

Diplomamunka

Iteratív CT rekonstrukciós eljárás kidolgozása GPU környezetben és alkalmazása mért adatsorokon

Molnár Balázs

Témavezető: Dr. Légrády Dávid
egyetemi docens
Nukleáris Technika Intézet
Budapesti Műszaki és Gazdaságtudományi Egyetem

BME
2015

Fizika MSc :: Diplomamunka :: Kiírás

Témavezető:

Neve: Légrády Dávid

Tanszéke: NTI

E-mail címe: legrady@reak.bme.hu

Telefonszáma: 1254

Azonosító: DM-2014-20	
Diplomamunka címe:	Iteratív CT rekonstrukciós eljárás kidolgozása GPU környezetben és alkalmazása mért adatsorokon
Melyik szakiránynak ajánlott?	"Orvosi fizika"
A jelentkezővel szemben támasztott elvárások:	GPU programozási ismeretek (CUDA), C programozási ismeretek, angol nyelvtudás, jó matematikai készség, numerikus matematika, képrekonstrukciós ismeretek
Leírása:	<p>A Nukleáris Technikai Intézetben fejlesztés alatt álló Komputertomográfiai (CT) platform fejlesztésébe kapcsolódhat be a hallgató a téma feldolgozásával. A rendelkezésre álló mikrofókuszos röntgenső és pixellált nagy felbontású rtg detektormodul, továbbá az összehangolt mechatronikai mérési rendszer által szolgáltatott tomográfias felvételtől a hallgatónak rekonstruálnia kell a vizsgált tárgy térbeli gyengítési együttható-térképét. A rekonstrukciós algoritmusok közül ki kell választani a lehető legtöbb fizikai információt (pl. spektrum és spektrum-változás) felhasználni képes módszereket, a praktikus megvalósítás és az adott mérőberendezés, illetve vizsgált tárgyak tulajdonságai jelentette optimum szerint. A szóba jövő algoritmusokat a fenti szempontok alapján össze kell vetnie, lehetőség szerint valós adatok segítségével is. A kiválasztott algoritmust implementálnia kell grafikus számítási környezetben (GPU) CUDA nyelven és demonstrálnia a rekonstrukció működőképességét, valamint elemezni tulajdonságait szimulált matematikai fantomokon és mérési adatokon is.</p>

Absztrakt

A tomográfias képrekonstrukciós eljárások kutatásában egyre nagyobb teret nyernek az iteratív algoritmusok. Az első iteratív komputertomográfias (CT) képrekonstrukció még Hounsfield nevéhez fűződik, aki az első CT készülék fejlesztésében részt vett. Bár jelenleg az orvosi diagnosztikában a szűrt visszavetítés az általánosan elterjedt rekonstrukciós eljárás, a szakirodalom szerint a legtöbb modalitás (például MRI, CT, PET, SPECT) esetében érdemes a jövőben áttérni az egy lépéses rekonstrukcióról iteratív technikákra. Az eddigi fejlesztések eredményei azt jósolják, hogy a számítástechnika fejlődésével hamarosan a klinikai gyakorlatban is megjelenhetnek az iteratív rekonstrukciók, mivel ezek alkalmazhatóságát általában csak a rekonstrukció ideje korlátozza. Alátámasztja ezt az elképzelést az egyik vezető, orvosi diagnosztikai eszközök (is), gyártó vállalat VEO¹ nevű terméke, ami az első piacra kerülő model-alapú iteratív rekonstrukciós eljárás (2011).

A diplomamunkámban egy CT képalkotásra alkalmas iteratív algoritmust ismertetek. Elméleti háttérét a Lange és Carson által transzmissziós tomográfiára javasolt iteratív ML-EM algoritmus adja [1]. Az iterációs sémának részletes elemzését és a megvalósítási lehetőségeket [2]-ben ismertettem, a diplomamunka elsősorban a grafikus kártyán történt implementációra, a módszer matematikai fantomok rekonstrukciójával végrehajtott konvergencia-vizsgálatára, paraméter optimalizációra, valamint a valós mérési adatokon való tesztelésre fókuszál.

A fejlesztés során elértem az algoritmus nagyfokú gyorsítását, kihasználva, hogy hagyományos CPU helyett Nvidia GTX 690 típusú grafikus kártyán programozva párhuzamos feldolgozás érhető el. Verifikáltam az előrevetítés eredményeit a Matlab² Radon-transzformáltjával összevetve. Kétdimenziós eloszlások rekonstrukcióját elemezve megállapításokat tettem a konvergenciára, a rekonstrukció szabad paramétereinek függvényében is. A képminőségre vonatkozólag bevezettem a kvantitatív elemzésre gyakran használt L2 normát, és vizsgáltam az elvégzett iterációk számának, futási időnek tekintetében. A valós mérési eredmények feldolgozásának problémáit, és lehetséges megoldásukat is ismertetem a 11. fejezetben, valamint bemutatom a végrehajtott rekonstrukciót is, összehasonlítva a szűrt visszavetítés eredményeivel.

¹ General Electric Company

http://www3.gehealthcare.ca/en-ca/products/categories/computed_tomography/discovery_ct750_hd/veo#

² The MathWorks, Inc.

Önállósági nyilatkozat

Alulírott Molnár Balázs, a Budapesti Műszaki és Gazdaságtudományi Egyetem fizikus MSc szakos hallgatója kijelentem, hogy a diplomamunkám meg nem engedett segédeszközök nélkül, önállóan, a témavezető irányításával készítettem, és csak a megadott forrásokat használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból vettem, a forrás megadásával jelöltem.

.....

Molnár Balázs

Tartalomjegyzék

1. Bevezetés	1
1.1 Általános bevezetés	1
1.2 Transzmissziós ML algoritmus Poisson approximációval	4
1.3 A megvalósított iterációs séma előállítása	6
2. Egyéb képrekonstrukciós algoritmusok és ML-EM interpretációs lehetőségek	9
3. A grafikus kártya programozása.....	10
4. Az ML-EM algoritmus implementációja.....	14
4.1 A szimuláció során alkalmazott közelítések, vizsgálati módszerek.....	14
4.2 A program felépítése	14
4.3 Az előrevetítés kernel	16
4.4 A visszavetítés kernel.....	17
5. Hatékony szabadúthossz mintavételezés Woodcock módszerrel	18
5.1 Woodcock módszer (delta-tracking)	18
5.2 Szóráscsökkentő eljárások.....	19
5.3 A szimuláció hatékonysága	21
5.4 Szabadúthossz-zsugorítás implicit befogással.....	22
6. Az előrevetítés verifikációja	27
7. A rekonstrukció működőképességének demonstrálása.....	29
7.1 Az L2 és CC norma	29
7.2 Konvergencia	29
7.3 Várakozások az ML-EM algoritmussal szemben.....	31
7.4 Shepp-Logan fantom rekonstrukciója jó minőségű mérési szimuláció esetén	31
7.5 Rossz jel/zaj arányú felvételek rekonstrukciója, dóziscsökkentés CT vizsgálatoknál...	35
7.6 Demonstráció további fantomokon	38
8. Paraméter optimalizáció	41
8.1 A képminőség függése az iterációk számától.....	43
8.2 A képminőség függése a rekonstrukcióban indított fotonok számától	44
8.3 A ray-marching algoritmusban használt lépésköz hatása a képminőségre	46
8.4 A visszavetítésben kijelölt LoR-ok számának hatása a képminőségre	47
8.5 Az úthossz nyújtás és zsugorítás hatása a képminőségre.....	49

9. Rekonstrukció korlátozott idő alatt	50
9.1 A korlátozott rekonstrukciós idő problémájának felvetése	50
9.2 A rekonstrukcióban indított fotonszám optimális beállítása korlátozott rekonstrukciós idő esetén.....	52
9.3 A ray-marching algoritmusban használt lépésköz optimális beállítása korlátozott rekonstrukciós idő esetén	54
9.4 A visszavetítő válaszegyenések számának optimális beállítása korlátozott rekonstrukciós idő esetén	55
9.5 Az úthosszat zsugorító illetve nyújtó paraméter optimális beállítása korlátozott rekonstrukciós idő esetén	56
10. A spektrumfelkeményedés és korrekciója.....	58
10.1 A spektrumfelkeményedés megjelenése a rekonstruált képen.....	59
10.2 Korrekció ML-EM algoritmussal.....	60
11. Valós mért adatsorok rekonstrukciója	63
11.1 A kísérleti elrendezés	63
11.2 A valós mérési adatokkal végzett rekonstrukció kihívásai	63
11.3 A rekonstrukcióhoz szükséges geometriai adatok	65
11.4 Elvárások a rekonstrukció minőségére vonatkozólag	66
11.5 Tesztesetek	67
12. Konklúzió, kitekintés	71
13. Irodalomjegyzék.....	73

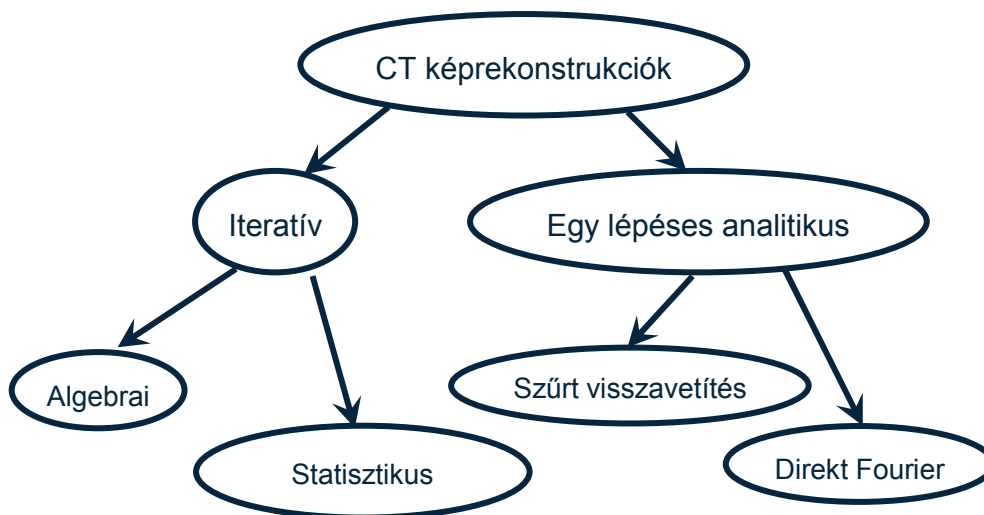


1. Bevezetés

1.1 Általános bevezetés

A röntgen komputertomográfia (továbbiakban CT) egy gyakran használt képalkotó modalitás az orvosi diagnosztikában. A CT vizsgálat során a betegről röntgen-felvétel készül különböző nézetekből (projekciók), amik matematikailag (ideális leképezést feltételezve) a röntgensugárzással szemben mutatott teljes gyengítési együttható eloszlás Radon-transzformáltjának (háromdimenziós esetben a sugártranszformáltjának³) felelnek meg. A képrekonstrukció egy inverziós feladat, a projekciókból ismeretében kell meghatározni az eredeti gyengítési együttható térképet. Az emberi szövetek gyengítési együtthatójának különbözősége kontrasztot ad a rekonstruált képen, ami segítséget nyújthat a helyes diagnózis felállításához.

A rekonstrukciós módszereket alapvetően két csoportba sorolhatjuk: iteratív és analitikus. Az iteratív rekonstrukciókon belül léteznek algebrai és statisztikus módszerek (itt található a diplomamunkámban ismertetett ML-EM módszer is), az analitikuson belül pedig többek közt a szűrt visszavetítést találjuk, ami a ma leginkább elterjedt technika a klinikai gyakorlatban. Áttekintést erről az 1. ábra, valamint [3] ad.



1. ábra
Képrekonstrukciós módszerek [3]

³ Nemzetközi irodalomban x-ray transform vagy john transform

Az általánosan használt szűrt visszavetítés egy jól működő, bevált módszere a kép rekonstrukciójának, azonban számos hátránya van más, például iteratív, módszerekkel szemben. Alapvető feltételezése ugyanis a módszernek, hogy a detektor által begyűjtött adatok a gyengítési együttható eloszlás radon-transzformáltjai (sugártranszformáltjai), ami a valóságban tökéletlen leképezés miatt nem teljesül maradéktalanul [4]. A rekonstruált képen ez műtermékek formájában jelentkezik, szélsőséges esetekben, például fém jelenléte esetén, a szűrt visszavetítés nem is alkalmazható (legalábbis korrekcióra szorul) a spektrumfelkeményedés és szórás okozta fém-műtermékek [5] megjelenése miatt. Egy statisztikus, iteratív képrekonstrukciós eljárás kevésbé érzékeny a detektor által mért adatok statisztikus zajára, hiányos adatra, nagy gyengítési együttható inhomogenitás jelenlétére, így egyes esetekben alkalmazása előnyösebb lehet. A legnagyobb lehetőség abban rejlik, hogy az inverzió megvalósításakor a teljes fizikai valóság beépíthető a modellbe, a rekonstrukció során figyelembe vehetjük például a röntgen sugárzás spektrumát, a szórást, nem-szokványos geometriákat.

A képszintézis elvi lehetősége a röntgen sugárzás anyaggal való kölcsönhatásán alapszik, ezért fontos kitérni ezen folyamatok legalább említés szintű tárgyalására. Az orvosi röntgen-diagnosztika esetében (sugárzás energiája, anyagi tulajdonságok) a legjelentősebb kölcsönhatások a rugalmas- és rugalmatlan szórás, valamint a fotoeffektus. A párkeltés jelensége nem játszik szerepet a CT vizsgálatok esetében, hiszen az alkalmazott röntgen sugárzás maximális energiája általában 100-200 keV.

A fotoeffektus során a röntgen-foton egy erősen kötött atomi elektronnal hat kölcsön, aminek eredményeképp a foton megszűnik, az elektron pedig kilökődik az atommag vonzásából. A folyamat szabad elektronnal nem játszódhat le, a kölcsönhatás létrejöttéhez kell egy harmadik résztvevő, az atommag. Az impulzus- és energia-megmaradás így teljesülhet (az atommag is meglökődik), a tömegviszonyok miatt viszont az elektron viszi el, mozgási energia formájában, a röntgen-foton energiájának és a kötési energia összegének közel 100%-át. A fotoeffektus valószínűsége a rendszámmal erősen növekszik, mivel az effektus megvalósulása szempontjából az erősen kötött elektronok számítanak. A foton energiájától függően ez a valószínűség (a fotoeffektus hatáskeresztmetszete), a rendszám 3.-5. hatványával arányos [6]. A hatáskeresztmetszet energiafüggését tekintve abszorpciós éleket tapasztalunk (rezonancia az elnyelési valószínűségében) az elektronok kötési energiáinál, ezektől távol a hatáskeresztmetszet az energia $-3,5$ hatványával arányos. A fotoeffektus folyamánya az atomi legerjesztődés, amely során történhet fluoreszcencia vagy keletkezhetnek auger-elektronok.

Rugalmatlan (inkohereus) szórás akkor játszódik le, ha a foton energiája és impulzusa sokkal nagyobb, mint az objektumé, amivel kölcsönhatásba kerül. Ilyen esetben Compton-effektus játszódik le a foton és jellemzően egy gyengén kötött elektron között. A Compton effektus hatáskeresztmetszetének rendszámfüggése lineáris, ugyanis a szóródás valószínűsége a szórócentrumok számával arányos. A fotonenergia növelésével a hatáskeresztmetszet gyengén csökken, nagyobb energián a fotoeffektussal szemben ez a domináns effektus.

Rugalmas (koherens) szórás akkor következik be, ha az előzőkkel ellentétben, az elektron kötési energiája nem hanyagolható el a fotoné mellett. Például ha a foton egy erősen kötött elektronnal lép kölcsönhatásba, és nem löki azt ki, a szórás lényegében rugalmasnak tekinthető (Rayleigh-szórás). Hiszen a foton által leadott energia az elektronton keresztül az egész atomra áterjed, a tömegviszonyok viszont nem engednek meg jelentős energiaátadást, így a foton energiája lényegében változatlan marad. A Rayleigh szórás hatáskeresztmetszete a fotonenergiával gyorsan csökken, és a rendszám második hatványával arányos. Összefoglalva tehát alacsony energián (<100keV) és/vagy nagy rendszámú anyag esetén a fotoeffektus a leggyakoribb kölcsönhatási forma, nagyobb energián (>100keV) viszont a Compton szórás domináns.

Modell-alapú iteratív algoritmusoknál lehetőség van az említett fizikai jelenségek figyelembevételére, és ezáltal egy realiztikus rekonstrukciót megvalósítani. Az ML-EM iterációs séma megköveteli az aktuálisan iterált gyengítési együttható eloszlás melletti röntgen-projekciók szimulációját, pontosabban a detektor-beütésszámok becslését. Az iteráció hajtóerejét ez adja, így ezt érdemes úgy implementálni, hogy a lehető legtöbb fizikai információt nyerhessük ki a rendszerből. Ezt a modellt Monte-Carlo részecsketranszport-szimulációval építettem fel, aminek alkalmazása ma már igen elterjedt az orvos-fizikai problémák körében is. Első közelítésben úgy kezeltem a problémát, hogy csak abszorpciót szimuláltam, vagyis szóródott fotonokat nem vettem figyelembe, így bár a Monte-Carlo projektor használata nem indokolt az analitikus vetítéssel szemben, mégis a fejlesztés szempontjából (később a fotonok spektrális eloszlásának figyelembevétele, szórás implementációja) ez a legjobb választás.

Az ún. visszavetítés algoritmus elemzése során gyakran fogok hivatkozni a Beer-Lambert törvényre, ugyanis a visszavetítő függvény lényegében egy azt felhasználó analitikus módszer. A Beer-Lambert törvényt az (1) alakban írhatjuk fel, ahol μ jelöli a (lineáris) gyengítési együtthatót, dr az infinitezimális elmozdulást, Γ az integrálandó tartományt (görbét), N a részecskék számát, N_0 pedig a kezdeti részecskék számát. $\mu = \mu(E, Z)$ energia és

rendsámfüggő, az említett folyamatok hatás keresztmetszetének összegét kell összeszorozni az atomsűrűséggel.

$$N = N_0 \cdot \exp\left(-\int_{\Gamma} \mu(\bar{r}) \overline{d\bar{r}}\right) \quad (1)$$

A Beer-Lambert törvény homogén közegben, egyenes mentén vett sugárgyengülés esetén (2) alakú.

$$N = N_0 \cdot \exp(-\mu l) \quad (2)$$

1.2 Transzmissziós ML algoritmus Poisson approximációval

A dolgozatomban bemutatott rekonstrukció a Maximum Likelihood (továbbiakban ML) alapötletét használja fel. Tekintsük egy transzmissziós kísérletnél Y -t a mért beütések számának valószínűségi változó vektoraként (továbbiakban vvv). Jelölje μ a rekonstruálandó skaláreloszlást (gyengítési együttható térkép), és vegyük figyelembe μ -t úgy, mint egy paramétert, ami mellett Y -t mérünk. Egy adott μ mellett mért Y vvv súlyfüggvénye legyen $g(y|\mu)$, aminek a maximumát keressük μ szerint, vagyis azt a skaláreloszlást, ami mellett y a legnagyobb valószínűséggel felelne meg a valódi mérésünk eredményének. Ez általában nem könnyű, ugyanis az Y adatok nem tartalmaznak elegendő információt a függvény maximalizálásához, hanem egy teljes adathalmaz vetületét adják csak. Ezt a teljes adathalmazt, amiből a rekonstrukciót maradéktalanul végre lehetne hajtani, jelölje X vvv, és legyen ennek súlyfüggvénye $f(x|\mu)$. Ha a mérendő térfogatot felosztjuk voxelekre⁴, akkor a legkézenfekvőbb választás, ha X tartalmazza az egyes voxelekbe belépő részecskék számát. Ekkor ugyanis egy projekció esetén a j . voxel μ_j abszorpciós értékét egyértelműen meg lehet határozni az X_j illetve az X_{j+1} , vagyis a belépő és kilépő fotonok számának, segítségével. Egy LoR⁵ által érintett voxelek száma legyen $m-1$, így igaz lesz az $Y = X_m$ egyenlőség (az utolsó voxelt elhagyó részecskék száma X_m). Tekintsük most a

$$H(\mu | \mu_n) = E(\ln f(X | \mu) | Y, \mu_n) - \ln[g(Y | \mu)] \quad (3)$$

függvényt. Megmutatható a Jensen egyenlőtlenség alkalmazásával, hogy H maximuma μ szerint éppen μ_n -nél van, vagyis $\max_{\mu} \{H(\mu | \mu_n)\} = H(\mu_n | \mu_n)$. A bizonyítás részleteit

⁴ Háromdimenziós pixel

⁵ Line of Response - válaszegyenes

mellőzve belátható (felírva a $H(\mu_n | \mu_n) - H(\mu | \mu_n)$ mennyiséget, és a logaritmus függvényre alkalmazva a Jensen egyenlőtlenséget), hogy ha úgy választjuk meg μ_n -t, hogy $f(x | \mu_n) > f(x | \mu)$, akkor H -nak valóban $\mu = \mu_n$ -nél van maximuma. Állítsunk elő tehát egy $\{\mu_n\}$ sorozatot, és minden lépésben maximalizáljuk $E(\ln f(X | \mu_{n+1}) | Y, \mu_n)$ értékét μ_{n+1} szerint. Ekkor (3)-at tekintve kimozdultunk $H(\mu_n | \mu_n)$ -ből $H(\mu_{n+1} | \mu_n)$ -be, így az előzőek alapján H kisebb lett. Ez maga után vonja azt, hogy $\ln[g(Y | \mu_{n+1})] > \ln[g(Y | \mu_n)]$, vagyis megvalósítottuk a Maximum Likelihood módszert. Az EM módszer összességében visszavezeti a likelihood becslést várható érték maximalizálásra, ahol a teljes adathalmaz Y melletti várható értékét kell optimalizálni a gyengítési együttható eloszlás szerint.

Egyetlen dolog hiányzik még a fenti módszer konvergenciáját tekintve: még nem mondtuk meg, hogy a log-likelihood függvénynek milyen szélsőértéke van. Természetesen azt szeretnénk, ha maximuma lenne, vagyis $\ln[g(Y | \mu)]$ μ -ben konkáv lenne. Ez pontosan akkor teljesül, ha második deriváltmátrixa negatív definit. A következőkben beláthatjuk, mikor is teljesül ez a szimulációkban.

Jelölje most Y_i az i . projekciónál a detektort elérő fotonok számának vv-ját. Modellezzük Y_i -t Poisson eloszlású vv-ként⁶, vagyis eloszlásfüggvénye legyen

$$g(y_i | \mu) = \frac{\lambda_i^{y_i}(\mu) \cdot e^{-\lambda_i(\mu)}}{y_i!} \quad (4)$$

egy adott μ paraméter mellett. Az Y_i valószínűségi változó λ_i várható értéke felírható a Beer-Lambert törvény alapján:

$$\lambda_i(\mu) = b_i \exp\left(-\sum_{j \in L_i} l_{ij} \mu_j\right) \quad (5)$$

ahol b_i az i . projekciónál a forrást elhagyó részecskék száma, l_{ij} pedig az i . projekcióban a j . voxelben a foton által befutott út hossza. Az összes projekcióra felírható Y együttes eloszlásfüggvény a $g(y_i | \mu)$ függvények szorzata lesz, hiszen az Y_i -k egymástól független vv-k.

⁶ Arról, hogy ez miért megfelelő feltételezés az esetünkben, és milyen feltételekkel, még lesz szó a későbbiekben

Ebből felírhatjuk a log-likelihood függvényt a következő alakban:

$$\begin{aligned} \ln[g(y|\mu)] &= \ln\left[\prod_i g(y_i|\mu)\right] = \sum_i \ln[g(y_i|\mu)] = \\ &= \sum_i \left\{ -b_i \exp\left(-\sum_{j \in L_i} l_{ij} \mu_j\right) - y_i \sum_{j \in L_i} l_{ij} \mu_j + y_i \ln b_i - \ln y_i! \right\} \end{aligned} \quad (6)$$

A konvexitás vizsgálatához venni kell a fenti függvény kétszeres parciális deriváltjait:

$$H_{kl} = \frac{\partial^2}{\partial \mu_k \partial \mu_l} \ln[g(y, \mu)] = -\sum_i a_{ik} b_i \exp\left(-\sum_{j \in L_i} l_{ij} \mu_j\right) a_{il} \quad (7)$$

ahol a_{ik} együtthatók az A mátrix elemei. A egy olyan mátrix, aminek sorai az egyes projekciókat indexelik, oszlopai pedig a voxeleket. A a_{ik} eleme azon úthossz, amit az i . projekcióban a k . voxelben befut a részecske, vagyis

$$a_{ik} = \begin{cases} l_{ik} & \leftrightarrow k \in L_i \\ 0 & \leftrightarrow k \notin L_i \end{cases} \quad (8)$$

Ahhoz, hogy $\ln(g)$ konkáv legyen, H-nak negatív definitnek kell lennie, vagyis minden v vektorra: $vHv < 0$ kell, hogy teljesüljön. Ez pontosan akkor áll fenn, ha az A mátrix teljes rangú $n \times m$ -es mátrix, ahol $n \geq m$. Tehát legalább annyi projekciót kell végeznünk, ahány voxelből áll a rekonstruálandó skaláreloszlás által elfoglalt térfogat.

A következőkben tegyük fel, hogy a konkávság feltétele teljesül (vagyis az ML-EM konvergens), és nézzük meg, hogyan lehet az iterációt előállítani.

1.3 A megvalósított iterációs séma előállítása

Az előzőekben már felírtuk a log-likelihood függvényt (6) alakban. Ugyanígy a teljes adathalmazra (X) is megtehetjük a log-likelihood függvény felírását. Észre kell vennünk ehhez, hogy a $j+1$. voxelbe jutó részecskék binomiális eloszlást követnek (feltéve, hogy a j . voxelbe beérkező részecskék száma fix, nem véletlen), vagyis $X_{j+1} | X_j \sim Binom(n, p)$, ahol a próbák száma $n = X_j$, valamint a siker valószínűsége $p = \exp(-\mu_j l_j)$. Tételezzük fel azt is, hogy az X_j , csakúgy, mint a detektort ért beütések száma, Poisson eloszlást követ λ_j paraméterrel ($X_j \sim Poi(\lambda_j)$). Az egyszerűség kedvéért bevezettük a $\lambda_j = b \exp\left(-\sum_{k=1}^{r-1} \mu_k l_k\right)$ jelölést. Így a teljes adathalmaz log-likelihood függvénye:

$$\ln[f(x|\mu)] = \sum_i -b_i + X_1 \ln b_i - \ln X_1! + \sum_{j=1}^{m-1} \ln \binom{X_j}{X_{j+1}} + X_{j+1} \ln(\exp(-l_{ij}\mu_j)) + (X_j - X_{j+1}) \ln(1 - \exp(-l_{ij}\mu_j)) \quad (9)$$

Láthatjuk, hogy ha ennek a mennyiségnek a várható értékét akarjuk maximalizálni X_m feltétel mellett, akkor még ki kell találnunk az $E(X_j | X_m)$ várható értéket.

Most fontos megemlíteni, hogy mikor is jó az a feltételezés, hogy mind X_j , mind $Y = X_m$ (a detektálható, vagyis az utolsó voxelt elhagyó részecskék száma) Poisson eloszlást követ. Mint láttuk a voxelt elhagyó részecskék a bemenő részecskék számától függő binomiálisak, az exponenciális gyengülés szerinti valószínűséggel. Ez több voxelen való áthaladásra is igaz,

a valószínűség változik meg: $p = \exp\left(-\sum_k \mu_k l_k\right)$ A binomiális eloszlás Poisson-

approximációját alkalmazhatjuk erre az esetre, ha a binomiális eloszlás n és p paraméterére: $n \rightarrow \infty$ ugyanakkor $p \rightarrow 0$, de úgy, hogy $np \rightarrow \lambda < \infty$. Ezt úgy teljesítjük, (visszanézve a fenti n és p paraméterekre) hogy sok fotont indítunk a szimulációban, valamint kellően abszorbens és/vagy nagyméretű közeget használunk (ezt a valóságban nem túl nehéz teljesíteni).

Ezek után kanyarodjunk vissza az eredeti célunk felé, ami szerint az $E(X_j | X_m)$ feltételes várható értékről szeretnénk információt kapni. X_m binomiális eloszlású, feltéve X_j -t,

a korábbi jelölésekkel $X_m | X_j \sim Binom\left(X_j, \frac{\lambda_m}{\lambda_j}\right)$. A Bayes-tétel segítségével a fordított

feltételt is kiszámíthatjuk, vagyis a kérdéses $P(X_j | X_m) = \frac{P(X_m | X_j) \cdot P(X_m)}{P(X_j)}$ kiszámítható,

hiszen mindegyikről tudjuk már, hogy milyen eloszlású (az egyenlet jobb oldala ismert súlyfüggvényekből áll). A részletes levezetés [1]-ben olvasható. Arra a következtetésre jutunk, hogy $X_j - X_m | X_m \sim Poi(\lambda_j - \lambda_m)$, amiből egyből következik, hogy $E(X_j - X_m | X_m) = \lambda_j - \lambda_m = E(X_j) - E(X_m)$, s ekképp:

$$E(X_j | X_m) = X_m + E(X_j) - E(X_m) \quad (10)$$

ami pontosan a maximalizálandó várható érték számolásához kell.

Most már elkezdhetjük a tényleges maximalizációt, vagyis deriváljuk le az $E(\ln f(X | \mu) | X_m, \mu)$ -t μ szerint, tegyük egyenlővé nullával. $\ln(f)$ adott a (9) formulában, a feltételes várható értékeket pedig éppen az előbb írtuk fel (10).

A deriválandó függvény tehát:

$$\sum_{j \in L_i} \left\{ E(X_{j+1} | X_m) \ln(\exp(-l_{ij}\mu_j)) + (E(X_j | X_m) - E(X_{j+1} | X_m)) \ln(1 - \exp(-l_{ij}\mu_j)) \right\} + R \quad (11)$$

ahol R tartalmazza az összes olyan paramétert, ami nem függ μ -tól. A szélsőérték keresést elvégezzük:

$$\begin{aligned} 0 &= -\sum_i E(X_{j+1} | X_m) l_{ij} + \sum_i (E(X_j | X_m) - E(X_{j+1} | X_m)) \frac{l_{ij} \exp(-l_{ij}\mu_j)}{1 - \exp(-l_{ij}\mu_j)} \\ &= \sum_i -E(X_{j+1} | X_m) l_{ij} + \sum_i (E(X_j | X_m) - E(X_{j+1} | X_m)) \frac{l_{ij}}{\exp(l_{ij}\mu_j) - 1} \end{aligned}$$

Ez a transzcendens egyenlet numerikusan oldható meg, ha $s = l_{ij}\mu_j \ll 1$:

$$\frac{1}{e^s - 1} = \frac{1}{s} - \frac{1}{2} + \frac{s}{12} + O(s^3) \quad (12)$$

(12) elsőrendű közelítését alkalmazva az iterációs képlet a (13) alakot ölti.

$$\mu_j^{n+1} = \frac{\sum_i (E(X_j | X_m) - E(X_{j+1} | X_m))}{\frac{1}{2} \sum_i (E(X_j | X_m) + E(X_{j+1} | X_m)) \cdot l_{ij}} \quad (13)$$

Behelyettesítve a feltételes várható értékeket (10)-ből, észrevehetjük, hogy a számlálóban csak a j . és a $j+1$. voxelre eső fotonok várható értéke marad, a nevezőben viszont megjelenik a detektort ért beütések száma, valamint ennek várható értéke is.

$$\mu_j^{n+1} = \frac{\sum_i (E(X_j) - E(X_{j+1}))}{\frac{1}{2} \sum_i (E(X_j) + E(X_{j+1}) + 2X_m - 2E(X_m)) \cdot l_{ij}} \quad (14)$$

A könnyebb érthetőség kedvéért alkalmazzunk más jelöléseket, amik már alkalmazkodnak az implementációhoz. (14)-ben mindkét összegzést az i -re hajtjuk végre, ami nem más, mint a LoR-okat számozó futóindex. Jelöljük most ezt úgy, hogy összegzünk minden olyan LoR-ra, ami átmegy a j . voxelen, aminek gyengítési együtthatóját épp rekonstruálni akarjuk (jelölés: $lor \in L_j$, ahol L_j az a halmaz, ami tartalmazza a j . voxelen átmenő válaszgyeneseket. Ezentúl a (14)-beli várható értékeket szimulálni fogjuk, a szimulációkban ezek tényleges fotonszámok lesznek, vagyis egy részecsketranszport-szimulációval megbecsüljük, hogy hány db részecske esett azon voxelre. Ezeket a becsléseket a továbbiakban $X_{lor[j]}$ alakban reprezentálom, valamint az utolsó m . voxel elhagyó részecskeszám esetén visszatérek az Y_{lor} jelölésre. Az i . LoR j . voxeljében befutott úthossz

pedig $L_{lor}[j]$ -vel fogom jelölni. Tekintsük tehát (15)-öt, ami megegyezik az implementált iteráció sémával.

$$\mu[j] = \frac{\sum_{lor \in L_j} (X_{lor}[j] - X_{lor}[j+1])}{\sum_{lor \in L_j} \left(Y_{lor} - \hat{Y}_{lor} + \frac{X_{lor}[j] + X_{lor}[j+1]}{2} \right) \cdot L_{lor}[j]} \quad (15)$$

(15) azt fogalmazza meg a matematika nyelvén, hogy a soron következő iterációban a j . voxel gyengítési együtthatóját úgy számolhatjuk, hogy a rajta átmenő összes válaszegyenest (LoR-t) tekintve összegezzük a benne elnyelődött fotonok számát (ez maga a számláló), majd elosztjuk egy olyan, szintén LoR-okra vonatkozó, összeggel, amiben a befutott úthosszakkal súlyozzuk a voxelbe bemenő, és abból kimenő részecskék számának átlagát, valamint a mért és a szimulált detektor-beütésszámokat. Hiszen a (14)-ben X_m éppen a detektort ért beütésszám vv-ja, valóságban ennek helyére X_m egy realizációja kerül, vagyis a mérési eredményünk: Y_{lor} (a lor által kijelölt detektorpixel beütésszáma). $E(X_m)$ pedig mérési eredmények várható értéke, amit itt, (15)-ben, \hat{Y}_{lor} -ral jelöltem.

2. Egyéb képrekonstrukciós algoritmusok és ML-EM interpretációs lehetőségek

A diplomamunkám elkészítésekor figyelembe vettem más képrekonstrukciós lehetőségeket is. Ahogyan az általános bevezetésben említettem, a szűrt visszavetítés egy gyakorlatban elterjedt módszer a képrekonstrukcióra, azonban a rekonstruált képen gyakran műtermékek jelennek meg. Az iteratív képrekonstrukció a szakirodalom alapján jobb tulajdonságokkal rendelkezik az elérhető képminőségét illetően, mivel a rekonstrukcióban a valóság pontosabb modellezését teszi lehetővé. A kutatás során a statisztikus, iteratív rekonstrukciókkal foglalkoztam.

A Bsc szakdolgozatomban [7] egy hasonló, iteratív képrekonstrukciós eljárást teszteltem, ami szintén az ML becslésből indult ki, azonban egy másik közelítő módszerrel, a teljes adathalmaz bevezetése nélkül, csupán a mérési eredmények, és a befutott úthosszak becsléséből számította ki az iteráció következő elemét. A rekonstrukciós módszerre az irodalomban gyakran ML-TR néven hivatkoznak. [1] szerint az EM algoritmus, és a

segítségével előállított formula, sokkal pozitívabb kilátásokkal kecsegtet, főleg a konvergencia, és a futási idő szempontjából.

Gyakorlatilag közvetlen előzménye ennek a dolgozatnak a 2013 évi TDK-munkám [2], amiben háromféle interpretációját vázoltam fel az ML-EM algoritmusnak, de ezeket csak CPU-n teszteltem. Azt a következtetést vontam le a vizsgálatuk során, hogy GPU-ra a [2]-ben „v3”-mal jelölt algoritmus a legalkalmasabb, ezért ennek fejlesztését végeztem el. Annak céljából, hogy ezen dolgozat olvasásakor ne vesszen el a lényeges információ a részletek közt, az interpretációk kiválasztásának problémáját nem fogom ismertetni a diplomamunkámban, a kiválasztott módszert viszont igen (4. fejezet).

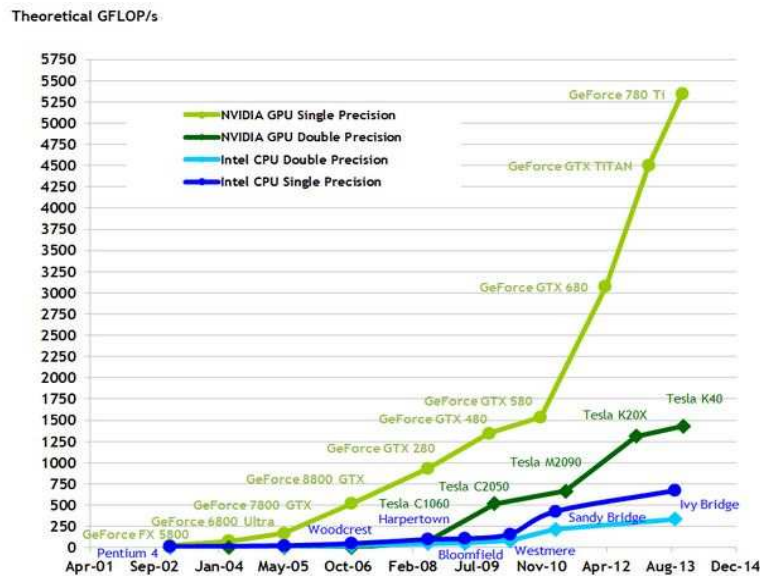
A TDK dolgozathoz képest újabb eredmények közé tartozik a teljes GPU-s implementáció, egy merőben más struktúrájú szoftver létrehozása. Ennek segítségével számos lehetőség tárul elénk, ugyanis a rekonstrukció sokkal gyorsabb, hatékonyabb lett, így tesztelését végre lehet hajtani más rekonstrukciós algoritmusokkal összevetve. Másrészt hatalmasat lépett előre ezzel a fejlesztés abban a tekintetben is, hogy valós mérési adatok rekonstrukcióját is létre lehet hozni segítségével, innentől pedig nyitva állnak a rekonstrukciós módszer kihasználásának lehetőségei. Mivel az ML-EM algoritmus GPU-n történő megvalósítása nem triviális programozási feladat, ezért a következőkben összefoglalom a grafikus kártyák szerepét az implementációban, ami talán elősegíti ezen előrelépés mikéntjének megértését.

3. A grafikus kártya programozása

Általánosan egy iteratív képrekonstrukciós megoldást választva a jobb képminőség elérésének reményében lemondunk a kis számításigényről, vagyis hatalmas mennyiségű műveletet végzünk annak érdekében, hogy az elért eredményeink precízebbek legyenek. Szerencsére, a párhuzamos feldolgozás adta lehetőségeket kihasználva csökkenteni tudjuk rekonstrukció (azonos körülmények között nagyságrendekkel hosszabb) futási idejét. Ahogyan azt [2]-ben is láthatjuk, az 1.3 fejezetben ismertetett (statisztikus, iteratív) ML-EM rekonstrukció CPU-n való futása igen lassú tud lenni. Célszerű kihasználni a grafikus kártyákban rejlő lehetőségeket az algoritmus gyorsítására. Mivel a NTI-ben lehetőségem van egy Nvidia GTX 690-es grafikus kártyát programozni, ezért az algoritmus implementációjának programozási nyelvét CUDA-nak választottam. Ez az Nvidia saját

programozási nyelve, alapvetően C/C++ alapú, kiegészítő parancsokkal, GPU specifikus memória fajtákkal, változókkal [8].

A grafikus kártyák kihasználása a párhuzamosan végrehajtható műveletek elvégzésére az utóbbi évtizedekben igen gyakorivá vált, mivel a jelenleg releváns programozási feladatok számításigényét már nem elégíti ki a hagyományos processzorok fejlődése (Moore-törvény). A grafikus kártyák és a processzorok fejlődését jól szemlélteti a 2. ábra.



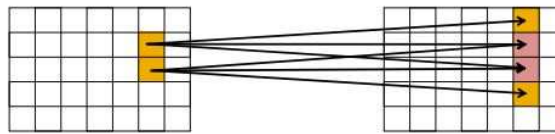
2. ábra

Az Nvidia kártyák számításkapacitása az Intel processzorokkal szemben [7]

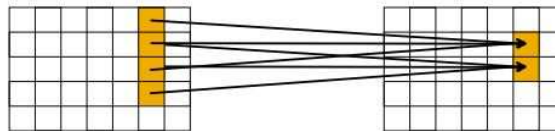
Fontos kiemelni, hogy bár a számítási kapacitás (GFLOP/s) általában több nagyságrenddel nagyobb a GPU-k esetében, mint a CPU-knál, ez természetesen hátrányokkal is együtt jár, amik általában az algoritmus struktúrájából fakadnak.

Az orvosi képalkotásban gyakran előfordul, hogy a feldolgozás/szimuláció sematikus szóró jellegű (ld 3. ábra), vagyis a kiinduló adatból az elvégzendő számítás több memóriahelyre tud (rész)eredményt biztosítani. Ez azt jelenti, hogy a párhuzamosan futó operációk között versenyhelyzet (race condition) alakulhat ki, aminek nem megfelelő kezelése információvesztést okoz. A grafikus kártyákon használható atomi műveletek erre megoldást nyújtanak, amik bár lassabbak, mint a hagyományos műveletek, de a párhuzamosan futó szálak között így nem alakul ki versengés.

Szórás (Scatter)



Gyűjtés (Gather)

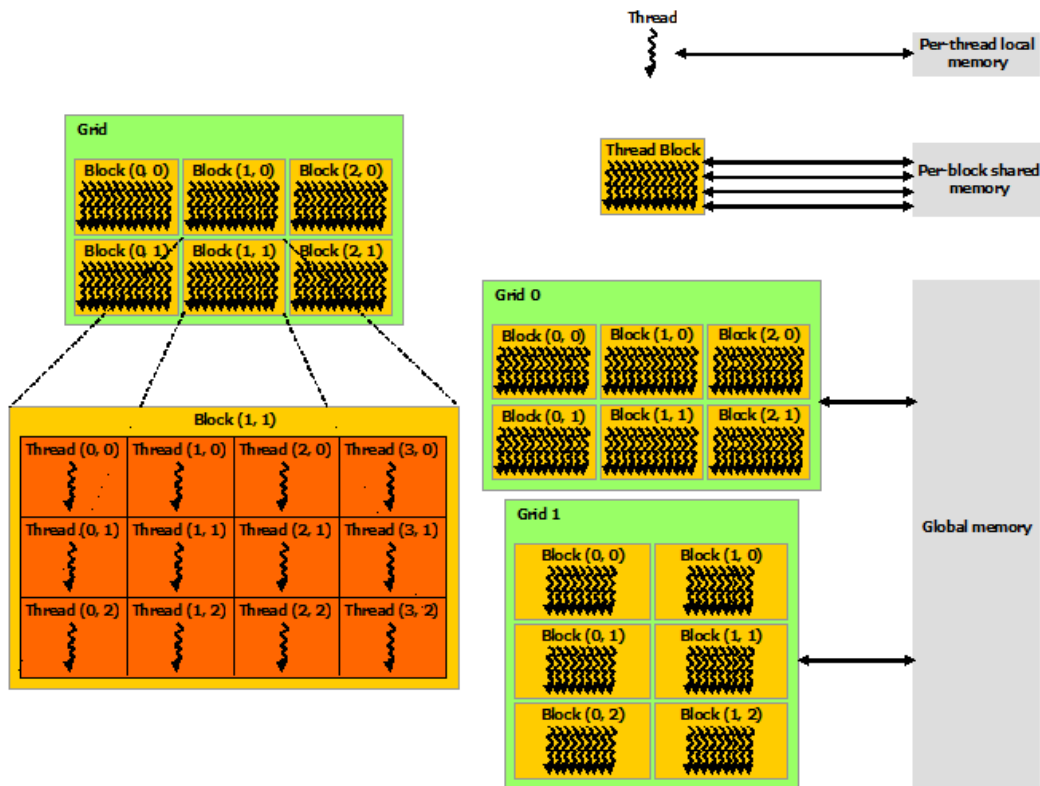


3. ábra
Párhuzamos sémák

A megoldandó probléma komplexitása szintén növeli a GPU-n való futás idejét, ezért alapvető követelmény a hosszú ciklusok elkerülése, valamint sok elágazás (if-else) használata a programozás során. A képrekonstrukciónál, és hasonló alkalmazásoknál, érdemes definiálni az aritmetikus intenzitás fogalmát (arithmetic intensity), ami annak mérőszáma, hogy milyen arányban vannak a forráskódban az aritmetikai műveletek a memória-elérési parancsokhoz képest. Minden GPU-s programozási feladatnál célszerű észben tartani, hogy a memória-hívások igen költségesek, főleg ezért is érvényes az, hogy az inherensen párhuzamos jellegű feladatok futása sokkal gyorsabb, mint a komplexebb problémáké. A GPU-ról, valamint annak orvosi képalkotásban való felhasználásáról kiváló cikket olvashatunk, és további hasznos információhoz juthatunk [9] megtekintésével.

A 4. ábra mutatja be, hogy a grafikus kártyához tartozó memória kezelése miként történik a CUDA architektúrán. Az egyes részfeladatokhoz (amiket párhuzamosan kívánunk elvégezni) a CUDA memóriaszalakat (továbbiakban thread-ek) rendel, amiket alapvetően két nagyobb struktúrába ágyaz. A thread-ek blokkokban (továbbiakban block) helyezkednek el, amik pedig egy rácsra (továbbiakban grid) vannak felhelyezve. A végrehajtás során megadhatjuk a block-ok által tartalmazott thread-ek számát, valamint azt, hogy egy grid hány block-ot tartalmazzon. Minden thread-nek van egy azonosítója (egy GPU specifikus változó), aminek használatával kioszthatjuk a párhuzamosan elvégzendő számításokat. A thread-ek rendelkeznek egy lokális memóriával, ami cache-elt, vagyis lényegesen gyorsabb elérésű, mint a globális memória. A block-ok is rendelkeznek megosztott (shared) memóriával, ami azonban lassabb, így a szálak közti kommunikáció kifejezetten hátrányos a futási időre nézve. Tehát ha olyan implementációt tudunk előállítani, amiben az egyes memóriaszalak egymástól függetlenül dolgozhatnak, és a memóriát megfelelően strukturáljuk (threads/block,

blocks/grid), akkor elérhetjük a párhuzamosítás maximális kihasználtságát. Ez a határ nyilván grafikus kártya-specifikus, ezért érdemes ismerni a kártya tulajdonságait, mielőtt elindítjuk a programunkat.



4. ábra
Memóriakezelés CUDA-ban [7]

A diplomamunkámban bemutatott képrekonstrukciót alapvetően két nagyobb programozási feladatra bonthatjuk: előrevetítés és visszavetítés. E kettőhöz egy-egy ún. kernel-függvény tartozik, ami szintén egy GPU specifikus, lényegében azt jelenti, hogy meghívásakor a GPU számításkapacitásának felhasználásával egyidőben végrehajtódik N db művelet (amit maga a kernel-függvény definiál), ahol N-et meghatározza a grafikus kártya paraméterei (például az ún. warp size), valamint a memória kiosztása (block-ok, grid-ek).

4. Az ML-EM algoritmus implementációja

4.1 A szimuláció során alkalmazott közelítések, vizsgálati módszerek

A fejlesztés jelenlegi stádiumában az algoritmus egyszerűsített formában működik a következő szempontokból:

- 1) A voxelmátrixon kívül a rekonstrukció vákuumot feltételez, itt nem szimulál kölcsönhatást
- 2) A kölcsönhatások során mindig abszorpció történik, vagyis szóródott fotonokat sem követ a rekonstrukció (nem szimulál szórást)
- 3) A rekonstruálandó térfogat vízzel van feltöltve, a tömegabszorpciós együtthatók energiafüggése a víznek megfelelő (ez a spektrummal történő rekonstrukciónál lényeges)
- 4) A detektálás hatásfokát nem veszi figyelembe
- 5) Egyszerű cone-beam geometriát feltételez a rekonstrukció (például a detektor síkja merőleges a főnyalábra)

A programban lehetőség van a bemenő mérési adatok szimulációjára, vagyis tesztelhetjük a rekonstrukciót matematikai fantomokon is. A fantommal ilyenkor egy Monte Carlo szimulációt végez az algoritmus, lefuttat egy szimulációt arra, hogy milyen projekciókat kapnánk, ha elvégeznénk a mérést a valóságban. Ez ugyanazzal a módszerrel történik, mint amit a rekonstrukció előrevetítés lépésében is használ a program, ezért lényeges ennek a szimulációnak a validálása, amiről a 6. fejezetben olvashatunk.

4.2 A program felépítése

Az alábbiakban az algoritmus felépítését fogom ismertetni általánosságban, csak főbb elemeit említve meg. A matematikai fantomok rekonstrukciójára fejlesztett forráskód és a valós mérési eredmények feldolgozására dedikált program lépései néhol eltérnek egymástól, ezeket ismertetni fogom.

Elsőként megtörténik az eredeti (később rekonstruálandó) kép bevitele. Ez egy önkényesen skálázott értékű elemekből álló mátrix, gondolhatunk alatta gyengítési együttható

eloszlásra, vagy, ha rendelünk minden voxelhez energiafüggő tömegabszorpciós együtthatót, amivel a számítások során megszorozzuk ezt az értéket, akkor sűrűségeloszlásra. Az eredeti képen az algoritmus mérési szimulációt végez, vagyis végrehajtja az előrevetítést, aminek eredménye a (15)-ben Y -nal jelölt beütésszám. A program a fejlesztés jelen fázisában nem szimulál szórást, a fotonnal csak abszorpció történhet. Az előrevetítést Monte-Carlo részecsketranszport módszerrel valósítja meg, a szabad úthosszt a Woodcock módszerrel mintavételezve. A valós mért adatok rekonstrukciója esetén a mérésből származó projekciók beolvasása, és a rekonstrukció előrevetítésével való összehangolása történik meg először. A program lehetőséget ad arra, hogy figyelembe vegyük a röntgensugárzás spektrumát, illetve a tömegabszorpciós együtthatók energiafüggését. A spektrumot illetve a tömegabszorpciós adatokat textúrában kezeli a szoftver a gyors memória-elérés érdekében.

Az algoritmus ezután elindítja az iterációs ciklust, amely során végrehajtódik a rekonstrukció. Az iterációkban még egy ciklus van beágyazva, ami a projekciókat lépteti. A fejlesztés során azt tapasztaltam, hogy az előrevetítést bár egyszerre is el lehet végezni az összes projekcióra, mégis érdekesebb projekciónként egy előrevetítést és utána közvetlenül visszavetítést elvégezni, hiszen ekkor nem szükséges tárolni csak az aktuális projekcióra vonatkozó eredményeket.

Egyetlen iteráció egy projekcióján belül mindig történik egy előre és egy visszavetítés. Az előrevetítésben az aktuálisan iterált gyengítési együttható eloszlás mellett hajtódik végre ugyanaz a szimuláció, ami a mérési eredményeket is szolgáltatja (matematikai fantomok esetén). Vagyis ekkor becsüljük a (15)-ben \hat{Y} -pal jelölt beütésszámokat. A visszavetítés-függvény végzi el az X, L változók becslését, vagyis a voxelekbe bejutott, és azokból kimenő részecskék számát, illetve a voxelekben befutott úthosszakot számítja ki. A visszavetítés struktúrája 3 fejezetben elmondottakhoz viszonyítva némileg ambivalens, mivel nem voxel-vezérelt (ez egy „gather” típusú, vagyis gyűjtő séma lenne), hanem LoR-vezérelt (szórás séma). Ez azt jelenti, hogy a visszavetítés során több LoR-hoz tartozó (párhuzamosan futó) memóriaszál is hozzájárul ez eredményekhez, és esetleg ugyanazon voxel memóriájához kíván írni. Ezért a visszavetítésben atomi műveleteket alkalmazok. Intuíció alapján a voxel-vezérelt (voxel-driven) visszavetítés gyorsabbnak tűnhet, de [3] szerint a LoR vezérelt (ray-driven) algoritmus jobb tulajdonságokkal bír, ezért is alkalmaztam inkább az utóbbi megoldást, ami könnyebben is implementálható.

4.3 Az előrevetítés kernel

Az előrevetítés elvégzéséhez Monte-Carlo részecsketranszportot alkalmaztam, a szabad úthosszat Woodcock módszerrel mintavételezve. A hozzá tartozó kernel függvény először kiszámítja, hogy melyik detektorpixelbe ütközik majd a foton, és hogy hol kerül be a rendszermátrixba, majd végiglépteti a részecskét a rendszeren, nyilvántartva a pozícióját, és a súlyát, ami egyszerű Woodcock módszer esetén bináris változó, 0, ha elnyelődött, 1, ha nem. Implementáltam szóráscsökkentő eljárásokat ennek a módszernek a fejlesztésére (implicit befogás, úthossz-mintavételezés torzítása), amiről az 5. fejezetben olvashatunk.

A GPU-memória kiosztása értelemszerűen történt: egy thread (memóriaszál) felelős egyetlen fotoncsomag szimulációjáért, a memóriaszálak között információ-áramlás nincs, így a párhuzamosítás nagy hatásokkal végezhető. Az előrevetítő kernel-függvény meghívódik minden egyes projekcióban. Olyan lehetőséget is vizsgáltam, ami ebben különbözik, vagyis az előrevetítést egyszerre végzi az összes projekcióra, azonban ekkor a memória-korlátok erősen befolyásolták a rekonstrukció minőségét. Gondoljunk bele abba, hogy már kétdimenziós esetben is mekkora memóriát igényel a rekonstrukció. A detektor felbontása, ahogy a 11. fejezetben is látni fogjuk, egyik irányban 1536 pixel. Ha legalább a detektorpixelek felének beütésszámát néhány százalékos relatív hibával akarjuk becsülni, akkor a Poisson-statisztika szerint átlagosan legalább néhány ezer foton kell, hogy egy pixelt érjen. Tegyük fel például, hogy $N=2500$ db részecske éri el a detektor egyes pixeleit átlagosan (ennek eléréséhez a forrásból már eleve többet kell indítanunk!), ekkor a relatív hiba $1/\sqrt{N} = 2\%$. Így egy projekcióban legalább $1536/2 \cdot 2500 \approx 2 \cdot 10^6$ db részecskét kell szimulálnunk ahhoz, hogy a látómezőnk felét kihasználjuk. Ez 360 projekció felvétele esetén már több százmillió db, ami memóriában (4 byte egy integer változó) máris több Gb-ot jelent. Ennyi memória általában nem áll rendelkezésre, illetve háromdimenziós esetben a voxelmátrix is Gb-os nagyságrendű helyet foglal, tehát itt ezt a megoldást el kell vetni. Ezért alkalmaztam a projekciónkénti előre- majd visszavetítési struktúrát, ekképp ugyanis az előző projekció előrevetítésének eredményére már nincs szükség.

4.4 A visszavetítés kernel

Miután a mérési szimuláció eredményei a kezünkben vannak, megpróbálunk következtetni a kérdéses μ eloszlásra, vagyis visszavetítést végzünk. A megvalósítási lehetőségek struktúráját [2]-ben ismertettem, most a GPU-s implementációra legalkalmasabb módszert foglalom össze. A (15) formula szerint a visszavetítésben becsülni az egyes voxelekben befutott úthosszakat (L), valamint a voxelekbe jutott részecskék számát (X). Egyszerű, gyors, és hatékony becslést kapunk a ray-marching módszerét alkalmazva, vagyis ha az L -et a voxelmátrixon való végiglépegetéssel becsüljük, X -et pedig eközben analitikusan a Beer-Lambert törvény alapján számoljuk. Választanunk kell egy lépésközt, valamint meg kell adjuk a LoR irányát, és ekkor a LoR és a voxelmátrix két döféspontja (bemenet és kimenet) között ezzel a lépésközzel végigléptethetjük a virtuális részecskéinket. Eközben minden voxelre regisztráljuk a bennük elkövetett lépések számát, ebből számoljuk az úthosszakat. A visszavetítést LoR-vezérelt struktúrával valósítottam meg, vagyis egy memóriaszál egy válaszegyenes mentén végezte el a kérdéses mennyiségek számítását. A futásidővel kapcsolatban a visszavetítés másképp reagál a voxelmátrix méretének változására, mint az előrevetítés. Míg az előrevetítés számításigénye lineárisan, vagy három dimenzióban maximum négyzetesen függ a voxelmátrix (lineáris) méretétől (hány voxelből áll egy oldala), a visszavetítés időigénye ennél magasabb hatvány szerinti függést mutat. Hiszen ugyanúgy, mint az előrevetítésnél három dimenzióban, a visszavetítő LoR-ok száma a rendszer lineáris méretével négyzetesen növekszik, de ezen visszavetítéseken belül az érintett voxelek száma is köbösen növekszik, így minden memóriaszálnak a műveletek sokszorososa jut a rendszer felbontásának növekedésével.

A futási idő szempontjából tehát esszenciális, hogy hány LoR-t jelölünk ki a visszavetítésben. A tapasztalatok szerint a képminőséget is nagyban befolyásolja (3D képalkotásnál), hogy hány válaszegyeneset és milyen szisztéma alapján jelölünk ki. Háromdimenzióban ugyanis, nem csak a konvergenciához szükséges visszavetítő egyenes megléte lényeges, hanem az is, hogy ezek eloszlása milyen, megfelelően vannak-e elosztva ahhoz, hogy minden voxelről információt kapjunk (esetleg van olyan voxel, amin nem megy át LoR!). Általában nem megfelelő választás az, ha a detektorpixelek geometriai közepét, és a forrást összekötő egyenesek mentén végezzük a visszavetítést. Ekkor ugyanis konzekvens módon kimaradhatnak voxelek a rekonstrukcióból, ha kicsi a detektor felbontása (bármely irányba) a voxelmátrixhoz képest. Ezért első közelítésképp, a visszavetítés

implementációjakor bevezettem egy olyan LoR mintavételezési algoritmust, ami egy pixelen belül választ véletlenszerűen K db pontot a válaszegyenesek kijelölésére. Ezzel elérhettem, hogy háromdimenziós esetben sem volt szisztematikus torzulás a rekonstruált képen, mivel közelítőleg minden voxelen ugyanannyi LoR haladt át.

Az alternatív mintavételező algoritmusok közül a leghatékonyabb a Poisson disk mintavételezés lehet, aminek lényege, hogy a síkon K db pontot jelöl ki úgy, hogy azok egymástól való távolsága a lehető legnagyobb legyen. Ez azért kedvező, mert a visszavetítés annál hatékonyabb, minél változatosabb útvonalakat jelölnek ki a válaszegyenesek, vagyis minél különbözőbb úton érintik az egyes voxeleket. Egy effektív Poisson disk algoritmust mutat be [10], ennek ML-EM rekonstrukcióban történő implementálása ajánlatos lehet a jövőben.

5. Hatékony szabadúthossz mintavételezés Woodcock módszerrel

Monte Carlo részecsketranszport módszerek során felmerülő probléma két ütközés közötti úthossz mintavételezése. Tudjuk, hogy fizikailag a részecskék szabadúthossza exponenciális eloszlású valószínűségi változó (vv.), ezért lehetséges egyszerűen inverz eloszlásfüggvény módszerrel sorsolni az úthosszakokat a transzport-szimulációban. Szakaszosan homogén közeg esetén az eloszlásfüggvény azonban nehezen kezelhetővé válik, elsősorban a közeget határok pozícióinak meghatározása szempontjából [11], ezért érdemes megfontolni más módszerek alkalmazását. Ebben a fejezetben egy ilyen módszert (Woodcock módszer) fogok ismertetni, szóráscsökkentő eljárásokat vizsgálva általánosan, és speciálisan a rekonstrukció hatékonysága szempontjából.

5.1 Woodcock módszer (delta-tracking)

A Woodcock módszer egy hatékony eljárás inhomogén közegben való részecsketranszport szabadúthossz-sorsolására. A módszer lényege, hogy a közeg hatáskeresztmetszeteit felülről becsülő majoráns hatáskeresztmetszet szerint sorsolunk úthosszt két ütközés között, majd az ütközést virtuálisnak tekintjük $1-P$ valószínűséggel (ilyenkor a pozíciót léptetjük, minden egyéb változatlan), P valószínűséggel valódinak. Ha P éppen az ütközési helyre jellemző hatáskeresztmetszet, és a majoráns hatáskeresztmetszet hányadosa,

akkor az úthossz sorsolás torzításmentes, amire a későbbiekben láthatunk bizonyítást. Ebből a szempontból az algoritmus egy szemléletes leírása, [11] alapján, a következő:

1. Válasszuk meg a torzított sűrűségfüggvényt $\Sigma_{maj} e^{-\Sigma_{maj} x}$ -nek, ahol $\Sigma_{maj} = \max_x \{\Sigma(x)\}$
2. Sorsoljunk szabadúthosszt a részecskének eszerint a sűrűségfüggvény szerint:
 $l = -\frac{\ln(r)}{\Sigma_{maj}}$, ahol r egy véletlen szám 0 és 1 között, és léptessük erre a helyre a részecskét
3. Képzeljük el Σ_{maj} -t a valódi, és egy virtuális hatáskeresztmetszet összegeként:
 $\Sigma_{maj} = \Sigma(x) + \Sigma_{virtuális}(x)$, és az aktuális pozícióban a valós ütközés arányában, vagyis $\frac{\Sigma(x)}{\Sigma_{maj}}$ valószínűséggel legyen valós ütközés, $\frac{\Sigma_{virtuális}(x)}{\Sigma_{maj}}$ valószínűséggel virtuális
4. Virtuális ütközés esetén térjünk vissza a 2. ponthoz, valós ütközés esetén pedig ennek megfelelően haladjunk tovább

Lényegében tehát a szabadúthossz eloszlásának torzítása mellett megfelelő kompenzációval (valódi ütközések csak a valós és majoráns hatáskeresztmetszetek arányában) érjük el a módszer torzításmentességét. Speciálisan, ha detektor-beütésszámokat akarunk becsülni egy olyan közeget tekintve, ahol csak elnyelődés történhet, a valós ütközés azt jelenti, hogy a részecske nem járul hozzá a beütésszámhoz (0-val járul hozzá), ha csak virtuális ütközéseket szenved, akkor 1-gyel járul hozzá (ennél finomabb skálát is tudunk konstruálni, ld. alább).

5.2 Szóráscsökkentő eljárások

A módszer hatékonysága alapvető jelentőségű a Monte Carlo részecsketranszport szimulációnál, mivel sok részecskét kell szimulálnunk ahhoz, hogy kielégítő pontossággal tudjunk becslést adni a meghatározandó mennyiségekre (általában valamilyen valószínűségi változó momentumai). A szóráscsökkentő eljárások lényege, hogy a részecskeszám emelése nélkül csökkentjük a becslés statisztikus bizonytalanságát, úgy, hogy a fázistér azon részét, ami a becslés szempontjából fontosabb, gyakrabban mintavételezzük. Erre többféle módszer létezik, aminek nagyrészt összefoglalja [12], két eljárást fogok az alábbiakban ismertetni: implicit befogás (implicit capture, survival biasing), és a torzított szabadúthossz mintavételezést. Ezeket a módszereket tiszta abszorpciós esetben, a detektor-beütésszámok

becslésére, Woodcock-féle szabadúthossz mintavétel alkalmazása mellett tárgyalom, mivel ebben formában használtam a rekonstrukcióban.

Az implicit befogás alapötlete a következő: vezessük be a részecskék súlyának fogalmát (mennyivel járul hozzá a becsléshez), minden részecskét indítsunk $w=w_0$ súllyal, biztosítsuk, hogy a részecskék sosem nyelődnek el, de a súlyukat csökkentsük az túlélés valószínűségének megfelelően, vagyis az i . virtuális ütközés után legyen

$$w_{i+1} = w_i \left(1 - \frac{\Sigma(x)}{\Sigma_{maj}} \right).$$

Így a részecskék hozzájárulása (súlya) a beütésszámokhoz egy finomabb

skálán változik mint az analóg esetben (ahol 0 vagy 1 volt), a súly a $[0, w_0]$ intervallumból kerül ki, és $K+1$ db lehetséges értéket vehet fel, ahol K a virtuális ütközések száma. K növelésével a skála tehát tovább finomítható, amit a Woodcock módszerrel úgy érhetünk el, hogy a sorsolt szabadúthosszt zsugorítjuk (torzítjuk a gyakori mintavétel irányába), vagyis nagyobb majoráns hatáskeresztmetszetet választunk. Ennek az eljárásnak a hatékonyságát a 5.4 fejezetben ismertetem.

A torzított szabadúthossz mintvételezést a Woodcock módszer mellett kétféleképp valósíthatjuk meg hatékony módon: az egyik az említett úthossz-zsugorítás implicit befogás mellett, nagyobb majoráns hatáskeresztmetszet választással, a másik az úthossz-nyújtás, (részlegesen torzított) analóg esetre. Ez utóbbi, az előző fejezetben ismertetett Woodcock módszer módosításával érhető el úgy, hogy a virtuális ütközéseket preferálva megnyújtjuk a valódi ütközésig befutott úthosszt, kompenzációt pedig a részecskék súlyának módosításával érünk el. Definiálj p^{-1} azt a paramétert, amilyen mértékben kívánjuk a valódi ütközésig befutott úthossz várható értékét növelni (úthossz nyújtásának mértéke), és legyen p ennek inverze ($p < 1$ esetben követünk el úthossz-nyújtást). A virtuális ütközéseket preferálva

módosítsuk a valódi ütközés valószínűségét $\frac{p \cdot \Sigma(x)}{\Sigma_{maj}}$ -re, és minden virtuális ütközésnél

szorozzuk meg a részecske súlyát egy korrekciós faktoral. A korrekció mértéke legyen a torzítatlan eset virtuális ütközési valószínűségének és a torzított eset virtuális ütközési valószínűségének hányadosa, vagyis az i . virtuális ütközés után legyen

$$w_{i+1} = w_i \left(1 - \frac{\Sigma(x)}{\Sigma_{maj}} \right) \left(1 - \frac{p \cdot \Sigma(x)}{\Sigma_{maj}} \right)^{-1}.$$

Ezzel elérhetjük, hogy csakúgy, mint az implicit befogás

esetében, a beütésszámokhoz való hozzájárulás skálája finomabb legyen, és p helyes megválasztásával hatékonyá tehetjük az algoritmust. Vegyük észre, hogy $p \rightarrow 0$

határesetben magát az implicit befogást kapjuk vissza, hiszen ekkor az elnyelés valószínűsége 0, a súlykorrekciós faktor pedig $w_{i+1} = w_i \left(1 - \frac{\Sigma(x)}{\Sigma_{maj}}\right)$, $p=1$ pedig nem változtat az eredeti sémán (0 vagy 1 súlyokkal).

5.3 A szimuláció hatékonysága

A Monte Carlo szimulációk hatékonyságát a célváltozóra adott becslés pontossága (statisztikai szórása) és a számítás elvégzéséhez szükséges idő határozza meg. Annál jobb a hatékonyság, minél rövidebb idő alatt érünk el pontos eredményt, illetve minél pontosabban tudunk becslést biztosítani adott idő alatt. A gyakorlatban használt mérőszáma a *FoM* (Figure of Merit), definíciója: $FoM = \frac{1}{r^2[X] \cdot t}$, ahol X a becslendő valószínűségi változó, aminek

relatív szórásnégyzete: $r^2[X] = \frac{D^2[X]}{(E[X])^2}$, t pedig a futási idő. A hatékony algoritmusnak tehát magas *FoM* értéke van.

A következőkben adjunk általános formalizmust *FoM* érték számolására egy olyan szimulációban, ahol a becslendő mennyiség a detektor-beütésszám várható értéke, amit a

következőképp becslünk: $M = \sum_{j=1}^N w^{(j)}$, ahol $w^{(j)}$ a j . részecske súlya, N pedig az indított

részecskék száma. Tekintsük $w^{(j)}$ -kre úgy, mint egy valószínűségi változókra (továbbiakban

vv), így $M = \sum_{j=1}^N w^{(j)}$ is vv, aminek kíváncsiak vagyunk a relatív szórására. (16)

összefüggésből láthatjuk, hogy ehhez ismernünk kell $w^{(j)}$ első és második momentumát:

$$r^2[M] = \frac{D^2[M]}{(E[M])^2} = \frac{D^2\left[\sum_{j=1}^N w^{(j)}\right]}{\left(E\left[\sum_{j=1}^N w^{(j)}\right]\right)^2} = \frac{\sum_{j=1}^N D^2[w^{(j)}]}{\left(\sum_{j=1}^N E[w^{(j)}]\right)^2} = \frac{N \cdot D^2[w^{(j)}]}{(N \cdot E[w^{(j)}])^2} =$$

$$= \frac{1}{N} \frac{D^2[w^{(j)}]}{(E[w^{(j)}])^2} = \frac{1}{N} \frac{E[(w^{(j)})^2] - (E[w^{(j)}])^2}{(E[w^{(j)}])^2} = \frac{1}{N} \left(\frac{E[(w^{(j)})^2]}{(E[w^{(j)}])^2} - 1 \right) = \frac{1}{N} \left(\frac{E[w^2]}{(E[w])^2} - 1 \right) \quad (16)$$

A fenti egyenlőségek alkalmazása helyénvaló, mivel $w^{(j)}$ -k független, azonos eloszlású vv-k, valamint emiatt (az egyszerűség kedvéért) jelölhetjük w -vel egy tetszőleges $w^{(j)}$ -t. Tovább csak akkor haladhatunk, ha ismert a súlykorrekció mértéke, ahogyan a következő alfejezetben például az implicit befogásnál látni fogjuk.

5.4 Szabadúthossz-zsugorítás implicit befogással

Ahogy korábban láttuk, a gyakoribb mintavételezést Woodcock módszerrel könnyen megvalósíthatjuk a majoráns hatáskeresztmetszet módosításával. A korábbi jelöléseket megtartva p jelölje az úthossz várható értékét módosító faktor inverzét, de most legyen $p > 1$, hiszen ekkor csökkentjük az úthosszt. Az úthosszt a torzítatlan esetben a $\sum_{maj} e^{-\sum_{maj} x}$ sűrűségfüggvényből sorsoljuk, aminek várható értéke $\frac{1}{\sum_{maj}}$, vagyis ha $\frac{1}{\sum_{maj}} \cdot p^{-1}$ -t szeretnénk átlagos szabad úthossznak, akkor a $p \cdot \sum_{maj} e^{-p \cdot \sum_{maj} x}$ sűrűségfüggvényből kell sorsolnunk. A kérdés az, hogy mi az átlagos úthossz megválasztásának optimuma, általános esetben, hol van a $FoM(p)$ függvény maximuma. Ez függ a futási idő p függésétől, illetve a relatív szórásnégyzet p függésétől. A kettő egymás ellen dolgozik, vagyis p növelésével nő a futási idő, a szórásnégyzet viszont csökken.

A futási idő tekintetében azt mondhatjuk, hogy általános számítási séma (soros feldolgozás) mellett p -ben lineáris, vagyis $t(p) = a + b \cdot p$. Az 5.2 fejezetben ismertetett implicit befogás algoritmus alapján belátható, hogy a számítás ideje (ideális esetben) az ütközések számával egyenesen arányos. Legyen K az ütközések száma. Az egyes ütközések Poisson folyamat szerint történnek, mert az ütközések között befutott úthosszak egymástól független $p \cdot \sum_{maj}$ paraméterű exponenciális eloszlást követnek. Ennek következményeképp K Poisson eloszlású $p \cdot \sum_{maj} \cdot L$ paraméterrel (várhatólag $p \cdot \sum_{maj} \cdot L$ ütközés történik L hossz alatt). Tehát ha a futási idő a K várható értékével egyenesen arányos, akkor $t \sim E[K] = p \cdot \sum_{maj} \cdot L$, vagyis p -vel is egyenesen arányos. Párhuzamos feldolgozás esetén ez nem feltétlenül igaz, vagyis lehet, hogy a lineáris összefüggés által jósoltnál jobb futási időket tapasztalunk bizonyos p -tartományon. Ez a grafikus kártyák esetén nem tűnik merész

elképzelésnek (ld 3. fejezet), alapos kivizsgálására azonban a diplomamunkában nem kerül sor, csupán magának a lehetőségnek a bemutatása.

Térjünk vissza a detektor beütésszámok becsléséhez: $M = \sum_{j=1}^N w^{(j)}$. M becslése esetén

a $FoM = \frac{1}{r^2 [M] \cdot t}$, ahol r^2 a relatív szórásnégyzet, t pedig a futási idő. Nyilván mindkét

változó p -nek függvénye, így kapjuk FoM(p)-t: $FoM(p) = \frac{1}{r^2 [M](p) \cdot t(p)}$. Szükség van

a futási idő mellett M relatív szórásnégyzetének meghatározására. Ehhez az (16) összefüggésből indulunk ki, és figyelembe vesszük, hogy az i . virtuális ütközés után

$w_{i+1} = w_i \cdot \left(1 - \frac{\Sigma(x_i)}{p\Sigma_{maj}}\right)$. Általános esetben a momentumok (17) szerint öltenek formát.

$$E[w] = E[E[w | K]] = \sum_{k=1}^{\infty} E[w | K = k] f_K(k) = \sum_{k=0}^{\infty} E\left[\prod_{i=0}^k \left(1 - \frac{\Sigma(x_i)}{p\Sigma_{maj}}\right)\right] f_K(k) \quad (17)$$

$$E[w^2] = E[E[w^2 | K]] = \sum_{k=1}^{\infty} E[w^2 | K = k] f_K(k) = \sum_{k=0}^{\infty} E\left[\prod_{i=0}^k \left(1 - \frac{\Sigma(x_i)}{p\Sigma_{maj}}\right)^2\right] f_K(k)$$

Itt $f_K(k) = \frac{(p\Sigma_{maj}L)^k}{k!} e^{-p\Sigma_{maj}L}$ annak az eseménynek a valószínűsége, hogy az ütközések száma

éppen k (ez a $p\Sigma_{maj}L$ paraméterű Poisson eloszlás súlyfüggvénye). Az egyszerűbb követhetőség kedvéért először a momentumokat számoljuk ki homogén esetre, vagyis amikor $\Sigma(x) = \Sigma$, $p\Sigma_{maj} = p\Sigma$:

$$E[w] = \sum_{k=0}^{\infty} \prod_{i=0}^k \left(1 - \frac{\Sigma}{p\Sigma}\right) f_K(k) = \sum_{k=0}^{\infty} \left(1 - \frac{1}{p}\right)^k f_K(k) \quad (18)$$

$$E[w^2] = \sum_{k=0}^{\infty} \prod_{i=0}^k \left(1 - \frac{\Sigma}{p\Sigma}\right)^2 f_K(k) = \sum_{k=0}^{\infty} \left(1 - \frac{1}{p}\right)^{2k} f_K(k)$$

(18)-ban $f_K(k) = \frac{(p\Sigma L)^k}{k!} e^{-p\Sigma L}$, mert homogén esetben K Poisson eloszlású $p\Sigma L$

paraméterrel, generátorfüggvényét (PGF: Probability Generating Function) a (19) formula adja meg.

$$g(s) = E[s^K] = \sum_{k=0}^{\infty} s^k f_K(k) = e^{p\Sigma L(s-1)} \quad (19)$$

Vegyük észre, hogy a w momentumait megkaphatjuk a generátorfüggvény segítségével (vö. (18) és (19)), és így kapjuk (20)-t.

$$\begin{aligned} E[w] &= g\left(1 - \frac{1}{p}\right) = e^{-\Sigma L} \\ E[w^2] &= g\left(\left(1 - \frac{1}{p}\right)^2\right) = e^{-\frac{2p-1}{p}\Sigma L} \end{aligned} \quad (20)$$

Észrevehetjük, hogy (20)-ban az első egyenlet éppen a Woodcock módszer torzítatlanságát igazolja (tekintsünk vissza a Woodcock módszert leíró 5.1 fejezetre, és képzeljünk p helyére a hatáskeresztmetszet és a majoráns hatáskeresztmetszet hányadosát!).

Most már behelyettesíthetünk (16)-ba:

$$r^2[M] = \frac{1}{N} \left(\frac{E[w^2]}{(E[w])^2} - 1 \right) = \frac{1}{N} \left(\frac{e^{-\frac{2p-1}{p}\Sigma L}}{(e^{-\Sigma L})^2} - 1 \right) = \frac{1}{N} \left(e^{\frac{2p-1}{p}\Sigma L} e^{2\Sigma L} - 1 \right) = \frac{1}{N} \left(e^{\frac{\Sigma L}{p}} - 1 \right) \quad (21)$$

Ha feltételezzük, hogy a futási idő lineáris p -ben, akkor a Figure of Merit érték számolása a futási idő és a relatív szórásnégyzet formulák alapján (22) szerint adódik, homogén közegre.

$$FoM(p) = \frac{1}{r^2[M](p)t(p)} = (r^2[M](p)t(p))^{-1} = \left(\frac{1}{N} \left(e^{\frac{\Sigma L}{p}} - 1 \right) \cdot (a + b \cdot p) \right)^{-1} \quad (22)$$

Ha $b \cdot p \gg a$ és $p \gg \Sigma L$ teljesül, akkor a FoM közelítőleg független lesz p -től, jelölje ezt az értéket FoM_{as} . Hagyjuk el a -t $b \cdot p$ mellett és elsőrendben közelítsük az exponenciális tagot, így kapjuk (23)-at.

$$FoM_{as} = \left(\frac{1}{N} \left(1 + \frac{\Sigma L}{p} - 1 \right) \cdot (b \cdot p) \right)^{-1} = \left(\frac{\Sigma L b}{N} \right)^{-1} = \frac{N}{\Sigma L b} \quad (23)$$

Ez természetesen megegyezik $FoM(p)$ függvény $p \rightarrow \infty$ -re vett határértékével:

$$\begin{aligned} \lim_{p \rightarrow \infty} FoM(p) &= \lim_{p \rightarrow \infty} \left(\frac{1}{N} \left(e^{\frac{\Sigma L}{p}} - 1 \right) \cdot (a + b \cdot p) \right)^{-1} = \lim_{p \rightarrow \infty} \left(\frac{b}{N} \left(e^{\frac{\Sigma L}{p}} - 1 \right) p \right)^{-1} = \\ &= \left(\frac{b}{N} \lim_{q \rightarrow 0} \frac{e^{\Sigma L q} - 1}{q} \right)^{-1} = \left(\frac{b}{N} \lim_{q \rightarrow 0} \frac{\Sigma L e^{\Sigma L q}}{1} \right)^{-1} = \left(\frac{b \Sigma L}{N} \right)^{-1} = \frac{N}{b \Sigma L} = FoM_{as} \end{aligned} \quad (24)$$

Ezek ismeretében, most tárgyaljuk a problémát általánosan, inhomogén esetben. Ekkor, (17)-ből kiindulva, a momentumokra a (25) összefüggést kapjuk.

$$E[w] = \sum_{k=0}^{\infty} E \left[\prod_{i=0}^k \left(1 - \frac{\Sigma(x)}{p\Sigma_{maj}} \right) \right] f_K(k) = \sum_{k=0}^{\infty} \prod_{i=0}^k \left(1 - \frac{E[\Sigma(x_i)]}{p\Sigma_{maj}} \right) f_K(k)$$

$$E[w^2] = \sum_{k=0}^{\infty} E \left[\prod_{i=0}^k \left(1 - \frac{\Sigma(x)}{p\Sigma_{maj}} \right)^2 \right] f_K(k) = \sum_{k=0}^{\infty} \prod_{i=0}^k \left(1 - 2 \frac{E[\Sigma(x_i)]}{p\Sigma_{maj}} + \frac{E[(\Sigma(x_i))^2]}{(p\Sigma_{maj})^2} \right) f_K(k) \quad (25)$$

Észrevehetjük, hogy w momentumait visszavezettük a közegre jellemző hatáskeresztmetszet-eloszlás momentumaira. Használjuk (26) jelöléseit:

$$\bar{\Sigma} = E[\Sigma(x_i)] \quad \text{és} \quad \bar{\Sigma}^2 = E[(\Sigma(x_i))^2] \quad (26)$$

Ekkor:

$$E[w] = \sum_{k=0}^{\infty} \left(1 - \frac{\bar{\Sigma}}{p\Sigma_{maj}} \right)^k f_K(k)$$

$$E[w^2] = \sum_{k=0}^{\infty} \left(1 - 2 \frac{\bar{\Sigma}}{p\Sigma_{maj}} + \frac{\bar{\Sigma}^2}{(p\Sigma_{maj})^2} \right)^k f_K(k) \quad (27)$$

Megismételve a homogén esetnél alkalmazott gondolatmenetet (generátorfüggvény felhasználása, behelyettesítés a (16)-ba):

$$r^2[M] = \frac{1}{N} \left(\frac{E[w^2]}{(E[w])^2} - 1 \right) = \frac{1}{N} \left(e^{\frac{\bar{\Sigma}^2 L}{p\Sigma_{maj}}} - 1 \right) \quad (28)$$

Láthatjuk, hogy a relatív szórásnégyzet továbbra is arányos $e^{\frac{c}{p}}$ -vel, ahol általános (inhomogén) esetben $c = \frac{\bar{\Sigma}^2 L}{\Sigma_{maj}}$, homogén esetben pedig $c = \Sigma L$. A FoM , és az aszimptotikus

FoM értékeket ezzel az észrevétellel könnyen megkaphatjuk:

$$FoM(p) = \left(\frac{1}{N} \left(e^{\frac{\bar{\Sigma}^2 L}{p\Sigma_{maj}}} - 1 \right) \cdot (a + b \cdot p) \right)^{-1} \quad (29)$$

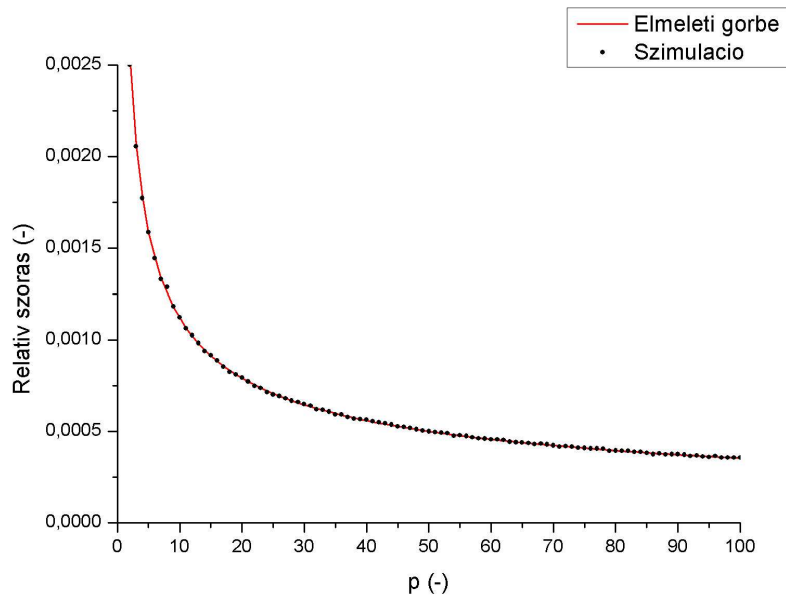
$$FoM_{as} = \frac{N\Sigma_{maj}}{\bar{\Sigma}^2 L b} \quad (30)$$

Láthatjuk, hogy a homogén esetet tartalmazza az általános (29) formula, hiszen homogén esetre épp (22)-tal ekvivalens. Tetszőleges hatáskeresztmetszet eloszlásra meghatároztuk tehát

a beütésszámok becslésének relatív szórásnégyzetét, a FoM értéket pedig p -vel egyenesen arányos futási időre. A (29) és (30) formulák azt mutatják, hogy a hatékonyság szempontjából legjobb választás ilyen esetben a $p = \infty$, vagyis gyakorlati szempontból (a diszkrét hatáskeresztmetszet eloszlás miatt) a determinisztikus mintavételezés. A térfogat diszkretizálása (voxelekre osztása) miatt ugyanis megtehetjük azt, hogy az egyenes (ami mentén vizsgáljuk a sugárgyengülést) és a voxelhatárok metszéspontjait kiszámítjuk, ebből ismert lesz a voxelekben befutott úthossz, majd a Beer-Lambert törvény alapján analitikusan megkapjuk a beütésszámok várható értékét. Ez a módszer egy 0 szórású, véges idő alatt végrehajtható eljárás, vagyis a $p = \infty$ -nél mindenképp hatékonyabb.

Ha lehetőségünk van olyan architektúrán végrehajtani a szimulációt, ami támogatja a párhuzamos feldolgozást, más eredményt kaphatunk. Ezért a vizsgálatot célszerű továbbvinni grafikus kártyán való implementáció szempontjából. Ezt megvalósítottam a diplomaunkámban, a rekonstrukcióban azonban nem a FoM értéket vizsgáltam, hanem az elérhető legjobb képminőséget adott rekonstrukciós idő alatt. A dolgozatban így csak a képrekonstrukcióban való alkalmazás kerül bemutatásra, a további vizsgálat pedig a jövő kutatási tervei közé sorolandó. Lényeges különbség, hogy a Figure of Merit érték helyett adott idő alatt elérhető képminőséget vizsgáltam, mert ez utóbbi több, mint pusztán a Monte Carlo projektor hatékonysága. Ahhoz, hogy elméletileg meghatározzuk az elvárt eredményeket (vagyis hogyan függ a képminőség a p paramétertől), a modellbe bele kellene vennünk a visszavetítés folyamatát, a változtatható paramétereket, rekonstrukciós időt befolyásoló beállításokat.

A (29) elméleti formula validálására készítettem egy egydimenziós részecsketranszport szimulációt, amiben az elnyelődés az egyetlen lehetséges fizikai folyamat. A p paraméter hatását vizsgáltam az egydimenziós közegen áthaladt részecskék becslésének (a fentiekben ismertetett M becselővel) relatív szórásnégyzetére. Az (16) képletben szereplő súly-momentumokat is, csakúgy, mint magát az áthaladt részecskék számát, statisztikai becsléssel közelítettem. A p függvényében felvettem a szórásnégyzetek grafikonját, ez látható az 5. ábrán. Az ábrán feltüntettem az elméleti görbét is, ami egyezik a szimuláció eredményével.



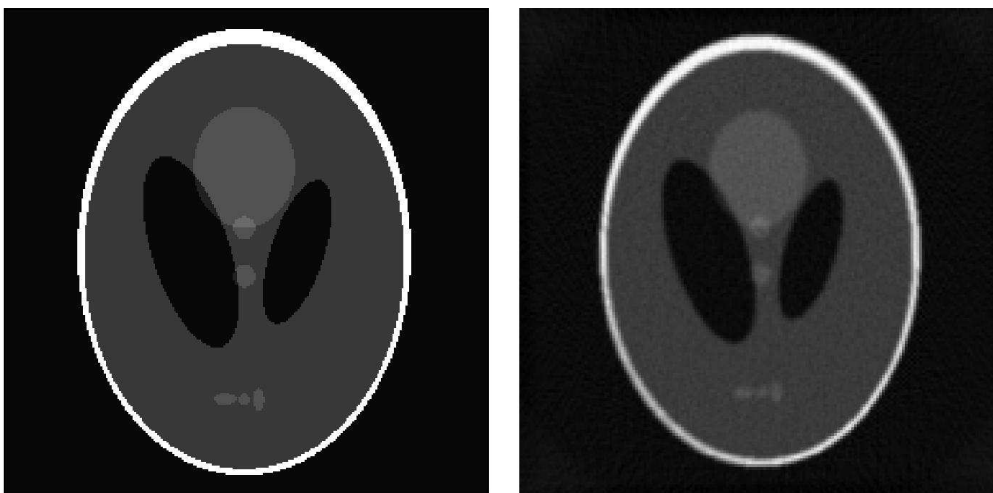
5. ábra
Relatív szórás (28) szerint (piros), szimuláció eredménye (fekete)

6. Az előrevetítés verifikációja

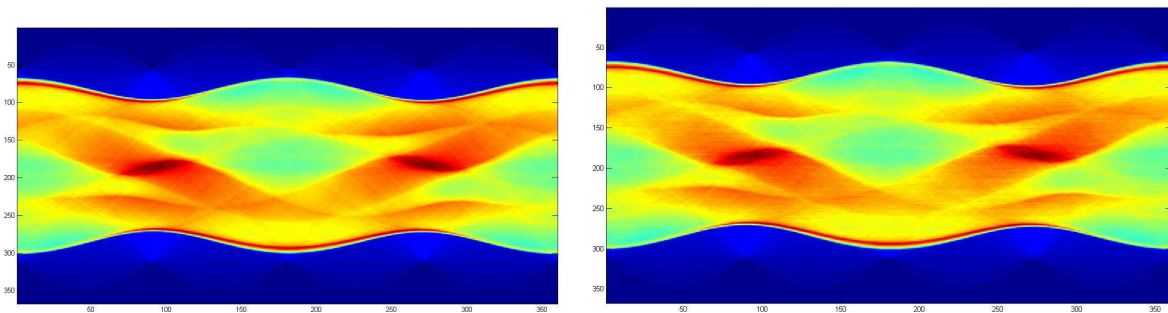
Alapvető elvárás a matematikai fantomokon elvégzett tesztek során, hogy a kiindulási alap, ami az iteráció „hajtóerejét” adja, lehetőleg minél kevesebb zajjal legyen terhelt. Vagyis a mérés szimulációjából származó szinogram lehetőleg közel tökéletesen egyezzen egy ideális mérés eredményével, vagyis épp a skaláreloszlásunk Radon-transzformáltja legyen. Ennek ellenőrzésére felhasználtam a Matlab *radon* nevű függvényét, amivel megkaptam a mesterségesen előállított fantom Radon-transzformáltját. Ezt vettem össze a saját szimulációm eredményével. Ez a gyakorlatban azt jelentette, hogy az előrevetítést elvégeztem az eredeti fantomon, összeszámoltam az egyes detektorpixelekre eső fotonok számát (a (15) képletben az \hat{Y} -al jelölt mennyiség), majd az ezen pixel irányába indított részecskék számával osztva vettem a hányados logaritmusát. Ez természetesen nem más, mint a gyengítési együttható eloszlás vonalintegráljának közelítése, a forrást a detektorpixellel összekötő LoR mentén (Beer-Lambert törvény). Ekképpen az előrevetítés egy becslést ad a kép szinogramjára, amit össze lehet hasonlítani a Matlab által készített szinogrammal⁷. Ezt megtettem a Shepp-Logan fantom esetén (6. ábra balra), ez látható a 7. ábrán. Azt tapasztaljuk, hogy kellően sok szimulált részecske esetén ($\sim 10^9$), a szinogramok szinte

⁷ A fan-beam geometria miatt a szimuláció szinogramja eltér a Matlab paralell-beam geometriára értelmezett Radon-transzformáltjától, így a `fan2para` parancsot kell alkalmazni a szimulált szinogramra

tökéletesen megegyeznek. A verifikációt meg lehet erősíteni azzal, hogy az előrevetítés eredményéből megpróbáljuk egy validált rekonstrukcióval visszaállítani az eredeti fantomot. Például alkalmazhatjuk a Matlab *iradon* függvényét a saját előrevetítem által generált szinogramra. Ezt a rekonstrukciót látjuk a 6. ábrán jobbra. Innen meggyőződhetünk arról, hogy a Monte Carlo projektor megfelelően működik, hiszen a Matlab által rekonstruált kép közelítőleg visszaadja az eredetit (a szűrt visszavetítésre jellemző műtermékek mellett nincs egyéb szisztematikus hiba), természetesen a tökéletlen leképezés, és a rekonstrukció pontatlansága miatt nem egyezik meg vele.



6. ábra
Shepp-Logan fantom (balra) és a saját előrevetítésből rekonstruált kép (jobbra)



7. ábra
A Matlab Radon-transzformáltja (balra) és a saját előrevetítem eredménye (jobbra) a Shepp-Logan fantomról

7. A rekonstrukció működőképességének demonstrálása

7.1 Az L2 és CC norma

Alapvetően az a célunk, hogy eldöntsük, mennyire hasonlít a rekonstruált kép az eredeti fantom gyengítési együttható eloszlására. Ennek vizsgálatára több módszer is létezik, az egyik legegyszerűbb ilyen az L2 norma kiszámítása. Az L2 normát a két kép, vagy képrészlet között az alábbi formában definiáljuk:

$$L2 = \frac{\sum_{i=1}^N (M_i^{eredeti} - M_i^{rekon})^2}{\sum_{i=1}^N (M_i^{eredeti})^2} \quad (31)$$

ahol M_i a kép(részlet) i . pixeljéhez tartozó érték (jelen esetben lineáris gyengítési együttható). Minél jobban hasonlít az eredeti eloszlásra a rekonstruált kép, annál jobb a képminőség, és annál kisebb az L2 norma.

Ugyanez vonatkozik a CC normára is, amit (32) szerint definiálok:

$$CC = 100 \cdot \left(1 - \left| \frac{\text{cov}(M^{eredeti}, M^{rekon})}{\sqrt{\text{cov}(M^{eredeti}, M^{eredeti}) \cdot \text{cov}(M^{rekon}, M^{rekon})}} \right| \right) \quad (32)$$

Eszerint a CC normát %-ban kapjuk meg. A CC norma invariáns a kép lineáris transzformációjára, olyan értelemben, hogy a képpixelek által felvett egyes értékek ugyanazzal a számmal való szorzása, vagy ezekhez ugyanazon szám hozzáadása nem változtat a CC normán.

Gyakran érdemes ábrázolni a CC vagy az L2 norma értékeit az elvégzett iterációk függvényében, annak érdekében, hogy a módszer konvergenciájáról kapjunk információt. Ezen függvény grafikonjára CC-görbéként vagy L2-görbéként fogok hivatkozni a következőkben.

7.2 Konvergencia

Az 1.2 fejezet rámutatott arra a pusztán matematikai tényre, hogy a maximalizálandó likelihood-függvénynek akkor van minimuma, ha több LoR-t jelölünk ki az iterációban, mint ahány voxelből áll a rendszer. A visszavetítésben ehhez mérten kell megválasztanunk a

válaszgyenesek számát. Ez szükséges, de nem elégséges feltétele a konvergenciának, hiszen a (15) iterációban valójában várható értékek szerepelnek, amiket csak (valamilyen hibával) becsülni tudunk. A módszer konvergenciájáról további elméleti megfontolásokat olvashatunk még [1]-ben, ahol megmutatták, hogy az EM algoritmus konvergens, ha az említett várható értékeket nagy pontossággal tudjuk becsülni (minden határt túllépő pontossággal), illetve a (12) sorfejtés érvényes. A gyakorlatban ez nem múlik máson, mint a szimuláció statisztikáján. Tekintve (15)-öt észrevehetjük, hogy az iteráció hajtóereje az előrevetítés eredménye, vagyis az, hogy mennyire tér el az aktuálisan iterált gyengítési együttható eloszlás mellett mérhető beütésszám a valós mérés beütésszámaitól. Ennek belátására tekintsük (15) iteráció fixpontját:

$$\mu[j] = \frac{\sum_{lor \in L_j} (X_{lor}[j] - X_{lor}[j+1])}{\sum_{lor \in L_j} \left(\frac{X_{lor}[j] + X_{lor}[j+1]}{2} \right) \cdot L_{lor}[j]} \quad (33)$$

A fixpontban ugyanis $Y_{lor} = \hat{Y}_{lor}$. A számlálóban a j . voxelben elnyelődött részecskék számának összege marad (ΔN), a nevezőben viszont a befutott úthosszakkal (l) súlyozott összegét kapjuk a j . voxelbe bejutott és abból kimenő fotonszámok átlagának (N). Ezekkel az egyszerűsített jelölésekkel a kérdéses fixpontot (34) alakba írhatjuk át.

$$\mu = \frac{\overline{\Delta N}}{\overline{Nl}} \quad (34)$$

A (34)-ben $\bar{\quad}$ -sal jelölt átlagot a LoR-okra kell elvégezni, vagyis ha feltesszük, hogy N értéke közel azonos a LoR-ok közt, vagyis minden válaszgyenes mentén hozzávetőlegesen ugyanannyi részecske halad, akkor N -et kihozhatjuk a várható értékből, és máris a (35) összefüggéshez jutunk.

$$\mu \bar{l} = \frac{\overline{\Delta N}}{\overline{N}} = 1 - \frac{N - \overline{\Delta N}}{N} \quad \Rightarrow \quad 1 - \mu \bar{l} = \frac{N - \overline{\Delta N}}{N} \quad (35)$$

(35) pedig nem más, mint a Beer-Lambert törvény közelítése $\mu \bar{l} \ll 1$ esetben.

Mindezzel arra igyekszem rávilágítani, hogy az algoritmus konvergenciája három dolog által van biztosítva:

- 1) Több LoR-t jelölünk ki, mint ahány voxelre bontjuk a rekonstruálandó térfogatot
- 2) A voxelméret (d) elég kicsi, a gyengítési együtthatóval vett szorzatra $\mu d \ll 1$ áll fenn
- 3) Az előrevetítés és a mérés/mérés-szimuláció statisztikája kielégítően pontos (hiszen az $Y_{lor} - \hat{Y}_{lor}$ különbség hajtja az iterációt a fixpontja felé!)

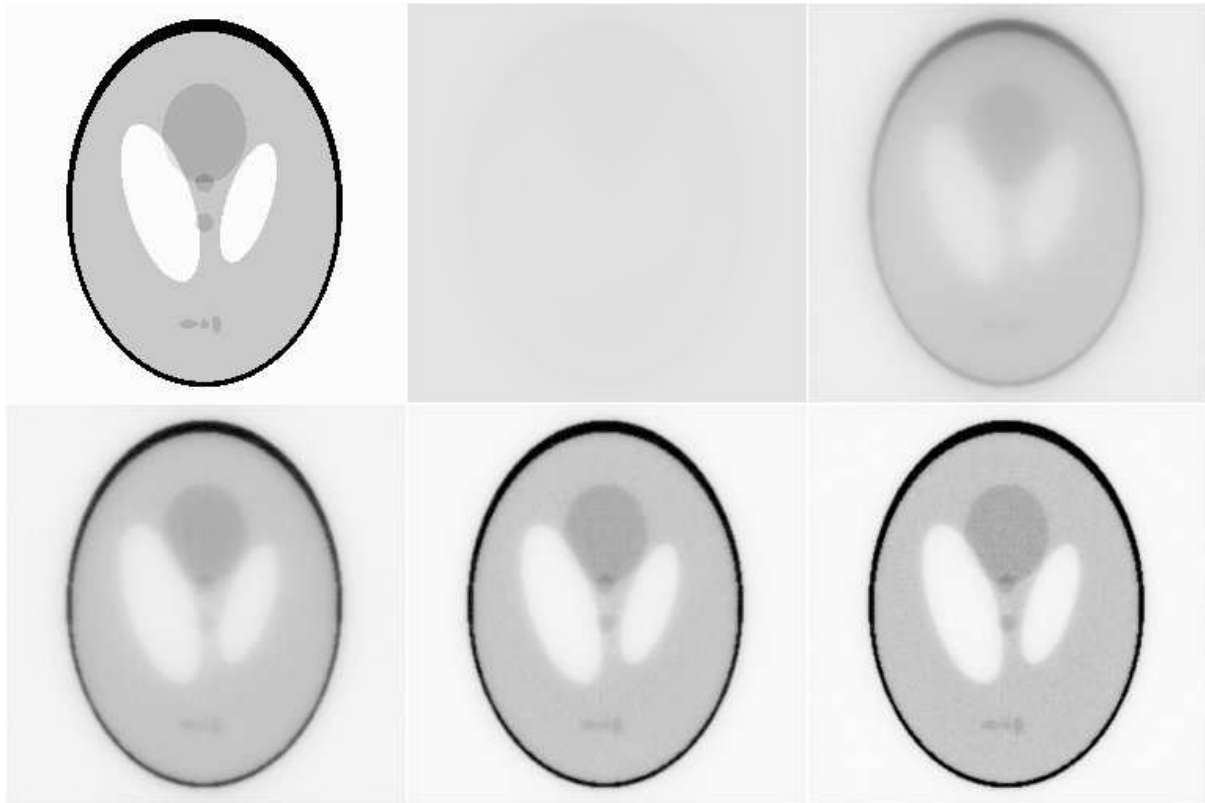
7.3 Várakozások az ML-EM algoritmussal szemben

A szűrt visszavetítéssel szemben egy iteratív, statisztikus képrekonstrukcióval az előállított kép minőségbeli javulását érhetjük el több számítási erőforrás felhasználásával. Az FBP-vel összehasonlítva jellemző, hogy jobb a rekonstruált kép kontrasztja, jobb a jel/zaj arány [5], valamint a háttér műtermékektől mentes. Az általam implementált ML-EM képrekonstrukció minősége azon múlik, hogy milyen precizitással tudjuk becsülni a (15) képletben szereplő változókat. Végtelen sok iterációt elvégezve, a rekonstruált kép ideális esetben a valós gyengítési együttható eloszlást adja vissza, de mivel a becsült változók zajjal terheltek, valamint a bemenő nyers adat (CT felvétel) sem tökéletes, így az iteráció nem fogja visszaadni a valódi eloszlást. Általában a numerikus pontatlanságok miatt az iterációk számának növelése egy idő után ront a képminőségen. Ezt a viselkedést lehet korrigálni például regularizációs eljárásokkal, ellenben ha nem alkalmazunk ilyet, akkor megállási feltételt kell szabni a rekonstrukciónak, vagyis az iterációt meg kell állítani az elérhető legjobb minőségű (például L2 normában az eredetihez legközelebb lévő) képnél. A következőkben gyakran fogok hivatkozni az elérhető legjobb L2 normára, ami alatt pontosan ezt, az iteráció során elérhető legjobb kép minőségét értem.

7.4 Shepp-Logan fantom rekonstrukciója jó minőségű mérési szimuláció esetén

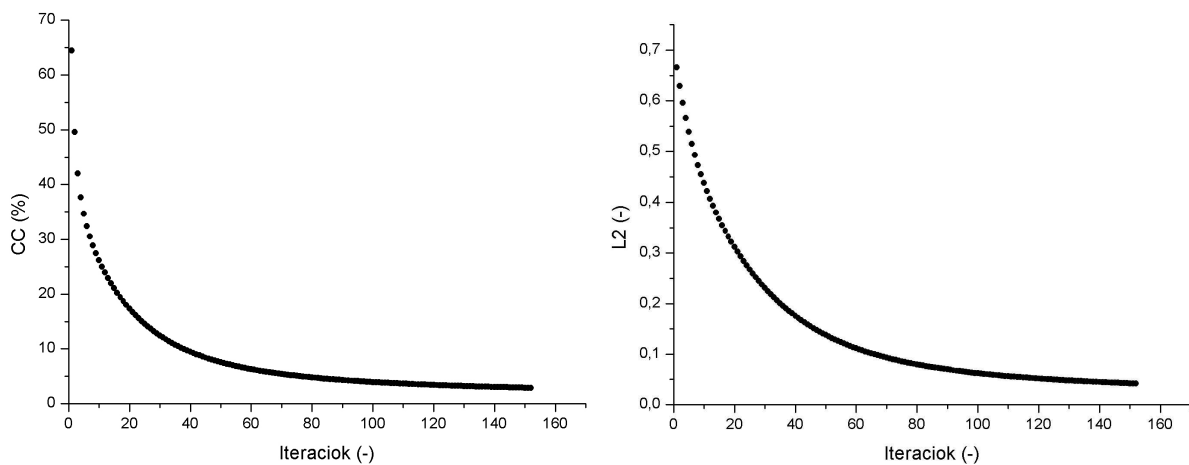
A következőkben a rekonstrukció működését szeretném demonstrálni, elsőként a gyakorlatban a rekonstrukciók minőségének ellenőrzésére előszeretettel használt [3] Shepp-Logan fantom rekonstrukciójával. Céлом ezzel az volt, hogy a paraméterek lehető legjobb beállítása mellett megmutassam az elérhető legjobb képminőségét, belátható idő alatt. Ezért projekciónként 10^9 nagyságrendben szimuláltam fotonokat a mérés szimulációjában, hogy a mérés minősége ne legyen korlátozó faktor, és a rekonstrukció szabad paramétereit is úgy választottam meg, hogy az egyes változókra adott becslés minél pontosabb legyen. Az egyes iterációk elvégzése után kapható képeket mutatja be (a szemléletesség kedvéért invertálva) a 8. ábra. Ábrázolásakor ugyanazt a szürkeskálát állítottam be mindegyik képnél, így a kontraszt javulása is megfigyelhető, valamint az eredeti képpel vizuálisan is torzításmentesen összehasonlítható a rekonstruált kép. Az első néhány iteráció után egy homályos képet kapunk, gyenge a kontraszt és életlen a kép, majd a kontraszt gyorsan javulni kezd (100.

iteráció), de a képélesség csak később kezd látványosan változni, jó minőségű képet például a 152. iterációnál kapunk. Mivel a mérés minősége, valamint a paraméterek megválasztása is megközelítette az ideálisat, ezért az iterációban számolt mennyiségek becslésének zajszintje nagyon alacsony volt. Az L2 görbe a 9. ábrán látható módon rajzolódott ki, nem tapasztaltam felzajosodási fázist (ld. 13. ábra), mivel ehhez akár több ezer iterációig is el kellett volna vinnem a vizsgálatot.



8. ábra

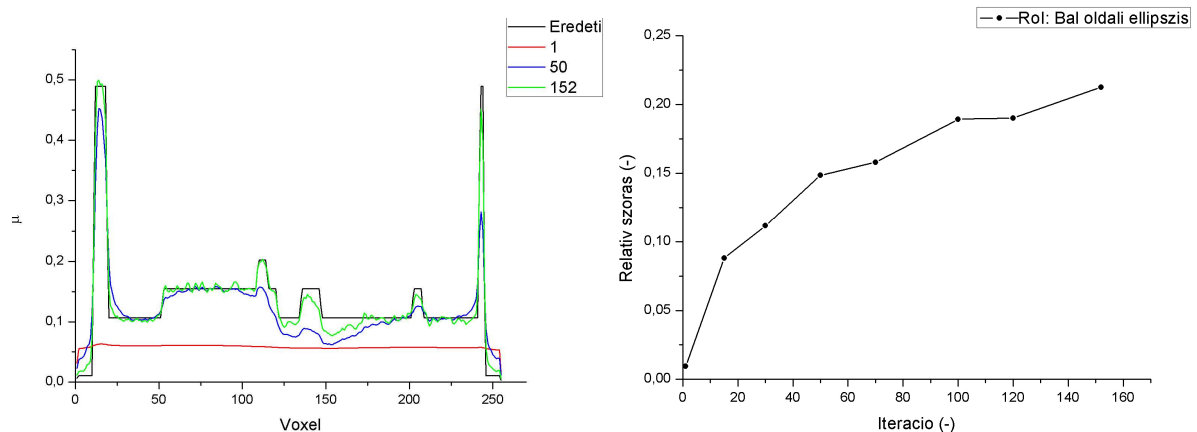
Balra, felül az eredeti kép, valamint a rekonstruált kép rendre 1, 15, 50, 100, 152 iteráció után



9. ábra

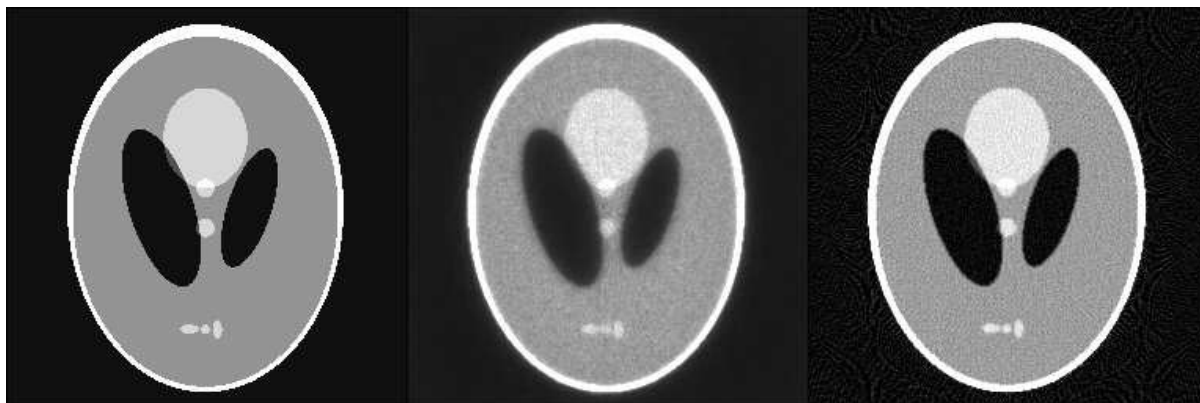
CC és L2 görbe a Shepp-Logan fantom rekonstrukciójánál

A rekonstrukció két fontos tulajdonságát demonstrálja a 10. ábra. Az iterációk során a kép kontrasztja és élessége egyre jobb lesz, azonban felerősödnek a pontatlan becslésekből eredő zajok. Az ábrán bal oldalt a függőleges középvonal mentén vett metszetet láthatunk, itt figyelhetjük meg a kontraszt és kontúr viszonyokat, a másik grafikonon pedig a bal oldali ellipszis tartalmazta voxelek gyengítési együtthatójának relatív szórása látható.



10. ábra
Shepp-Logan fantom függőleges középvonalán vett metszetek különböző iterációk után (jobbra), valamint a bal oldali ellipszis voxeleinek relatív szórása (balra)
Az ML-EM képrekonstrukció iterációs lépései során a kontraszt és képélesség jellemzők javulnak, de a jel/zaj arány romlik.

Összehasonlításra teremt lehetőséget a 11. ábra, ahol a Matlab *iradon* függvénye által rekonstruált kép mellett ábrázoltam a saját rekonstrukcióm, szintén ugyanazon a szűrkeskálán. Szembetűnő, hogy Matlab által rekonstruált kép háttérén látható műtermék az ML-EM rekonstrukcióban nem jelenik meg, viszont a kép életlenebb, ami szintén jellemző az iteratív rekonstrukcióra. Globális képminőségi jellemzőket a 1. táblázatban találunk, ami feltünteti a keresztkorrelációs normát (ld. (32) definíciót). A szűrt visszavetítést a Matlab által felkínált összes szűrővel elvégeztem, és feltüntettem a CC normát minden esetben. A legjobb képminőséget adó szűrő ebben az esetben a Shepp-Logan szűrő volt (ezt, a legjobb rekonstruált képet tüntettem fel a 11. ábrán).



11. ábra
Shepp-Logan fantom (balra), ML-EM rekonstrukció - 152. iteráció (középen),
Matlab rekonstrukciója – Shepp-Logan szűrő (jobbra)

	RamLak	Cosine	Hann	Hamming	SheppLogan	ML-EM
CC(%)	4,518	4,579	4,855	4,791	4,491	2,870

1. Táblázat
Rekonstruált képek CC normája: Matlab rekonstrukció különböző szűrőkkel,
ML-EM rekonstrukció 152 iterációval

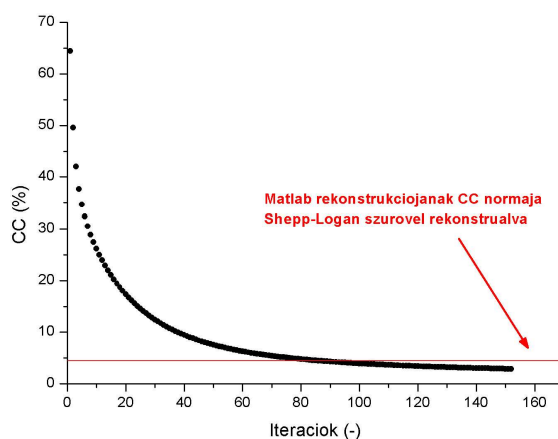
Grafikus információ formájában az 12. ábrán azt láthatjuk, hogy a Matlab által rekonstruált kép CC normában mért minősége elmarad az ML-EM rekonstrukciótól, ~100 iteráció elvégzése után jobb képet kapunk, mint a szűrt visszavetítéssel. Ugyanígy az L2 normákat tekintve az mondhatjuk, hogy elérhető ~30%-os javulás képminőségben a Matlab szűrt visszavetítéséhez képest (ML-EM L2: 0,0417, Matlab L2: 0,0604), úgy, hogy nem értük el a lehető legkisebb értéket az iteráció során (láthatjuk, hogy 100 iteráció fölött a konvergencia már nagyon lassú, a későbbiekben látunk majd olyan esetet, ahol eléri a görbe a minimumát).

A 2. táblázat a kontrasztról és jel/zaj arányról ad felvilágosítást. A vizsgálat során két RoI-t jelöltem ki: a felső nagy, világos kört, és a bal oldali sötét ellipszist. A Matlab szűrt visszavetítése a kis elnyelésű képrészeket alul, a nagy elnyelésűeket felülbecsülte, így jobb a két RoI közötti kontraszt, azonban a kis elnyelésű részekben jóval nagyobb zajjal terhelt, mint az ML-EM rekonstrukció. Valószínűleg ebből a zajból ered az eredeti képtől L2 normában való eltérése is.

	Átlagos érték			Relatív szórás		
	Eredeti	ML-EM	Matlab	Eredeti	ML-EM	Matlab
Felső kör	0,155	0,154	0,159	0	0,052	0,044
Bal oldali ellipszis	0,011	0,019	0,008	0	0,211	0,750
Kontraszt	0,867	0,780	0,904			

2. Táblázat
Kontraszt és jel/zaj arány a Shepp-Logan fantom rekonstrukciójánál

Összefoglalva azt mondhatjuk, hogy a Matlab rekonstrukció jobb kontrasztot ad ebben az esetben, de sokkal zajosabb, és ahogy azt az L2 norma is mutatja, jobban eltér az eredeti fantomtól. ML-EM rekonstrukcióval mintegy 30%-os javulás érhető el L2 értékben. Ha nem „tökéletes” mérésből indulunk ki, vagyis a szinogram generálásra nem számunk ennyi számítási időt, akkor a várakozásainknak megfelelően a szűrt visszavetítéssel szemben még többet nyerhetünk. A fenti esetben a mérést Monte Carlo alapú előrevetítéssel generáltam, mintegy 10^9 foton szimulációjával projekciónként, a következőkben szándékosan gyengébb minőségű mérésből indultam ki.



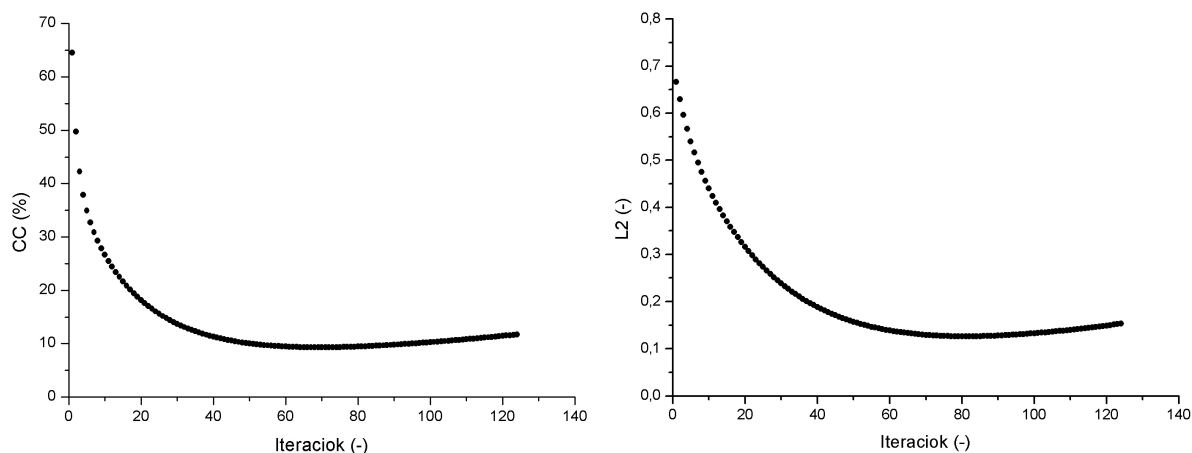
12. ábra
CC görbe a Shepp-Logan fantom rekonstrukciójánál, feltüntetve a Matlab által rekonstruált kép CC normája

7.5 Rossz jel/zaj arányú felvételek rekonstrukciója, dóziscsökkentés CT vizsgálatoknál

Mivel a CT vizsgálatok dózisterhelése egy nagyságrenddel meghaladja egyéb orvosi képalkotó eljárásokat, fontos célja és kihívása az iteratív rekonstrukciós eljárásoknak a minél alacsonyabb dózisu felvételekből történő képszintézis (low dose CT). [13]-ban megmutatták, hogy a mellkasi CT felvételeknél érdemes alacsonyabb anódáram mellett elvégezni a mérést, ezáltal csökkentve a páciens sugárterhelését. Ez elsősorban magas kockázatú betegeknél jelentős eredmény (pl. dohányzóknál). Alacsony dózisu felvétel rekonstrukciója természetesen a rekonstruált kép minőségének romlásával jár, [13]-ban elsősorban a kép felzajosodását említették meg. Statisztikus, iteratív rekonstrukcióval végzett kutatásról számolnak be [14]-ben (hasi CT vizsgálatok), a hagyományos szűrt visszavetítéshez rendelt mérési protokoll helyett alacsony dózisu felvételekből ASIR (Adapted Statistical Iterative

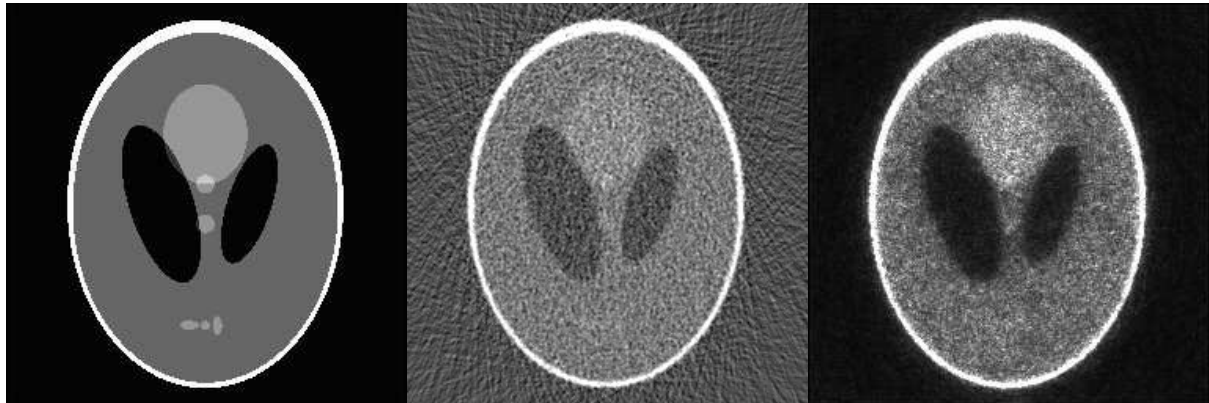
Reconstruction) technikával végeztek rekonstrukciót. Ezáltal 23-66%-os dóziscsökkentést lehetett elérni a pácienseknél, míg a képminőség diagnosztikai szempontból megfelelő maradt (egyedül a képélesség romlását emelik ki). Érdekes tehát megvizsgálni az ML-EM képrekonstrukciót is gyengébb minőségű mérés esetén. Az általam írt előrevetítéssel szimuláltam „zajos” mérési eredményeket, azaz a Monte Carlo alapú projektorral kisebb fotonszámmal generáltam szinogramot, növelve ezzel az eddigiekben látott rekonstrukciókhoz képest a kiinduló mérési adatok statisztikai szórását (jó minőségű mérési szimulációban 11000beütés/pixel, zajos „mérésnél” 40/pixel statisztika volt jellemző).

A gyengébb minőségű detektoradatok az iteráció konvergenciáját módosítják, az L2 görbe minimumhelye eltolódik az alacsonyabb iterációs szám felé, értéke nagyobb lesz, és azt elérve sokkal jobban felzajosodik rekonstrukció, mint precíz mérési adatok esetén (vö. 9. ábra és 13. ábra). Az elért eredményeket ugyanazon rekonstrukciós paraméterekkel értem el, mint a jó minőségű mérés szimuláció esetében, a rekonstruált képet szintén a Matlab szűrt visszavetítésével vetettem össze, ami megtekinthető a 14. ábrán.



13. ábra
CC és L2 görbék zajos mérési adatok esetén, Shepp-Logan fantomnál
A rekonstrukció azonos paraméterekkel történt, mint a jó minőségű mérési szimuláció esetén

Az ML-EM módszerrel rekonstruált kép láthatóan sokkal kevésbé terhelt zajjal, mint a szűrt visszavetítéssel rekonstruált, elsősorban a háttér zajossága szembetűnő. Kvantitatív eredmények is ezt demonstrálják (3. Táblázat), ismét két RoI-t jelöltem ki: a felső, világos kört, és a bal oldali ellipszist.

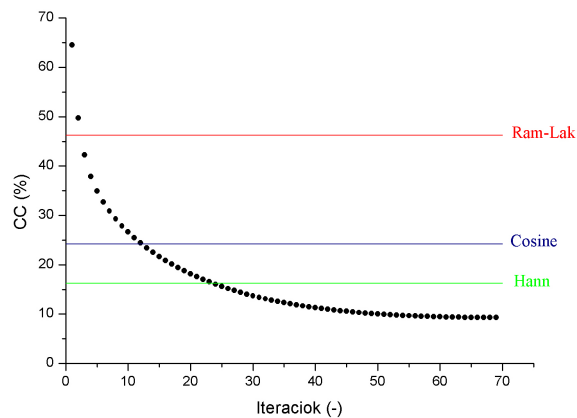


14. ábra
Shepp-Logan fantom rekonstrukciója zajos mérési adatok esetén
Eredeti fantom (balra), Matlab szűrt visszavetítése Hann szűrővel (középen),
ML-EM rekonstrukció (jobbra)

	Átlagos érték			Relatív szórás		
	Eredeti	ML-EM	Matlab	Eredeti	ML-EM	Matlab
Felső kör	0,155	0,15	0,16	0	0,253	0,381
Bal oldali ellipszis	0,011	0,025	0,008	0	0,28	6,625

3. Táblázat
A rekonstruált kép statisztikai jellemzői két RoI esetén, zajos mérési adatok mellett

A Matlab rekonstrukció elvégzésekor megvizsgáltam különböző szűrők esetén a rekonstruált kép CC normáját (4. Táblázat), és összehasonlítottam az ML-EM rekonstrukcióval (ez a 4. Táblázat mellett az 15. ábrán is látható).



15. ábra
CC görbe zajos mérési adatok rekonstrukciójánál, színessel a szűrt visszavetítés által rekonstruált képek CC normája feltüntetve

	Cosine	Hann	Hamming	Ram-Lak	Shepp-Logan	ML-EM
CC (%)	24,226	16,202	17,963	46,248	38,822	9,3

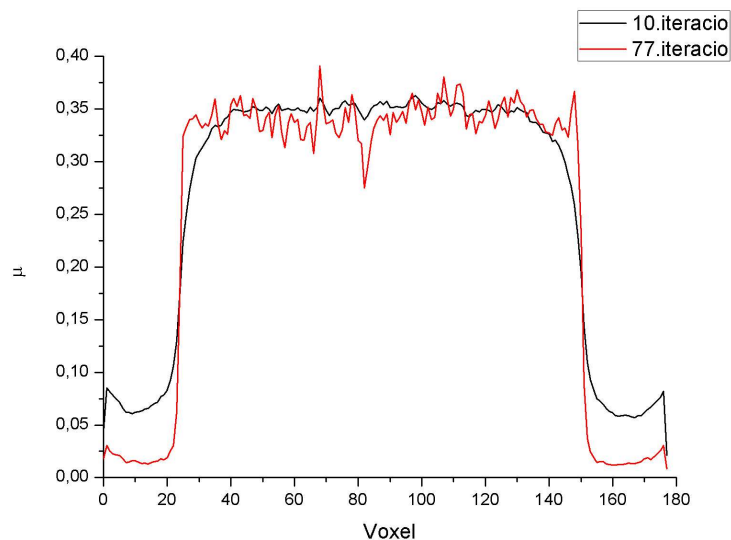
4. Táblázat
Rekonstrukció minősége zajos mérési adatokból kiindulva

A Matlab szűrt visszavetítésének (Hann szűrővel) L2 normája 0,247 volt, míg ML-EM rekonstrukció esetén a 77. iteráció 0,126-os L2 értéket adott, vagyis ~kétszeres javulást ért el az iteratív rekonstrukció. Az 15. ábrán megfigyelhetjük, hogy már 25 iteráció is jobb CC normában, mint a szűrt visszavetítés, rossz szűrőválasztás esetén pedig kifejezetten rossz minőségű képeket kapunk.

A diplomamunka kereteit meghaladta a problémakörnek részletes tárgyalása, de sikerült demonstrálni, hogy érdemes a jövőben ilyen irányban további tesztet végrehajtani, fejleszteni az algoritmust zajos mért adatok rekonstrukciójának szempontjából.

7.6 Demonstráció további fantomokon

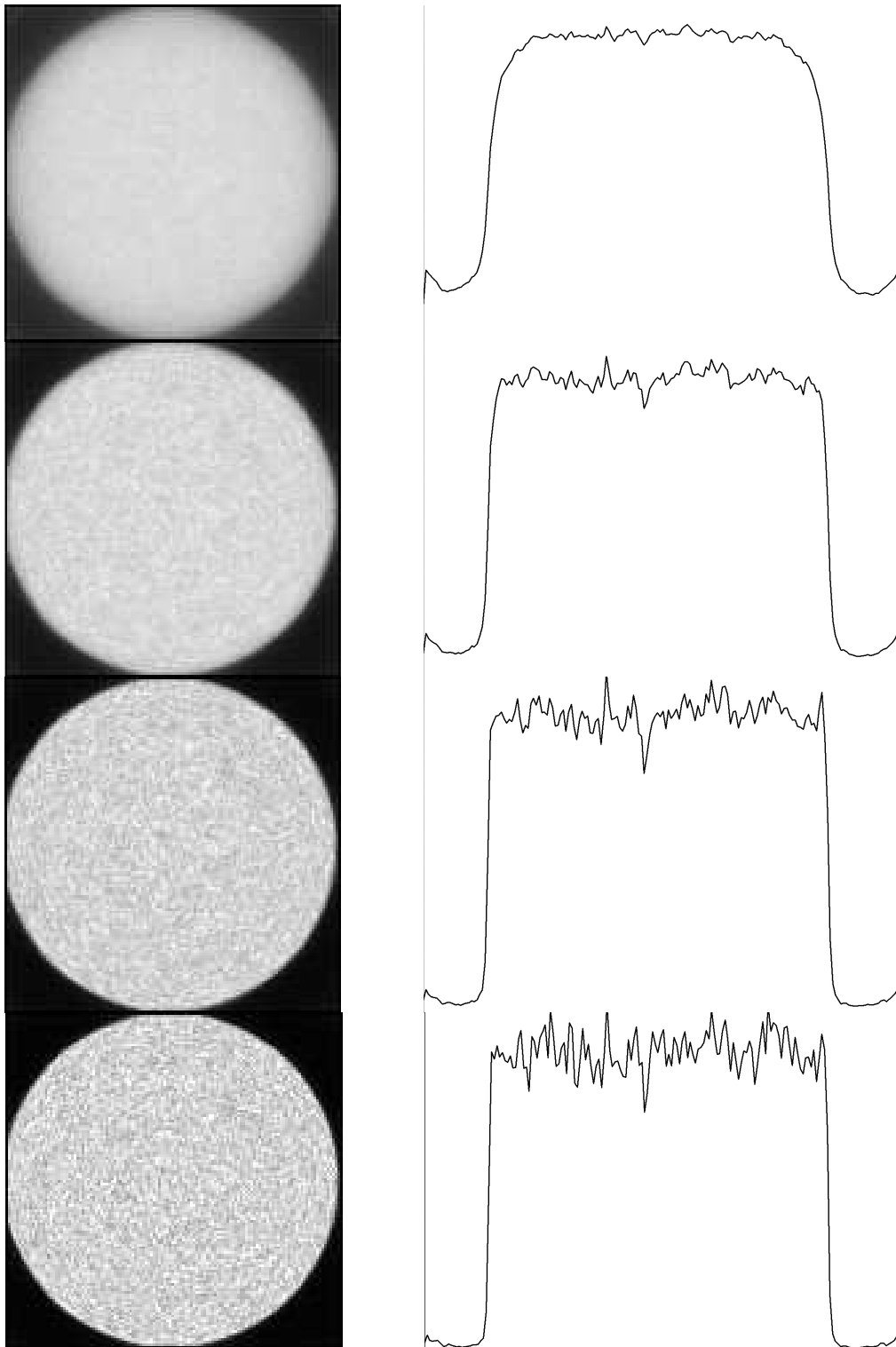
További két fantom rekonstrukcióját mutatom be a következőkben. A képekkel és grafikonokkal főleg az eddig levont következtetéseket kívánom alátámasztani más fantomok rekonstrukciójával is, valamint a rekonstrukció értékhelyessége irányában végzett vizsgálatot ismertetem. Az 17. ábrán a „kör fantom” rekonstrukciója látható, néhány iterációnál ábrázolva a rekonstruált kép, valamint a képátlón vett metszet. Az ábra a Shepp-Logan fantom eseténél jobban szemlélteti az ML-EM rekonstrukció tulajdonságait: az iterációk számának növelésével javuló kontraszt és képélességet, valamint a növekvő zajszintet, a 10. iterációnál és a 77. iterációnál vett metszeteket külön ábrázoltam az 16. ábrán, amin ez még inkább megfigyelhető.



16. ábra

A 10. és 77. iteráció után vett metszet

Kontraszt (szélek és kép közepének különbsége) és képélesség (meredekebb lefutás) javulása, zajok felerősödése az iteráció során



17. ábra

**Rekonstruált kép és képtől vett metszet rendre 10, 30, 77 150 iteráció után
A legjobb L2 normája a 77 iterációval rekonstruált képnek volt**

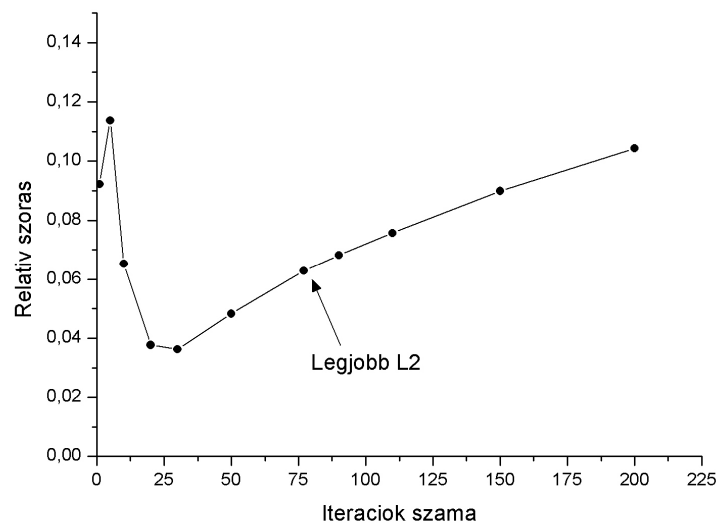
A kör rekonstrukcióról összesítő statisztikai adatokat az 5. Táblázat mutat.

Iteráció	Átlag	Szórás	Minimum	Maximum	Relatív szórás
Eredeti	0,3419	0	0,3419	0,3419	0
1	0,1311	0,01208	0,1034	0,1501	0,0922
5	0,3301	0,0375	0,2211	0,3848	0,1137
10	0,3356	0,0219	0,2450	0,3727	0,0653
20	0,3384	0,0127	0,2745	0,3799	0,0377
30	0,3397	0,0123	0,2893	0,3916	0,0362
50	0,3404	0,0164	0,2765	0,4165	0,0483
77	0,3408	0,0214	0,2614	0,4428	0,0628
90	0,3408	0,0232	0,2596	0,4459	0,0681
110	0,3408	0,0258	0,2569	0,4569	0,0757
150	0,3409	0,0307	0,2361	0,4741	0,0899
200	0,3410	0,0358	0,2238	0,4955	0,1043

5. Táblázat

A kép kör részéről vett statisztikák különböző iterációknál, **legjobb L2 normát a 77. iteráció adott**

Ha vizsgáljuk a homogén kör alakú területen a voxelek gyengítési együtthatójának relatív szórását, akkor azt tapasztaljuk, hogy alacsony iterációs számnál csökken, majd ~50. iteráció után elkezdi növekedni, a legjobb L2 normát a 77. iterációnál érjük el, de ekkor a rekonstruált képen már kicsit romlott a jel/zaj arány. Ezt szemlélteti az 18. ábra.

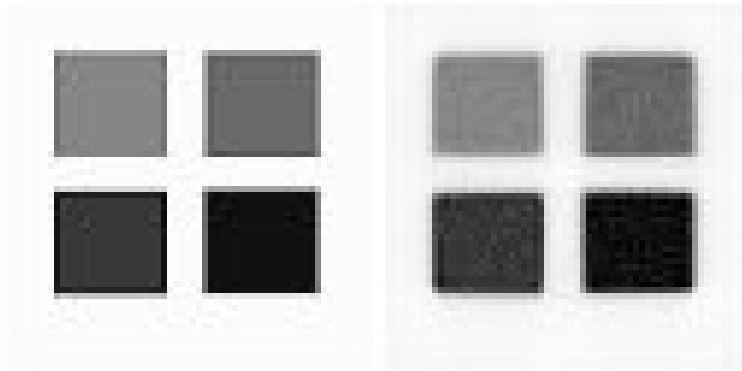


18. ábra

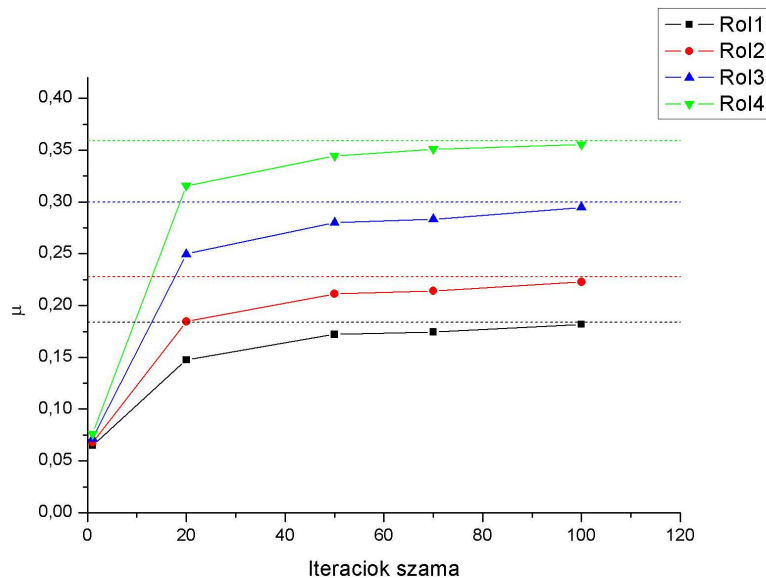
Relatív szórás a kör alakú fantomrészről az iterációk számának függvényében

Különböző gyengítési együtthatójú homogén képrészletek rekonstrukcióját is elvégeztem („négyzetek fantom”). Az eredeti és rekonstruált kép látható a 19. ábrán. A négyzeteket, mint RoI-kat jelöltem ki bal felsőtől a jobb alsóig 1-4 számozásban, és ezeken számoltam statisztikai mutatókat. Az 20. ábrán láthatjuk, hogy a rekonstrukció értékhelyes, vagyis az

iterációk számának növelésével a rekonstruált kép(részletek) átlagban az eredeti kép(részlet)hez tartanak.



19. ábra
Négyzetek fantom és rekonstrukciója



20. ábra
A négyzetek átlagos gyengítési együtthatójának konvergenciája az eredeti fantom értékeihez

8. Paraméter optimalizáció

A rekonstruált kép minőségének elemzése során a legnagyobb kihívást az algoritmus szabad paramétereiben való optimalizálás jelenti. Ez egy sokdimenziós paramétertérben történő szélsőérték keresést jelent, aminek során a képi minőséget az L2 vagy CC normával jellemezzük, az állítható paraméterek pedig például az iterációszám, a szimulált fotonok száma, a válaszegyenések száma a visszavetítésben, a ray-marching algoritmus során

alkalmazott lépésköz, a választott voxelméret (és a voxelek száma), detektort elérő fotonok számának statisztikája. Ezeken kívül az optimum még magától a rekonstruálandó gyengítési együttható eloszlástól is függhet. Egy teljes körű optimalizáció során a paraméterteret kell feltérképezni meglehetősen sok mérési pont felvételével (a program többszöri futtatásával), ami önmagában hatalmas feladat, ráadásul praktikatlan, hiszen különböző fantomok esetén különböző eredményt kapunk az optimumra. Érdemes először a rekonstrukció minőségét, mint többváltozós függvényt, a paramétertér bizonyos egyenesei mentén mintavételezni, például leírni a viselkedését egyetlen paraméter változásának esetén. A fejezetben egy tipikus fantomra, ezen egyenesek mentén végzett vizsgálatokat mutatok be, vagyis mindig egyetlen paraméter képminőségre gyakorolt hatását vizsgálom, az összes többi paramétert fixen tartva. Azt kívánom elérni, hogy a teljes optimalizációhoz útmutatást adjak, feltárjam a rekonstrukció minőségének a paraméterek beállítására való érzékenységet, annak érdekében, hogy egy konkrét alkalmazás esetén (az alkalmazásra jellemző fantomokra) az optimalizáció lényegesen könnyebb legyen. Láthatjuk a későbbiekben, hogy bizonyos esetekben lehet csökkenteni a dimenziószámot, ha például két paramétert azok arányával helyettesítünk, és észrevesszük, hogy képminőség csak ettől az arányszámtól függ (a többi paramétertől eltekintve). A paraméterteret a következő mennyiségek változtatásával mértem fel (a számozás az alfejezeteknek megfelelően):

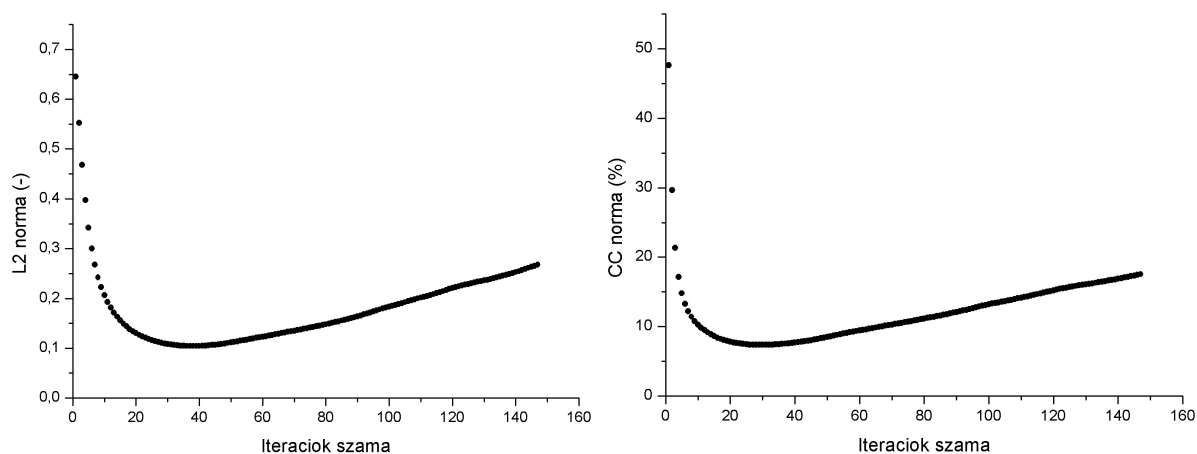
1. Az elvégzett iterációk száma
2. A rekonstrukció előrevetítésében indított fotonszám és a mérési szimulációban indított fotonok számának aránya
3. A visszavetítés ray-marching algoritmusában használt lépésköz és a voxelméret aránya
4. A visszavetítésben kijelölt válaszegyenesek száma és a voxelek számának aránya
5. Az előrevetítésben használt Monte Carlo projektor szabadúthossz mintavételezését torzító paraméter (úthossz-zsugorítás illetve úthossz-nyújtás mértéke)

Gyakorlatias okok miatt érdemes azt is vizsgálni, hogy a fent említett paraméterek közül melyikre legérzékenyebb a rekonstrukció a paramétertér egy adott pontjában. Ezzel útmutatást kaphatunk ugyanis, hogy az optimalizálandó többváltozós függvényünknek merre mutat a gradiense, vagyis mely paraméter(ek) változtatásával lehet a legtöbb javulást elérni L2 normában. Ez a paramétertér minden pontjában különböző lehet. Ahhoz, hogy minden pontban meg tudjuk mondani, melyik paraméter beállítására érzékenyebb a rekonstrukció, érdemes definiálni egy olyan referencia-értéket, ami alapján eldönthetjük, hogy a gradiensnek

az adott egyenes irányába mutató komponense jelentős-e, valamint kijelölhetünk ezáltal egy olyan pontot, aminél az érzékenység ugyanolyan viselkedést mutat különböző paramétereknél. Két eset lehetséges (ezek szerint különböző definícióval kell élnünk): amikor a paraméter növelése okoz képminőség javulást, illetve amikor a csökkentése. Előbbi esetben a referencia-értékre a definíció legyen a következő: válasszuk az adott paraméternek a legnagyobb olyan értékét, aminek megduplázása hatására az L2 norma 10%-nál többet nem csökken. Utóbbi esetben pedig: válasszuk az adott paraméternek a legkisebb olyan értékét, aminek felezése hatására az L2 norma 10%-nál többet nem csökken. Ez összehasonlítási alapot ad a különböző paraméterekre való érzékenységre: azt mondjuk, hogy az adott paraméterre érzékeny a képminőség, ha a rekonstrukcióban használt értéke kisebb/nagyobb, mint a referencia-érték, vagyis duplázás/felezés esetén több, mint 10%-ot javulna a képminőség. Ezt tisztán fogjuk majd látni az egyes konkrét esetekben.

8.1 A képminőség függése az iterációk számától

Az iterációk számának vizsgálati módszere némileg eltér a többi paraméter esetétől, illetve a levonható következtetések is másak, így logikus először ezt az elemzést megismerni. Ahogy a működőképesség demonstrációjánál láthattuk, néhány esetben (ideálisat megközelítő mérési eredmények, kis szórású becslések) nagyon sok iterációt végezhetünk el úgy, hogy az L2 norma monoton csökken. A numerikus pontatlanságok azonban azt okozzák, hogy az L2 görbe egy bizonyos iterációs szám után növekedni kezd, vagyis romlik a képminőségünk. Ez a tipikus eset, egy ilyen L2 és CC görbe látható a 21. ábrán.



21. ábra
A rekonstrukció minőségének függése az elvégzett iterációk számától

Láthatjuk, hogy az iterációs szám függvényében van optimális érték, aminél a rekonstrukció a legjobb (a többi paraméter esetében ilyet nem tapasztalunk). A 21. ábrán megfigyelhetjük, hogy a 30. iterációig az L2 görbe monoton csökken, majd ezután monoton nő, ezt nevezzük felzajosodási fázisnak. A minimum, vagyis az elérhető legjobb L2 norma értékét és helyét a rekonstrukció egyéb paramétereinek megválasztásával befolyásolhatjuk, így az iterációs szám függés egy kicsit külön kezelendő a többi paramétertől, a többi esetben ugyanis, értelemszerűen, az elérhető legjobb L2 normát fogom vizsgálni, így függetlenül a konklúziókat az iterációs számtól.

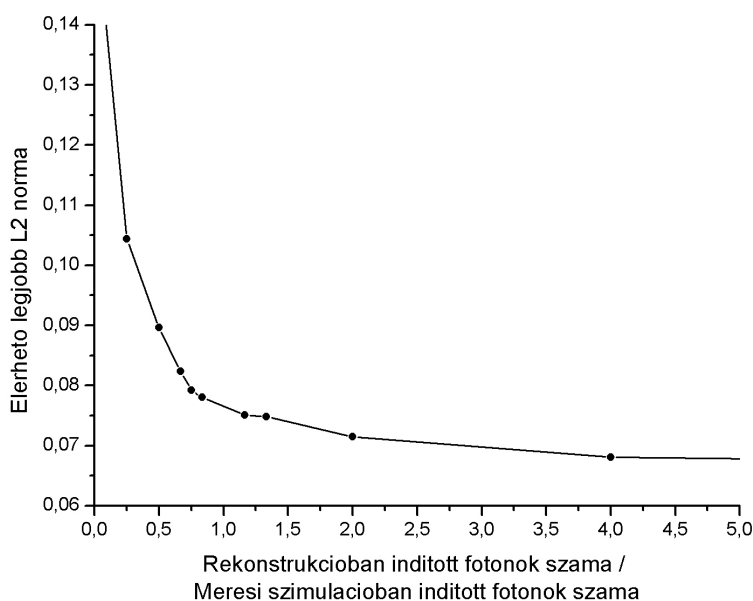
8.2 A képminőség függése a rekonstrukcióban indított fotonok számától

A tesztelés során összegyűlt tapasztalatok arra mutatnak rá, hogy a legfontosabb szerepe a paraméterek közül az indított fotonok számának van, ami nem meglepő, hiszen az iteráció hajtóereje a becsült beütésszámok statisztikai szórásán múlik. Ebben a fejezetben ennek képminőségre gyakorolt hatását részletezem, az összes többi paramétert fixen tartva.

Érdemes megemlíteni, hogy a matematikai fantomokon történő rekonstrukció során magát a mérést is szimuláljuk, így, mint a 7.5 fejezetben láttuk, annak minősége is befolyásolja az iteráció konvergenciáját. Ezért a rekonstrukcióban szimulált fotonok és a mérésben szimulált fotonok számának arányától való függését tártam fel a képminőségnek. A tapasztalatok alapján különböző statisztikájú mérési szimulációval végzett rekonstrukciókat is össze lehet hasonlítani így, vagyis a képminőség érzékenysége az arányszám változtatására független a fotonok szám abszolút mennyiségétől.

Az képminőség-fotonok szám-arány függvény feltérképezése a gyakorlatban a következőképpen zajlik: a szimulált fotonok számának változtatása során minden fotonok szám mellett fel kell venni a rekonstruált kép L2 normáját az egyes iterációk után, és meg kell keresni a legjobb L2 értéket. Ezzel a módszerrel függetlenítjük a képminőség mutatókat az iterációs számtól, hiszen minden esetben az eredeti fantomtól elérhető legkisebb L2 távolságot tekintjük. Ezeket az L2 normákat kell megnézni tehát a mérési fotonok szám – rekonstrukció fotonok szám arány függvényében. Ez grafikusan a 22. ábrán látható. Minél több foton szimulálunk, annál pontosabb becslést kapunk a beütésszámokra, és annál inkább megközelítjük a rekonstruált képpel az eredetit, így értelmezendő a görbe monoton csökkenése. A legjobb eredményt akkor érjük el, ha a szóbanforgó arányszámot a lehető legnagyobbra választjuk, habár a görbe aszimptotikus viselkedése arra mutat rá, hogy

szignifikáns javulást nem érünk el nagyon nagy arányszámok esetén. A grafikonon a szemléletesség kedvéért az arányszámok csak egy bizonyos intervallumán ábrázoltam a legjobb L2 értékeket, de a mintavételezést ezen kívül lévő arányszámoknál is elvégeztem. Például 50-szeres aránnyal elvégzett rekonstrukció minősége kevesebb, mint 5%-kal haladta meg az 5-szörös aránnyal rekonstruált képét.

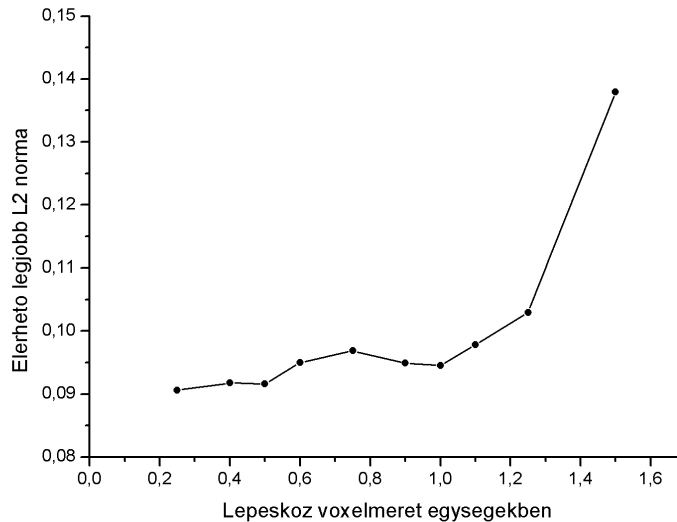


22. ábra
Elérhető legjobb L2 norma a fotonszám függvényében

A korábbi definíciót felhasználva azt mondhatjuk, hogy a rekonstruált kép minősége nagyon érzékeny a fotonszámok arányának beállítására, ha ez az arányszám $\sim 2/3$ -nál kisebb, ennél az aránynál van ugyanis az a pont, amikor a részecskeszám megduplázása már nem okoz 10%-nál több javulást. Ebben az esetben a referencia-érték tekinthető egy javaslatnak a fotonszám megválasztására, hiszen ennél az értéknél kezd aszimptotikussá válni a képminőség grafikonja (22. ábra). Ezt azért fontos megállapítani, mert egy konkrét esetben ennek alapján eldönthetjük, hogy mennyire érdemes emelni a szimulált fotonok számát jobb képminőség reményében más paraméterek változtatásával szemben. Az 22. ábrán például azt is láthatjuk, hogy a mérés statisztikai minőségét (beütésszámok szórását) nem érdemes a szimulált fotonok számának növelésével meghaladni, mert lényeges információt már nem tudunk kinyerni az adatokból, viszont a számítás ideje többszörösére is nőhet. Nyilvánvaló, hogy gyakorlati szempontból, néhány százalékos L2 norma csökkentésért nem érdemes 10-szer, vagy még ennél is több időt szánni a rekonstrukcióra. Előállhat olyan helyzet, mikor más paraméterek állításával sokkal többet nyerhetünk képminőségben.

8.3 A ray-marching algoritmusban használt lépésköz hatása a képminőségre

A rekonstrukció a visszavetítési lépésében az egyes voxelekbe bejutó fotonok számát, valamint a befutott úthosszakat becsli ray-marching módszerrel. Nem triviális feladat megmondani, hogy milyen lépésközt használjunk ennek során, vagyis milyen felbontással mintavételezzük a voxelekben befutott úthosszt, és a voxelek abszorpcióját. Mivel a válaszegyenesek száma, amik mentén a ray-marching algoritmust alkalmazzuk, még a voxelek számánál is több kell legyen, ezért előfordulhat, hogy elég kis pontossággal becsülni ezen változókat egy válaszegyenes esetében, hiszen a LoR-okra vonatkozó átlagot tekintve, a várható értéket így is elegendően megközelítjük. Más szavakkal: a rekonstrukció minősége nem ugyanolyan érzékeny a paraméter megváltoztatására a paramétertér különböző helyein. A 3D rekonstrukció esetében a visszavetítés számításigénye meghaladhatja az előrevetítését, így jelentős javulást érhetünk el hatékonyságban a lépésköz ésszerű megválasztásával. Az 23. ábrán láthatjuk, hogy a lépésköz növelésével az elérhető legjobb képminőség a voxelméret fölött monoton nő, mivel ekkor ritkán mintavételezzük a voxeleket, lesznek olyanok, amik kimaradnak a visszavetítésből, vagy legalábbis kevesebb LoR-ból gyűjtünk információt róluk. A legjobb L2 norma relatív változása nem szignifikáns, vagyis kevésbé érzékeny a képminőség a lépésköz beállítására, ha a voxelméretnél kisebb lépésközökről van szó. Némi numerikus bizonytalanságtól eltekintve (a 23. ábrán a voxelméret 0,6-0,8 részénél látható anomális viselkedés) ezen az intervallumon is azt mondhatjuk, hogy minél kisebb lépésközt választunk, annál jobb képminőséget érhetünk el. Az érzékenységbeli referencia-érték, amit a fotonszámoknál is meghatároztam, itt a voxelméretre tehető, vagyis a voxelméret az a legnagyobb lépésköz, amit ha megfelezzük, és úgy végezzük el a rekonstrukciót, a képminőség 10%-nál kevesebbet javul. A referencia-érték a fotonszámokkal összehasonlítási alapot teremt: a lépésköz voxelméretre felére csökkentése hozzávetőlegesen ugyanolyan hatással van a képrekonstrukció minőségére, mint a fotonszámok arányának 2/3-ról 4/3-ra növelése.



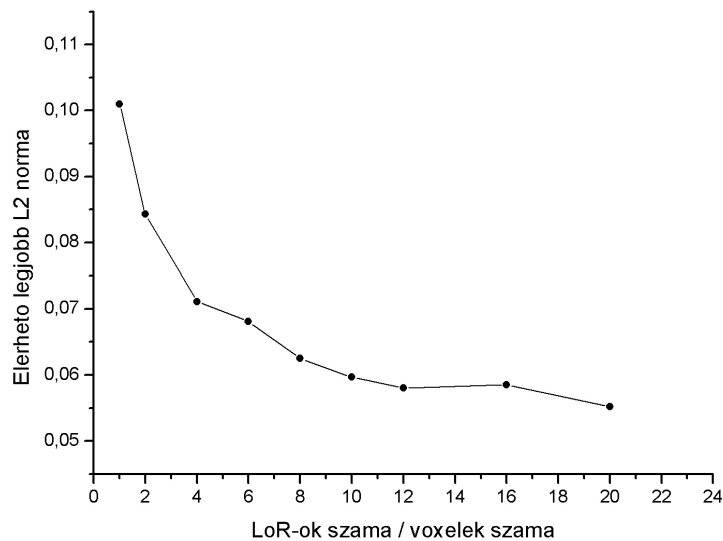
23. ábra
Az elérhető legjobb L2 norma a ray-marching algoritmus során alkalmazott lépésköz függvényében

Annak feltárása, hogy a képminőség lépésköz-változtatásra vonatkozó érzékenysége hogyan viselkedik, gyakorlati szempontból talán a fotonszámok esetén is fontosabb, hiszen 3D rekonstrukciónál a lépésköz harmadik hatványával csökken a visszavetítéshez szükséges számítási idő, így ennek ésszerű megválasztásával rengeteget nyerhetünk a rekonstrukció hatékonyságában. Például, ha adott egy rekonstrukciós protokoll, ami a paramétereket úgy határozza meg, hogy az indított fotonok száma legyen annyi, amennyit a mérés-szimulációban is indítottunk, a lépésköz pedig legyen a voxelméret kétszerese, és el kell döntenünk, hogy melyik paramétert változtassuk ahhoz, hogy minél nagyobb javulást érijünk el képminőségben, akkor egyértelműen a lépésköz csökkentését kell választanunk, hiszen a paraméterternek ezen pontjában a képminőség-gradiens leginkább a lépésköz-csökkentés irányába mutat.

8.4 A visszavetítésben kijelölt LoR-ok számának hatása a képminőségre

Az 1.2 fejezetben láttuk, hogy az iteráció konvergenciájának szükséges feltétele, hogy több válaszgyenest vegyünk fel, mint ahány voxelből áll a rekonstruálandó térfogat. Felmerülhet a kérdés, hogy érdemes-e ennél többet kijelölni, és ha igen, közelítőleg mennyi LoR mentén érdemes elvégezni a visszavetítést. Ahogyan az előzőkben, itt is érdemes relatív paramétert bevezetni, és voxelszám egységekben végezni a vizsgálatokat. Konkrét esetben ez azt jelenti, hogy az adott rekonstrukciós feladatnál megkövetelt felbontás meghatároz például egy elérendő $256 \times 256 \times 256$ -os felbontást a rekonstruálandó térfogatra, vagyis 256^3 db

voxelünk lesz, és kérdés, hogy hányszor ennyinek válasszuk a visszavetítő LoR-ok számát, ahhoz, hogy minél jobb minőségű képet kapjunk. Ebben a példában a fenti követelmények úgy fogalmazhatók meg, hogy egy iterációs lépésben kell lennie minimálisan 256^3 db válaszegyenesnek, amik mentén ray-marching algoritmust végzünk, és becsüljük a voxelekbe érkező fotonok számát, illetve a befutott úthosszakat. Egy projekció esetén a válaszegyenesek minimális számát úgy kapjuk, ha 256^3 -t elosztjuk a projekciók számával. Ennyi LoR-t kell kijelölnünk, azaz ennyi pontot kell felvennünk a detektoron minimálisan. A tesztelés során ezt úgy tettem meg, hogy minden detektorpixelen belül választottam véletlenszerűen K db pontot, így kaptam $K \cdot \#\{\text{hasznos detektorpixelek}\}$ db pontot, amiket a forrással olyan LoR-ok kötnek össze, amik átmennek a voxelmátrixon (ezt értem hasznos detektorpixel alatt). Az 4.4 fejezetben olvashatunk bővebben arról, hogy milyen alternatív algoritmusok jöhetnek számításba ezen módszer helyett (Poisson disk sampling). Ezt a K paramétert változtattam a különböző rekonstrukciók között, és felvettem a legjobb L2 norma–LoR-ok és voxelek számának aránya függvényt, ami látható a 24. ábrán.



24. ábra
Elerhető legjobb L2 norma a visszavetítésben kijelölt válaszegyenesek számának függvényében

Az előzőkhöz hasonlóan itt is meghatározható a referencia-érték, aminél fennáll az, hogy kétszer annyi LoR kijelölésével nem javíthatunk az L2 normán 10%-nál többet. Ez a 8 aránynál adódik, ezzel ismét összehasonlíthatunk például két paraméterre mutatott érzékenységet: ha a LoR-ok számának voxelszámra vonatkoztatott arányát 8-ról a 16-szorosára emeljük, hasonló változást érünk el, mintha a lépésközt voxelméretről annak felére csökkentettük volna. Egy olyan esetben, ahol a rekonstrukció lépésköz paraméterét a

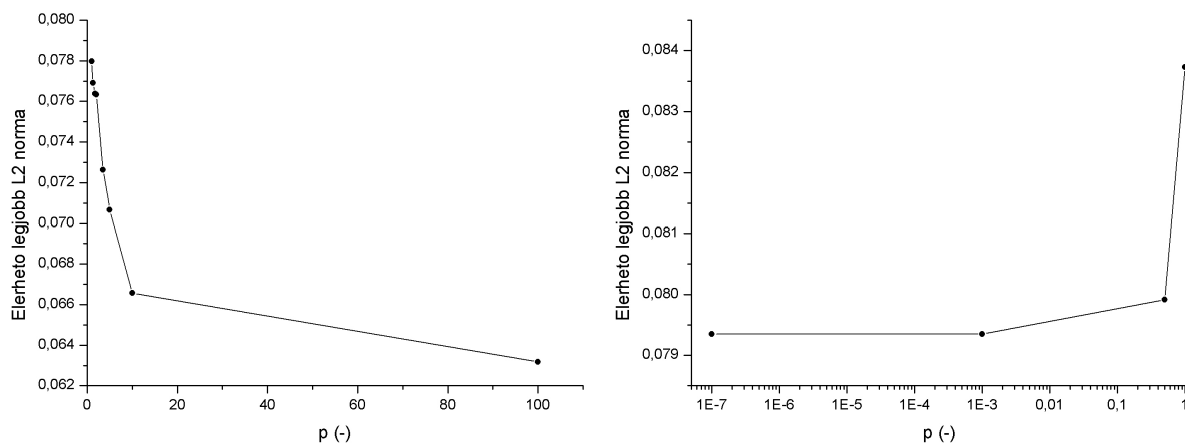
voxelméret töredékének, a LoR-ok számát viszont csak a voxelszám néhány szorosának választjuk, érdemes az utóbbi paramétert növelni, ezzel hatékonyabban fogjuk csökkenteni az L2 normát.

8.5 Az úthossz nyújtás és zsugorítás hatása a képminőségre

Ahogy az előrevetítés működésének részletezésekor láttuk, rekonstrukcióban lehetőség van a Monte Carlo projektor Woodcock módszer szerinti szabadúthossz-sorsolásának torzítására. Az említett kétféle módszer a következő:

1. Úthossz nyújtás: a 0-1 súlyokkal történő Woodcock-féle szimulációt módosítjuk úgy, hogy rákényszerítjük a részecskéket virtuális ütközésre, megnyújtjuk az úthosszat a valós ütközések között. A p -vel jelölt paraméter skálazza az átlagos úthossza inverzét valós ütközések között, $p=1$ esetben a 0-1 súlyokat kapjuk vissza, $p=0$ eset pedig csak virtuális ütközéseket eredményez, az implicit befogást kapjuk vissza
2. Úthossz zsugorítás: az implicit befogást tovább fokozzuk úgy, hogy még több virtuális ütközésre kényszerítjük a fotonokat, a p paraméter itt is az átlagos úthossz inverzét skálazza

Az 25. ábrán láthatjuk, hogy az elérhető legjobb L2 normát csökkenthetjük ezekkel a módszerekkel, azonban a p paraméterre nem olyan érzékeny a rekonstrukció, mint az előbbi esetekben. Erre az mutat rá, hogy ha a korábban definiált referencia-értéket megkeressük, akkor azt találjuk, hogy a p értékének megduplázása (zsugorításnál) illetve felezése (nyújtásnál) nem okoz az értelmezési tartományán sehol sem 10%-nál jobb javulást.



25. ábra
Úthossz-zsugorítás (balra) és úthossz-nyújtás (jobbra) hatása a képminőségre

Mivel az úthossz-zsugorítás illetve a nyújtás interpretációja némileg eltér, a két módszer által kapott eredményeket megtévesztő közös grafikonon ábrázolni. Ugyanis a zsugorításnál a p paraméter a virtuális ütközések közötti átlagos szabadúthosszat skálázza (itt eleve csak virtuális ütközés van, implicit befogás), úthossz-nyújtásnál viszont a valós ütközések közöttit. A valós ütközések közti úthossz nyújtása $p=0$ határértékben (végtelen nagy nyújtás) visszaadja az implicit befogást a zsugorító módszer $p=1$ értékénél (nincs zsugorítás a virtuális ütközések között). Az 25. ábrán jobbra láthatjuk, hogy a Woodcock módszer 0-1 súlyokkal adja a legpontatlanabb becslést a beütésszámokra, és emiatt gyenge képminőséget kapunk. Viszont ha a virtuális ütközéseket egyre inkább preferáljuk (úthossz nyújtás), jobb képminőséget érhetünk el, majd ezen az úton továbbmenve a maximális nyújtást (csak virtuális ütközés van) érjük el. Ez közelítőleg megegyezik az 25. ábrán balra látható $p=1$ ponttal. Itt tovább javul az elérhető legjobb L2 norma, ha a virtuális ütközések közti úthosszat zsugorítjuk, vagyis ha még több virtuális ütközést eszközölünk. A virtuális ütközések tehát hasznosak a rekonstrukció minősége szempontjából, hiszen ekkor gyakran mintavételezzük a voxelmátrixot, ráadásul finomabb lesz az a skála, amin a lehetséges súlyok helyezkednek el.

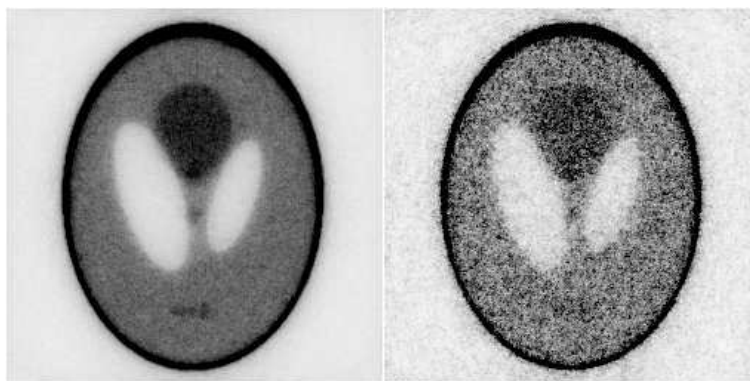
9. Rekonstrukció korlátozott idő alatt

Az előző fejezetben láthattuk, hogy a paraméterek változtatása milyen hatással van a rekonstrukcióra, közelítőleg sikerült felmérni a paramétertérben való képminőség-változást, valamint az L2 norma érzékenységet a paraméterek beállítására. Ebben a fejezetben arra adok útmutatást, hogy korlátozott rekonstrukciós idő alatt hogyan válasszuk meg az egyes paramétereket, adott idő mellett mi az optimális beállítás. Ezt úgy fogom megtenni, hogy ismételten nem a teljes paraméterteret járom be, hanem néhány beállítást (az adott időkorlát mellett közel elérhető legyen a legjobb L2 norma) vizsgálva megállapítom, hogy ezek közül melyik az optimális, az adott rendelkezésre álló idő mellett.

9.1 A korlátozott rekonstrukciós idő problémájának felvetése

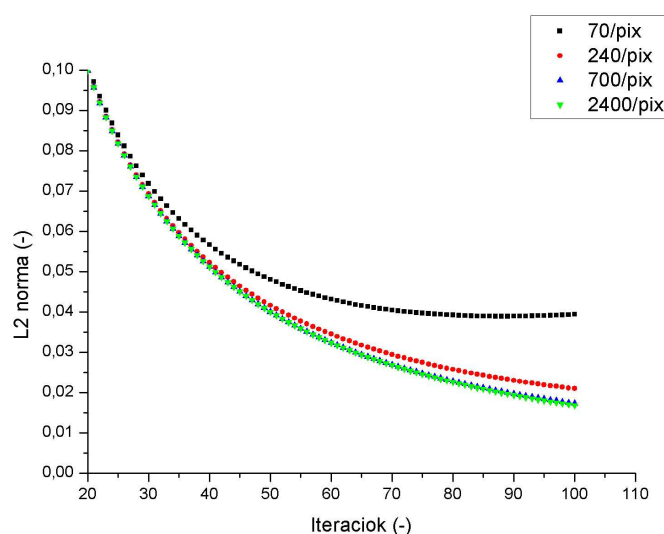
A korlátozott idő alatt végzett rekonstrukció problémáját a fotonszámok esetén mutatom be, de az elmondottak igazak más paraméterek esetén is. A fotonszám esetén a becslés statisztikáját fogom feltüntetni, pontosabban azt, hogy az egyes detektorpixelekre

hány foton jut a szimulációban. Ahogy azt az előbbi fejezetben láttuk, és a 26. ábrán is megfigyelhetjük, kevesebb foton szimulálása a kép minőségének romlásához vezet. A legjobb választás az lenne, ha végtelen sok fotont szimulálnánk. Korlátozott idő alatt azonban egyáltalán nem biztos, hogy ez az ésszerű választás, hiszen ugyanazon időn belül több foton szimulálásával végrehajtott rekonstrukcióban például az iterációk száma is kevesebb lesz.



26. ábra
Rekonstruált kép különböző rekonstrukcióban szimulált fotonsszámmal
2,5 milliárd/iteráció (balra), 360 millió/iteráció (jobbra)

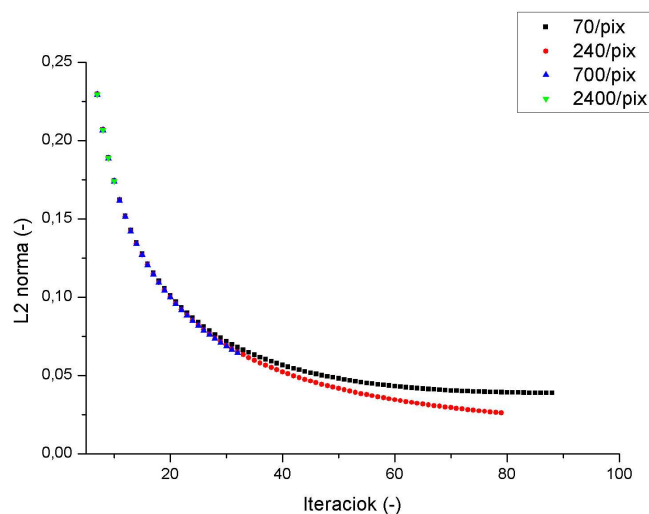
Az L2 görbe lefutására jellemző, hogy kevesebb foton szimulációja esetén a minimumhely eltolódik a kisebb iterációs szám felé, azonban távol a minimumhelytől (alacsony iterációs számoknál) a különböző fotonsszámú rekonstrukciók konvergenciájának sebessége közel azonos (27. ábra). Vagyis lényegesen jobb statisztikájú szimulációt csak akkor érdemes végezni, ha elegendő idő áll rendelkezésünkre ahhoz, hogy akár több száz iterációt is elvégezzünk. Ha korlátozott idő áll rendelkezésre, érdemes kompromisszumot kötni a futási idő és a szimulált részecskék száma között, ahogy azt a 28. ábrán is láthatjuk.



27. ábra
Az L2 norma iterációs szám-függése, különböző szimulált fotonsszámnál
A különböző színű görbék különböző beütésszám/detektorpixel statisztikájú szimulációkat szemléltetnek

9.2 A rekonstrukcióban indított fotonszám optimális beállítása korlátozott rekonstrukciós idő esetén

Ha megtekintjük azt, hogy 10 perces rekonstrukciós idővel milyen képminőséget tudunk elérni, akkor azt tapasztaljuk, hogy nem érdemes túl gyenge statisztikát adó szimulációt végezni, annak érdekében, hogy már 10 percen belül elérjük az L2 görbe minimumát, de túl sok fotont sem éri meg szimulálni, hiszen ekkor kevesebb iterációra elég az idő. A 28. ábra alapján a legjobb választás a „négyzetek fantom” rekonstrukciója esetén a 240/pixeles statisztika, vagyis ennek megfelelő fotonmennyiséget kell indítani a szimulációban. Ez az eredmény általában nem terjeszthető ki más esetekre, hiszen függ a rendelkezésre álló számítási kapacitástól, magától a rekonstruálandó gyengítési együttható eloszlástól is. Például, ha több grafikus kártya áll rendelkezésre, vagy a jövőben akár nagyságrendekkel gyorsabb hardveres környezetben tudjuk realizálni a rekonstrukciót, akkor nyilván az egyes iterációk elvégzéséhez szükséges idő lecsökken, és a 28. ábra egészen más eredményt mutathat.

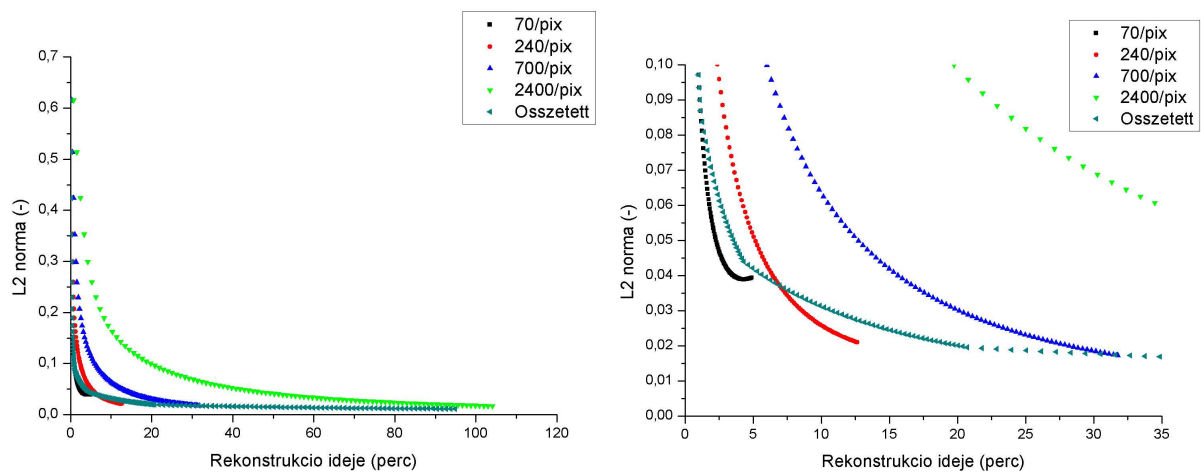


28. ábra
Elérhető legjobb L2 norma korlátozott rekonstrukciós idő (10 perc) alatt

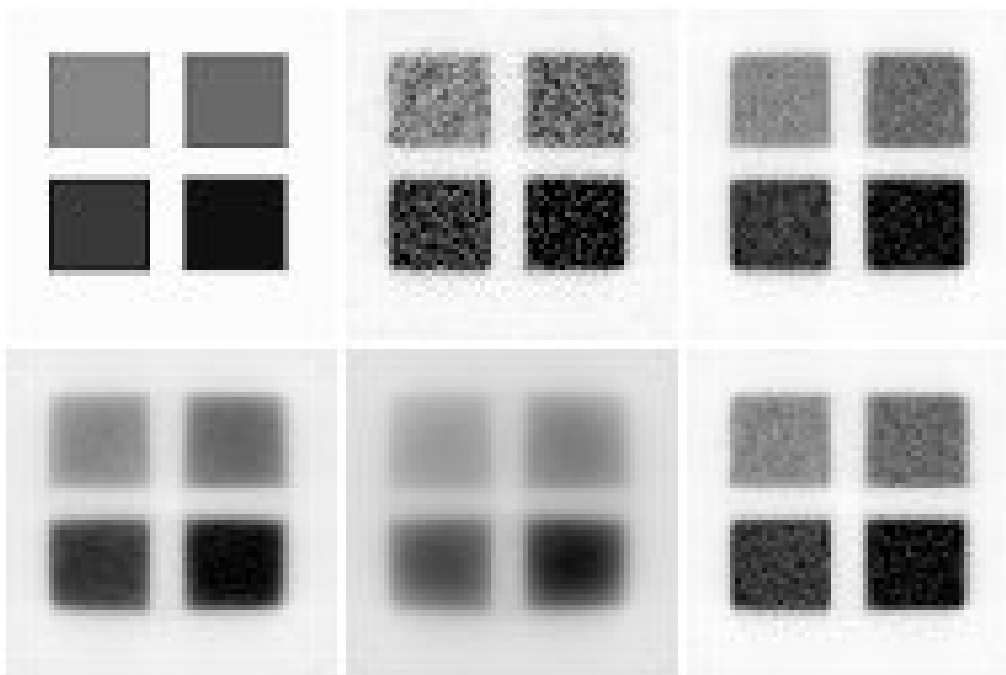
A fentiek alapján jó lehetőséget láthatunk abban, hogy a szimulált fotonok számát iterációs szám-függővé tegyük. Vagyis megtehetjük például, hogy az iteráció elején kisebb foton számot választunk, „gyorsan” csökkentjük az L2 normát, majd amikor a rekonstrukció hajtóereje lecsökken (például a 27. ábrán ez a fekete adatsor esetén ~25 iterációnál van), akkor elkezdünk több fotont szimulálni. Implementáltam ennek megvalósítására egy próbarekonstrukciós sémát, ami a 27. ábrán szereplő statisztikák között váltogatja a szimulált fotonok számát, a 25. iterációig 70/pixel beütésszámmal rekonstruál, majd megemeli

240/pixeles statisztikára, és így tovább (50. és 100. iterációnál emelkedik még a fotonszám). Az tesztet a „négyzetek fantomon” valósítottam meg, amit szemléltetnek a 29. ábra grafikonjai („Összetett” jelzésű adatsor), és a 30. ábrán látható rekonstruált képek.

A 29. ábrán már a problémának egy árnyaltabb bemutatását látjuk, hiszen itt nem az iterációk számának függvényében, hanem a rekonstrukciós idő függvényében ábrázoltam az elért L2 normákat. Innen megállapíthatjuk, hogy tetszőleges rendelkezésre álló idő alatt melyik paraméterbeállítás a számunkra legkedvezőbb.



29. ábra
L2 norma a rekonstrukció idejének függvényében, különböző fotonszámú szimulációk esetén
Teljes görbe (balra), nagyított grafikon (jobbra)

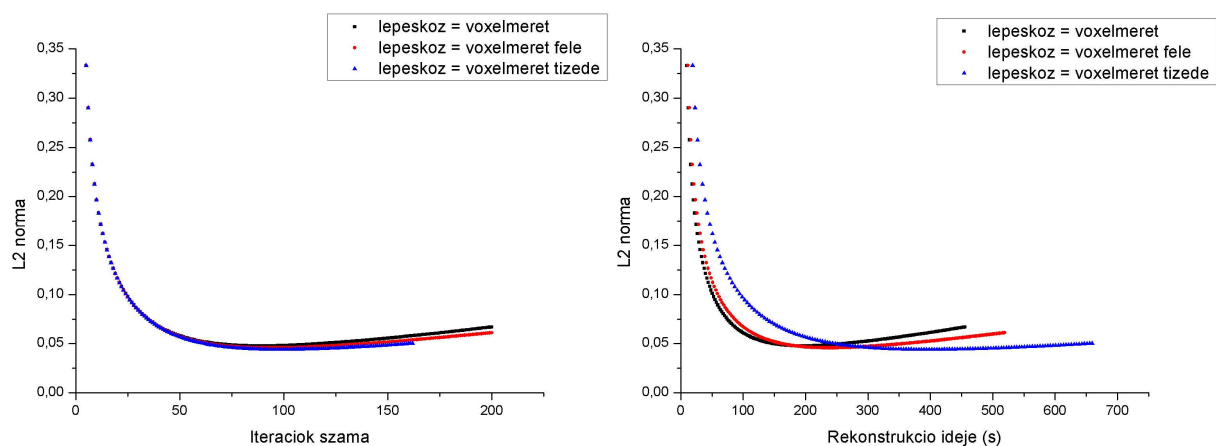


30. ábra
L2 normában elérhető legjobb képek 10 perc alatt, különböző fotonszám statisztikákkal
Bal felső – jobb alsó sorrendben: eredeti fantom, 70/pixel – 88 iteráció, 240/pixel – 80 iteráció, 700/pixel – 30 iteráció, 2400/pixel – 10 iteráció, összetett séma – 80 iteráció

Az összetett séma láthatóan gyorsabban ad jobb minőségű képet, mint a magasabb fotonszámú rekonstrukciók, azonban alkalmazása nem előnyös a 240/pixel statisztikájú szimulációhoz képest. Ahogy a rekonstruált képeken is feltűnik, a rekonstrukció egy lokális minimumba „befagy”, a zajok a képen megmaradnak, éppen azért, mert az iteráció elején gyenge statisztikával rekonstruáltunk. Ezért nem célszerű ennek a konkrét sémának a megválasztása, talán a paraméterek finomabb hangolása javíthat ezen, mindenesetre teljes bizonyossággal azt lehet állítani, hogy korlátozott rendelkezésre álló idő alatt érdemes odafigyelni a fotonszámok helyes megválasztására.

9.3 A ray-marching algoritmusban használt lépésköz optimális beállítása korlátozott rekonstrukciós idő esetén

Mint azt a 8. fejezetben láttuk, a képminőség a lépésköz megválasztásának monoton függvénye, minél kisebb a lépésköz, annál jobb a képminőség. Ezt szemlélteti a 31. ábra bal oldali grafikonja. Korlátozott rekonstrukciós idő esetén érdemes az L2 normákat az elérésükhöz szükséges idő függvényében ábrázolni (31. ábrán jobbra). Az alábbi grafikonok azt szemléltetik, hogy a paraméter megválasztása, a vizsgált esetben, kis hatással van a képminőségre, azonban annál többet nyerhetünk futásidőben, ha helyesen választjuk meg azt.



31. ábra

A ray-marching algoritmus során alkalmazott lépésköz megválasztása az L2 görbét kis mértékben befolyásolja, rekonstrukciós időben viszont jelentős különbségeket látunk

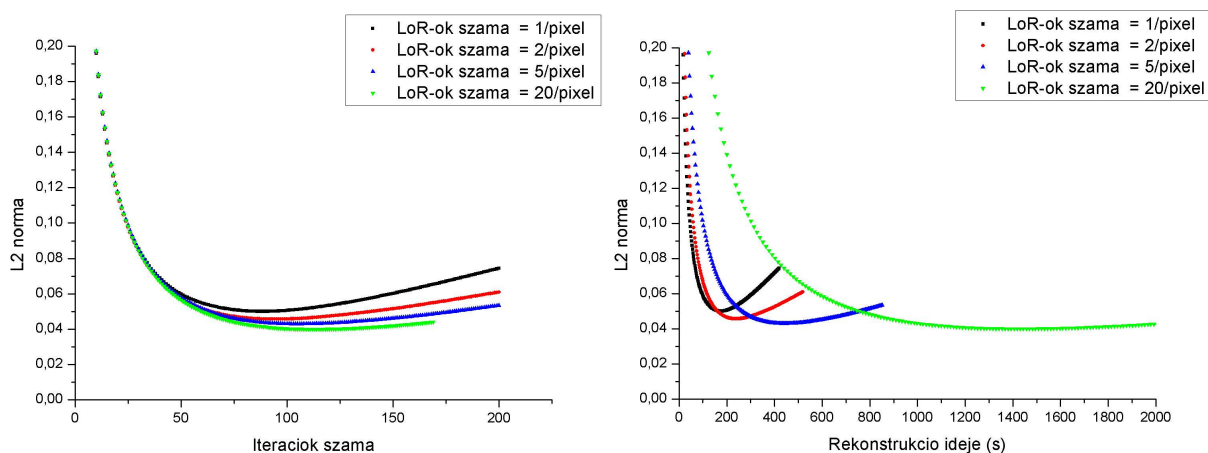
A 31. ábra alapján azt mondhatjuk, hogy 10 perc rendelkezésre álló idő alatt a lépésközt érdemes a voxelméret tizedének választani, habár láthatjuk, hogy milyen kevéssel haladja

meg az elérhető legjobb L2 norma a voxelméret felének megfelelő lépésköz esetét. Ez összhangban van a 8. fejezetben elmondottakkal.

Más időkövetelmények mellett más következtetéseket vonhatunk le, például 1-2 perc rekonstrukciós idővel a legjobb választás a voxelmérettel egyenlő lépésköz lenne. Továbbmenve, fantomtól sem független az optimális választás, tekinthetjük a fejezetben bemutatott módszert egy javaslatnak, ami alapján el lehet dönteni a konkrét alkalmazásra jellemző tipikus fantomok esetén, hogy miként kell a paraméterek beállításának optimumát megkeresni.

9.4 A visszavetítő válaszegyenesek számának optimális beállítása korlátozott rekonstrukciós idő esetén

A kijelölt LoR-ok számától a vizsgált fantomnál és a vizsgált paramétertartományon szintén kevésbé függött a képminőség, így ismét érdemes az elérhető L2 normákat az elérésükhöz szükséges idő függvényében tekinteni (32. ábra jobb oldalt). 10 perc rekonstrukciós idő alatt a legjobb választás ezen esetben az 5 kijelölt válaszegyenes volt pixelenként. Láthatjuk, hogy ha ennél kevesebb idő áll rendelkezésre, akkor érdemes a válaszegyenesek számát csökkenteni a visszavetítésben, hiszen ekkor hamarabb érhetünk el magas iterációs számot, és így a legjobb L2 normát.



32. ábra

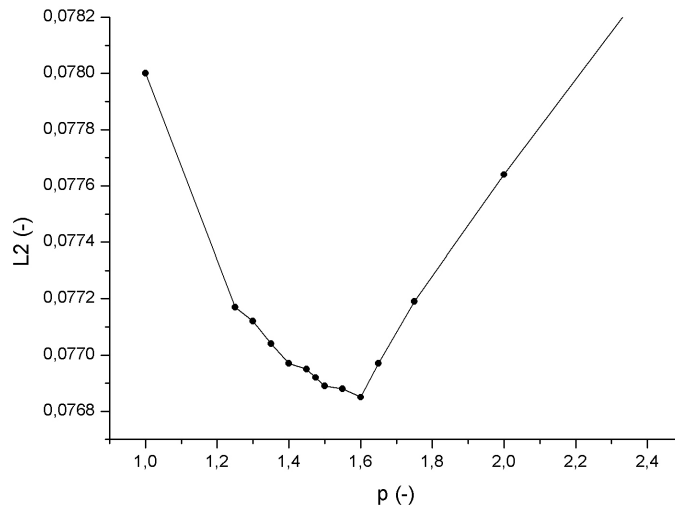
A visszavetítésben kijelölt válaszegyenesek száma kis hatással van az elérhető legjobb L2 normára (bal oldali grafikon), rekonstrukció idejét viszont csökkenteni tudjuk helyes választással (jobbra)

9.5 Az úthosszat zsugorító illetve nyújtó paraméter optimális beállítása korlátozott rekonstrukciós idő esetén

A rekonstrukció előrevetítés függvénye a 4.3 fejezetben leírtaknak megfelelően működött, Woodcock módszerrel választva szabadúthosszat, valamint többféle lehetséges szóráscsökkentő eljárással implementálva. A következőkben ezek összehasonlítását fogom megtenni két különböző típusú fantomra, amiknél különböző módszert alkalmazva találjuk meg az optimális beállítást korlátozott rekonstrukciós idő alatt.

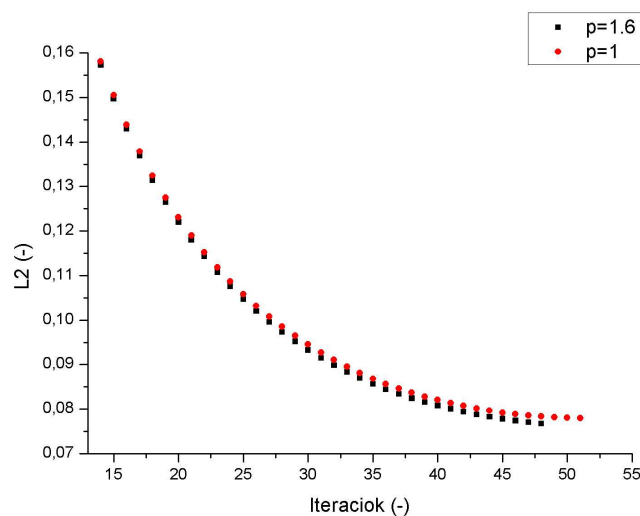
Az összehasonlítás a séma hatékonysága alapján történt, amit a rekonstrukció esetében az adott idő alatt elérhető L2 norma határoz meg. Vagyis az összehasonlítás során az az eljárás lesz optimális, ami a legkisebb L2 normát adja egy fix idő mellett. A fentiekben bemutatott elméleti megfontolások alapján, végtelen idő alatt az implicit befogás, $p = \infty$ mellett a leghatékonyabb, hiszen így tudjuk a legkisebb szórást produkálni. Ellenben ez a legnagyobb számításigényű szimulációs séma, így elképzelhető, hogy korlátozott idő alatt p véges értékénél találunk minimális L2 normát. Ugyanígy, a futási időben leginkább kedvező 0-1 súlyokkal végzett Woodcock-szimuláció is felveheti hatékonyságban a versenyt, valamint az úthossz-nyújtás megengedő súlykompenzációs módszer, ami gyakorlatilag a 0-1 súlyok, és az implicit befogás között skálázza a számítási sémákat a p paraméterrel. Korábban, például a fotonszámok beállításánál láthattuk, hogy a lehető legjobb L2 norma eléréséhez, korlátozott rekonstrukciós idő alatt, nem érdemes a lehető legtöbb fotont szimulálni, hiszen így kevesebb iterációt tudunk elvégezni. Most is ugyanez érvényes: meg kell találni az optimumot a módszerek között egy adott futási időre, amivel a legjobb L2 normát tudjuk elérni.

A tapasztalatok alapján a rekonstruálandó gyengítési együttható eloszlás szempontjából érdemes két esetet vizsgálni: az egyik az erősen abszorbens fantom, másik a gyengén elnyelő eloszlás. Erősen abszorbeáló közeg esetén az implicit befogás használata nem kedvező, hiszen a számítás ideje nagyon megnő, a mérés-szimuláció statisztikus zajához képest kis javulást kapunk a becsült beütésszámok szórásnégyzetében. Gyengén elnyelő esetben azonban érdemes implicit befogást és úthossz-zsugorítást végezni (vagyis a ritka közeget sűrűbben mintavételezni), az optimális választás a bemutatott esetben $p = 1,6$ -ra adódott (33. ábra).



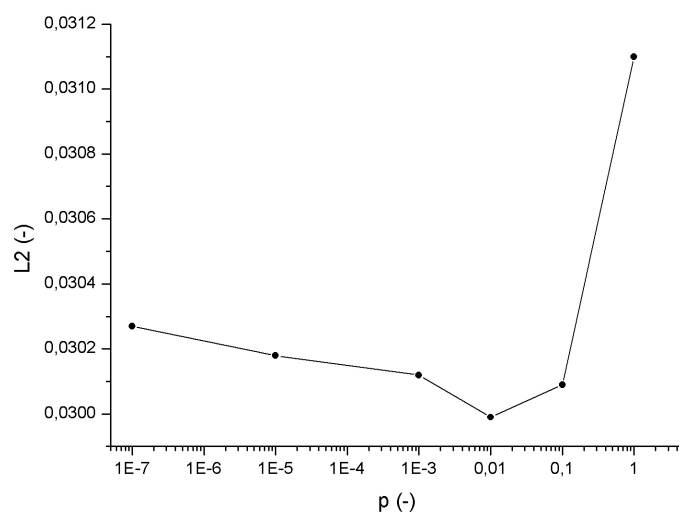
33. ábra
1 perc alatt elérhető L2 norma, úthossz-zsugorítás esetén, ritka közegben

A teszt során fix rekonstrukciós idő alatt (1 perc), rögzített fotonszámokkal és egyéb paraméterek mellett vizsgáltam az elérhető legjobb L2 normát p néhány értékénél. Az L2 normákat úthossz-nyújtás esetén is megfigyeltem, azonban szignifikánsan rosszabb értékeket kaptam (ritka közeg esetén), így ezt a 33. ábrán nem tüntettem fel. A legjobb L2 normához tartozó p paraméter 1,6-nak adódott, vagyis ebben az esetben érdemes volt 1,6-od részére zsugorítani az úthosszak várható értékét a hatékonyság növelése érdekében. Az L2 görbe lefutását mutatja be a 34. ábra $p=1$ és $p=1,6$ esetben. Látható, hogy a rekonstrukció minőségében néhány %-os (egész pontosan 1,5%-os) javulás érhető el.



34. ábra
L2 görbe implicit befogás $p=1$ és $p=1,6$ esetben

Ritka közeg esetében az úthossz-nyújtás nem vezetett célra, nem növelte a hatékonyságot. Mivel az eredeti 0-1 súlyokkal végzett eljárás akkor lehet hatékony, ha nagyon elnyelő a közeg, ezért megnéztem, milyen eredményeket kapok sűrű fantom esetén. Ilyenkor az implicit befogáshoz képest a 0-1 súlyokkal történő szimuláció még mindig kevésbé volt hatékony, azonban az úthossz-nyújtással kisebb L2 normákat tudtam elérni. Ezt demonstrálja a 35. ábra, amin a $p = 0$ eset az implicit befogás sémájának felel meg, $p = 1$ pedig a 0-1 súlyokkal végzett szimulációnak.



35. ábra
Úthossz-nyújtás alkalmazásának hatása a képminőségre, sűrű közeg esetén

Következtetésként elmondhatjuk, hogy nagy elnyelésű közeg esetén érdemes a sűrű közeget ritkábban mintavételezni, az általam vizsgált esetben a legjobb választás p -re a $p = 0,01$ volt.

10. A spektrumfelkeményedés és korrekciója

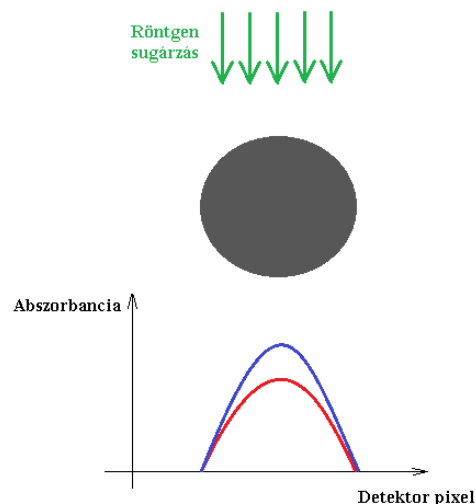
A spektrum felkeményedése (beam hardening - nyalábkeményedés) a fizika több területén előforduló jelenség (reaktorfizika, kozmikus sugárzás, orvosi képalkotás). CT képalkotás esetén ez, a rekonstruálandó eloszlástól függő szisztematikus torzulást okoz rekonstruált képen, előfordulhat, különösen nagy abszorpciós tényezővel rendelkező fantomok esetén, hogy a kép értékelhetetlenné válik.

A röntgen-forrásból kilépő fotonok spektruma a néhány keV-os tartománytól a használt gyorsítófeszültségtől függően akár néhány száz keV-ig is terjedhet, emellett már figyelembe kell vennünk az anyaggal való kölcsönhatások során a fotoeffektus valószínűségének

energiafüggését is. Mivel a kis energiájú röntgen-sugárzás nagyobb valószínűséggel nyelődik el, mint a nagyobb energiájú, ezért az anyagon való áthaladás során a forrás spektruma eltolódik a nagyobb energiák felé. Ez alatt, pontosabban azt értjük, hogy a fotonok energia szerinti eloszlásának várható értéke a fantom abszorbanciájának monoton növekvő függvénye.

10.1 A spektrumfelkeményedés megjelenése a rekonstruált képen[7]

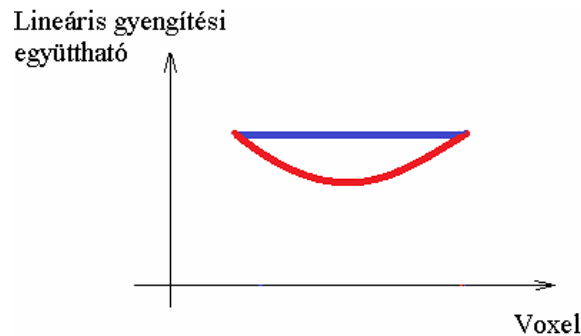
Egy tomográfias képalkotásnál a spektrumfelkeményedés hatása akkor jelenik meg, ha nagy tömegabszorpciós együtthatójú anyag található a vizsgált fantomban, és a fantom relatív méretei (a fotonok szabad úthosszához képest) elegendően nagyok. A rekonstrukciós eljárás során ugyanis, egy ilyen anyag felülethez közeli tartományáról összegyűjtött információ (vagyis az elnyelés mértéke) jóval különbözik a belsejéről származóval, mivel ha a sugárzás hosszú utat tesz meg a szóban forgó fantom-részben, a spektrum jelentősebb mértékben eltolódik. Interpretálhatjuk ezt úgy is, hogy míg a felületen az eredeti spektrumot érzékeli, az erősen elnyelő anyag belső régióit egy felkeményedett spektrummal „világítjuk át”. Egy ilyen esetben, ha egy homogén, hengeres szimmetriájú fantom egyetlen projekcióját vizsgáljuk, akkor a 36. ábrán látható jelenséget tapasztaljuk.



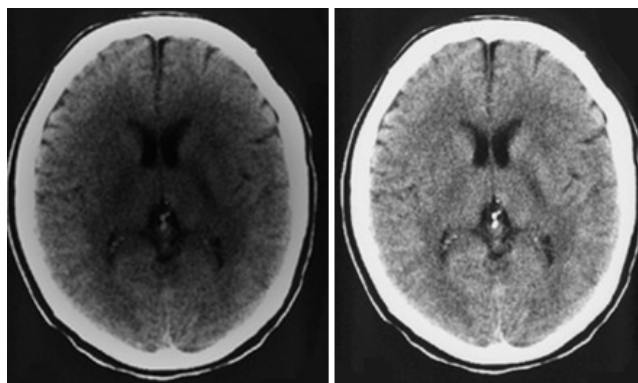
36. ábra
A felkeményedés hatása a mért eloszlásra
Kék vonal: monokromatikus projekció, piros vonal: mért abszorbanciák

Ha ezekből a projekciókból rekonstruálni kívánjuk az eredeti eloszlást (korrekciók nélkül), akkor azt fogjuk tapasztalni, hogy a képen a valóságnál kisebb abszorpciós együtthatókat

kapunk középen, mint széleken. A 38. ábrán erre láthatunk példát, itt a koponyacsontok miatt fellépő kisenergiás foton-elnyelés miatt az agyban a vártnál kevesebb foton nyelődik el, így módon a valóságosnál kisebb gyengítési együtthatókat kapunk a rekonstrukció során (a 38. ábrán ez úgy jelenik meg, hogy a kép közepe sötétebb, mint a spektrum figyelembevétele esetén). Ugyanennek a jelenségnek egy homogén fantom rekonstruált képének metszetén való megjelenését a 37. ábra szemlélteti.



37. ábra
 Spektrumfelkeményedés megjelenése a rekonstruált kép egy metszetén, homogén fantomban: rekonstrukció spektrummal (kék), spektrum nélkül (piros)



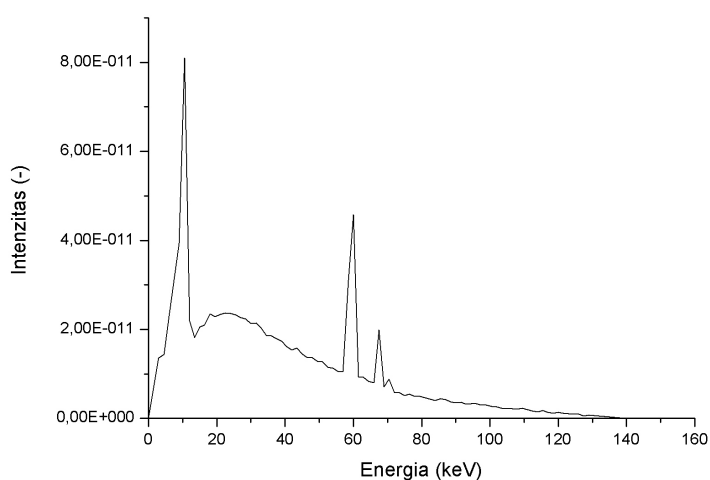
38. ábra
 Horizontális agyi metszetek CT képalkotással [15]
 Spektrumfelkeményedés-korrektció nélkül (balra), korrekcióval (jobbra)

10.2 Korrekció ML-EM algoritmussal

A spektrumfelkeményedés által okozott műtermék nem jelenik meg a rekonstruált képen, ha figyelembe vesszük a röntgen fotonok spektrális eloszlását, illetve a fotoeffektus hatáskeresztmetszetének energiafüggését. Az iteratív ML-EM képrekonstrukció erre egy egyszerűen implementálható lehetőséget ad: az előrevetítés során ki kell bővíteni a szimulációt a fizikai valóság modellezésének ezen aspektusával. A MC projektornak

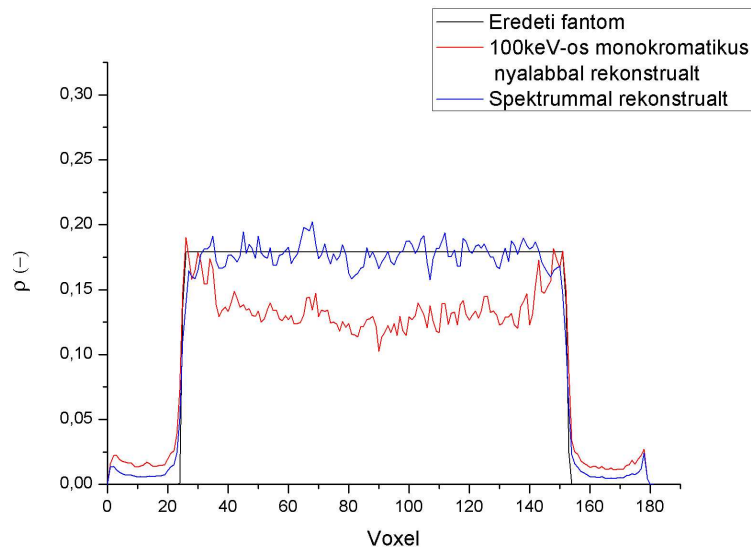
képesnek kell lennie a fotonok energiájának adott eloszlásból való sorsolására, valamint a részecsketranszport során a voxelekhez hozzárendelt energiafüggő hatáskeresztmetszetekkel való számolásra.

A spektrumfelkeményedés korrekciójára egy olyan tesztet végeztem el, amiben a 39. ábrán látható spektrum szerint sorsolt fotonenergiákkal végeztem egy mérés-szimulációt, majd a rekonstrukciót elvégeztem az előrevetítésben 100keV energiát használva (monokróm nyaláb), és az eredeti spektrum szerint is. A teszthez a „kör fantomot” használtam (41. ábra balra), így hasonló eredményeket vártam, mint a korábbiakban tárgyalt esetben (37. ábra).

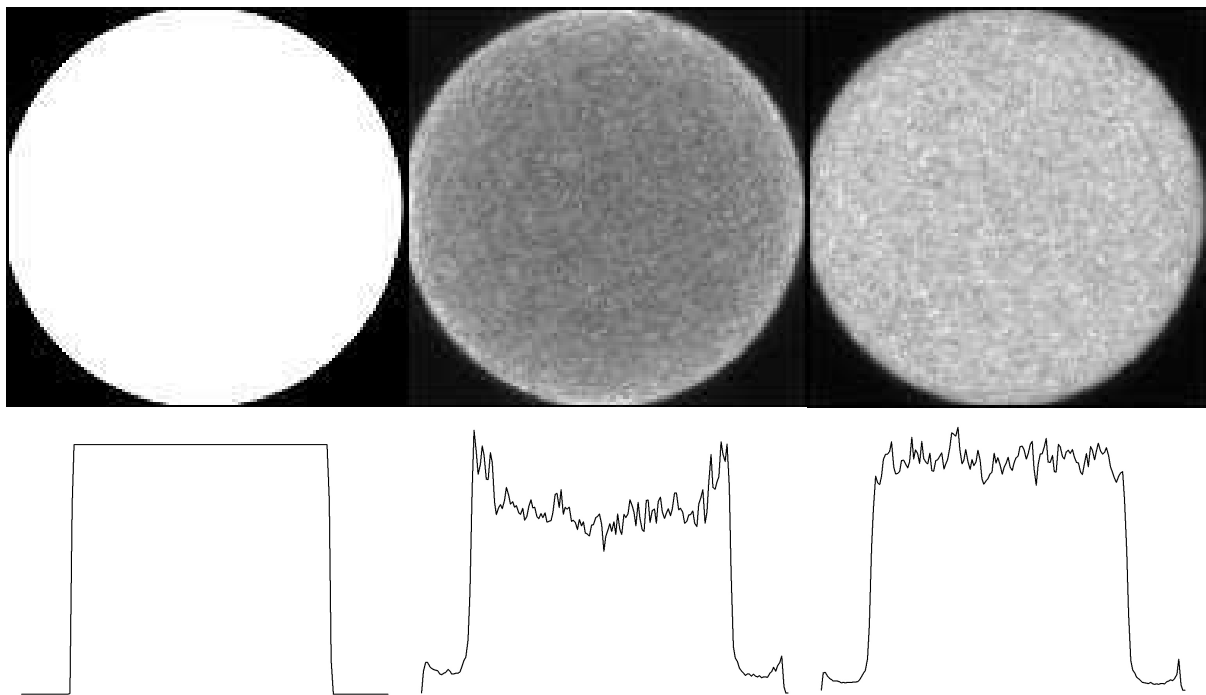


39. ábra
A szimulációhoz felhasznált röntgenspektrum

A szimuláció végrehajtásánál a víz tömegabszorpciós együtthatójának energiafüggését vettem figyelembe, vagyis a fantomot egy homogén, hengeres vízoszlopnak feltételeztem (ennek axiális szelete az ábrázolt fantom). Ha figyelembe vesszük a lineáris gyengítési együttható energiafüggését, akkor valójában a sűrűségeloszlást rekonstruálhatjuk. A fantom sűrűségét önkényesen választottam meg, csakúgy, mint eddig a lineáris gyengítési együtthatók, egy dimenzió nélküli számként kerül ábrázolásra, például a 40. ábrán. Itt a rekonstruált kép átlóján vett metszetek láthatók. A 100keV energiával rekonstruált kép természetesen nem érték helyes, emiatt a gyakorlatban a Hounsfield-korrekciónak alkalmazzák, esetünkben ezt elégséges szemléltetni úgy, hogy a skálázást önkényesen választjuk meg. Érdeemes megfigyelni a csészealj-szerű (cupping artefact) bemélyedést a metszeten (csakúgy, mint [16]-ban), amit a spektrumfelkeményedés okoz, valamint ennek hiányát, ha spektrummal rekonstruálunk. Ugyanezt demonstrálja a 41. ábra is, ahol maguk a rekonstruált képek is láthatók.



40. ábra
Képtálon vett metszetek közös grafikonon



41. ábra
A spektrumfelkeményedés műterméke és korrekciója: rekonstruált képek és képtálon vett metszetek
Eredeti fantom (balra), monokróm nyalábbal rekonstruált (középen), spektrummal rekonstruált (jobbra)

Láthatjuk, hogy az ML-EM képrekonstrukció képes a spektrumfelkeményedés okozta műtermék korrekciójára, a mérési eredmények és a rekonstrukcióbeli előrevetítés összehangolásával. Ha a projekciók szimulációinál figyelembe vesszük, hogy nem monokróm nyalábbal végeztük a mérést, akkor a rekonstrukció visszaadja a valós sűrűségeloszlást, valamint a rekonstruált képen nem jelenik meg csészealj-műtermék (cupping-artefact).

11. Valós mért adatsorok rekonstrukciója

A képrekonstrukciós szoftver fejlesztésében új kihívásokkal szembesülünk, ha valós mérési adatokon teszteljük azt. Ezzel egy lépést tehetünk annak irányába, hogy az algoritmus felhasználása bekerülhessen az orvosi gyakorlatba. Lehetőségem volt a rekonstrukciót kipróbálni az NTI Oktatóreaktorában működő CT készülékkel. A mérési adatok kinyerése, a kalibrációs eljárások elvégzése nem volt része a diplomamunkámnak, ezeket Kleizer Gábornak és Tölgyesi Botondnak köszönhetem.

11.1 A kísérleti elrendezés

A kísérleti elrendezés egy hagyományos kúpsugaras (cone-beam) CT geometriát valósít meg, azzal a különbséggel, hogy a mintát lehet forgatni egy tengely körül, nem pedig a forrás-detektor rendszert. A mérési eredményeket szolgáltató CT berendezés az alábbi egységekből állt:

- Source Ray Inc. SB-80-1k mikrofókuszos röntgenső
- Dexela 1207-as számú nagyfelbontású röntgendetektor
- Phi Instruments lineáris és rotációs mozgatómodulok

A röntgendetektor rekonstrukciós szempontból lényeges tulajdonságait a 6. táblázatban külön kiemelném, mivel ezek minden rekonstrukciónál állandó paraméternek számítottak.

Pixelméret	74,8 μ m
Felbontás	1536x864 pixel

6. Táblázat
A Dexela 1207 röntgendetektor tulajdonságai

A berendezés további specifikációiról bővebben [17]-ben olvashatunk.

11.2 A valós mérési adatokkal végzett rekonstrukció kihívásai

A dolgozatban eddig a pontig olyan rekonstrukciókat mutattam be, amik bejövő adatként a mérés-szimuláció eredményeit kapták meg. Ezt a rekonstrukcióban is használt

előrevetítéssel állítottam elő, ezért egyszerű volt a „mérés” és a szimuláció összehangolása. A valós mérés során előállított projekciók esetében azonban felmerülnek olyan problémák, amik eddig nem kerülhettek elő.

- (i) Először is tudnunk kell, hogy miként egyeztessük össze a detektorból kiolvasható adatokat az iterációs sémával. Tekintsük meg újra az 1.3 fejezetben szereplő (15) iterációt:

$$\mu[j] = \frac{\sum_{lor \in L_j} (X_{lor}[j] - X_{lor}[j+1])}{\sum_{lor \in L_j} \left(Y_{lor} - \hat{Y}_{lor} + \frac{X_{lor}[j] + X_{lor}[j+1]}{2} \right) \cdot L_{lor}[j]}$$

A bevezetőben tisztáztuk, hogy az Y és \hat{Y} mennyiségek rendre a valós mérési eredményeket, és a rekonstrukcióban aktuális iterált gyengítési együttható eloszlás mellett szimulált mérés eredményeit jelentik. Pontosabban ezek részecskeszámokat jelentenek, ahogy az X -szel jelölt mennyiségek is, a voxelekbe bejutott részecskék számára vonatkoznak. Továbbá a szimuláció során, a visszavetítésnél ismernünk kell, hogy hány részecske indult el a forrásból az (egyébként önkényesen kijelölt) LoR mentén. A részecskeszámokat azonban nem tudjuk megmérni, így másképp kell következtetni ezekre a mennyiségekre.

- (ii) Egy másik jelentős probléma a röntgenforrásból származó fotonok energia szerinti eloszlása. A leképezést egy polikromatikus nyalábbal hajtjuk végre, aminek egy, a röntgen-csőre jellemző spektruma van. A röntgen sugárzás lineáris gyengülését jellemző μ gyengítési együttható azonban nemcsak anyag (rendszer), hanem energia-függő is. Ehhez hozzájárul az még, hogy a detektor is spektrálisan érzékeny (esetünkben a gadolínium-oxid kristály energiafüggő hatáskeresztmetszete játszik szerepet). Vagyis ahhoz hogy értékhelyes rekonstrukciót kapjunk, az iteráció során végzett szimulációkban (előrevetítésekben) ugyanazt a spektrumot kell használnunk, mint ami a valódi mérés során jellemezte röntgenforrást.
- (iii) A harmadik lényegi különbség az eddigiekhez képest, hogy a mérési geometria nem ismert, vagy legjobb esetben is, csak véges pontossággal ismert. Ezalatt arra gondolok, hogy a mérés-szimuláció során (amikor matematikai fantomokon végzem el az előrevetítést) alkalmazott geometriát egzaktul felhasználhatjuk a rekonstrukcióban alkalmazott előrevetítésekben (nincs diszkrépancia a „mérés” és a rekonstrukció között). Természetesen a valós mérési geometriát meg lehet határozni geometriai kalibráció

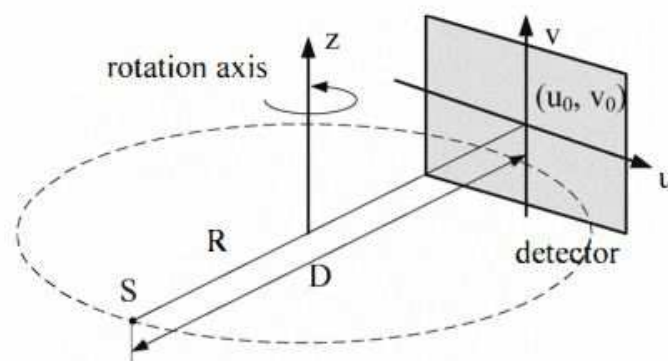
segítségével, azonban a becsült paraméterek hibával terhelték lesznek. Ezen hibák rekonstrukció minőségére vonatkozó hatását meg kell vizsgálni a tesztelés során.

- (iv) Érdemes még megemlíteni az egyes projekciók felvételekor érvényes geometriai viszonyok problémáját. Ha a mérési adatok gyűjtése folyamatos üzemmódban történik (a forgatás nem áll meg az expozíció idejére sem), akkor utólag nem triviális meghatározni azt, hogy az egyes projekciókat milyen szögelfordulások jellemzik, illetve, hogy hányadik projekció jelenti a félfordulatot (180°), vagy éppen a teljeset.
- (v) Végül kiemelném még a mérést terhelő zajokat, amiknek többségét szintén kalibráció segítségével lehet csökkenteni. Az előrevetítési szimuláció során azért nem kapunk tökéletes Radon-transzformáltat a fantomról, mert a részecskék száma Poisson-statisztikát követ, aminek relatív hibája N részecske esetén $1/\sqrt{N}$, viszont más zajforrás nincs. A valóságban a detektorunk nem tökéletes, gondolhatunk például ún. halott pixelekre (dead pixel), vagy arra, hogy a detektálásnak 1-től különböző hatásfoka van.

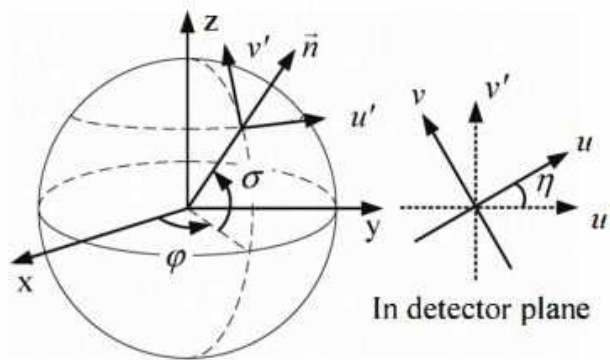
Összefoglalva tehát a rekonstrukció fő kihívásai: részecskeszámok becslése a detektor által mért adatokból, röntgen-spektrum, geometriai kalibráció, folyamatos üzemmódu adatgyűjtés, és a zaj hatása a rekonstrukció minőségére.

11.3 A rekonstrukcióhoz szükséges geometriai adatok

A leképezés során érvényes elrendezést geometriai kalibrációval tudjuk meghatározni. A 42. és 43. ábrán láthatjuk, hogy milyen paramétereket tudunk becsülni az eljárás során. Az előbbi ábrából az egyes elemek elhelyezését jellemző távolságokat, valamint a főnyaláb irányát ismerhetjük meg, utóbbi pedig a detektor dőlésszögeit jellemzi.



42. ábra
A CT berendezés geometriája [18]



43. ábra
Detektor elhelyezkedését jellemző koordináták [18]

Mivel a detektor beállítása megfelelő pontossággal merőleges lett a főnyalábra, így első közelítésben elhanyagolhatjuk a detektor dőlésszögeit. Ekképp a rekonstrukcióhoz szükséges geometriai paraméterek lényegében a 42. ábrán látható mennyiségek. Ezeket a 7. táblázatba gyűjtöttem ki, ahol megadtam a kalibráció során számolt értéküket is, amiket a 11.5 fejezetben látható rekonstrukciókban használtam (a geometriai kalibráció elvégzését [17] mutatja be).

Paraméter jele	Definíciója	Értéke (11.5 fejezet rekonstrukciójában)
D	Forrás-detektor távolság	43,2 cm
R	Forrás-forgástengely távolság	16,5 cm
u_0	Főnyaláb és detektor metszéspontjának vízszintes koordinátája	802 pixel
v_0	Főnyaláb és detektor metszéspontjának függőleges koordinátája	432 pxel

7. Táblázat
A 42. ábra paraméterei és értékei

11.4 Elvárások a rekonstrukció minőségére vonatkozólag

A rekonstrukciótól az eddigi tesztek alapján máris lehetnek várakozásaink, főleg a képminőséget illetően. Láttuk ugyanis (pl. 7. fejezet), hogy a kontúrok elmosódása egy

jellemző tulajdonsága a rekonstrukciónak, de a műtermékek megjelenése egy szűrt visszavetítéshez képest nem számottevő. A rekonstruált kép jel/zaj arányában is várhatunk különbségeket, elsősorban a kis elnyelésű térrészeknek megfelelő képszegmensek lesznek kevésbé zajosak, mint a szűrt visszavetítésnél (ld. 7.4 fejezet). Kontrasztban a Matlab rekonstrukcióját nem sikerült felülmúlni a matematikai fantomok esetében, így nem várunk el kiemelkedően jó kontrasztot (ellenben a valós adatok rekonstrukciója esetén éppen ez változik meg látványosan). Ezekon kívül számíthatunk még esetleg spektrumfelkeményedés okozta műtermékekre is, hiszen jelenleg még nem áll rendelkezésre adat a röntgen-cső spektrumáról.

11.5 Tesztesetek

A 11.2 fejezetben áttekintettük, hogy milyen kihívásokkal szembesülünk a valós mérési adatok rekonstrukciója során. Az indítandó fotonszám esetében a gain-korrigált projekciókra könnyű megoldást találni. A gain-korrektió után ugyanis az egyes detektorpíxelek, amik irányában nem volt elnyelés, a maximális $2^{14}-1$ értéket veszik fel. Meg kell tekinteni, hogy a rekonstrukcióban hány detektorpixel irányában kívánunk szimulációt végrehajtani (esetenként nem akarjuk az egész térfogatot rekonstruálni), és ennyiszor ($2^{14}-1$)-hez kell igazítani a részecskeszámokat. Vagyis akkor, ha M db fotont akarunk szimulálni, de a mérési adatokból számolt részecskeszám N db, akkor egy fotont vegyünk figyelembe N/M súllyal a szimulációban.

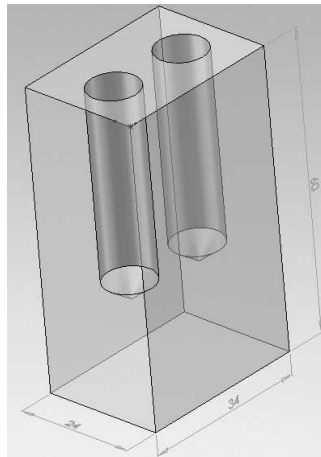
A röntgencső spektrumának kérdése a diplomamunka elkészüléséig még nyitott maradt, ugyanis jelenleg nem áll rendelkezésre adatsor a spektrumra vonatkozólag. A spektrumot kétféleképp lehet meghatározni:

1. Kísérletileg, aminek mérés technikája igen bonyolult
2. Szimulációval (pl. MCNP), ehhez azonban ismerni kell a forrás felépítését

A jelenlegi algoritmusban a röntgensugárzásra vett effektív energiát használtam a spektrum helyettesítésére, vagyis lényegében egy monokromatikus nyalábot indítottam a szimulációban a spektrumban szereplő legnagyobb energia 1/3 részét véve. [19] szerint ugyanis az effektív energiát jól becsli a maximális energia (gyorsító-feszültségnek megfelelő) 1/3-ad része.

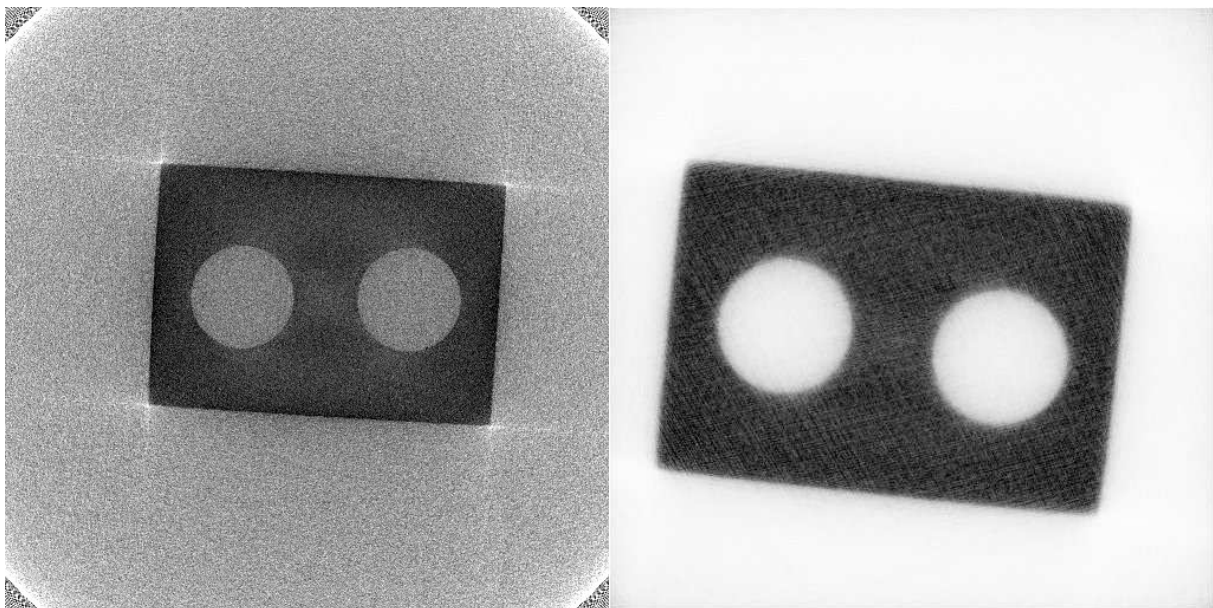
A 11.3 fejezetben említett további kalibrációk elvégzése [17]-ben olvasható, ezek eredményeit (7. táblázat) felhasználva a 44. ábrán látható fantom egy szeletét rekonstruáltam

(amelyik a főnyaláb síkjába esett). A fantom plexiből készült, benne két, egymással és a szimmetriatengellyel párhuzamos furat található.



44. ábra
A plexi tesztfantom [16]

Referencia rekonstrukcióként Deli G. diplomamunkájában [20] bemutatott algoritmus által rekonstruált képet használtam fel. Ez a 45. ábrán balra látható rekonstrukció, ugyancsak GPU-n futó, szűrt visszavetítési algoritmus eredménye. Az általam fejlesztett szoftver, pedig a 45. ábrán jobb oldalt látható képet adta. A rekonstrukciót ~50millió foton/projekció szimulációjával, 100 iterációval hajtottam végre, ami ~1órát vett igénybe.



45. ábra
Szűrt visszavetítés [20] (balra), ML-EM rekonstrukció (jobbra)

	FBP		ML-EM	
	átlag	relatív szórás	átlag	relatív szórás
plexi	32265	0,00577	0,29915	0,36457
furat	32870	0,00555	0,03509	0,18638
kontraszt	0,00930		0,79003	

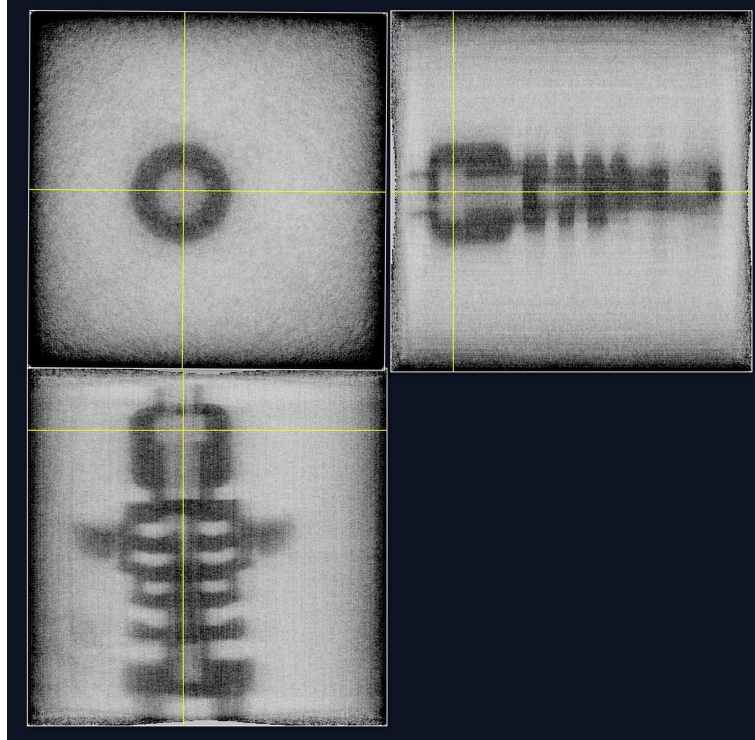
8. Táblázat

Kontraszt és jel/zaj arányok a 45. ábra rekonstruált képein

A voixelértékek máshogy skálázódtak, ezért az átlagokban látható különbségek, azonban a relatív szórás, és a kontraszt is független a skálától

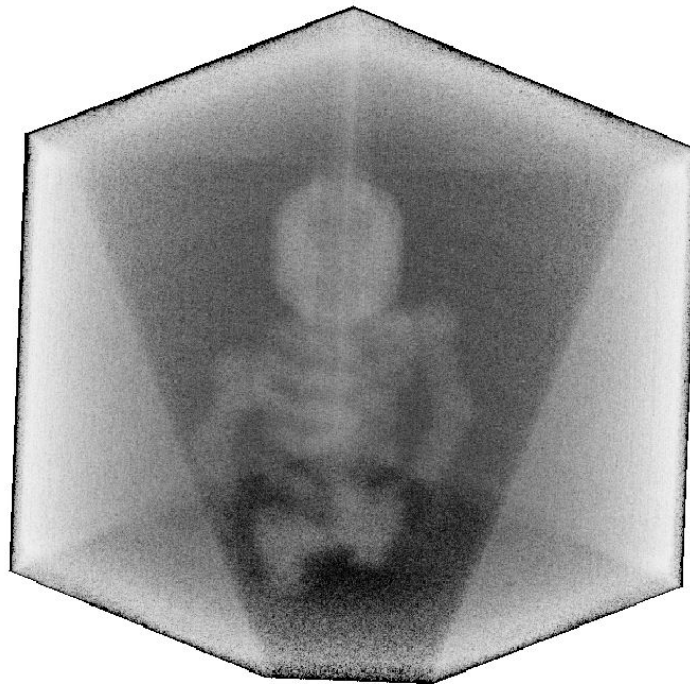
A rekonstruált kép minősége többnyire visszaadja az 11.4 fejezetben leírt várakozásokat. Két, az eddigi tapasztalatoktól való, eltérést emelnék ki: az ML-EM rekonstrukció zajossága, és a meglepően jó kontrasztja. A jel/zaj arányokról és a kontrasztról ad felvilágosítást a 8. Táblázat. Láthatjuk, hogy az ML-EM rekonstrukción megjelenő csíkos műtermék (ami afféle durva felület érzetét kelti) erősen rontja a jel/zaj arányt. A műtermék megjelenésének oka a jövő kutatási részét képezheti. A kontraszt tekintetében lényegesen jobb eredményt ér el az ML-EM képrekonstrukció, mint a szűrt visszavetítés, valamint a háttérzajossága is jelentősen kisebb mértékű. Nem jelennek meg rajta a plexi sarkainál látható csillogás műtermékek sem, viszont a három sávszerű műtermék, ami mintha a furatokat összekötné, előjön mind a két rekonstrukción. Ez utóbbi valószínűleg a spektrumfelkeményedés okozta műtermék lehet. Ennek felderítésére jó módszer lenne egy MCNP szimulációt végrehajtani a röntgencső spektrumának közelítésére, majd ezzel a spektrummal rekonstruálni, ahogy az a 10. fejezetben látható szimulált spektrummal történt.

Végrehajtottam rekonstrukciót egy csontváz alakú lego-figuráról készült felvételtől is. A rekonstrukció a rendelkezésre álló memória által megengedett legnagyobb felbontásban készült (400x400x400). Ez tehát egy teljes 3D inverzió volt, ami alatt azt értem, hogy bár eddig is cone-beam geometriában végeztem az előre-és visszavetítéseket, de a térfogatnak csak egy kiválasztott szeletét rekonstruáltam. Ebben a példában a minta által elfoglalt teljes teret rekonstruáltam, a lego-figura néhány szeletét az 46. ábrán tüntettem fel. Ezeken a képeken is észrevehetjük a csíkos műterméket, aminek eredetére nem derült fény.



46. ábra
Rekonstruált képszeletek a csontváz lego-figuráról (felül, oldal, és előnézet)

A rekonstrukciót három dimenzióban is megjelenítettem (47. ábra), ennek során azonban transzformálnom kellett a rekonstruált nyers adatokat, így ez csupán szemléltető célzatú.



47. ábra
A rekonstruált lego figura 3D vizualizációja

12. Konklúzió, kitekintés

A diplomamunkám során a Lange és Carson által javasolt transzmissziós ML-EM algoritmus egy lehetséges megvalósítását hoztam létre. A lehetséges interpretációk közül kiválasztottam a grafikus kártyán történő implementáció szempontjából a leghatékonyabbat. A GPU-n implementáltam a rekonstrukcióhoz szükséges két kernel-függvényt: előre- és visszavetítés. Utóbbit analitikusan, a Beer-Lambert törvény alapján, az előrevetítést pedig Monte Carlo projektorral valósítottam meg, mivel így a fizikai valóság nagymértékben figyelembe vehető, és ezáltal egy realiztikus rekonstrukcióra (műtermékektől mentes) van lehetőségünk.

A részecsketranszport hatékonyságának maximalizálása céljából szóráscsökkentő eljárásokat dolgoztam ki, bemutattam, hogy a képrekonstrukció során is javasolt a használatuk. A dolgozatban ismertettem az ML-EM módszer előnyeit, hátrányait, alkalmazhatóságát, több esetben összehasonlítva a szűrt visszavetítéssel. Matematikai fantomokon számos szempontból megnéztem a működésének viselkedését, és azt tapasztaltam, hogy jobb képminőség érhető el, mint a szűrt visszavetítéssel. Ez elsősorban a jobb jel/zaj aránynak volt köszönhető. Gyengébb, zajos felvételekből történő rekonstrukciónál az ML-EM módszer lényegesen felülmúlta a szűrt visszavetítés minőségét, kontrasztot, jel-zaj arányt és értékhelyességet (L2 normát) tekintve is. Az ML-EM rekonstrukció ezen tesztek alapján alkalmasnak bizonyulhat alacsony dózisu CT felvételek jóminőségű rekonstrukciójára. További vizsgálatokat is elvégeztem a konvergenciára, szabad paramétereinek optimális megválasztására, képminőségre vonatkozólag. Részlegesen felmértem a paraméterteret, vizsgáltam a képminőség érzékenységét az állítható paraméterekre, útmutatást adtam a korlátozott idő alatt végzett rekonstrukció optimalizálására. Kimutattam a spektrumfelkeményedés jelenségét a rekonstruált képen, majd korrigáltam azt. Háromdimenziós valós adatsorok rekonstrukcióját is elvégeztem, a rekonstruált képeken csíkozódásos műterméket fedeztem fel. A műtermék oka valószínűleg a rekonstrukció és a valóság közti diszkrepancia (és nem a mért adatsorok, kalibrációk hibája), ennek javítása viszont a jövőbeli kutatásra maradt. Ezen hibát leszámítva, a rekonstrukció minősége mind kontraszt, mind jel/zaj arányban meghaladja a szűrt visszavetítését.

A jövőre vonatkozólag az alábbi fejlesztéseket javaslom:

- A valós mért adatok rekonstrukciójánál megjelenő csíkozódás-műtermék okának felderítése és javítása
- Paraméter optimalizáció különböző fantomok esetén
- Az előrevetítés optimalizációja GPU-s hatékonyság növelésével: kevesebb if-else elágazás, hatékonyabb memóriaszál struktúra
- A rekonstrukció futási idejének gyorsítása: implementáció több grafikus kártyán / nagyobb teljesítményű kártyán, OSEM (Ordered Subset EM) implementálása
- A visszavetítésben kijelölt válaszegyeneselek mintavételezése a detektoron Poisson disk sampling módszerrel
- Az Oktatóreaktorban használt röntgen-cső spektrumának mérése/szimulációja, felhasználása a rekonstrukcióban
- A detektálás hatásfokának figyelembevétele
- Szórás-modellezés (Compton, Rayleigh) az előrevetítésben
- Fém-műtermékek megjelenése és korrekciója
- Detektor geometriának nagyobb szabadsági fokú implementációja (ami megengedi a detektor dőlését különböző irányokban)
- Nem pontszerű röntgen-forrás modellezése, első közelítésben Gauss-eloszlás szerinti indulási hely a részecsketranszportban
- Felbontás meghatározása modulációs transzfer-függvény (MTF) segítségével
- Regularizációs eljárások (például TV -Total Variation) implemetációja

13. Irodalomjegyzék

- [1] K. Lange, R. Carson : "*EM Reconstruction Algorithm for Emission and Transmission Tomography*", **Journal of Computer Assisted Tomography**, 8(2):306-316
- [2] Molnár B. : "*Iteratív ML-EM CT képrekonstrukció GPU-ra*", TDK dolgozat (Budapesti Műszaki és Gazdaságtudományi Egyetem), 2013
- [3] G. Jakab, T. Huszár, B. Csébfalvi: "*Iterative CT Reconstruction on the GPU*", Sixth Hungarian Conference on Computer Graphics and Geometry, Budapest, 2012
- [4] B. De Man : „Iterative Reconstruction for Reduction of Metal Artifacts in Computed Tomography”, Doctoral Thesis (Katholieke Universiteit Leuven), 2001
- [5] F. E. Boas, D. Fleischmann : "*CT artifacts: Causes and Reduction Techniques*", **Imaging Med.** (2012) 4(2), 229-240
- [6] M. N. Wernick, J. N. Aarsvold: "*Emission Tomography*", 2004, Pages 70-75
- [7] Molnár B. : "Iteratív képrekonstrukciós eljárás gamma tomográfiához: tesztelés és verifikáció", Szakdolgozat (Budapesti Műszaki és Gazdaságtudományi Egyetem), 2013
- [8] NVIDIA CUDA™, „NVIDIA CUDA C Programming Guide”, 2012
- [9] Guillem Pratx, Lei Xing : „*GPU computing in medical physics: A review*”, **Medical Physics** **38 (5)**, May 2011
- [10] M. McCool, E. Fiume : "*Hierarchical Poisson Disk Sampling Distributions*", Proceedings of the conference on Graphics Interface '92, Pages 94-105,
- [11] I. Lux, L. Koblinger: "*Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*", 1991, 39-41
- [12] P. Saidi, M. Sadeghi, C. Tenreiro: "*Variance Reduction of Monte Carlo Simulation in Nuclear Engineering Field*", 2013, Theory and Applications of Monte Carlo Simulations, Chapter 7
- [13] D. P. Naidich, C. H. Marshall, C. Gribbin, R. S. Arams, D. I. McCauley : "*Low-Dose CT of the Lungs: Preliminary Observations* ", **Radiology** 175:729-731, 1990
- [14] Y. Sagara, A. K. Hara, W. Pavlicek, A. C. Silva, R. G. Paden, Q. Wu: "*Abdominal CT: Comparison of Low-Dose CT With Adaptive Statistical Iterative Reconstruction and Routine-Dose CT With Filtered Back Projection in 53 Patients*", **American Journal of Roentgenology**: **195**, September 2010
- [15] <http://www.dotmed.com/news/story/18437>, 2013. május 6.

- [16] Jakab G. : "*Hibrid Monte Carlo CT szimuláció a GPU-n*", KÉPAF 2013 - a Képfeldolgozók és Alakfelismerők Társaságának 9. országos konferenciája
- [17] Tölgyesi B. : "*Röntgen CT készülék építése*", Diplomamunka (Budapesti Műszaki és Gazdaságtudományi Egyetem), 2014
- [18] Dufan Wu, Liang Li, , Li Zhang, Yuxiang Xing, Zhiqiang Chen, Yongshun Xiao: „*Geometric calibration of cone-beam CT with a flat-panel detector*”, 2011 IEEE Nuclear Science Symposium Conference Record, MIC I2.M-79
- [19] Resource manual for compliance test parameters of diagnostic x-ray systems, Diagnostic Devices Branch, Division of Enforcement I, Office of Compliance
- [20] Deli G. : "*Általános tomográfiás rekonstrukciós szoftver fejlesztése GPU-ra*", Diplomamunka (Budapesti Műszaki és Gazdaságtudományi Egyetem), 2014