

CELLULÁRIS AUTOMATA (SEJTAUTOMATA)

Számítások randevú típusú szinkronizálással

RÖVID ÖSSZEFOGLALÁS

A sejtautomata leggyakoribb formája: egy négyzetrácsban (a *sejttérben*) helyezkedik el, a négyzetrácsok által közrefogott cellákat sejteknek nevezzük. A sejteknek különféle állapotaik lehetnek (véges sokféle). Ahogy az idő telik, a cellák változtatják állapotukat, általában saját és más sejtek, például néhány szomszédjuk előző időpillanatbeli állapotától függően.

Burian Sándor , AWXYHE

Szoftverfejlesztés párhuzamos architektúrákra

2020-2021, első félév

Óbudai Egyetem, Neumann János Informatikai kar

Tartalom

Bevezetés, a probléma bemutatása	2
Az élet játéka	2
Párhuzamosítás kérdése	4
Megvalósítás	5
Egy szál szimuláció megvalósítás	6
Mérési eredmények	8
Több szál szimuláció megvalósítás	11
A szinkronizáció kérdése	13
Mérési eredmények ⁵	13
Következtetés:	15
Forrásjegyzet	16

Bevezetés, a probléma bemutatása

“A feladatteret cellákra osztjuk. Minden cella véges állapotok közül pontosan egyet vesz fel. A cellákra a szomszédjai valamilyen hatással vannak bizonyos szabály szerint, és minden cella egy „generációs” szabállyal is rendelkezik, amely szimultán fut. A szabályok újra és újra alkalmazásra kerülnek minden generációs lépésben, így a cellák fejlődnek, vagy megváltoztatják állapotukat generációról generációra. A legismertebb celluláris automata John Horton Conway cambridge-i matematikus „Élet játéka” (“Game of Life”).”¹

Az élet játéka

A táblás játék elméletben végtelen méretű kétdimenziós tömbökkel. Minden tömb cellákat tartalmaz. Tehát minden cellának 8 szomszédja van.

Pl:

		↖	↗
tömb	...	Cella	Cella	Cella	...	
tömb	...	Cella	Cella	Cella	...	
tömb	...	Cella	Cella	cella	...	
		↘	↙

A cellákban amikben egy-egy organizmus lehet. Kezdetben néhány cella foglalt, a következő szabályok szerint, amiket Conway fektetett le¹:

1. Minden organizmus aminek 2-3 szomszédja van életben marad a következő generációra.
2. Minden olyan organizmus amelynek 4+ szomszédja van meghal a túlnépesedés miatt.

¹ https://elearning.uni-obuda.hu/main/pluginfile.php/274206/mod_resource/content/1/P%C3%A1rhuzamos_programoz%C3%A1s_javasolt_feladatok.pdf hozzáférés dátuma: 2020 szeptember 28

3. Mindegyik olyan organizmus amelyiknek 2-nél kevesebb szomszédja van, tehát maximum egy, meghal az izoláltság miatt.
4. Minden üres cellában amelynek pontosan három nem üres szomszédja van egy új organizmus születik.

A példa mögé több mindent is beleláthatunk, a cellák tartalmi könnyen tovább gondolhatóak nyulakra és rókákra egy szigeten, vagy három dimenziós tömbökkel halakra és cápákra, az alábbiak szerint ¹:

Halak:

1. Ha van üres szomszédos cella odaúszik, ha több cella is üres akkor véletlenül választ egyet
2. Ha nincs üres szomszédos cella akkor helyben marad
3. Ha mozog és eléri egy természetes értéként megadott nemzési idejét akkor egy újabb hal kerül a szomszédos, üresen maradt cellába
4. Adott idő után a hal elpusztul

Cápák:

1. Ha szomszédos cellában hal van, akkor oda lép és megeszi, ha több szomszédos cellában is hal van akkor véletlen szerűen választ egyet.
2. Ha nincs szomszédos cellában hal akkor egy szomszédos üres cellába mozog, mint egy hal, ha több szomszédos cella is üres akkor véletlenül választva.
3. Ha eléri a nemzési idejét akkor akár egy hal új cápa kerül egy megüresedő cellába.
4. Ha a cápa adott generáción keresztül nem eszik akkor elpusztul.

Hasonlóan mögé lehet látni folyadék vagy gáz dinamikai folyamatokat, biológiai fejlődést, légáramlás vagy eróziós folyamatok modellezését.

Párhuzamosítás kérdése

Mivel egy nagy területen szükséges sokszor ugyanazt a műveletet elvégeznünk *(történetesen, hogy a különböző szomszédjaiban mik az állapotok egy-egy cellára nézve)* ezért a párhuzamosítás használata egyértelműen hasznos. Ugyanakkor az is egyértelmű, hogy időnként szinkronizációra van szükség. Elképzelésem szerint a teret felosztom nagyjából egyenlő részekre, annyira ahány magra optimalizálva számítom a feladatot². Innentől csupán az egymással szomszédos területeket számoló folyamatok kell megvárják egymást, és csak azok kell szinkronizálódjanak, így, globális szinkronizációra nincs is igazán szükség, hiszen a terület két teljesen eltérő sarkában számolt értéknek nem sok értelme van, hogy egymásra várjon, mert nem sok közük van egymáshoz, nem befolyásolja egyik a másikat, a szinkronizálás pedig alapvetően egy költséges művelet [2], így a minél kevesebb szinkronizálás a praktikusabb a gyorsaság szempontjából.

“Az egyes folyamatok csak arra várnak, hogy az alapgráfbeli összes szomszédjuk *helyi_szint*-je legyen legalább k . Így ezt az összes folyamat maga felderítheti a *helyi_szint* minden egyes növekedése után a szomszédba vezető éleken küldött üzenettel. [...] időbonyolultságának felső határa:

$$O(n \log n(1 + d))$$

A kommunikációs bonyolultság azonban rosszabb a mindenegyres szinten használt szinkronizációs üzenetek miatt. Ez most $O(|E| \log n)$ ”[3]

Ezek persze csak a várt eredmények. A szükséges tárhely előre nem leszögezhető, hiszen a feladat elviekben végtelen tereket használ, így végtelen tárhelyre is volna szükség, ami persze jelenleg nem lehetséges. Előre tervezetten vagy a felhasználó által beadható, vagy a teljes memóriát hkihasználó tárhelyre lehetne felkészülni mint kényelmes-praktikus számítási modell.

² Ez akár dinamikusan is lekérhető, így a program konkrét processzortípustól függetlenül is optimalizált lehet [1]

Megvalósítás

A beadandómban csak a 2D szimulációt fogom figyelembe venni, halak-cápák felosztást tekintve. Az egy és a több szálas megvalósítás között sok közös vagy hasonló elem, függvény, osztály megtalálható, ami egyfelől az összevethetőséget támogatja, másfelől az újrahasznosítást elősegíti.

A program megvalósításához szükség lehet a főprogram osztályán kívül még két osztályra. Az egyik a szimulációs tér mátrix absztrakciójának elemeit valósítja meg, azaz, hogy az adott mátrix elem hal, cápa vagy víz-e, és ennek megfelelően a szükséges segédfüggvények, segédváltozókat állítja be. A másik a szimulációs tér kezelésére egységéldobjektum, ami Descartesi koordináta szerűen tudja kezelni a matrix elemeit, hogy azokra könnyebb legyen hivatkozni, például parameter átadás során.

A méréseket két processzoron is elvégeztem, melyeknek paraméterei az alábbiak:

	Intel Core i7-4510U	AMD Ryzen 3 3200U
Socket Type	FCBGA1168	FP5
CPU Class	Laptop	Laptop
Clockspeed	2.0 GHz	2.6 GHz
Turbo Speed	Up to 3.1 GHz	Up to 3.5 GHz
# of Physical Cores	2 (Threads: 4)	2 (Threads: 4)
Max TDP	15W	15W
Yearly Running Cost	\$2.74	\$2.74
First Seen on Chart	Q2 2014	Q2 2019
# of Samples	3217	546
Cross-Platform Rating	5482	6494
Single Thread Rating	1557	1927
CPU Mark	2602	4101

Mérésekhez használt processzorok adatai [4]

Egy szál szimuláció megvalósítás

Alapvetően három fő funkció van implementálva a programba:

- 1) Futtathatjuk úgy, hogy a konzolra kirajzolja az aktuális futási állapotot, ez minden lény módosításánál frissül, azaz, nem minden ciklus után, hanem minden változás után rajzolja ki az állapotokat. Ez nyilván sokkal lassabb, de látványosabb szimulációs módszer.
- 2) Futtathatjuk úgy, hogy a képernyőre, a konzolra naplózza ki az eseményeket, ez a mérési módszer az előzőnél valamivel gyorsabb lesz, és a naplózás alapján elég pontosan követhetőek így is az események.
- 3) Végül szimulálhatunk úgy is, hogy statsztikai adatokat szeretnénk látni, azaz fájlba mentse³ a mérési eredményeket a kiinduló és a végső egyedszámokkal a futási sebességgel együtt. Ezt megtehetjük a Conway féle szabályok alkalmazásával és azok elhagyásával is. Utóbbi esetben csak a kiinduló populáción múlik, hogy mi lesz a végeredmény, hiszen nem generálódnak újelemek a szabad helyekre és nem halnak meg izoláltság miatt sem. A méréseket az utóbbi módszerrel végeztem.

Az implementálás szempontjából kulcskérdés, hogy szimulációs tér mátrixának frissítését hogyan oldjuk meg. Ez mind a tárhely mind a processzorkihasználtság szempontjából fontos kérdés, valamint mivel nem szeretnénk, csak nagyon kevés módosítást elvégezni a többszálúsításhoz, ahogy méréseink pontosabbak legyenek, ezért a többszálúsított programunk esetére is fel kell készülnünk, amennyire csak lehetséges.

Két irányba indulhatunk el. Egyfelől a szimulációs mátrixban minden egyes elem objektumánál külön-külön is eltárolhatjuk, hogy adott ciklusban lekezeltek-e már, vagy sem, ekkor viszont megnő a futási idő, mivel minden egyes alkalommal mikor az éppen vizsgált objektum (hal/cápa) körül levő többi cella objektumát (víz/hal/cápa) ellenőrizzük, nem csak azt kell figyelembe vennünk, hogy élőlény-e vagy sem, hanem, hogy ebben a fordulóban lépett-e vagy sem. Pl: ha van egy cápa ami egy egyel lentebbi sorba lép, akkor a ciklus később újra megvizsgálná, és újra léptetné, növelné a vemhessége idejét, ha tudná etetné, csökkentené a hátralevő életidejét stb. Ugyanígy egy hal esetében is. Más művelet esetén, pl ha azt vizsgáljuk, hogy van-e ehető hal a környéken, vagy mi a szabad pozíció, vagy hasonló akkor nem okoz gondot, hiszen nem befolyásolja, ahogy az a hal már lépett, vagy lépni fog-e még, mert úgymond, "a jelenben kerül elfogyasztásra" a cápa által. Tehát, a futási költség az összesen annyival növekszik, hogy amikor megvizsgáljuk, hogy az adott elem amit vizsgálunk az víz-e vagy sem, amennyiben nem víz úgy megvizsgáljuk, hogy lépett-e már vagy sem. Ugyanakkor minden módosítás(léptetés, szülés, evés, stb) után egy plusz művelet, hogy be kell állítani az aktuális ciklus ciklusszámlálóját mint utolsó módosítási ciklus jelölőt, de mivel ez vagylagos, ezért ez csak egy művelet pluszba minden ciklus esetén. A többi műveletbe ez a lépés nem kell

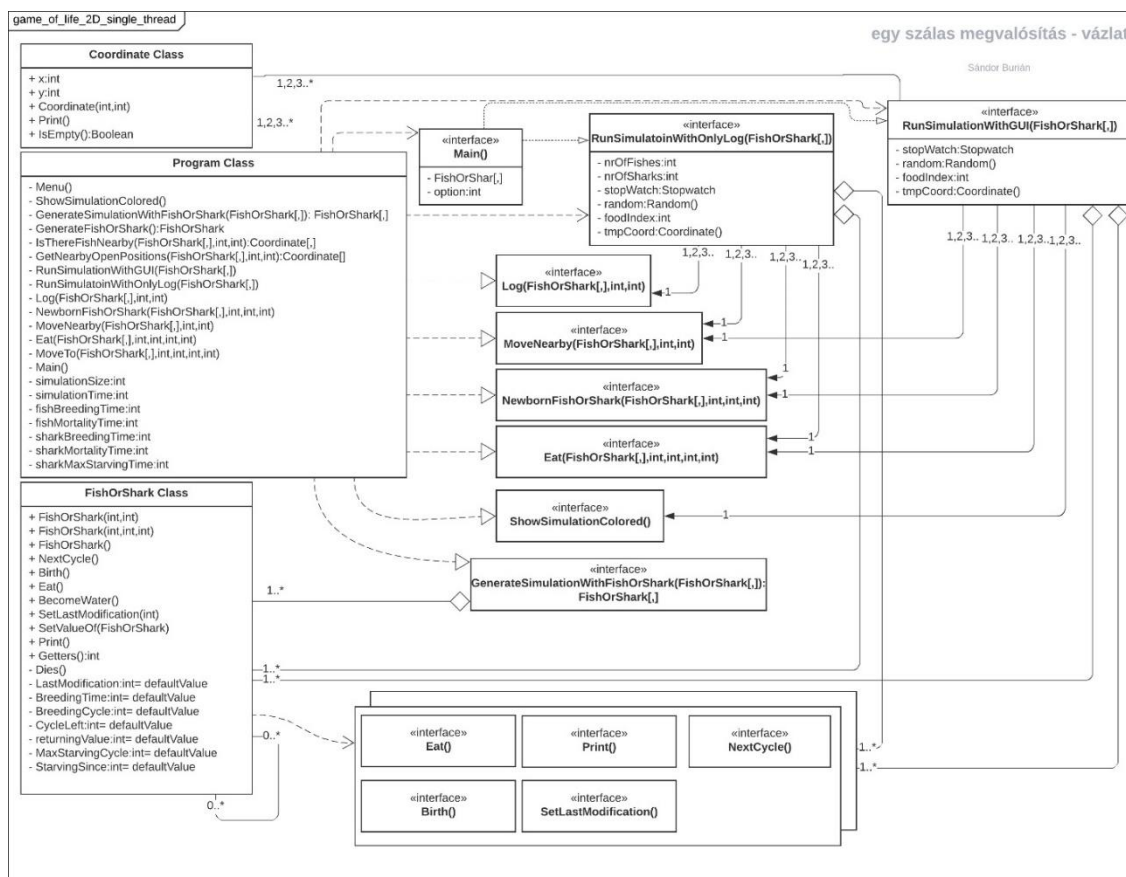
³ Ekkor a `..\game_of_life_2D-single_thread\game_of_life_2D-single_thread\bin\Debugstat.txt` és a `..\game_of_life_2D-single_thread\game_of_life_2D-single_thread\bin\Debugstat_with_conway.txt` állományokról van szó.

beimplementálódjon, a fentebb kifejtettek okán. Tehát, összegzésképp összesen két plusz lépést kapunk, egy érték vizsgálatot és egy értékadást.

Másfelől készíthetünk egy másolatot a szimulációs mátrixról, amiben eltároljuk, az adott ciklusban már lekezelt elemeket. Ebben az esetben ahhoz, hogy csökkentsük a költségeket tehetünk úgy, hogy minden páros ciklusban a másolatot vizsgáljuk, és "léptetjük", majd a változtatott elemeket az "eredeti" szimulációs mátrixba másoljuk át, és minden páratlan ciklusban fordítva. Ugyanekkor ez később gond lehet, a léptetés megoldása, hiszen egyszerre két részmátrixban kell vizsgálni a szabad cellákat. Ez számszerűleg több futtatást igényel, mivel két léptető ciklus helyett négyet kell futtatnunk minden egyes esetben, és mindegyikben egy érték összehasonlítást jelent pluszban.

Ez alapján arra következtethetünk, hogy jobb választás a háttértár szempontjából is és a futási sebesség szempontjából is egy változóban tárolni az objektum oldalán, hogy az adott ciklusban vizsgáltuk-e már az adott objektumot.

Alább az egyszálú program UML diagrammjának vázlata látható, amiben csak a fontosabb részek kerültek megjelenítésre, a jobb átláthatóság végett.



Vázlatos UML diagram az egyszálú megvalósításhoz⁴

⁴ Készült a <https://lucid.app> segítségével. A vázlat nem tartalmaz sok segédfüggvényt, és a nagyon hasonló működésű függvényeket sem!

Mérési eredmények⁵

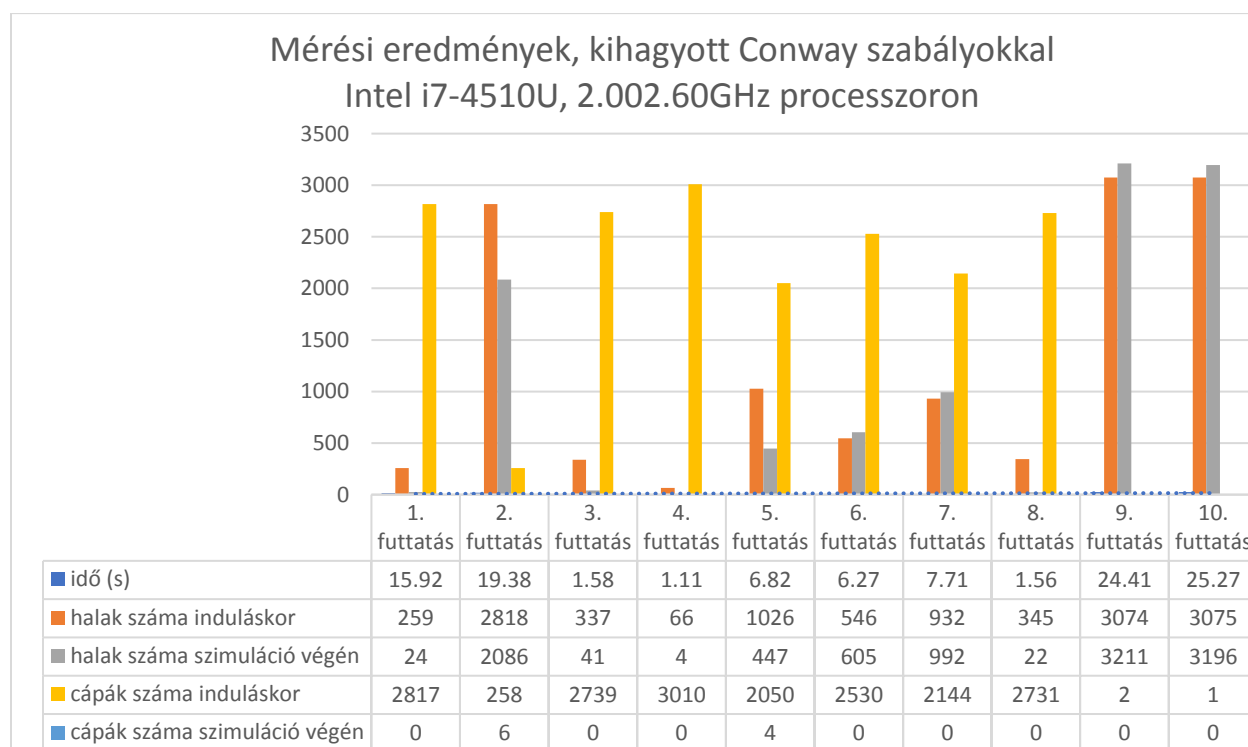
A kritériumfeltételek minden mérési esetben a következők:

Szimuláció ciklusainak száma	100
Szimuláció mérete ($n \times n$)	100
Halak halálzási ideje	10
Halak nemzési ideje	5
Cápák halálzási ideje	15
Cápák nemzési ideje	10
Cápák maximum éhezési ideje	5

Megadott szimulációs paraméterek

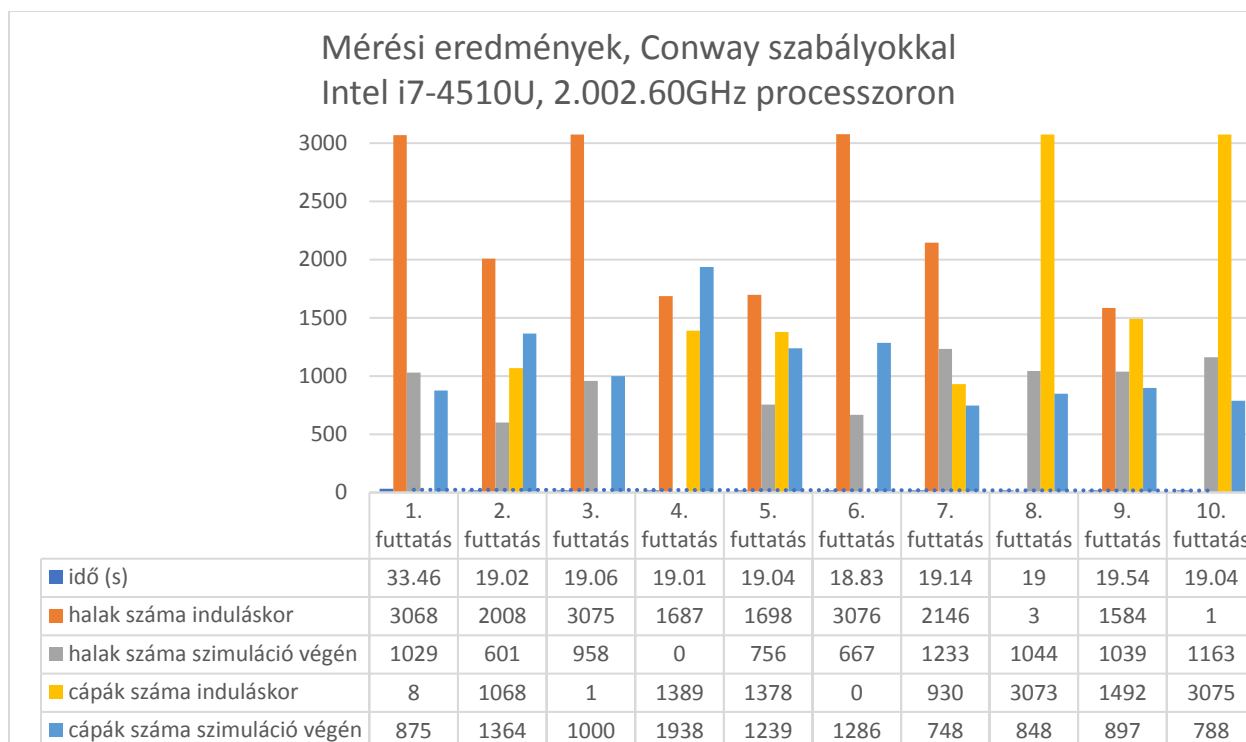
Első esetben azt teszteltem le, hogy alakul a szimulációban a populáció, ha nem alkalmazzuk Conway eljárásából azt a kritériumot, hogy:

1. Minden organizmus aminek 2-3 szomszédja van életben marad a következő generációra.
2. Minden olyan organizmus amelynek 4+ szomszédja van meghal a túlnépesedés miatt.
3. Mindegyik olyan organizmus amelynek 2-nél kevesebb szomszédja van, tehát maximum egy, meghal az izoláltság miatt.
4. Minden üres cellában amelynek pontosan három nem üres szomszédja van egy új organizmus születik

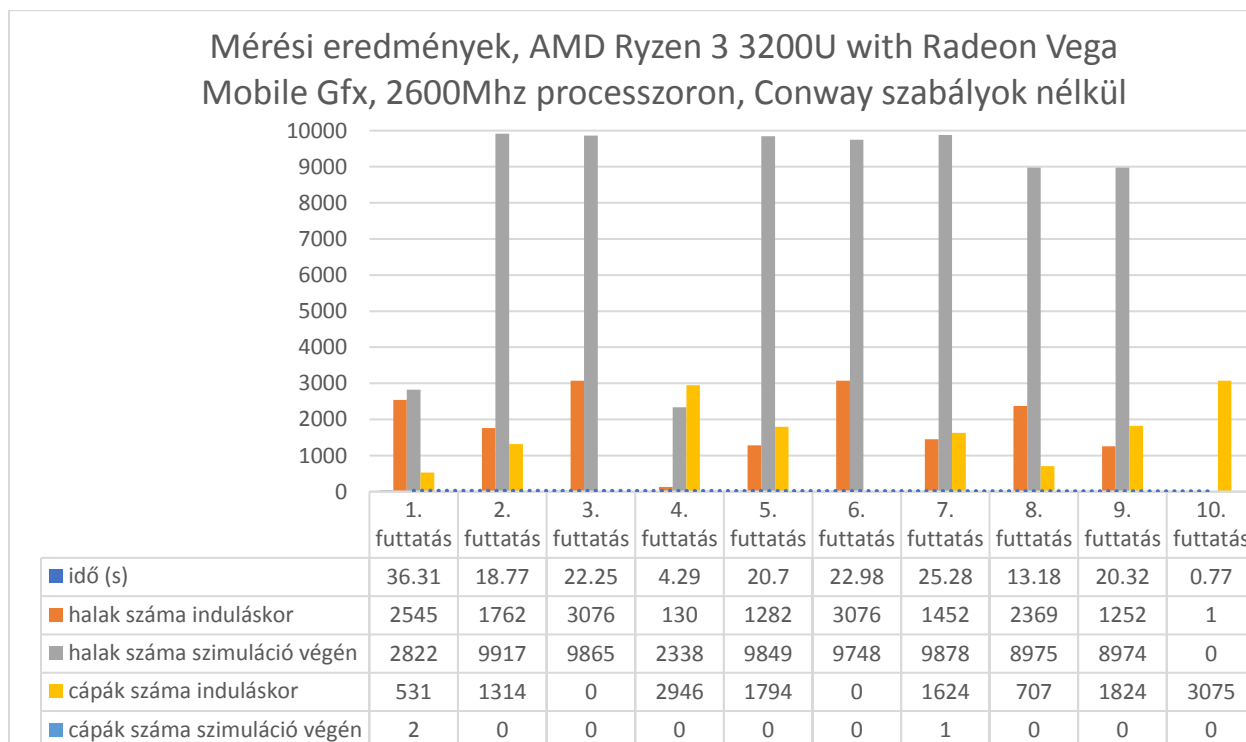


Mérési eredmények ábrázolása kihagyott Conway szabályokkal intel processzor

⁵ A mérések nem *debug* módban történtek, és csak a lényegi részt mértem, tehát pl a mátrixgenerálást nem!

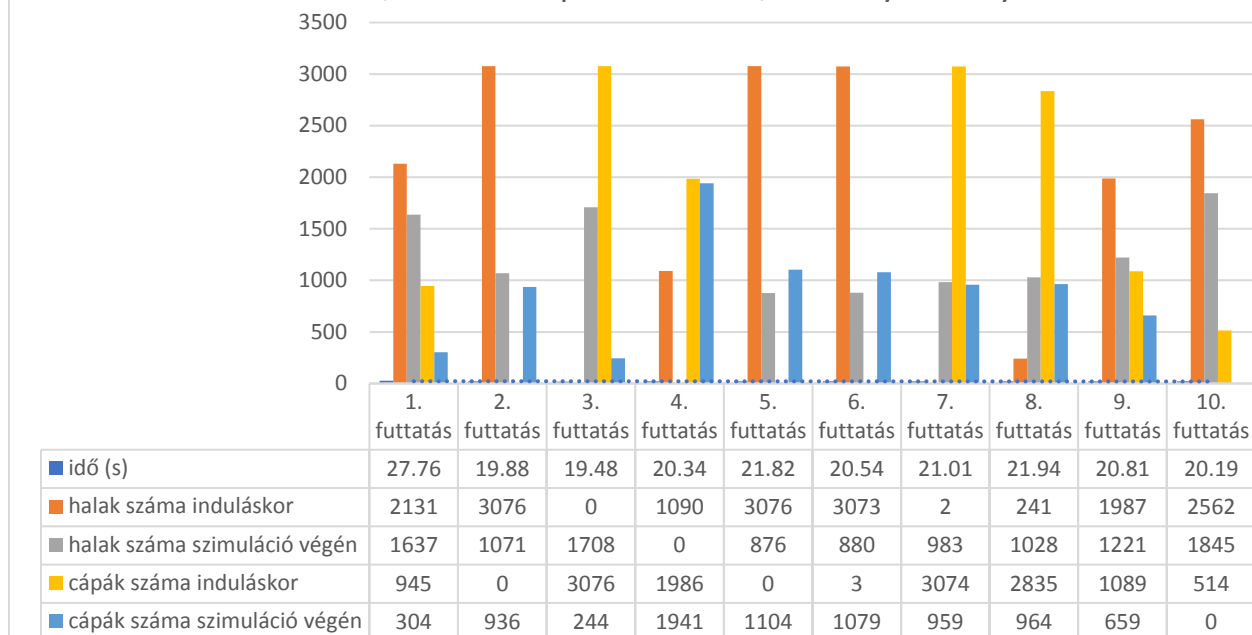


Mérési eredmények Conway szabályok alkalmazásával, intel proceszoron

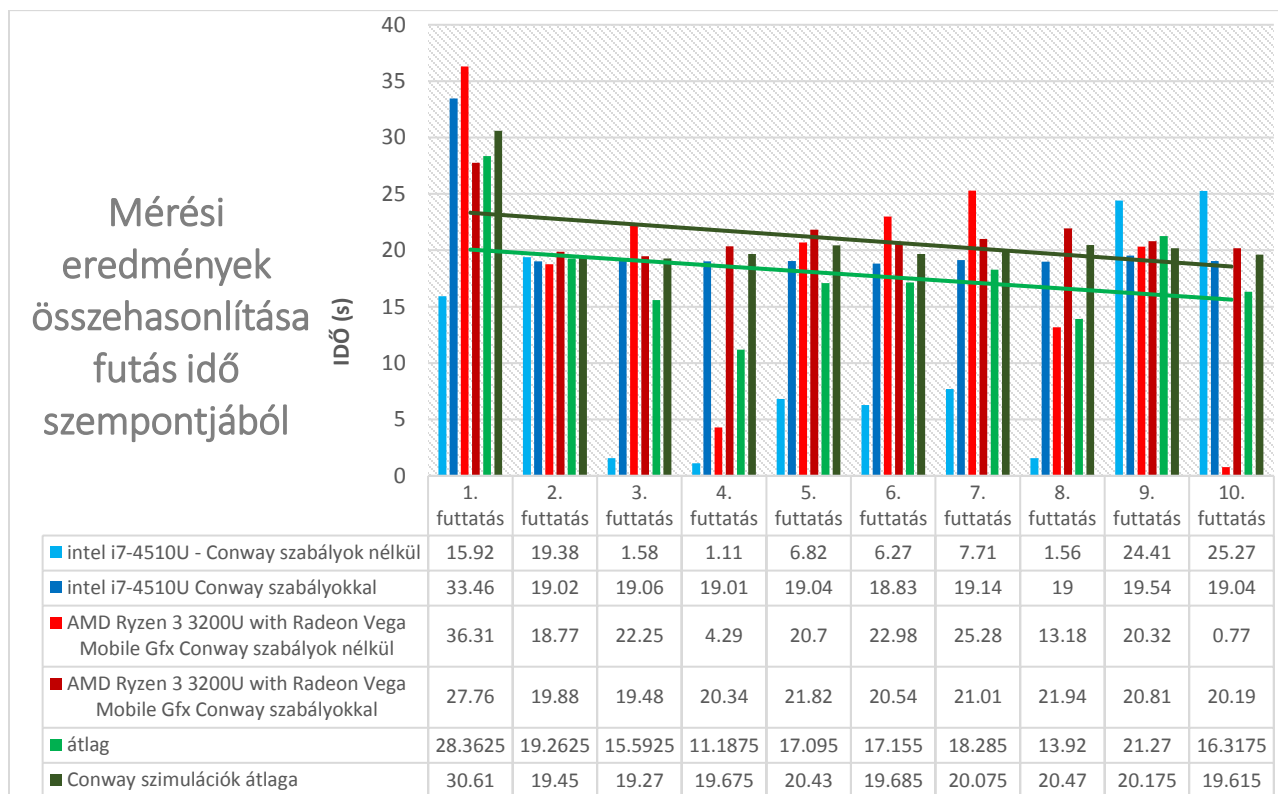


Mérési eredmények Conway szabályok alkalmazása nélkül AMD processzoron

Mérési eredmények, AMD Ryzen 3 3200U with Radeon Vega Mobile Gfx, 2600Mhz processzoron, Conway szabályokkal



Mérési eredmények Conway szabályok alkalmazásával AMD processzoron



Mérési eredmények futási idejének összehasonlítása 1 szál futtatás esetén

Több szálas szimuláció megvalósítás

A több szálas futtatás több kérdést is felvet, a matrix kezelésével kapcsolatban. Például, mekkora területeket kezeljenek a szálak? Hány szál legyen? Egymás alatti vagy egymás melletti területeket is feldolgozzanak a szálak? Hogy osszuk fel a szálak által feldolgozandó területet, úgy, hogy a legkevesebb szinkronizációs időre legyen szükség? És ehhez hasonló.

Alább az egymás alatti szálakra bontással példa:

1. Szál
2. Szál
3. Szál
4. Szál

Elemek beszínezve azon szálak színével amivel fel lesznek dolgozva:

11	12	13	14	15	16	17	18
21	22
31	...	33
41	...		44
51	...			55
61	...				66
71	...					77	...
81	...						88

Ezzel a megközelítéssel nyilvánvalóan a legrosszabb esetben egy szál két másik szállal, legjobb esetben egy másik szállal kell szinkronizálnon. Tehát négy szál esetén három szinkronizációs várakozásra van szükség. Ezzel szemben ha egymás mellé teszünk két szálat és ezt megimsételjük a következő sorban akkor már rögtön hat szinkronizációs pont szükséges ha átlósan is léphetnek a halak vagy cápák és négy ha nem léphetnek átlósan. *Megj: a fenti szimulációban engedélyezett az átlós lépés!*

Ugyanakkor ha eredetileg egy 100x100-as mátrixról van szó akkor az előbbi esetben egy szál egy kb 20 x 100-as részmátrixon dolgozik, és egy (vagy két) 100 hosszú vektor szakaszon lehet érintett szinkronizációs kérdésben⁶. Az utóbbi esetben viszont egy 25 x 25-ös részmátrixon dolgozik egy szál, és egy 49⁷ elemű vektor

⁶ A teljes eredeti mátrix szélességén.

⁷ 25 sorban, 25 oszlopban mínusz egy, hogy a sarkot ne számoljuk kétszer.

szakaszán lehet érintett szinkronizációs kérdésben, ugyanakkor ez nem egy teljes szakasz, hanem ha átlós lépéseket is engedélyezünk (*mint a szimulációkban*) akkor három másik szállal kapcsolódó szinkronizációs pontok, amiből a legkritikusabb a sarokpont ami be kell várja mind a három másik szálat, hiszen mindegyikkel kapcsolásban áll.

Ha egy négyszet alakba helyezzük el a szálat, az előző módszert alkalmazva az alábbi módon ábrázolhatjuk a második esetet:

1. Szál				2. Szál			
3. Szál				4. szál			
11	12	13	14	15	16	17	18
21	22
31	...	33
41	...		44
51	...			55
61	...				66
71	...					77	...
81	...						88

Jól látszik tehát, hogy ezt a módszert használva konstans három másik szállal kell kommunikáljon egy-egy szál.

Ezek alapján azt figyelembe véve, hogy ajánlott, hogy ugyanakkora szeleteket dolgozzanak fel a szálak, ezért az előbbi verziót választottam megvalósításra, hiszen ahhoz szükséges kevesebb szinkronizációs idő, tehát úgy kaphatunk gyorsabb futást.

Azért, hogy a programot könnyebben lehessen tesztelésre használni négyüzem módot építettem be a többszálú programba:

1. Szimuláció többszöri futtatási móddal a program által az adott CPU-n megtalált szálak számára optimalizálva, a Conway szabályok mellőzésével.
2. Szimuláció többszöri futtatási móddal a program által az adott CPU-n megtalált szálak számára optimalizálva, a Conway szabályok alkalmazásával.
3. Szimuláció többszöri futtatási móddal a felhasználó által megadott szálak számára optimalizálva, a Conway szabályok mellőzésével.
4. Szimuláció többszöri futtatási móddal a felhasználó által megadott szálak számára optimalizálva, a Conway szabályok alkalmazásával.

A szinkronizáció kérdése

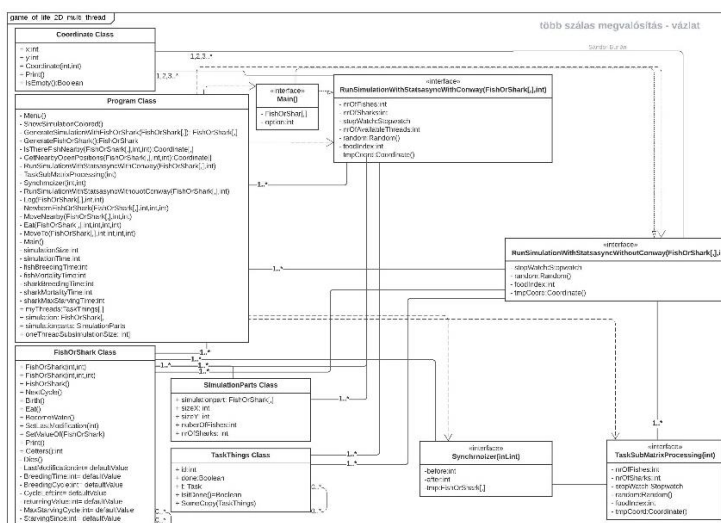
A program futtatása érdekében létrehozunk annyi taskot amennyi szálunk van. Ezek feloldása és lefoglalása fogja jelezni A szálakat egy segéd típusban kezeltem. Létrehozok egy akkora tömböt mint a szálak mennyisége, a tömb minden eleme ez a segédobjektum típusú. Ha egy szál végzett a ciklusával, akkor könnyen szinkronizálható az előtte-utána levő szálakkal mivel a tömbbeli sorszámmal tudnánk könnyen ihavatkozni egymásra.

Mérési eredmények⁵

A méréseket a könnyebb összehasonlíthatóság végett ugyanazokkal a paraméterekkel futtattam:

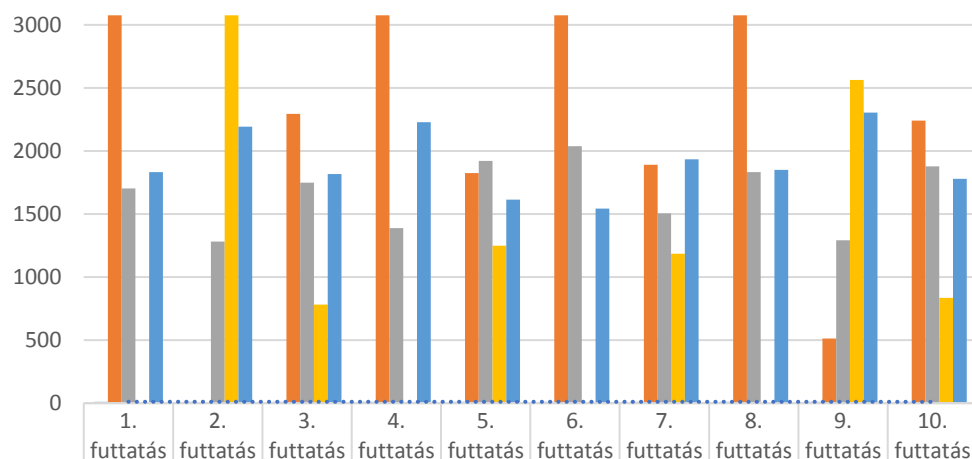
Szimuláció ciklusainak száma	100
Szimuláció mérete ($n \times n$)	100
Halak halálozási ideje	10
Halak nemzési ideje	5
Cápák halálozási ideje	15
Cápák nemzési ideje	10
Cápák maximum éhezési ideje	5

A méréseket jelen esetben csak Conway módszerrel végeztem, mivel eddig is tapasztaltuk, hogy mennyire jól tudja normalizálni a szimulációt. Felmerülhet a kérdés, hogy mit várhatunk ettől a megoldástól, egyértelműen több műveletet végzünk, de azt nem egy szálon, becsléseim szerint egy-egy szál legalább $sz * n + 2 * n$ objektum műveletet⁸ végez, és legfeljebb $(sz * n) * 2 + n$ műveletet végez ahol n az eredeti matrix mérete, és sz a szálszám; a szinkronizációhoz pluszba az egy szálas megoldáshoz képest. Ugyanakkor a szinkronizációs műveletek egyszerre két szálat is érintenek, így várható, hogy felgyorsul a futás. Alább a program lényegi részéről készült UML vázlat:



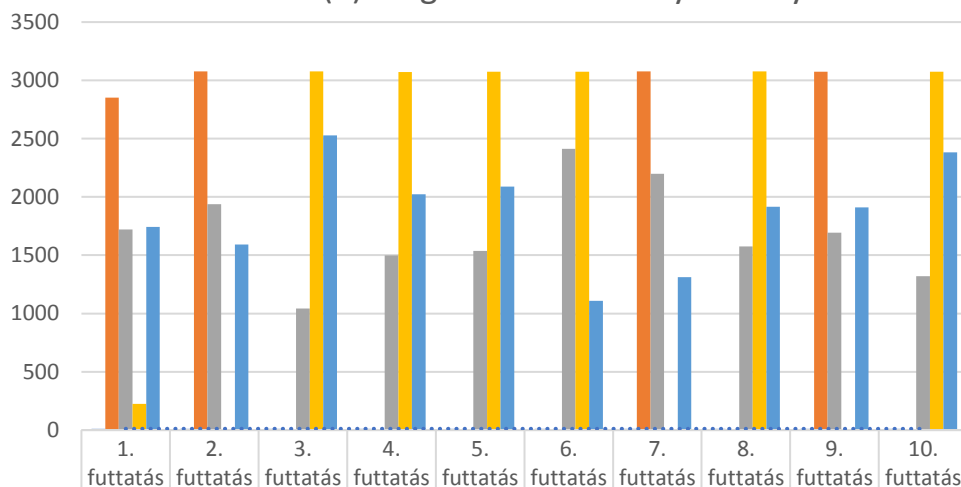
⁸ Köztük komplexebb műveletek is, nem csak iteráció az objektumon, szűrés, keresés stb

Mérési eredmények, Conway szabályokkal
Intel i7-4510U, 2.002.60GHz processzoron, a processzor által
virtualizálható szálszámnak(4) megfelelően



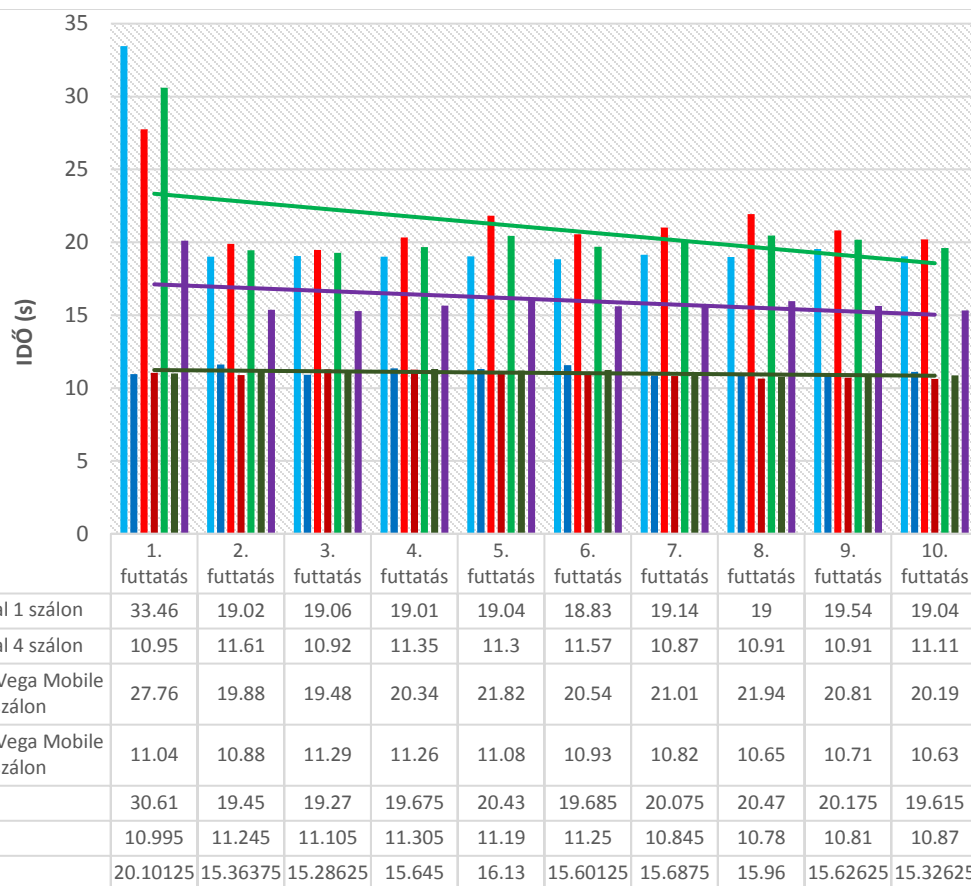
■ idő (s)	10.95	11.61	10.92	11.35	11.3	11.57	10.87	10.91	10.91	11.11
■ halak száma induláskor	3076	0	2294	3076	1826	3076	1890	3076	514	2242
■ halak száma szimuláció végén	1703	1283	1749	1389	1921	2037	1504	1833	1291	1877
■ cápák száma induláskor	0	3076	782	0	1250	0	1186	0	2562	834
■ cápák száma szimuláció végén	1832	2193	1818	2229	1613	1544	1933	1850	2304	1778

Mérési eredmények, AMD Ryzen 3 3200U with Radeon Vega
Mobile Gfx, 2600Mhz processzoron, a processzor által
virtualizálható szálszámnak(4) megfelelően Conway szabályokkal



■ idő (s)	11.04	10.88	11.29	11.26	11.08	10.93	10.82	10.65	10.71	10.63
■ halak száma induláskor	2851	3076	0	4	2	1	3076	0	3075	1
■ halak száma szimuláció végén	1722	1939	1043	1499	1538	2413	2198	1575	1694	1321
■ cápák száma induláskor	225	0	3076	3072	3074	3075	0	3076	1	3075
■ cápák száma szimuláció végén	1743	1593	2527	2022	2088	1108	1312	1917	1910	2383

Mérési
eredmények
összehasonlítása
futás idő
szempontjából
több szálon végzett
számításokkal



Következtetés:

Hiába a szálak egyenként több művelete a szinkronizációk miatt, a futásuk egyértelműen gyorsabb, már-már konstans idővel számolhatunk a futás időt tekintve.

Forrásjegyzet

Borítókép: [Online] Available: https://regi.tankonyvtar.hu/hu/tartalom/tamop412A/2011-0063_13_parhuzamos_algoritmusmodellek/images/fig3-9.jpg. [Hozzáférés dátuma: 2020 Szeptember 28]

[1] Environment.ProcessorCount Property [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/api/system.environment.processorcount?redirectedfrom=MSDN&view=netcore-3.1#System.Environment.ProcessorCount> [Hozzáférés dátuma: 29 09 2020]

[2] TÖBBSZÁLÚ/TÖBBMAGOS PROCESSZORARCHITEKTÚRÁK PROGRAMOZÁSA [Online] Available: https://elearning.uni-obuda.hu/main/pluginfile.php/274151/mod_resource/content/1/0053_Tobbszalu_Tobbmagos_Processzorarchitekturak_Programozasa.pdf [Hozzáférés dátuma: 29 09 2020]

[3] Nancy Ann Lynch: Oszott algoritmusok, ISBN: 963-9301-03-5 Kiskapu kiadó, 2020 április 18, 479. Oldal

[4] Intel Core i7-4510U @ 2.00GHz vs AMD Ryzen 3 3200U [Online] Available: <https://www.cpubenchmark.net/compare/Intel-i7-4510U-vs-AMD-Ryzen-3-3200U/2248vs3431> [Hozzáférés dátuma: 2 11 2020]

UML diagrammokat a <https://lucid.app> segítségével készítettem.