, Academic Press, Inc. New York and London 1968

Drommerné Takács Viola (szerk.): Sejtautomaták. Gondolat 1978 (18 cikk és részletes irodalomjegyzék)

T. Toffoli: On the Large-Scale Implementation of Cellular Spaces by Means of Integrated-Circuit Arrays. CNR Istituto per le Applicazioni del calcolo 1972

T. Toffoli, N. Margolus: Cellular Automata Machines – A new environment for modeling. The MIT Press Cambridge, Massachusetts; London England 1987

Legendi T. és munkatársai tíz cikke sejtprocesszorok tervezéséről, programozásáról és alkalmazásáról. Magyar Elektronika III. évfolyam 6. szám 1986

Katona E.: Sejtalgoritmusok (a projekt öt résztvevőjének, köztük a szerzőnek az algoritmusai alapján) az NJSZT és az MTA SZTAKI kiadásában a Párhuzamos Számítási Rendszerek Szakcsoport 1981/5 kiadványa

Vollmar, R.: Algorithmen in Zellularautomaten. B. G. Teubner, Stuttgart, 1979

Vollmar, R. and Legendi T.: Cellprocessors and cellgorithms – Reports on a common research project. Volume I. 1981. Volume II. 1985. Technische Universität Braunschweig

Parcella'84 (edited by) W. Händler, T. Legendi, G. Wolf Akademie Verlag Berlin (benne a projekt résztvevőinek 9 cikkével)

Parcella'88 (editors) G. Wolf, T. Legendi, U. Schendel Springer Verlag Fourth International Workshop on Parallel Processing by Cellular Automata and Arrays Berlin, October 1988 Proceedings

Parcella'90 Proceedings edited by G. Wolf, T. Legendi, U. Schendel Akademie – Verlag Berlin (itt a projectet 3 előadással és 4 működő sejtprocesszorral képviseltük egy kis párhuzamos kiállítás keretében)

Tartalom

Előszó	3
Az első számítógép alkalmazásával megjelenő numerikus problémák (Obádovics J. Gyula)	5
Bevezetés	5
1. Mátrixok	8
2. Determinánsok	12
3. A négyzetes mátrix inverze	17
4. Lineáris egyenletrendszerek	22
5. Gram–Schmidt-féle ortogonalizálási eljárás	32
6. A gyengén meghatározott lineáris egyenletrendszerek megoldásáról	36
Irodalom	40
Rejtett paraméterek, avagy a statisztikus fizika és a kvantummechanika matematikai megalapozása (Dávid Gyula)	41
1. Neumann János és a fizika	41
2. Lökéshullámok és az atombomba	45
3. Az ergodikus tétel	47
4. A kvantummechanika matematikai alapjai	63
5. Rejtett paraméterek	87
Neumann János hozzájárulása a játékelmélethez és a matematikai közgazdaságtanhoz (Forgó Ferenc–Zalai Ernő)	99
1. A kétszemélyes, zérusösszegű játékok egyensúlya létezésének bizonyítása	101
2. Az egyensúlyi gazdasági növekedés Neumann-féle modellje	107
3. A Neumann-modell és az általános gazdasági egyensúly korai modelljei	117
4. Neumann hozzájárulása a játékelmélet további fejlődéséhez	128
Felhasznált irodalom	135
Neumann János jelentősége a meteorológiában (Götz Gusztáv)	138
1. Az időjárás-előrejelzések készítésének kezdetei	138
2. Az időjárás objektív előrejelzésének első kudarca	139
3. A kudarc okainak elhárítása	142
4. Sikeres előrejelzési kísérletek az ENIAC-en	143
5. Determinisztikus valószínűségi időjárás-előrejelzések	150
6. Neumann gondolatai a természeti környezet megóvásáról	154
7. Epilógus	156

Gondolatok Neumann János <i>First Draft of a Report on the EDVAC</i> című, 1945 júniusában megjelent tanulmányáról (Kovács Győző–Szelezsán János)	157
1. Út a „First Draft...”-ig. A Neumann János előtti informatikatörténelem	157
2. A First Draft	167
3. Egy utasításrendszer, egy „magas szintű” nyelv	199
4. Epilógus	203
Neumann mint a sejtautomata megalkotója, elméleti és gyakorlati követői (Legendi Tamás)	207
1. Neumann elméleti követői	211
2. Neumann sejtautomatájának gyakorlati követői (Sejtautomata számítógép, emulátor, sejtprocesszor létrehozása és alkalmazása)	217
3. Hungarian cellprocesszor project (magyar sejtprocesszor projekt)	221
4. A megépített sejtprocesszorok	223
Irodalom	230
Tartalom	231

Nemzeti Tankönyvkiadó Rt.
 A kiadásért felel: Pálfi József vezérigazgató
 Raktári szám: 53375
 Felelős szerkesztő: Dr. Koreczné Kazinczi Ilona
 Műszaki igazgató: Babicsné Vasvári Etelka
 Műszaki szerkesztő: Wéber Andrea
 Grafikai szerkesztő: Megyeriné Kovács Katalin
 Terjedelem: 20,74 (A/5) ív
 Első kiadás, 2003
 Formakészítés: Nemzeti Tankönyvkiadó Rt., Stúdió
 Nyomtatás és kötés a Szekszárdi Nyomda Kft.-ben készült
 Felelős vezető: Vadász József
 03.0913

Ki volt
igazából **Neumann**
János?

**A könyvben szereplő
tudományterületek és szerzők**

Matematika

Dr. Obádovics J. Gyula a matematika tudomány kandidátusa

Számítástechnika

Kovács Győző

Dr. Szelezsán János a matematika tudomány kandidátusa

Automaták elmélete

Dr. Legendi Tamás

Fizika

Dr. Dávid Gyula

Meteorológia

Dr. Götz Gusztáv a földrajz tudomány doktora

Közgazdaságtan

Dr. Forgó Ferenc a közgazdaság-tudomány kandidátusa

Dr. Zalai Ernő akadémikus, a közgazdaság-tudomány doktora

