



LUND
UNIVERSITY

Master of Science Thesis
2026

Comparison of Detectability Between Digital LYSO and BGO-Based PET/CT Scanners: A Phantom Study

Marielle Sohlman

Supervisors

Sofia Kvernby and Anna Stenvall

Medical Radiation Physics, Lund
Medical Physics Programme
Faculty of Science
Lund University, Sweden
www.msf.lu.se

Populärvetenskaplig sammanfattning

Positronsemissionstomografi (PET) är en bildtagningsmetod där man med hjälp av radioaktiva läkemedel undersöker funktion i kroppens organ. En vanlig diagnos som undersöks med PET är spridning av cancer, där det oftast användas radioaktiva läkemedlet är fluorodeoxyglukos (FDG), även kallat radioaktivt socker. Eftersom cancerceller kräver mer socker än celler i normalvävnad tas FDG i större utsträckning upp i cancercellerna i jämförelse med de normala cellerna. Den radioaktiva delen i läkemedlet är fluor-18, vilken är en fluorisotop som sönderfaller med positronstrålning. Från positronen som skickas ut från sönderfallet skapas två fotoner som sänds ut i en 180° vinkel. Från många positronsönderfall skapas alltså lika många fotonpar och det är dessa fotonpar som undersökningen går ut på att samla in. Insamlingen av fotonparen görs med en PET kamera, där fotonerna bromsas upp och avger energi som sedan kan utläsas som en elektrisk signal. Från den insamlade signalen skapas en avbildning av patienten i kameran, där områden med högt upptag av läkemedlet syns tydligare än områden med lågt upptag.

Materialet i PET-kameran som bromsar upp fotonerna är en kristall, med specifika egenskaper som gynnar den återskapade bilden på olika sätt. Kristallen ska önskvärt kunna samla in så många fotonpar som möjligt samt vara bra på att omvandla fotonparet till en optisk signal. Något som forskningen har strävat efter är att kristallen ska vara snabb med att sända ut den optiska signalen från att fotonen har avgett sin energi i kristallen. Om kristallen är tillräckligt snabb i detta steg är det möjligt att mäta tidsskillnaden mellan de två fotonerna i fotonparet. I sin tur ger detta en mer säker position i bilden, på vart positronsönderfallet skedde i kroppen. Möjligheten att kunna mäta tidsskillnaden mellan fotonerna kallas time-of-flight (TOF). Ett PET system med TOF har därför bättre förutsättningar för att noggrannare kunna markera ut var exempelvis små tumörer är lokaliserade i jämförelse med ett PET system utan TOF. Nyligen installerades en PET kamera i Lund (OMNI Legend) där TOF inte är möjligt, då kristallerna är för långsamma. I stället har denna kristall andra fördelaktiga egenskaper, så som god förmåga att fånga fotonerna samt att den är relativt billig. I kompensation till förlusten av TOF har företaget som skapade kameran tränat en AI modell som ska efterlikna sann TOF, och på så vis kan kristallens andra egenskaper utnyttjas utan att resultatet ska försämras av de långsamma kristallerna.

I detta arbete undersöktes skillnaden i hur väl små lesioner, eller tumörer, kan uppfattas eller detekteras mellan de två olika PET systemen (OMNI Legend och Discovery MI) för att ta reda på om de presterar lika bra under lika förutsättningar. En viktig parameter i projektet var att undersöka hur patientstorleken påverkar detekterbarheten, då lesioner är svårare att korrekt avbilda i mer utmanande patientgeometrier, så som i större patienter. Som representation av patienter användes fantom, vilket är en form av vattentät plastbehållare. Ett mindre cylindriskt fantom och ett 3D-printat skal till det mindre fantomet användes för att representera olika patientstorlekar. I det mindre fantomet placerades sfärer i olika storlekar, som skulle motsvara olika stora lesioner. Fantomen och sfärerna fylldes sedan med utspädd 18F-FDG, där sfärerna hade tio gånger högre koncentration än de stora volymerna, och placerades sedan i de olika kamerorna för avbildning.

Resultatet från mätningarna visar på att detekterbarheten av sfärerna minskar med ökad fantomstorlek och att mindre sfärer är svårare att se, trots samma aktivitetskoncentration. Detekterbarheten hos det mindre fantomet visade ingen större skillnad mellan kamerorna, medan för det större fantomet visade sig detekterbarheten vara bättre på kameran utan sann TOF (OMNI Legend). Detta är enligt teorin inte väntat, då fördelen av sann TOF förväntas vara större för större patient. Utvärderingen av de applicerade AI modellerna visade ingen större förbättring i jämförelse med datan utan AI modell.

Abstract

Background and Aim: A new PET/CT scanner using BGO crystals was recently introduced in the clinic. Unlike scanners with LSO/LYSO crystals, this scanner does not include the time-of-flight (TOF) property, due to low timing resolution of BGO crystals. To compensate for the lack of TOF, the reconstructed images can be filtered with an AI model, called Precision DL (PDL), which has been trained on images with and without TOF. It still remains unknown how well the AI model can enhance small lesions in a patient compared to a system with TOF, especially in more challenging conditions, such as larger patients. The aim of this project is to compare detectability of small lesions between a PET/CT system with true TOF (GE HealthCare's Discovery MI) and a PET/CT system with AI based TOF (GE HealthCare's OMNI Legend).

Materials and Methods: To compare detectability of small lesions between the two systems, a Jaszczak phantom and a 3D printed XL shell to Jaszczak were used. The XL phantom was divided into two cases, the shell filled with and without background activity. Six hot spheres in different sizes were placed in Jaszczak, in a sphere-to-background ratio (SBR) of 10:1, where the spheres had an activity concentration of 12 kBq/ml at start of scan. The PET radiopharmaceutical used was ^{18}F -FDG and the phantoms were scanned sequentially on the two PET systems (DMI and OMNI). Images were acquired in listmode and rebinned into ten 90-s frames per phantom and PET-system, with one bed position and matrix sizes of 256×256 and 384×384 . Phantom images were reconstructed using OSEM + PSF together with a post filter of 3 mm and BSREM with β value of 500. The reconstructions were made with and without TOF for the DMI scanner, whereas the BSREM reconstructions for the OMNI scanner were combined with the three PDL levels. Recovery coefficients (RC) were calculated from peak values in the spheres and the calculated activity concentrations in the spheres at the start of scan, and signal-to-noise ratio (SNR) was determined in the background for each phantom. A peak-to-peak (P2P) value was calculated as the ratio of the peak in a sphere to the peak in the background and was evaluated as a detectability value. Non-blinded readers did a visual evaluation of a sample of the images to visually estimate the lesion detectability.

Results: The results showed that both reconstruction methods with TOF (OSEM and BSREM) with the DMI scanner gave higher RC than all reconstructions from the OMNI scanner across all three phantom cases. The TOF effect was clearly shown for the DMI scanner, while the PDL filters seemed to give little to none effect on the recovery. A decrease in the background SNR was seen in reconstructions with TOF compared to those without TOF, while the PDL models had an increasing effect on the SNR. The P2P values for the DMI tend to be higher for reconstructions with TOF compared to those without. For the Jaszczak phantom, the DMI scanner reaches higher P2P values than the OMNI scanner, while the opposite trend is seen for the XL phantom. In general, higher P2P values were reached for Jaszczak compared to the XL phantom, which means that more spheres were thought to be seen in the images for the smaller phantom. This result was confirmed through the visual evaluation, which also resulted in higher means of detected spheres for the BGO system. The results of the P2P values for both systems also showed a large spread of the data points within the ten time frames of the XL phantom.

Conclusion: The P2P values together with the visual evaluation conclude that a smaller phantom has greater detectability than a larger one. The P2P values show that the benefit of true TOF is greater for the Jaszczak phantom than the XL phantom, which contradicts the theory. The benefits of the BGO system were also shown to be larger in the phantom case with no activity in the XL shell, which is the most patient like phantom case. Further studies are needed to confirm the effect of true TOF versus the sensitivity of the BGO crystals for larger patients.

Abbreviations

- BGO - Bismuth Germanate Oxide
- BSREM - Block Sequential Regularised Expectation Maximisation
- CT - Computed Tomography
- DL - Deep Learning
- DMI - Discovery MI
- FDG - Flourodeoxyglucose
- FWHM - Full Width at Half Maximum
- LMM - Linear Mixed Model
- LOR - Line Of Response
- LYSO - Lutetium Yttrium Orthosilicate
- ML-EM - Maximum Likelihood-Expectation Maximisation
- NECR - Noise Equivalent Noise Rate
- OS-EM - Ordered Subset-Expectation Maximisation
- P2P - Peak-to-peak
- PDL - Precision Deep Learning
- PET - Positron Emission Tomography
- PL - Penalized Likelihood
- PSF - Point Spread Function
- RC - Recovery Coefficient
- RDP - Relative Difference Penalty
- ROI - Region Of Interest
- SBR - Sphere-to-Background Ratio
- SiPM - Silicon Photomultiplier
- SNR - Signal-to-Noise Ratio
- TOF - Time OF Flight
- VOI - Volume Of Interest

Contents

1 Introduction	1
1.1 Aim	1
2 Background	2
2.1 The PET scanner	2
2.1.1 Detected events	2
2.1.2 Performance parameters	2
2.1.3 Time-of-flight	3
2.1.4 Effect of patient size	4
2.2 Scintillation crystals	5
2.2.1 LYSO	5
2.2.2 BGO	6
2.3 Reconstruction of PET images	6
2.3.1 Penalized likelihood - Q.Clear	6
2.4 Deep learning to replace time-of-flight	7
2.5 ^{18}F -FDG	7
3 Materials and Methods	8
3.1 PET/CT systems	8
3.2 Phantom structure	8
3.2.1 Jaszczak phantom	8
3.2.2 Phantom printing	9
3.3 Measurement preparation	10
3.4 Scanner measurements	10
3.5 Image reconstruction	11
3.6 Data analysis	12
3.6.1 Extraction of data values from VOI and ROI	12
3.6.2 Recovery coefficients	13
3.6.3 SNR in phantom background	13
3.6.4 Peak-to-peak in spheres and background	13
3.6.5 Visual evaluation of reconstructed images	13
3.7 Statistical analysis	14
4 Results	14
4.1 Recovery coefficients	14
4.2 Signal-to-noise ratio in the background	16
4.3 Peak-to-peak as detectability value	17
4.4 Visual evaluation of reconstructed images	18
4.5 Statistical analysis	19
5 Discussion	25
5.1 Recovery coefficients	25
5.2 Signal-to-noise ratio in the background	26
5.3 Peak-to-peak as detectability value	26
5.3.1 Peak-to-peak in relation to RC and SNR	28
5.4 Visual evaluation	28
5.4.1 Visual evaluation in relation to peak-to-peak	29
5.5 Variation in repeated measurements	29
5.6 Limitations and future aspects	30

6 Conclusion	30
Acknowledgements	31
References	31

1 Introduction

In present healthcare positron emission tomography (PET) plays a valuable role in detecting and characterising tumours and is an important factor in treatment strategies over a wide range of applications^[1]. Even though it has a well-defined role in oncology, another important aspect of PET imaging lies in its ability to diagnose neurologic disorders and evaluating patients with coronary artery disease. The detection sensitivity of PET compared to other imaging modalities is significantly higher, in the field of tracers with low concentrations. This makes PET a useful tool in detection of lesions in early stages. Together with computed tomography (CT), hybrid PET/CT systems merge the functional information from the PET image with the morphologic data from the CT, creating a basis for a more accurate diagnosis^[2].

Almost all PET scanners today are purchased as hybrid PET/CT systems, where the CT system not only provides anatomical localisation, but also attenuation correction. These two features have significantly improved the performance of nuclear imaging, but there are also various parameters associated with the PET scanner that influence the system performance. Improvements in a PET scanner's spatial resolution, system sensitivity, timing resolution, and energy resolution largely depend on the choice of scintillation crystal and the scanner geometry^[2]. Historically, the electronic readout on PET systems have been based on scintillation crystals coupled to photomultiplier tubes. Development has led to new applications, offering opportunities for other photodetectors, such as silicon photomultipliers (SiPMs). The SiPMs have shown great potential to be the leading photodetector for most PET systems^[3]. In recent years, most PET scanners available have been manufactured with lutetium-(yttrium) oxyorthosilicate (L(Y)SO) crystals, chosen for their high light output and exceptional timing resolution^[1] ^[2]. One of the limitations with the L(Y)SO crystals is the higher cost, compared to other crystals such as the bismuth germanate oxide (BGO)^[2]. The BGO was the choice of crystals during the 1980s and 1990s, with its strong stopping power, but due to the limited light yield, the systems resulted in relatively poor energy resolution. Therefore, it was a natural next step to search for more suitable crystals, in hope of increasing the system's performance and image quality^[4].

A great way to increase the performance of a PET system is the use of the time-of-flight (TOF) property, which is enhanced by crystals with good timing resolution. A system with L(Y)SO crystals can improve image quality further with the TOF capability, while the slow BGO crystals cannot. Thanks to artificial intelligence (AI) new possibilities have been introduced in PET research. This has lead to GE HealthCare recently introducing a PET/CT system with digital BGO detectors and AI modelled TOF. In this way the excellent stopping power is combined with the detection capabilities of SiPMs, while still compensating for the lack of TOF. It also has created new possibilities to lengthen the axial field of view (AFOV), which produces a clear increase in system sensitivity^[5].

1.1 Aim

The aim of this project was to compare detectability of small lesions between a PET/CT system with true TOF (GE HealthCare's Discovery MI) and a PET/CT system with AI based TOF (GE HealthCare's OMNI Legend). The project included different phantom sizes, representing different patient body types, to investigate how size influences detectability and whether the two systems respond to the size parameter in different ways. Various reconstruction methods were also included to emphasize different aspects of the PET systems and to create various approaches when comparing the two.

2 Background

2.1 The PET scanner

2.1.1 Detected events

Positrons emitted by PET radiotracers results in two back-to-back 511 keV annihilation photons. If the photons are of correct energy and detected in coincidence, the event of the annihilation can be assumed to have occurred along the line between two detector elements, the so called line of response (LOR). If the photons are indeed from the same annihilation event, the event is a true coincidence [6]. However, the measurement will contain a large fraction of incorrect coincidences, due to limited energy and timing resolution of the PET scanner. Incorrect coincidences can either be scattered coincidences, due to Compton scattering of one or two photons, or random coincidences, from photons from two different annihilation sites. The different coincidences are shown in Figure 1. These events introduce incorrect LORs and give false information about the scanned object. Whether the coincidence is a true, scattered or random coincidence, it is registered as a coincidence if the two photons are detected within a given time coincidence window, which later must be compensated for [4]. A large fraction of the detections are not detected in coincidence, the so called singles. However, the singles are still a problem since they contribute to the count rate seen by the system, which becomes a problem because too high count rates can saturate the detectors [7].

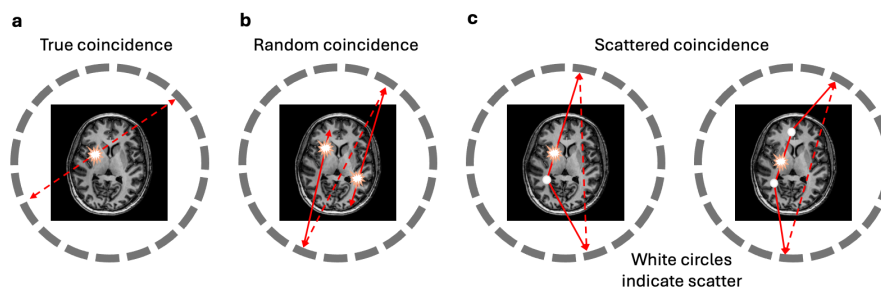


Figure 1: The images show illustration for the different detection events: a) Detection of two annihilation photons from the same decay in detector elements opposite from each other, creating a true coincidence and a true LOR going through the decay site. b) Detection of two annihilation photons from two different decays within the time window, creating a random coincidence and a false LOR. c) Detection of either one or two Compton scattered photons, where the white circles indicate a scatter, in both cases creating a false LOR. Illustration inspired from Magdy M. Khalil (Ed.) [2]

2.1.2 Performance parameters

The performance of a system depends on various parameters such as system sensitivity, noise equivalent count rate (NECR), coincidence timing window, spatial resolution and time-of-flight (TOF). The last parameter will be further discussed in the next subsection. The system sensitivity is a parameter that highly effect the SNR and image quality and is a measure of the amount information that can be gather during a scan. The sensitivity can be divided in two components: geometric and intrinsic efficiency. The former is depending on the geometric design of the detector components and how close together the detector elements are packed. The geometric efficiency can be increased by decreasing the diameter of the detector rings or by increasing the AFOV, i.e. by extending the tunnel. Although decreasing the ring diameter would benefit the sensitivity, it would at the same time negatively effect the spatial resolution, due to parallax

errors. Increasing the AFOV is instead related to the cost of detector material, since more scintillation crystals, readout channels and electronic circuitry are needed. The second component, intrinsic efficiency, depends on the scintillation crystal material and thickness, which will be further discussed in a later subsection.

The need for a high count rate ability is important from the statistical point of view, which can be obtained by either higher administered activities or longer acquisition times. These two ways are limited by radiation protection and patient inconvenience respectively. If one were to ignore the former, another problem would arise, in terms of count rate saturation and dead time losses. The rate of random coincidences increases quadratically with the injected activity, which therefore differ between patients and type of scan. A convenient metric to measure the system's count rate performance is the NECR. This value measures true events while accounting for dead time, scatter and random events. The count rate performance is directly depending on the electronic timing window, which sets an acceptance boundary for coincident events. The window width plays a significant role in which events that are included and rejected, where it is desired to reject more false than true events. Shortening the timing window increases the peak NECR, and therefore improves the NECR measurement.

The spatial resolution of a PET system is a crucial parameter for image evaluation, and a lot of factors affect the resolution limitation. First, it is limited by the positron range in tissue, which is in the fundamental nature of positron annihilation. Hence, the annihilation site is not the same as the decay site. Luckily, for most usable positron emitters, the positron range is relatively small. The second limiting factor is acollinearity between the two annihilation photons, which one normally would expect to be antiparallel. Due to variation in momentum of the positron, an angular uncertainty in the direction of the photons occur. This effect increases with increasing detector ring, but can be reduced by including the acollinearity in the iterative scheme[2]. A third important factor is the intrinsic spatial resolution of the detector, which often is determined by the crystal width d . The resolution of a coincidence detector-pair is often measured as the full width at half max (FWHM) of the point spread function (PSF) from convolution between the two detectors. The spatial resolution of a system, by including parameters effecting the resolution, can be approximated to

$$\text{FWHM} = \sqrt{(d/2)^2 + b^2} + (0.002D) + r^2 \quad (1)$$

where b account for secondary losses in photon detection, $(0.002D)$ is a factor compensating for acollinearity and D is the diameter of the detector ring, and r^2 relates to blurring effects due to positron range[2]. A last limiting parameter is the parallax error, which is introduced due to the uncertainty in depth of the photon interaction in the crystal. If the photon enters the crystal in a sideways direction, the photon may not even be detected in the first detector element it encountered, thus leading to incorrect LOR.

2.1.3 Time-of-flight

To increase a PET system's performance one approach is to find more timing information about the coincidence. We know that the annihilation occurred somewhere along the LOR, but not closer on where. The two photons are detected with a certain time difference. With this information it is possible to estimate the spatial location of the annihilation along each LOR, and is used in image reconstruction to improve the image quality[4]. This technology goes under the name time-of-flight (TOF). For systems without the TOF capability, the annihilation location arises equally likely along any point on the LOR[2], while for TOF system the emission location along each LOR is estimated by TOF resolution[8]. The time difference in arrival at

the detectors can be expressed in relation to the position of the positron source according

$$\Delta t = \frac{2\Delta x}{c} \quad (2)$$

where Δt is the time difference, Δx is the distance from decay site to the midpoint between the two detectors and c is the speed of light. The concept of TOF is illustrated in Figure 2, showing how the LORs differ between systems with and without TOF. The timing resolution of TOF is in practice not so perfect as it in theory can be. The annihilation detections of the two photons only create an approximation of the position of the source. Not only is the timing accuracy measured from the full-width at half maximum (FWHM) around a mean time difference, but it also depends on crystal properties, such as size, and electronic timing circuits [9].

The introduction of TOF detectors has been a major driving factor for improvements made in the imaging performance of clinical PET systems. The detectors, that initially were based on photomultiplier tubes, later SiPMs, creates a system of high spatial resolution and detection efficiency [6]. The TOF property can be achieved by scintillator crystals with high efficiency and with a short decay constant, often called "fast scintillators". A scintillator with a relatively long decay time has a worsening response time, making it impossible to determine the time difference between the two detections from a decay [2].

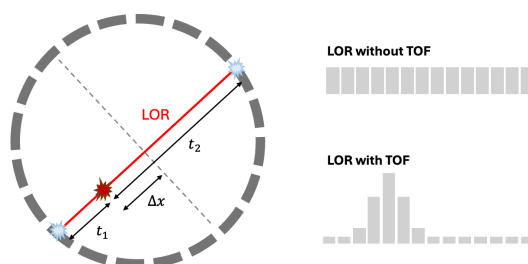


Figure 2: Illustration of annihilation photons being detected in two opposite detector elements, at time t_1 and t_2 . Depending on whether the PET/CT system has the TOF ability, the probability along the LOR either is evenly distributed or pushed toward a more probable location. The distance, Δx , between the decay site and the midpoint between the detector elements can for a TOF system be calculated from the time difference of the two detected events. Illustration inspired from Jansen et al. [10].

2.1.4 Effect of patient size

In the process of evaluating PET images, good spatial resolution is important for good contrast and more accurate estimation of pharmaceutical uptake in small lesions. Another important factor for a system's performance is the sensitivity, which can be improved by going from 2D to 3D scanners and increasing the AFOV. The sensitivity influences the number of detected true, scattered and random coincidences, and hence also affects the noise level. Although image quality increases with a more sensitive system, the difference between a small and large patient also increases. The degradation in image quality results from more attenuation and the increase in scattered and random coincidences in larger patients [11]. Noisier images can reduce the visual detection of lesions [12] and therefore lead to uncertain diagnosis [13]. By introducing TOF PET imaging of 3D systems, the difference between light and heavy patients are believed to decrease [11].

It has been shown that for a simple and uniform activity distribution the relative SNR gain

of TOF compared to a non TOF PET system is

$$\frac{\text{SNR}_{\text{TOF}}}{\text{SNR}_{\text{nonTOF}}} = \sqrt{\frac{D}{\Delta x}} \quad (3)$$

where D is the size of the object being imaged and Δx is the distance from decay site to the midpoint between the two detectors [9] [14]. This could be interpreted as if the patient is larger, the extra information about the location is more valuable than for a small patient, where the extra information has lower impact. Since the benefit of TOF is higher for heavier patients, the TOF addition in clinic leads to a more uniform performance over all patient sizes [15].

The SNR gain received with TOF, expressed in Equation [3], was derived for data acquired in 2D and for reconstructions made with filtered back projection. Today, scanners operate in fully-3D mode and reconstructions algorithms use iterative methods. The increased sensitivity creates a different impact of scatter and random coincidences, and the choice of number of iterations impacts the contrast and noise in a non-linear way. Therefore, assigning a value of the SNR gain from TOF becomes impossible and Equation [3] cannot be used directly. However, the benefit of TOF of current PET systems is still in agreement with the previous, in terms of the fact that the SNR gain is higher for larger patients [15].

2.2 Scintillation crystals

The most essential component in detection of the annihilation photons is the scintillation crystal. The purpose of the crystal is to convert the energy from the incident photons into light quanta [2], i.e. a signal giving the position, energy and timing of the event [10]. When a photon hits the crystal it deposits its energy in the crystal, mostly by photoelectric emission and Compton scattering. Electrons released from these processes excite other electrons in the scintillator material, which in turn decay and emit scintillation light. The choice of composition of the scintillation crystal is important in the performance of the detector, and at some point crystal qualities must be traded at the cost of others [2]. Different parameters of the crystals used in the two scanners that are examined in this project are shown in Table [1].

Table 1: Various crystal properties of the two scintillator materials used in the two PET systems (BGO and LYSO:Ce) [2].

Scintillation crystal properties	BGO	LYSO:Ce
Density (gm/cm ³)	7.13	7.1
Effective atomic number (Z)	74	64
Linear attenuation coefficient (1/cm)	0.92	0.86
Light yield (% NaI[Tl])	15	75
Decay time (ns)	300	41
Emission maximum (nm)	480	420
Hygroscopic	No	No
Photoelectric effect (%)	40	33
Refractive Index	2.15	1.81

2.2.1 LYSO

A common choice of scintillation crystal in PET scanners is cerium doped lutetium yttrium orthosilicate (LYSO:Ce or LYSO). LYSO is, thanks to its relative high effective Z , high density

and high linear attenuation constant, an efficient scintillator for stopping the annihilation photons. LYSO also has an attractive property in short decay time (41 ns), or fast response, which makes this crystal popular for TOF applications[2].

2.2.2 BGO

For early PET systems, bismuth germanate oxide(BGO) was an attractive crystal thanks to its high photofraction and high density, higher than for LYSO[5]. BGO also has a higher stopping power relative to other crystal alternatives, resulting in a higher sensitivity and bigger reduction of scatter inside the crystal[10]. The BGO crystal have due to its low light yield and its poor response time and energy resolution become inferior to the LYSO crystal. The poor response time causes an increase in dead time and random coincidences, and makes the TOF application not possible[2].

2.3 Reconstruction of PET images

The choice of reconstruction method plays a huge role in the outcome of image quality and enhancement of important structures. Reconstruction of images within the field of nuclear medicine concern the pixels or voxels values that are used to define the radiotracer distribution in the scanned object. The best method to do this is to use iterative methods. Compared to non-iterative methods, such as filtered back projection, iterative reconstruction methods more accurately model the discretization of the radiotracer distribution, the mean and the noise in the collected data[4]. The iterative methods started with the expectation maximization (EM) algorithm, which is a technique to determine the maximum likelihood (ML) estimates[16]. The combined ML-EM algorithm (first proposed by Shepp and Vardi[17]) is nowadays well-known and has become a ground for development of even better emission algorithms. The ML-EM algorithm works in a series of steps to calculate the estimate of the radioactive distribution $\mathbf{F}$ from a given set of projections $\mathbf{P}$ and iterate the steps to improve this estimate[4]. Positron emission is characterized by a Poisson distribution[18] and for this kind of data, the maximization of the statistical probability can mathematically be described as the logarithm of the likelihood

$$L(\mathbf{F}) = \ln[\text{pr}(\mathbf{P}|\mathbf{F})] = \sum_{i=0}^{m-1} ((p_i \ln[\mathbf{A}\mathbf{F}]_i - [\mathbf{A}\mathbf{F}]_i - \ln(p_i!)) \quad (4)$$

where $\mathbf{A}$ is the system matrix defined by $\mathbf{A}\mathbf{F}=\mathbf{P}$. The equation seeks the activity distribution $\mathbf{F}$ that maximises the log-likelihood. Since the ML-EM algorithm uses all measured data in every iteration to maximise the log-likelihood, the algorithm is rather slow. By splitting the data in smaller sets and only use a fraction of them in each iteration, the algorithm can be accelerated. This method is call ordered subsets expectation maximization (OSEM)[4]. A problem with this approach is the increased noise level after each iteration and due to this the algorithm is stopped before full convergence can be reached. In clinical applications, post-filtering is applied, but the optimal filtering parameters and number of iterations are for a patient population difficult to determine[19], which becomes limiting factors in the ability to detect small lesions[20].

2.3.1 Penalized likelihood - Q.Clear

An alternative to the OSEM algorithm is to use an reconstruction algorithm with Bayesian penalized likelihood (PL). The PL method contains a penalty function that controls the noise and resolution and is optimized until convergence is accomplished. The PL algorithm can, in contrast to OSEM, be customized for desired image properties[19]. A PL algorithm that has been clinically introduced is Q.Clear (GE HealthCare), which includes a penalty term, optimized by the image noise level. The penalty in the Q.Clear algorithm provides activity dependant noise

control by using Relative Difference Penalty (RDP), which preserves edges while suppressing background image noise [21]. It is given by:

$$\beta R(x) \equiv \beta \sum_{j=1}^{n_v} \sum_{k \in N_j} w_j w_k \frac{(x_j - x_k)^2}{(x_j - x_k) + \gamma |x_j - x_k|}, \quad (5)$$

where β is a regularization strength parameter, n_v and N_j indicate the number of voxels and the neighbours of voxel j , respectively, w_j and w_k are penalty weights for different function components and γ tunes the edge-preservation [21] [22]. To maximise the PL, Q.Clear uses the Block Sequential Regularization Expectation Maximization (BSREM) algorithm, which guaranties full convergence for every voxel in the image [21]. Results from Miwa et.al. [23] show that the PL with a high γ preserve the edges, while a low γ promoted image smoothing. However, the RDP in Q.Clear has a fixed γ value of 2, and therefore cannot be user controlled. The β value can be user optimized, and results from Lantos et.al. [22] have showed that a value of at least 350 may be appropriate, especially for low count statistical data.

2.4 Deep learning to replace time-of-flight

During the latest years, the scintillation crystals in most PET scanners on the market have been LYSO-based, chosen for their excellent timing resolution. However, this has not stopped researchers from trying to increase a PET system's spatial resolution and overall sensitivity. For this, BGO seemed as a good candidate, as its high stopping power has been shown to increase sensitivity to LYSO [1]. To compensate for the bad timing resolution of BGO, GE HealthCare has developed a deep learning (DL) algorithm, called Precision DL (PDL). This model used deep convolution neural network that learned and imitated the properties of TOF images. The training of Precision DL has been made on pairs of TOF and non-TOF images, to learn TOF features. The used patient images were from ^{18}F -FDG datasets from GE HealthCare's Discovery MI scanner (DMI), reconstructed with GE HealthCare's own BSREM reconstruction algorithm, Q.Clear, with and without TOF. Precision DL comes with different degrees of TOF strengths, for the user's choice of balance between contrast and noise in the image. Three different levels - "low" (L), "medium" (M) and "high" (H), were trained on images with different beta values, which tend to emphasize contrast and noise differently. The L model was trained on images with higher β values and emphasizes noise reduction, while the H model was trained on images with lower β values and emphasizes contrast. To complement the PDL models, the user can control the beta value of the input image, in the BSREM reconstruction. Instead of reconstruct images with a high beta value to reduce noise, the L model can compensate for this effect [10]. According to Jansen et.al [5] the choice of the BGO scintillator was made to highlight sensitivity over timing resolution. Phantom measurement with GE HealthCare's DMI 25 cm system was compared to the OMNI Legend 32 cm system, which showed that the Q.Clear reconstruction with OMNI Legend was more sensitive than the DMI system with TOF, and with applied PDL filter even the CNR was higher than the TOF system [8]. Worth noting is that the PDL was developed to enhance the image quality of non-TOF images, in which lesions are already present. Images without any signs of lesions can hence not become "better" in terms of lesion detectability.

2.5 ^{18}F -FDG

Fluorodeoxyglucose(^{18}F), or ^{18}F -FDG, is a diagnostic radiopharmaceutical used for tumor detection. FDG is a glucose analog, i.e. a chemical compound that structurally resembles glucose. This compound is accumulated in all cells that use glucose as its primary energy source, which is why FDG is accumulated in tumors with high glucose metabolism. In healthy patients FDG is well distributed in the whole body, especially to the brain and heart, and in less extent to the lungs and liver [24].

Fluorine-18, or ^{18}F , is the most popular radionuclide for positron imaging, thanks to some favorable properties. Because of fluorine’s physical properties such as high electronegativity and the ability of forming strong bonds with carbon, the products of fluorine are highly stable. This property is favorable for radiotracers, making the radiopharmaceuticals produced with ^{18}F highly stable. The decay of ^{18}F is essentially by positron decay (97%), and the rest by electron capture (3%). The positron emission is of low energy, with a maximum β^+ energy of 0.635 MeV and a maximum range in water of 2.4 mm, which is favorable for the image resolution. The half-life of ^{18}F is 109.8 minutes, which is a good half-life considering the radiation dose to the patients, and also that it makes it possible to transport the radiopharmaceuticals to other nuclear medical centers [25].

Together with other PET radionuclides, ^{18}F has become very important for cancer detection and treatment monitoring. The radionuclides are also used in PET for studying the brain physiology and pathology, as well as being significant in cardiology. [26].

3 Materials and Methods

3.1 PET/CT systems

The two different PET/CT systems used in this project were the GE HealthCare’s Discovery MI (DMI) with LYSO crystals, and the GE HealthCare’s OMNI Legend with BGO crystals. Both scanners have a digital PET detector technology, using the SiPM block design. Different system parameters are stated in Table 2.

Table 2: Performance specifications for the two PET/CT systems used in this work.

PET Performance Specifications	Discovery MI	OMNI Legend
Scintillation material	LYSO	BGO
Scintillation dimensions	3.95 mm $\times$ 5.3 mm $\times$ 25 mm	4.1 mm $\times$ 4.1 mm $\times$ 30 mm
Number of detector rings	36	72
Axial field of view	20 cm	32 cm
Transaxial field of view	70 cm	70 cm
NEMA sensitivity	13.5 cps/kBq	46 cps/kBq
Scatter fraction	41 %	35 %
Clinical NECR	53 kcps @ 2.4 kBq/ml	200 kcps @ 2.4 kBq/ml
Coincidence window	4.9 ns	6.9 ns
Spatial resolution	4.7 mm	4.2 mm
(Axial @ 10 cm, VUE Point*)		

* VUE Point (or VUE Point HD) is an iterative reconstruction algorithm used in GE HealthCare PET/CT scanners.

3.2 Phantom structure

3.2.1 Jaszczak phantom

During the course of this project, two different phantoms were used. The first was a Jaszczak phantom, a cylindrical phantom with an inner diameter and height of 21.6 cm and 18.6 cm respectively, giving a volume of 6.82 dm³. In the bottom plate of the cylinder, 8.8 cm long rods could be placed on which spheres were be attached. In this study six different hot spheres and one cold sphere were used, the diameter and volume of the spheres are shown in Table 3.

Table 3: Dimensions and volumes of the used spheres. The cold sphere is not given a volume since it is solid and therefore only an outer diameter is presented. The hot spheres are hollow and therefore both their inner and outer diameter are presented, together with their corresponding volume. All numbers are approximate.

Dimensions and volumes (Approx.)				
	Label	Volumes (μL)	Inner diameter (mm)	Outer diameter (mm)
Cold	S0			15.9
Hot	S1	31	4	6
Hot	S2	63	5	7
Hot	S3	125	6.2	8
Hot	S4	250	7.9	10
Hot	S5	500	9.9	12
Hot	S6	1000	12.4	14

3.2.2 Phantom printing

Since the aim of this project was to investigate the detectability of lesions in more extreme conditions, such as larger patients, the Jaszczak phantom was considered too small in diameter to be a reasonable representation. Also, for comparison between phantom sizes, a larger phantom was needed to be created. The idea was to create an outer shell to the Jaszczak phantom. A question was asked to Skåne University Hospital’s 3D centre to see if it was possible to 3D print a shell as requested. The shell itself needed to be hollow and capable of being filled with a solution without leakage. The staff at 3D centre thought this was possible to achieve, but in four different pieces that could be puzzled together instead of one ring. What needed to be tested was whether either of the two possible materials from which the shell could be printed had the ability to adsorb the radiopharmaceuticals. The ^{18}F -FDG do not tend to be adsorbed on plastic materials[27]. However, a test of the new materials would prove this statement true in this case as well. Two small test tubes were printed. For the test, a solution of ^{18}F -FDG and water was mixed to a concentration of approximately 100 kBq/ml, which is approximately a concentration five times stronger than the maximum concentration that will be used in the actual measurements. The test tubes were placed in the solution, so that the solution covered both the outside and the inside of the bottom half of the tubes. A couple of hours later, the tubes were dipped in a cup of water, to remove the radioactive solution without affecting the eventual adsorbed radiopharmaceuticals, and put on a paper plate together with a syringe containing the solution. A five minute scan provided images showing a distinct syringe and no visible tubes. Therefore, it was concluded that ^{18}F -FDG did not stick to the walls of the tubes, and both possible materials would work in terms of adsorption.

The four pieces, labelled A-D, was printed with 4 mm thick walls. The total thickness of each piece was 100 mm, with a height of 184 mm. Each piece had one opening with a screw-cap and o-ring, making the printed pieces leakage safe. The pieces were weighted while still dry, then filled with water and weighted again. In this way, the volume of each piece was obtained, which are shown in Table 4. The circumference of the phantom was measured to 132 cm. From now on the printed phantom will be named as the Jaszczak XL phantom or just the XL phantom. An imagined illustration and the printed XL phantom with Jaszczak are shown in Figure 3.

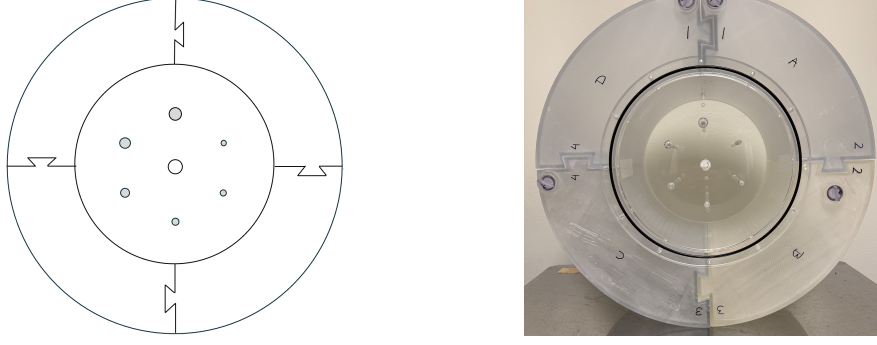


Figure 3: An illustration and an image of the Jaszczak phantom and the 3D printed XL phantom ring, with attached spheres in different sizes.

Table 4: The by weight measured volumes of each 3D printed part of the XL phantom.

Printed phantom volumes	
Volume A	3882 mL
Volume B	3879 mL
Volume C	3854 mL
Volume D	3884 mL

3.3 Measurement preparation

The measurements of the Jaszczak phantom were planned to have a SBR of 10:1 and were divided in three parts: measurements with only the Jaszczak phantom and the Jaszczak phantom with the 3D-printed XL phantom, with and without activity in the XL phantom. The SBR was chosen based on measurements from article by Oddstig et al. [28] and based on experiments following the specifications of the EARL PET/CT program [29]. The activity concentration was chosen to be 12 kBq/ml in the spheres at the start of the measurement.

To fill the spheres a solution of 200 ml of the aimed concentration was made, and a solution of 100 ml was made for the background. Hence, two syringes from the FDG vial were drawn, one for the sphere solution and one for the background solution. The activity in the different syringes was drawn according to the start of the the first scan, which worked as a reference time, t_0 , for all measurements. The two syringes were emptied in the two cups of water, with 200 and 100 ml respectively, and carefully blended. The syringes were measured in an activity meter afterwards to assure minimal activity was left in the syringes. The spheres were directly filled from the sphere solution, while the solution volume to the background parts were calculated according to the total volume of each part. To minimize error, it was decided to add a rather large volume, 10 ml, of the background solution to part A, and add a solution volume to the other parts relative to part A. The volume of the solution added to the Jaszczak phantom was therefore approximately 17.6 ml. Four syringes of approximately 10 ml each were prepared for part A-D, for later use.

3.4 Scanner measurements

The phantoms were scanned sequentially on the two scanners in the order of: Jaszczak, Jaszczak XL without activity and Jaszczak XL with activity. The phantom were carefully centred in

the three dimensions, according to the systems' lasers. This was to make the images from all phantoms and scanners take place under as similar circumstances as possible and therefore create the best basis for comparison. The phantoms were placed so the spheres were located in the middle slice of the AFOV, for both system, which is where the axial sensitivity profile is most sensitive. The data were collected in list mode and the acquisition time was set according to the acquisition time used in the clinic, 90 seconds, and was then repeated ten times. Instead of doing many smaller measurements, one long measurement for each scanner and phantom was taken, which could later be re-binned into shorter frames. Hence, the acquisition time was decided according to ten frames à 90 s, i.e. 15 minutes. To shorten the acquisition time the activity concentration was increased according to the following relation:

$$T_{new} \cdot C_{new} = T_{old} \cdot C_{old} \quad (6)$$

where T_{new} would be the new shorter acquisition time received with the new higher activity concentration C_{new} . The relation is linear, meaning that the acquisition time is halved if the concentration is doubled. This is the acceleration that was chosen. A 7.5 minute scan would be the acquisition time for the first measurement at the reference time, and with an activity concentration of 24 kBq/ml in the spheres. The following measurements would then be delayed relative to this reference time and the phantoms would have a lower concentration. The concentration at the start of the following measurements was calculated according to:

$$C(t) = C_0 e^{-\lambda t} = C_0 e^{-\ln 2 / T_{1/2} (t - t_0)} \quad (7)$$

where $C(t)$ is the activity concentration at the new time t , C_0 is the activity concentration at t_0 and $T_{1/2}$ is the physical half-life of ^{18}F .

To be able to compare the measurements, the counts received from the phantoms must be the same and since the concentration is lower for each measurement, the acquisition time must be extended, according the relation in Equation [6]. In this case the new time T_{new} would instead be longer than T_{old} since C_{new} will be lower than C_{old} . This calculation holds for all measurements.

3.5 Image reconstruction

Images from the measurements on both the DMI scanner and the OMNI scanner were reconstructed using GE HealthCare's own reconstruction software. The different reconstructions for the two scanners are presented in Table [5], together with some previous stated scanner properties and with short notations that later will be used. All reconstructions were made with two different matrix sizes: 256×256 and 384×384 . The OSEM reconstructions used a post filter of 3 mm and the BSREM reconstructions had a β value of 500.

All nine reconstruction methods were made with attenuation correction, using CT for an attenuation map. The different measurements were reconstructed as a dynamic scan, creating shorter time frames of the full acquisition. The scans were modified in such a way that each time frame represent 90 seconds of the full acquisition time, with a total of ten back-to-back frames per scan. Due to the short half-life of ^{18}F , the decay of radionuclides during one scan cannot be considered negligible. Therefore, a time correction for each frame was performed, using Equation [6] where C_{old} and T_{old} is the activity concentration and scan time, respectively, for the first time frame in one measurement. C_{new} in this case is the activity concentration in the beginning of each new frame. These concentrations were calculated according to:

$$C(t_x) = C_{x-1} e^{-\lambda T_{x-1}} = C_{x-1} e^{-\ln 2 / T_{1/2} T_{x-1}} \quad (8)$$

Table 5: The total of nine reconstruction methods and their short notations, shown for each scanner together with a few scanner properties.

PET/CT system	Scintillation material	NEMA sensitivity	NECR	Reconstructions (2 it $\times$ 33 or 34 subs*)		Short notations
DMI	LYSO	13.5 cps/kBq	53 keps	OSEM	w/ TOF + post filter	mi osem-tof
					w/o TOF + post filter	mi osem
				BSREM	w/ TOF	mi bsrem-tof
					w/o TOF	mi bsrem
OMNI	BGO	46 cps/kBq	200 keps	OSEM	w/o TOF + post filter	om osem
				BSREM	w/o TOF	om bsrem
					w/o TOF + PDLL	om bsrem-l
					w/o TOF + PDLM	om bsrem-m
					w/o TOF + PDLH	om bsrem-h

* The number of subsets are 33 for OMNI and 34 for DMI.

where $C(t_x)$ is the concentration at the start time of frame x , C_{x-1} is the concentration at the start of the previous frame and T_{x-1} is the scan time of the previous frame. The time per frame was by the reconstruction software limited to seconds, making the times T_x rounded to full seconds. For the shorter time frames, the increase in scan time became so small that all ten frames in one measurement are very close in length of time. This difference becomes smaller the shorter the frames are modified, which also depends on how long the full measurement was.

3.6 Data analysis

3.6.1 Extraction of data values from VOI and ROI

The data to be extracted from the measurements were the peak values of the spheres and the background, as well as mean value and standard deviation in the background. To extract the peak values together with the mean and standard deviation, volumes and regions of interest (VOIs and ROIs) were placed in the image series of the full measurements of the three phantom geometries, and for each matrix size. The VOIs and ROIs were drawn in Hermia Hybrid Viewer, version 7.0.2 (HERMES Medical Solutions AB, Stockholm, Sweden). The full measurements were chosen because the spheres were most visible in these. A VOI, with a diameter of 1.5 cm, was placed over each hot sphere in the middle slice of the image series, in the order from the biggest to the smallest sphere. The VOIs were placed in combination with the CT image, to control that the VOIs were actually placed over the spheres. A large ROI, with a diameter of 8 cm, was placed in the middle, over the cold sphere. This ROI was used to extract a peak value in the background. The mean value and standard deviation in the background were instead extracted from a small ROI, with a diameter of 3 cm, placed between the two smallest hot spheres and the cold sphere, in a relatively homogenous region. The placement of the VOIs and ROIs were saved as xml files, which later could be implemented to all the other dicom files, i.e. the ten repeated measurements, or the ten time frames, from the two camera systems. From these xml files, the peak values were extracted by a home written Python script, and saved in Excel files. Hence, one excel file contained peak values from six VOIs and two ROIs for ten time frames.

The peak values were extracted by calculating the mean value in a 1 cm^3 cube around each voxel in the VOI, and the cube with the highest mean became the peak value. Peak values in the background were extracted in the same way, but instead of each voxel in a VOI only the

pixels in a ROI were used as a representation of the background. The voxel dimensions were not the same for the two matrices, neither were the slice thickness' for the two scanners, and the cube volumes were therefore needed to be calculated by interpolation between cubes with a whole number of voxels.

3.6.2 Recovery coefficients

A well established method to see how well the activity concentration in the reconstructed data reflect the activity concentration at the start of image acquisition determined from the phantom preparation is to calculate the recovery coefficient (RC) and can be defined as:

$$RC = \frac{\text{measured activity concentration}}{\text{activity concentration determined using activity meter}}. \quad (9)$$

Since the different phantom measurements were performed at different times, the actual concentrations in the phantoms were different. This means that the denominator is different for the different measurements. The activity concentration using activity meter was calculated based on the amount of activity that was drawn from the vial and the measured activity that was left in the syringe after being added to the sphere solution. The activity meter used for determination of the activities was calibrated with traceability to primary standard for ^{18}F . The decay time correction was calculated at the start of each measurement.

3.6.3 SNR in phantom background

In complement to the recovery, the signal-to-noise ratio (SNR) in the background was determined and calculated as:

$$SNR = \frac{\text{mean value in background ROI}}{\text{standard deviation in background ROI}}. \quad (10)$$

This ratio gives a measure for how noisy the background is, where a high SNR implies a less noisy background, and hence a lower probability of finding a high peak value in the background due to noise. The mean value and standard deviation were extracted from the small ROI, with a diameter of 3 cm, placed between the two smallest hot spheres and the cold sphere.

3.6.4 Peak-to-peak in spheres and background

The goal of this project was to investigate the detectability of lesions between the two PET/CT systems. It was thought that a possible value of detectability could be peak-to-peak (P2P), which is defined as:

$$P2P = \frac{\text{peak in sphere}}{\text{peak in background}}. \quad (11)$$

Theoretically a P2P value of ten would be assumed since the SBR is 10:1, but due to partial volume effect this is not likely. A P2P value of one would theoretically mean that the spheres cannot be detected, since the image measured activity concentration in the sphere is estimated to be the same as the concentration in the background. Due to noise in the images the peak values in the background might not be a good representation of what visually can be seen, which must be considered when evaluating the results. However, the P2P metric is believed to better reflect the distinguishability of small hot spheres in a noisy background, as for standard contrast-to-noise ratio.

3.6.5 Visual evaluation of reconstructed images

The images from five of the nine reconstruction methods were visual evaluated in Hermes Medical Software, by three readers. The evaluated methods were "mi osem-tof", "mi bsrem-tof", "om osem", "om bsrem" and "om bsrem-m". This means that only one version of the AI filtered

images (the M filtered) were visually evaluated. For simplicity, the images were ordered by scanner, phantom and reconstruction method, with the ten repeated frames after one other. The readers were due to this not blinded. The images were ordered according to time frame, making the OSEM-TOF reconstruction and BSREM-TOF alternating through all images. The same held for the OMNI scanner, where the first three images showed the same time frame for OSEM, BSREM and BSREM-PDLM respectively. The images from the DMI scanner, as the name shows, include the TOF application. The third reconstruction method for the OMNI scanner (BSREM-PDLM) includes the medium level of AI-filter. During the evaluation, each reader marked each image with a number from 1 to 6, depending on how many spheres that could be distinguished from the background. In advance, two criteria were decided: 1) if a noise cluster in the same size as the sphere is seen, then the sphere is classified as not detected and 2) only the number of spheres are counted, even if a smaller sphere but not a bigger is seen.

3.7 Statistical analysis

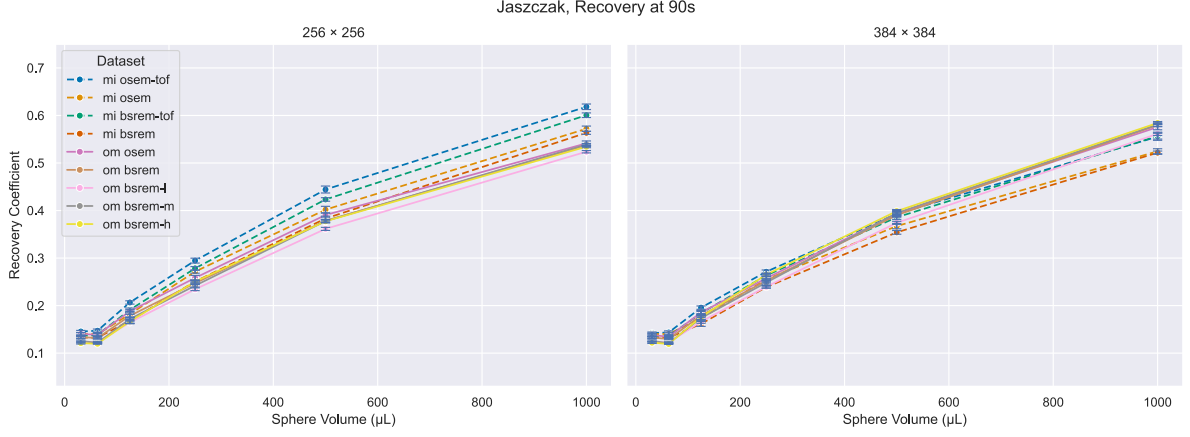
Statistical analysis was performed in SPSS with Linear Mixed Model (LMM) analysis, with P2P as the dependant variable. Fixed effects were the phantoms, reconstruction methods and sphere numbers, while time frames were set as a random effect. Only the data from the three largest spheres were used in the tests. This was chosen because the three smallest spheres all show very low P2P values. The statistical analysis were separated in two parts. In the first part, the 256×256 matrix was chosen for both scanners, which also was the case in the visual evaluation. In the second part, the clinical used matrix sizes were chosen, which is 256 on DMI and 384 on OMNI. Different tests were run for each phantom, to individually see the differences between the reconstruction models, but also a test with all three phantoms, to see statistical differences in a small population. By not separating the phantoms in the test makes it more similar to a patient study, with patients of different sizes. All tests were corrected using Bonferroni correction, to compensate for multiple comparison tests. From the tests, differences between means and signification values between the phantoms and the reconstruction methods were extracted.

4 Results

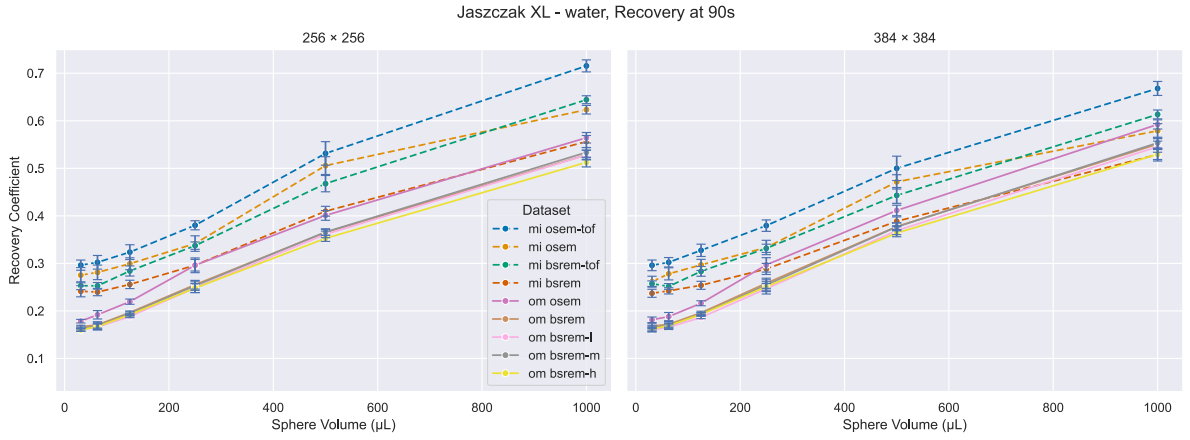
4.1 Recovery coefficients

The calculated recovery coefficients as a function of sphere volume are for the different phantoms, matrices and reconstruction methods shown in Figure 4a, 4b and 4c, with an acquisition of 90 seconds per frame. Each figure show nine different reconstructions, four from the DMI scanner and five from the OMNI scanner (named as "mi" and "om"). The name follows with either "osem" or "bsrem" (OSEM or BSREM reconstruction) and with "tof" if TOF is activated. Reconstructions seven to nine come with the addition of either "-l", "-m" or "-h", which implies the degree of the applied AI filter (low, medium or high). The data points on the curves show the mean value and standard deviation of the ten time frames. For all phantoms, the RC increased with increasing sphere size. One can also note that Jaszczak XL with water in general reaches the highest RC values, and that the Jaszczak XL with activity reaches the lowest values.

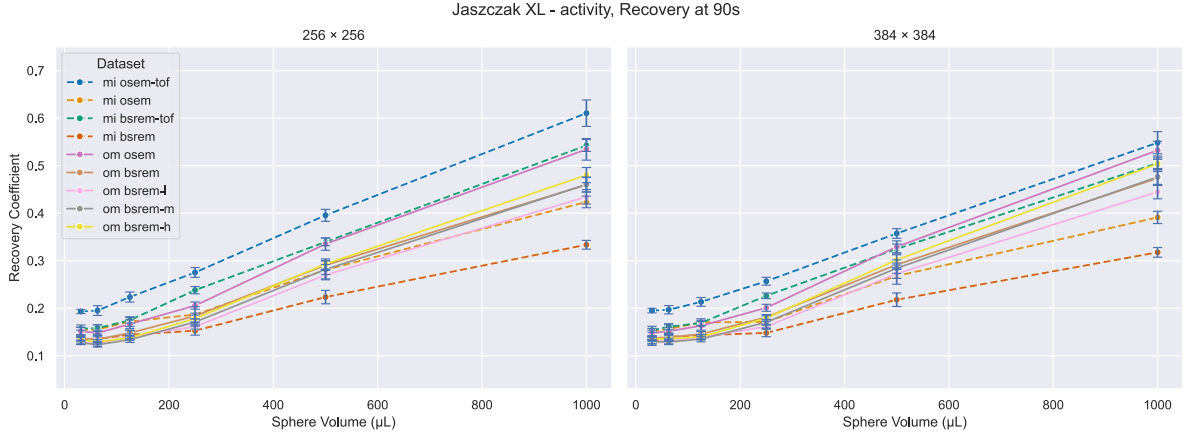
Figure 4a shows the RC curves for the Jaszczak phantom, and here it can be seen that the smallest spheres demonstrate similar RC regardless of reconstruction method, while toward the larger spheres the spread in RC between the methods increases. The RC curves for the Jaszczak XL - water phantom and Jaszczak XL -activity phantom are shown in Figure 4b and 4c respectively. In these cases the RC values are more dependent on the reconstruction method, even for the smaller spheres, where reconstruction methods are more separated from each other. For RC with the 256 matrix, it can be observed that the "mi osem-tof" gives the highest RC for all spheres and phantoms. For the 384 matrix, "mi osem-tof" only show highest RC for the XL



(a) Recovery as a function of sphere volume for Jaszczak phantom.



(b) Recovery as a function of sphere volume for Jaszczak XL - water phantom.



(c) Recovery as a function of sphere volume for Jaszczak XL - activity phantom.

Figure 4: The three sub figures (a), (b) and (c) show the RC for Jaszczak, Jaszczak XL - water and Jaszczak - activity respectively. The data points on the curves show the mean value and standard deviation of the ten time frames. The name of the reconstruction methods tells which scanner the data is from (mi for DMI and om for OMNI) and if the method is OSEM or BSREM. The name comes with an additional term if the data has added TOF or AI model (-tof for true TOF and -l, -m or -h for PDLL, PDLM and PDLH respectively). The panel to the left shows the nine reconstruction for each phantom reconstructed with a matrix size of 256×256 . The panel to the right shows the corresponding with a matrix size of 384×384 .

phantoms. In these five cases, most of the "mi" reconstructions receive higher RC than the "om" reconstructions. However, it can be noted that all RC values from the DMI scanner are slightly higher with the 256 matrix than the 384 matrix. The opposite trend is seen for the OMNI data, where an increasing effect is seen especially in the Jaszczak case, but slightly in the XL cases as well. The way the matrix size affects the two systems, makes, for the Jaszczak phantom, the order of the curves to change when going from the 256 to the 384 matrix. For both scanners, the OSEM reconstruction is generally higher than the corresponding BSREM reconstruction (with or without TOF), but the difference between OSEM and BSREM is seen to be larger for the XL phantoms. The OMNI data with BSREM reconstructions were for all phantoms close together, but, like the DMI data, the OSEM was most often higher than the BSREM reconstructions. Also, the PDL filters do not seem to have that increasing effect on the RC as true TOF does.

4.2 Signal-to-noise ratio in the background

The SNR in the background ROI is in Figure 5 plotted for each phantom and reconstruction. The SNR from the 256 matrix and the 384 matrix is shown in the left and right panel respectively. It can be noted that higher SNR is overall reached for the smaller phantom, while between the two XL cases no clear difference is seen. The outliers seen for especially the XL water phantom, makes the boxes hard to compare to the XL activity phantom. For both the OSEM and BSREM reconstructions it can be seen that the SNR is lower for the reconstructions with TOF, compared to the corresponding reconstruction without TOF. One can also note that the OSEM without TOF on the DMI system reaches about the same value as the OSEM on the OMNI system. The same holds for the BSREM reconstructions without TOF, which show similar SNR values on both systems, for all three phantoms. When applying the PDL filters a clear increase in SNR can be seen, especially for the smaller phantom and the XL phantom with activity. Within the PDL filtered reconstructions a trend can be seen, where the low degree of AI filter tend to give the highest SNR and the high degree give the lowest SNR. This reflects the properties of the different filters, where the low filter has the most smoothing effect. The SNR between the two matrix sizes does not show a certain difference.

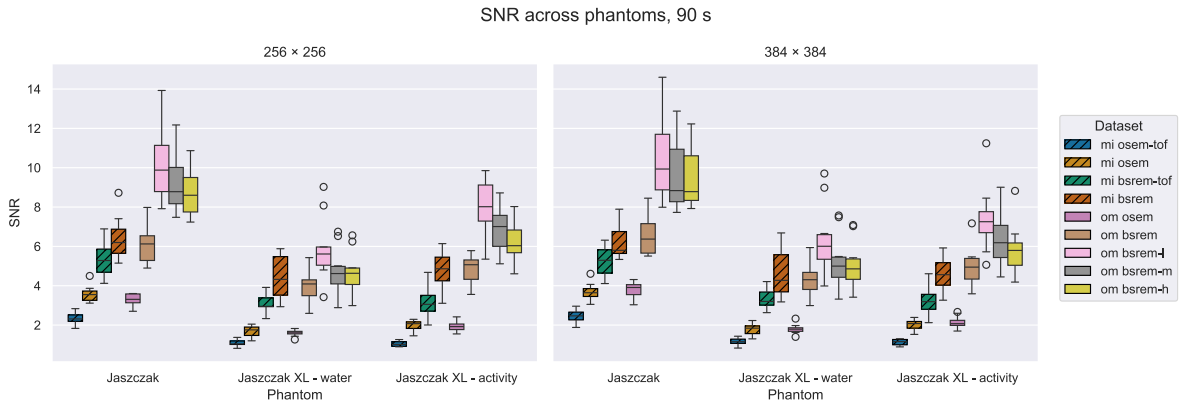


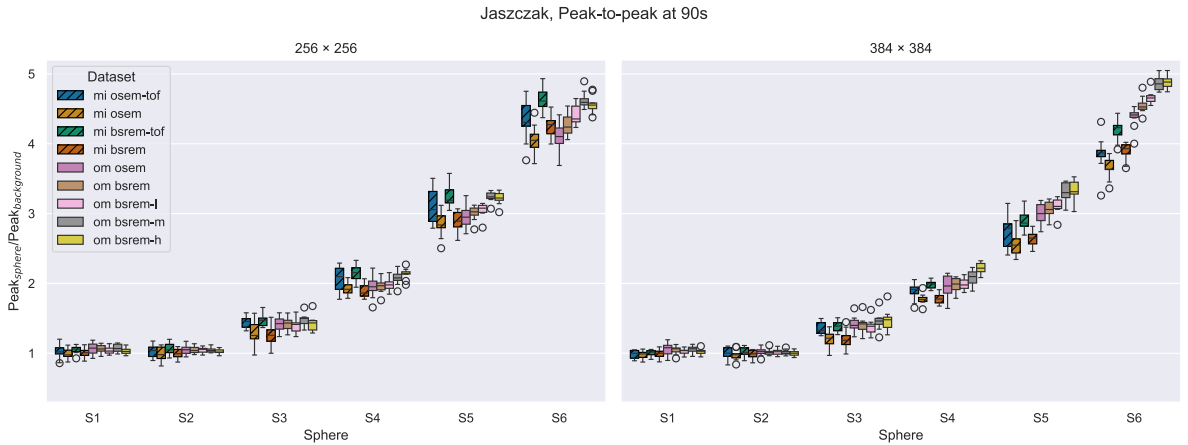
Figure 5: Signal-to-noise ratio defined as mean value over standard deviation, which were extracted from a 3 cm ROI placed in the background of Jaszczak, in the middle slice of each image series. The boxes show the distribution of SNR values from the ten time frames for each reconstruction method. The name of the reconstruction methods tells which scanner the data is from (mi for DMI and om for OMNI) and if the method is OSEM or BSREM. The name comes with an additional term if the data has added TOF or AI model (-tof for true TOF and -l, -m or -h for PDL, PDLH and PDLH respectively). The panel on the left hand side shows the nine reconstruction for each phantom, reconstructed with a matrix size of 256×256. The panel on the right hand side shows the corresponding with a matrix size of 384×384.

4.3 Peak-to-peak as detectability value

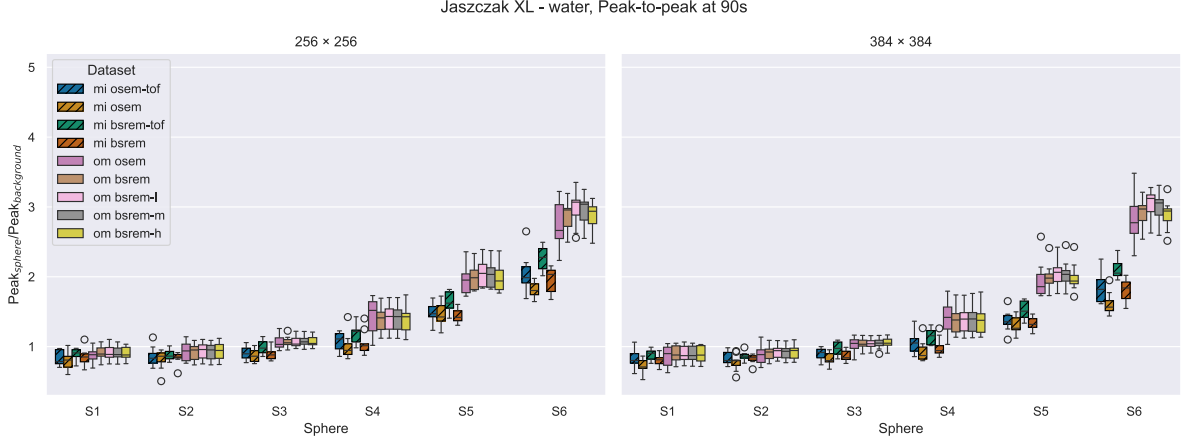
The calculated detectability values, P2P ($\text{Peak}_{\text{sphere}}/\text{Peak}_{\text{background}}$), for OSEM and BSREM reconstructions and all spheres are shown in Figure 6a, 6b and 6c, for each phantom respectively. Each figure show the nine different reconstructions and with acquisition time of 90 s per frame. The x-axis shows the different spheres, where S1 is the smallest and S6 is the largest. The first thing that can be noticed is that the detectability decreases with smaller sphere size. For the smaller phantom (Jaszczak) it can be seen that the P2P value is around one for the two smallest spheres, S1 and S2. For the larger phantom (Jaszczak XL - water/activity) the P2P value starts to end up around one for S3, even S4 for some reconstructions. A P2P value of one means that the peak value in the sphere is the same as in the background. This would in turn mean that the sphere cannot clearly be distinguished from the background. The XL phantom reaches the value one for larger spheres than the small phantom and it can also be seen that lower P2P values are in general reached for the XL phantom with activity, and even lower P2P values without activity.

If comparing the two matrix sizes one can observe that the P2P values for the DMI show higher values for the 256 matrix. The trend could be seen in the XL phantoms, but it seems to be more affected in the case of a smaller phantom. In the XL cases, the OMNI data is rather similar for both matrices, but do also, like the DMI, show a different behavior for the small phantom. However, instead of a decrease with the 384 matrix, an increase is seen compared to the OMNI data with the 256 matrix. The trend within the reconstructions for the Jaszczak phantom therefore changes when changing matrix size. This trend was also seen in the recovery curves.

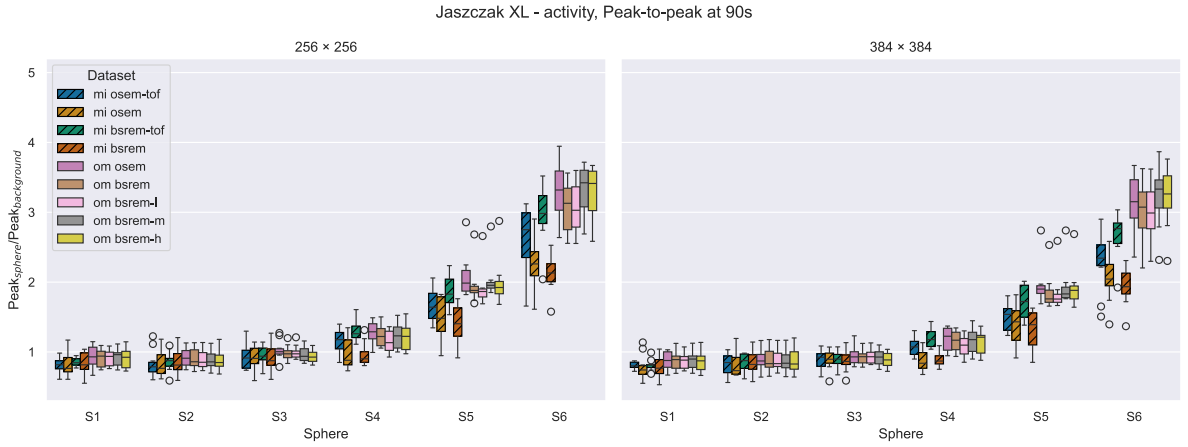
If comparing all reconstructions for the Jaszczak phantom, both matrices, the "om bsrem-m" and "om bsrem-h" reaches the highest median P2P, even though "mi bsrem-tof" show a high maximum value with the 256 matrix. For the Jaszczak XL - water phantom, the highest median is reached with "om bsrem-l", but the results show just small differences between the OMNI reconstructions, with rather large spread in data points. In the case of the Jaszczak XL - activity, the "om bsrem-m" gives the highest median, but again the spread in data is worth noting. Even though the distinction between DMI data and OMNI data not is as large as for the water case, one could still see that the OMNI reconstructions in general produce higher P2P than the DMI.



(a) P2P for each sphere in the Jaszczak phantom.



(b) P2P for each sphere in the Jaszczak XL - water phantom.



(c) P2P for each sphere in the Jaszczak XL - activity phantom.

Figure 6: The three sub figures (a), (b) and (c) show the P2P for Jaszczak, Jaszczak XL - water and Jaszczak - activity respectively. The boxes show the distribution of P2P values from the ten time frames for each reconstruction method. The name of the reconstruction methods tells which scanner the data is from (mi for DMI and om for OMNI) and if the method is OSEM or BSREM. The name comes with an additional term if the data has added TOF or AI model (-tof for true TOF and -l, -m or -h for PDLL, PDLM and PDLH respectively). The panel to the left shows the nine reconstruction for each phantom, reconstructed with a matrix size of 256×256 . The panel to the right shows the corresponding with a matrix size of 384×384 .

4.4 Visual evaluation of reconstructed images

To put the P2P values into relation with something visual, a visual reader study was designed. The ten time frames for each phantom from five different reconstruction methods ("mi osem-tof", "mi bsrem-tof", "om osem", "om bsrem" and "om bsrem-m") were visually analysed and given a number from 1 to 6, representing the number of visible spheres. The mean value for each phantom and reconstruction method from three non-blinded readers are shown in Table 6. It can be seen that for each phantom the highest score was given to the BSREM from OMNI. Also, it can be noted that for the small phantom the OSEM reconstruction gave a similar score for both scanners, while the scores for the larger phantom clearly varied between OSEM-TOF on DMI and OSEM on OMNI. The mean values of the XL phantoms can roughly be translated to that one sphere can be detected with the OSEM-TOF on the DMI and two spheres can be detected with the OSEM on the OMNI.

Table 6: Mean values of detected spheres from each of the five evaluated reconstruction methods for each of the three phantoms. The mean values are determined from the ten repetitive time frames and three non-blinded readers.

	DMI		OMNI		
	OSEM-TOF	BSREM-TOF	OSEM	BSREM	BSREM-PDLM
Jaszczak	4.4	4.7	4.8	5.1	4.7
Jaszczak XL - water	0.7	2.7	2.4	3.3	3.3
Jaszczak XL - activity	1.1	3.0	2.2	3.0	2.9

A subset of the visualised images is shown in Figure 7. When comparing the images from the different measurements it was noted that the different time frames could differ quite a bit, which is also seen in the variation of the P2P values. For the visualisation in Figure 7 the upper median peak value of the 10 time frames of the largest sphere was chosen as a representative for the ten repetitions. In some cases the upper median was not the same time frame for all reconstructions. Therefore, the time frame was determined based on the BSREM reconstruction with the matrix size 256×256 , for both scanners. Hence, when comparing images within the same phantom it is the same under laying data. Due to the variation in time frames, these images do not necessarily correspond to the result in Table 6.

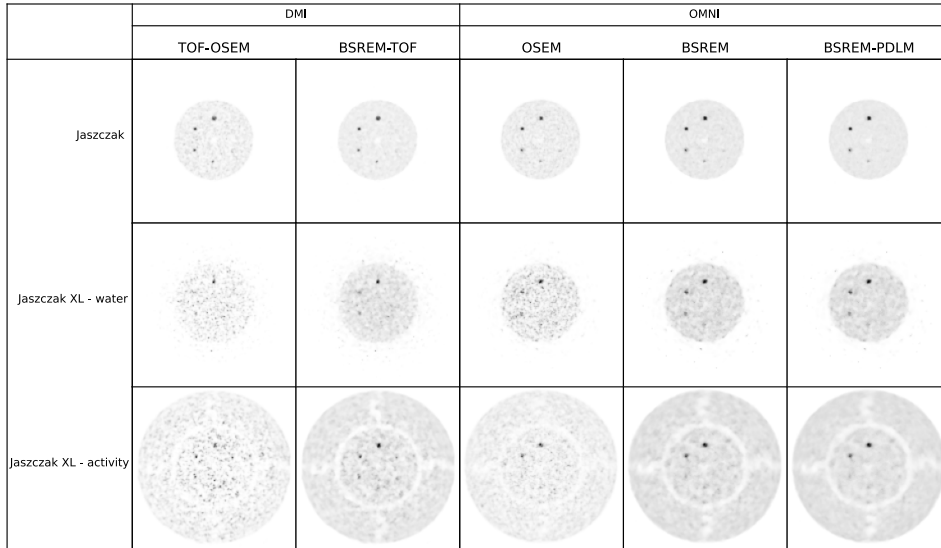


Figure 7: A sample of the images that were visually evaluated, for the three phantoms and the five reconstruction methods. The images in this figure show the middle slice of each image series. Note that the time frame for each measurement (each phantom on each scanner) is chosen based on the upper median of the P2P values in the largest sphere, and is just a representation of the results shown in Table 6.

4.5 Statistical analysis

Statistical evaluation was performed on the P2P values with LMM analysis. The analysis was divided in two main cases; the first using the 256 matrix on both scanners (256/256) and the second using the 256 matrix on DMI and the 384 matrix on OMNI (256/384), which are the clinical used matrix sizes. One separate test were then run for all phantoms together and all phantoms individually. A total of eight test were hence made. The values extracted from the tests were the difference in mean between the compared variables and the significant value of

the difference in mean. From the tests containing all three phantoms, the difference in mean between the phantoms were obtained and are shown in Table 7 and 8. Each row show how the P2P values of that phantom is in relation to the others, either positive or negative. Also, it is shown if the difference (above) between the phantoms are significant (below).

Table 9-16 show the pairwise comparison between all the reconstruction methods with the 256 matrix on both systems and with 256 on DMI and 384 on OMNI. The tables are pairwise shown in the following order: all phantoms, Jaszczak, Jaszczak XL - water and Jaszczak XL - activity. All tables show the pairwise comparison between the nine reconstruction methods, where each cell show the difference in mean (above) with its corresponding significance value (below). Each row represent one reconstruction method and the values in that row show how that reconstruction method is in relation to each of the reconstruction methods in the columns. If the P2P is higher for the variable in the row than in the compared column, the difference in mean is positive and is coloured green. If the P2P is lower for the row variable than for the column variable, the difference in mean is negative and is coloured red. These values are colour scaled to how positive or negative the values are. If the difference in mean is significant, <0.05 , the value is coloured blue, and if it is not significant the value is not coloured.

Table 7: Pairwise comparison of differences in means between P2P values from the three phantoms. Each cell shows if the difference in mean between the compared variables is positive or negative and whether it is significant. The matrix size for the DMI data is 256 and for the OMNI data it is 256.

Differences between means and sig. (256/256)			
	Jaszczak	Jaszczak XL - water	Jaszczak XL - activity
Jaszczak	-	1.301 <.001	1.188 <.001
Jaszczak XL - water	-1.301 <.001	-	-.113 <.001
Jaszczak XL - activity	-1.188 <.001	.113 <.001	-

Table 8: Pairwise comparison of differences in means between P2P values from the three phantoms. Each cell shows if the difference in mean between the compared variables positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 384.

Differences between means and sig. (256/384)			
	Jaszczak	Jaszczak XL - water	Jaszczak XL - activity
Jaszczak	-	1.362 <.001	1.300 <.001
Jaszczak XL - water	-1.362 <.001	-	-.062 .096
Jaszczak XL - activity	-1.300 <.001	.062 .096	-

Table 9: Pairwise comparison of differences in means between P2P values for the nine reconstruction methods. The data is from all three phantoms. Each cell show if the difference in mean is positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 256.

Differences between means and sig. for all phantoms (256/256)									
	mi osem-tof	mi osem	mi bsrem-tof	mi bsrem	om osem	om bsrem	om bsrem-l	om bsrem-m	om bsrem-h
mi osem-tof	-	.182 .008	-.194 .003	.182 .022	-.258 <.001	-.249 <.001	-.283 <.001	-.379 <.001	-.352 <.001
mi osem	-.182 .008	-	-.376 <.001	-.014 1.000	-.440 <.001	-.431 <.001	-.465 <.001	-.561 <.001	-.533 <.001
mi bsrem-tof	.194 .003	.376 <.001	-	.362 <.001	-.064 1.000	-.055 1.000	-.089 1.000	-.185 .006	-.158 .046
mi bsrem	-.168 .022	.014 1.000	-.362 <.001	-	-.426 <.001	-.417 <.001	-.451 <.001	-.547 <.001	-.520 <.001
om osem	.258 <.001	.440 <.001	.064 1.000	.426 <.001	-	.009 1.000	-.025 1.000	-.121 .477	-.094 1.000
om bsrem	0.249 1.000	.431 <.001	0.055 1.000	0.417 <.001	-0.009 1.000	-	-0.034 1.000	-0.130 0.282	-0.103 1.000
om bsrem-l	.283 <.001	.465 <.001	.089 1.000	.451 <.001	.025 1.000	.034 1.000	-	-.096 1.000	-.068 1.000
om bsrem-m	.379 <.001	.561 <.001	.185 .006	.547 <.001	.121 .477	.130 .282	.096 1.000	-	.028 1.000
om bsrem-h	.352 <.001	.533 <.001	.158 .046	.520 <.001	.094 1.000	.103 1.000	.068 1.000	-.028 1.000	-

Table 10: Pairwise comparison of differences in means between P2P values for the nine reconstruction methods. The data is from all three phantoms. Each cell show if the mean difference is positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 384.

Differences between means and sig. for all phantoms (256/384)									
	mi osem-tof	mi osem	mi bsrem-tof	mi bsrem	om osem	om bsrem	om bsrem-l	om bsrem-m	om bsrem-h
mi osem-tof	-	.182 .011	-.194 .004	.168 .030	-.254 <.001	-.271 <.001	-.301 <.001	-.393 <.001	-.389 <.001
mi osem	-.182 .011	-	-.376 <.001	-.014 1.000	-.436 <.001	-.453 <.001	-.483 <.001	-.575 <.001	-.571 <.001
mi bsrem-tof	.194 .004	.376 <.001	-	.362 <.001	-.060 1.000	-.077 1.000	-.107 1.000	-.199 .003	-.195 .004
mi bsrem	-.168 .030	.014 1.000	-.362 <.001	-	-.422 <.001	-.439 <.001	-.469 <.001	-.561 <.001	-.557 <.001
om osem	.254 <.001	.436 <.001	.060 1.000	.422 <.001	-	-.017 1.000	-.047 1.000	-.139 .205	-.135 .261
om bsrem	.271 <.001	.453 <.001	.077 1.000	.439 <.001	.017 1.000	-	-.030 1.000	-.122 .550	-.118 .683
om bsrem-l	.301 <.001	.483 <.001	.107 1.000	.469 <.001	.047 1.000	.030 1.000	-	-.092 1.000	-.088 1.000
om bsrem-m	.393 <.001	.575 <.001	.199 .003	.561 <.001	.139 .205	.122 .550	.092 1.000	-	.004 1.000
om bsrem-h	.389 <.001	.571 <.001	.195 .004	.557 <.001	.135 .261	.118 .683	.088 1.000	-.004 1.000	-

Table 11: Pairwise comparison of differences in means between P2P values for the nine reconstruction methods. The data is from the Jaszczak phantom only. Each cell show if the difference in mean is positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 256.

Differences between means and sig. for Jaszczak (256/256)									
	mi osem-tof	mi osem	mi bsrem-tof	mi bsrem	om osem	om bsrem	om bsrem-l	om bsrem-m	om bsrem-h
mi osem-tof	-	.222 <.001	-.175 <.001	.157 .002	.163 .001	.086 .960	.016 1.000	-.148 .006	-.142 .011
mi osem	-.222 <.001	-	-.397 <.001	-.066 1.000	-.059 1.000	-.136 .019	-.206 <.001	-.370 <.001	-.364 <.001
mi bsrem-tof	.175 <.001	.397 <.001	-	.331 <.001	.338 <.001	.261 <.001	.191 <.001	.027 1.000	.033 1.000
mi bsrem	-.157 .002	.066 1.000	-.331 <.001	-	.006 1.000	-.070 1.000	-.140 .012	-.305 <.001	-.298 <.001
om osem	-.163 .001	.059 1.000	-.338 <.001	-.006 1.000	-	-.077 1.000	-.147 .007	-.311 <.001	-.305 <.001
om bsrem	-.086 .960	.136 .019	-.261 <.001	.070 1.000	.077 1.000	-	-.070 1.000	-.234 <.001	-.228 <.001
om bsrem-l	-.016 1.000	.206 <.001	-.191 <.001	.140 .012	.147 .007	.070 1.000	-	-.165 .001	-.158 .002
om bsrem-m	.148 .006	.370 <.001	-.027 1.000	.305 <.001	.311 <.001	.234 <.001	.165 .001	-	.007 1.000
om bsrem-h	.142 .011	.364 <.001	-.033 1.000	.298 <.001	.305 <.001	.228 <.001	.158 .002	-.007 1.000	-

Table 12: Pairwise comparison of differences in means between P2P values for the nine reconstruction methods. The data is from the Jaszczak phantom only. Each cell show if the difference in mean is positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 384.

Differences between means and sig. for Jaszczak (256/384)									
	mi osem-tof	mi osem	mi bsrem-tof	mi bsrem	om osem	om bsrem	om bsrem-l	om bsrem-m	om bsrem-h
mi osem-tof	-	.222 <.001	-.175 <.001	.157 .006	.056 1.000	-.033 1.000	-.087 1.000	-.252 <.001	-.313 <.001
mi osem	-.222 <.001	-	-.397 <.001	-.066 1.000	-.167 .002	-.256 <.001	-.310 <.001	-.474 <.001	-.535 <.001
mi bsrem-tof	.175 <.001	.397 <.001	-	.331 <.001	.230 <.001	.141 .023	.087 1.000	-.077 1.000	-.138 .031
mi bsrem	-.157 .006	.066 1.000	-.331 <.001	-	-.101 .509	-.190 <.001	-.244 <.001	-.408 <.001	-.469 <.001
om osem	-.056 1.000	.167 .002	-.230 <.001	.101 .509	-	-.089 1.000	-.143 .020	-.307 <.001	-.368 <.001
om bsrem	.033 1.000	.256 <.001	-.141 .023	.190 <.001	.089 1.000	-	-.054 1.000	-.218 <.001	-.279 <.001
om bsrem-l	.087 1.000	.310 <.001	-.087 1.000	.244 <.001	.143 .020	.054 1.000	-	-.164 .003	-.225 <.001
om bsrem-m	.252 <.001	.474 <.001	.077 1.000	.408 <.001	.307 <.001	.218 <.001	.164 .003	-	-.061 1.000
om bsrem-h	.313 <.001	.535 <.001	.138 .031	.469 <.001	.368 <.001	.279 <.001	.225 <.001	.061 1.000	-

Table 13: Pairwise comparison of differences in means between P2P values for the nine reconstruction methods. The data is from the Jaszczak XL - water phantom only. Each cell show if the difference in mean is positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 256.

Differences between means and sig. for Jaszczak XL - water (256/256)									
	mi osem-tof	mi osem	mi bsrem-tof	mi bsrem	om osem	om bsrem	om bsrem-l	om bsrem-m	om bsrem-h
mi osem-tof	-	.104 1.000	-.159 .113	.053 1.000	-.513 <.001	-.548 <.001	-.614 <.001	-.591 <.001	-.546 <.001
mi osem	-.104 1.000	-	-.263 <.001	-.052 1.000	-.617 <.001	-.652 <.001	-.718 <.001	-.695 <.001	-.650 <.001
mi bsrem-tof	.159 .113	.263 <.001	-	.212 .003	-.354 <.001	-.389 <.001	-.455 <.001	-.432 <.001	-.387 <.001
mi bsrem	-.053 1.000	.052 1.000	-.212 .003	-	-.565 <.001	-.601 <.001	-.667 <.001	-.644 <.001	-.599 <.001
om osem	.513 <.001	.617 <.001	.354 <.001	.565 <.001	-	-.035 1.000	-.101 1.000	-.079 1.000	-.033 1.000
om bsrem	.548 <.001	.652 <.001	.389 <.001	.601 <.001	.035 1.000	-	-.066 1.000	-.043 1.000	.002 1.000
om bsrem-l	.614 <.001	.718 <.001	.455 <.001	.667 <.001	.101 1.000	.066 1.000	-	.023 1.000	.068 1.000
om bsrem-m	.591 <.001	.695 <.001	.432 <.001	.644 <.001	.079 1.000	.043 1.000	-.023 1.000	-	.045 1.000
om bsrem-h	.546 <.001	.650 <.001	.387 <.001	.599 <.001	.033 1.000	-.002 1.000	-.068 1.000	-.045 1.000	-

Table 14: Pairwise comparison of differences in means between P2P values for the nine reconstruction methods. The data is from the Jaszczak XL - water phantom only. Each cell show if the difference in mean is positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 384.

Differences between means and sig. for Jaszczak XL - water (256/384)									
	mi osem-tof	mi osem	mi bsrem-tof	mi bsrem	om osem	om bsrem	om bsrem-l	om bsrem-m	om bsrem-h
mi osem-tof	-	.104 1.000	-.159 .181	.053 1.000	-.521 <.001	-.563 <.001	-.627 <.001	-.607 <.001	-.551 <.001
mi osem	-.104 1.000	-	-.263 <.001	-.052 1.000	-.625 <.001	-.667 <.001	-.731 <.001	-.711 <.001	-.655 <.001
mi bsrem-tof	.159 .181	.263 <.001	-	.212 .008	-.362 <.001	-.403 <.001	-.467 <.001	-.448 <.001	-.392 <.001
mi bsrem	-.053 1.000	.052 1.000	-.212 .008	-	-.574 <.001	-.615 <.001	-.679 <.001	-.659 <.001	-.603 <.001
om osem	.521 <.001	.625 <.001	.362 <.001	.574 <.001	-	-.041 1.000	-.106 1.000	-.086 1.000	-.030 1.000
om bsrem	.563 <.001	.667 <.001	.403 <.001	.615 <.001	.041 1.000	-	-.064 1.000	-.044 1.000	.012 1.000
om bsrem-l	.627 <.001	.731 <.001	.467 <.001	.679 <.001	.106 1.000	.064 1.000	-	.020 1.000	.076 1.000
om bsrem-m	.607 <.001	.711 <.001	.448 <.001	.659 <.001	.086 1.000	.044 1.000	-.020 1.000	-	.056 1.000
om bsrem-h	.351 <.001	.655 <.001	.392 <.001	.603 <.001	.030 1.000	-.012 1.000	-.076 1.000	-.056 1.000	-

Table 15: Pairwise comparison of differences in means between P2P values for the nine reconstruction methods. The data is from the Jaszczak XL - activity phantom only. Each cell show if the difference in mean is positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 256.

Differences between means and sig. for Jaszczak XL - activity (256/256)									
	mi osem-tof	mi osem	mi bsrem-tof	mi bsrem	om osem	om bsrem	om bsrem-l	om bsrem-m	om bsrem-h
mi osem-tof	-	.218 .226	-.248 .071	.295 .009	-.425 <.001	-.285 .014	-.252 .059	-.398 <.001	-.367 <.001
mi osem	-.218 .226	-	-.466 <.001	.076 1.000	-.643 <.001	-.504 <.001	-.471 <.001	-.616 <.001	-.585 <.001
mi bsrem-tof	.248 .071	.466 <.001	-	.542 <.001	-.177 .962	-.038 1.000	-.005 1.000	-.150 1.000	-.119 1.000
mi bsrem	-.295 .009	-.076 1.000	-.542 <.001	-	-.719 <.001	-.580 <.001	-.547 <.001	-.693 <.001	-.662 <.001
om osem	.425 <.001	.643 <.001	.177 .962	.719 <.001	-	.139 1.000	.172 1.000	.026 1.000	.057 1.000
om bsrem	.285 .014	.504 <.001	.038 1.000	.580 <.001	-.139 1.000	-	.033 1.000	-.113 1.000	-.082 1.000
om bsrem-l	.252 .059	.471 <.001	.005 1.000	.547 <.001	-.172 1.000	-.033 1.000	-	-.146 1.000	-.115 1.000
om bsrem-m	.398 <.001	.616 <.001	.150 1.000	.693 <.001	-.026 1.000	.113 1.000	.146 1.000	-	.031 1.000
om bsrem-h	.367 <.001	.585 <.001	.119 1.000	.662 <.001	-.057 1.000	.082 1.000	.115 1.000	-.031 1.000	-

Table 16: Pairwise comparison of differences in means between P2P values for the nine reconstruction methods. The data is from the Jaszczak XL - activity phantom only. Each cell show if the difference in mean is positive or negative and whether it is significant. The difference is the horizontal row minus the vertical column. The matrix size for the DMI data is 256 and for the OMNI data it is 384.

Differences between means and sig. for Jaszczak XL - activity (256/384)									
	mi osem-tof	mi osem	mi bsrem-tof	mi bsrem	om osem	om bsrem	om bsrem-l	om bsrem-m	om bsrem-h
mi osem-tof	-	.218 .224	-.248 .070	.295 .009	-.297 .008	-.218 .229	-.189 .645	-.320 .003	-.303 .006
mi osem	-.218 .224	-	-.466 <.001	.076 1.000	-.516 <.001	-.436 <.001	-.407 <.001	-.539 <.001	-.522 <.001
mi bsrem-tof	.248 .070	.466 <.001	-	.542 <.001	-.049 1.000	.030 1.000	.059 1.000	-.072 1.000	-.055 1.000
mi bsrem	-.295 .009	-.076 1.000	-.542 <.001	-	-.592 <.001	-.512 <.001	-.483 <.001	-.615 <.001	-.598 <.001
om osem	.297 .008	.516 <.001	.049 1.000	.592 <.001	-	.079 1.000	.108 1.000	-.023 1.000	-.006 1.000
om bsrem	.218 .229	.436 <.001	-.030 1.000	.512 <.001	-.079 1.000	-	.029 1.000	-.102 1.000	-.086 1.000
om bsrem-l	.189 .645	.407 <.001	-.059 1.000	.483 <.001	-.108 1.000	-.029 1.000	-	-.132 1.000	-.115 1.000
om bsrem-m	.320 .003	.539 <.001	.072 1.000	.615 <.001	.023 1.000	.102 1.000	.132 1.000	-	.017 1.000
om bsrem-h	.303 .006	.522 <.001	.055 1.000	.598 <.001	.006 1.000	.086 1.000	.115 1.000	-.017 1.000	-

5 Discussion

The aim of this thesis was to investigate if the detectability of small lesions differed between two PET/CT systems and to evaluate how the detectability was affected by different phantom sizes, as an approximation of patient sizes. Different phantom sizes were obtained by using a Jaszczak phantom and by a 3D printed XL shell to Jaszczak. The project was design as a three case study; Jaszczak, Jaszczak XL - water (without activity in the XL part) and Jaszczak XL - activity (with activity in the XL part). Each phantom was scanned on the two scanners and the full measurements were rebinned into ten 90 seconds time frames. Images were reconstructed with two matrix sizes (256×256 and 384×384) and using a total of nine different reconstruction methods (four for DMI and five for OMNI). From the reconstructed data, P2P values, RC and SNR were extracted, which together with a visual evaluation of a subset of the reconstructed images, are the basis for whether detectability differs between the two systems and if P2P is a good metric for representing detectability.

5.1 Recovery coefficients

The recovery coefficient is a metric that tells how well the measured activity concentration in the image is reproduced in relation to the activity meter determined. It gives a quantitative value of the concentrations in the spheres, but does not say anything about how well the spheres are visualised from the background. However, the RC from the different phantoms and reconstructions can give an idea if a high P2P value depends on a high peak in the sphere or a low peak in the background, and vice versa for a low P2P value.

The trends seen in the RC curves for each phantom are rather different between the phantoms, where for example, the range of RC values for each sphere across reconstructions was larger for the XL phantom than for the Jaszczak phantom. It can be noted that curves representing the various/different reconstruction methods for the OMNI stay close together considerably independently of phantom, while for the DMI the curves representing the various/different methods stay relatively close for the small phantom but clearly diverge for the large phantom. This can be interpreted as the DMI data is more stable on a small phantom, while on a large phantom the data is relatively unstable and dependant on the reconstruction method. Also, the fact that the curves for the DMI are closer together, could mean that the use TOF is limited on a small phantom, while for the large phantom it is clear that the TOF benefit the RC. This corresponds to Equation [3](#), which implies that the gain of TOF is greater for an object with larger diameter. The OMNI data do not in the same way show this dependency of reconstruction method, for either phantom. The DMI scanner therefore seems to be more affected by phantom size, and consequently by the patients size, and also by the increased noise level that occurs from more scattering material.

Another interesting observation is the difference between the two matrix sizes. The trends look rather the same, but a slight difference is seen in the numerical values. The DMI data show a somewhat decrease for all phantoms, while the OMNI data instead show an increase, at least for the small phantom. The DMI scanner seems more adapted for the 256 matrix, with larger voxels, which could depend on the higher noise sensitivity, which is emphasized in smaller voxels. Despite the more averaging that occurs when the signal is distributed in larger voxels, the peak value still becomes higher than when the signal is distributed in smaller voxels. The peak value when using the 384 matrix includes more voxels with a greater variation in values, where this variation emerges from noise, and for the DMI scanner the noise seems to outweigh the sampling advantage of smaller voxels. The OMNI scanner, which does not seem sensitive to noise in the same way, can instead use the sampling of smaller voxels to reach higher peak values in the spheres.

5.2 Signal-to-noise ratio in the background

From the small ROI that was placed in the background, SNR values were obtained for each phantom and reconstruction. Differences between the small and large phantom were expressed by higher SNR for the Jaszczak phantom. A smaller phantom, with less attenuating and scattering material, creates more true LORs, which can suppress the noise in higher degree than the XL phantom, and hence give higher SNR. When comparing the SNR in the two cases of the XL phantom one could note a trend that the XL phantom with only water obtained lower SNR than the XL phantom with activity. With the many outliers one could argue that this possibly is pure chance, and if ten new measurements were performed the SNR could end up being the same for both XL cases. Very little difference in SNR was seen between the two matrices, which indicates that the matrix size does not affect the mean value and standard deviation as it does with the peak value. It was also noted that the applied AI filters in general increased the SNR in the background. Since the AI filters smooth the data to different degrees, it is reasonable that the dataset with a smoothing filter generates data with less noise. The AI filters therefore seem to operate as smoothing filters, while still maintaining the peak in the spheres, seen from the RC curves.

In Figure 5 it was seen that the DMI reconstruction methods with TOF had a lower SNR in the background than the corresponding reconstructions without TOF. This implies that the added TOF makes the background more noisy, and is an expected result since the signal is pushed toward a more probable position, rather than blended over the full LOR. The higher SNR in non-TOF images arises from increased spatial averaging, as events are effectively distributed along the full line of response rather than localized, leading to greater smoothing. Smoothing of the signal also occurs due to parallax effects, due to uncertainties in the positions of the detections. The distribution of the signal increases for an object more peripherally localised, blending the signal over a larger volume, which leads to reduction in resolution and a possible change in the noise characteristics. This is the case for the XL activity phantom, where activity is more peripherally localised, and hence the background could assumably be smoother than the other phantoms. However, the parallax effects are not assumed to affect central signal, but the possibility that the SNR values from the XL activity phantom are more affected by parallax effects cannot be rejected.

Regarding the data without TOF on the DMI, it seems to have a similar SNR as the OMNI data without AI filter. This holds for both OSEM and BSREM, for all three phantom cases and for both matrices, and does imply that the SNR in these cases does not explain the difference seen in P2P. Due to the difference in noise characteristics when adding TOF, images with the same β value in the clinical choices of reconstruction methods will look rather different in terms of noise. Since the OMNI scanner does not have TOF, its images will be smoother than the images from the DMI. To match the noise in DMI images, the OMNI reconstructions have in the clinic a lower β value, and in future work it would therefore be interesting to add reconstructions with various β values to the evaluation.

5.3 Peak-to-peak as detectability value

For the various phantoms, spheres and reconstruction methods, P2P values have been calculated from the SUV peak in the sphere and the SUV peak in the background. What hopefully can be answered is if P2P is a good metric of detectability, and hence can be used to establish which PET system that better can detect small lesions. Box plots with P2P values were created for each phantom, which were individually and collectively evaluated, both visually and statistically.

The first thing noted was the decrease in P2P when decreasing the sphere size. This clearly

implies that size is a vital factor in how detectable the lesion is, for lesions with the same activity concentration. It was noted that general lower P2P values were seen for the XL phantom compared to the Jaszczak phantom, which also implies that lesions are less detectable in larger patients compared to smaller patients. This is supported by the significant mean difference between Jaszczak and both XL phantoms, for the two matrix cases, seen in Table 7 and 8. This is reasonable, because a larger volume creates more material for the photons to interact with. This generates fewer detections and more random coincidences, leading to a lower signal in the spheres and more noise.

The fact that the smallest spheres, S1 and S2, receive a P2P value around one is because the spheres are too small and together with the chosen activity concentration, the partial volume effect makes them disappear in the background. However, a P2P of one does not have to mean that the spheres cannot be observed in the 3D stack of images, since the peak value of the sphere could still be high enough to be separated from the noisy background. The high peak value in the background implies, though, that there are voxels with signal, due to noise, in the same order as S1 and S2, making it possible to think that spheres are in these positions as well. Knowing the location of the spheres hence creates a bias when studying the images, but this was taken into account in the visual evaluation.

The P2P median for the DMI scanner continuously shows a lower value for the 384 matrix (both OSEM and BSREM reconstructions), even though the large spread in data points reduces the possible difference between the matrices. For the OMNI scanner not much difference is seen in the XL phantoms, while in the Jaszczak phantom it can be seen that the 384 matrix gives higher values. This is the same trend that was seen in the RC curves, and probably depend on the same parameter. If the DMI is more sensitive to noise, the signal can be helped by the larger voxels. Smaller voxels for the DMI might lead to low signal over many spread out voxels, instead of gathering the signal to a specific place.

From the statistical analysis of all phantoms together, shown in Table 9 and 10, the OMNI reconstructions show significantly higher values of P2P than all DMI, except from the "mi bsrem-tof" which not show significantly lower values compared to the OMNI reconstructions. This holds for both 256/256 and 256/384 (with one exception in 256/256). This implies that the OMNI reconstructions for this phantom population produces higher P2P, but that the "mi bsrem-tof" is the candidate from DMI that is in the better range. In the specific case of the data from Jaszczak with 384 matrix, one saw that the OMNI reconstructions gave higher P2P values than DMI. However, for all reconstructions of Jaszczak, both 256 and 384, it was hard to see which reconstruction that gave highest value, and if there was a significant difference. By comparing the mean difference in P2P shown in Table 11 and 12, it can be seen that the "mi bsrem-tof" row is all green in the 256/256 case, while in the 256/384 case the "mi bsrem-tof" becomes red for two of the OMNI reconstructions. One also see that in the 256/256 case the "mi bsrem-tof" is not significant different from the "om bsrem-m" and "om bsrem-h", while it is significant lower than the "om bsrem-h" in the 256/384 case. This says that the P2P for "om bsrem-h" on Jaszczak, increased enough with the 384 matrix to become significant higher than the best DMI reconstruction. For the Jaszczak XL - water phantom, the difference between means of all OMNI reconstructions are significantly larger than all DMI reconstructions, and holds for both 256/256 and 256/384 (Table 13 and 14). For this phantom the "mi bsrem-tof" no longer stand a chance against the OMNI reconstructions. The fact that the statistical tests showed significantly larger P2P for OMNI than DMI could be assumed from the distinct difference in the P2P results in Figure 6b. The values for Jaszczak XL - activity (Table 15 and 16) show that the "mi bsrem-tof" in this case can compare to the OMNI reconstructions, and is not significantly lower than any of them.

5.3.1 Peak-to-peak in relation to RC and SNR

A question that was wished to be answered was if the P2P value can be considered a good metric for detectability. It was previously stated that P2P has drawbacks since the peak in the background highly depends on image noise, rather than actual activity concentration. The benefit of using the peak in the background is that it in fact captures possible noise clusters, that visually could be mistaken for a hot sphere, which is not information that standard contrast-to-noise ratio could give. To answer whether P2P is a representable metric, other parameters were included in the analysis. The RC and SNR can help with the interpretation of the P2P values, and together they can possibly explain why the P2P values show the results they do.

It was previously stated that similar SNR was reached between the DMI and OMNI data in the cases without TOF or AI based TOF. Since the recovery in these cases is higher for the DMI data, the SNR cannot explain why the P2P values for the OMNI data is higher than for the DMI data in the XL phantom. The in general lower P2P values for all DMI reconstructions compared to the OMNI reconstructions seen in Figure 6b and 6c, show that the OMNI data should have a higher sphere peak or a lower background peak. From the RC curves it is known that the peak values from the spheres are higher for the DMI data. It is therefore reasonable to assume that the noise must be higher in the DMI data for the XL phantom, giving higher background peaks and making the P2P values drop below the OMNI values. The assumed higher noise in the DMI data is however not seen in Figure 5, since the "mi osem" and "mi bsrem" show very similar values as "om osem" and "om bsrem" respectively. It therefore must be another parameter controlling the noise that the SNR cannot represent.

Previously, it was stated that the recovery for the Jaszczak XL - water phantom was higher than for the Jaszczak XL - activity phantom. It was also stated that the Jaszczak XL - activity obtained higher P2P values than Jaszczak XL - water. Since the water case give higher peak in the spheres it would be reasonable to think that the P2P also would be higher. As this is not the case, the peak in the background must play a vital role, and for this to add up, the background peak must be higher in the water case than the activity case. Why this is the case is nothing that can be explained from the data produced in this project. Both XL cases involve the same amount of attenuating and scattering material, since it is the same phantom. What differs is the amount and distribution of total collected signal. In the XL water case, the central signal from Jaszczak challenges the attenuation correction harder to reproduce the counts in a correct way. The XL activity phantom also produces a larger number of true LORs. In all reconstruction methods without TOF, the signal will be equalised along the full LORs. Signal from the ring is therefore smoothed over the full object, and will give a spill-in effect on Jaszczak. This spill-in effect does not happen for the XL water phantom and could possibly be a factor to the difference in background peak. These two possible factors; a more challenging attenuation correction and spill-in effect from activity in the XL phantom, are hypothesis' that must be further investigated to say why the peak value in the background is higher in the XL water case.

5.4 Visual evaluation

The visual evaluation for the 256 matrix generated an overall higher score for the OMNI scanner than the DMI scanner. For the Jaszczak phantom there is not a big difference in the mean value of the scores for both OSEM, 4.4 (DMI) vs 4.8 (OMNI), and BSREM, 4.7 (DMI) vs 5.1 and 4.7 (OMNI), reconstructions. The BSREM reconstructions for the XL phantom also do not show a significant difference, 2.7 (DMI) vs 3.3 (OMNI). However, the values still show an advantage for the OMNI scanner. What can be seen for the XL phantom is the clear difference between the OSEM reconstructions. The XL - water phantom on the DMI gives a mean of 0.7 compared to the mean of 2.4 from the OMNI. The XL - activity phantom shows a slightly lower difference, but

still distinct, with values of 1.1 vs 2.2. This implies that the hardware of OMNI is preferred for a larger phantom, and that the better sensitivity of this scanner possibly plays a big role in these results. What can also be noted for the OMNI scanner is the small difference between BSREM and BSREM-PDLM. The applied AI filter does not show an increase in mean, rather a slight decrease. This means that the AI filter does not improve the visual detectability of small lesions in terms of the number of lesions. This implies that the filter cannot enhance signal that is not present in the first place, and is reasonable since the AI model otherwise must have to create new data from no ground signal. The question of whether the detected lesions are more visible when the filter is applied is not answered through the mean values, since the score was given in terms of whether the sphere was seen or not, not in terms of how well the sphere was emphasized in the reconstructions. It is therefore possible that the filter improved the detectability in the already detected spheres, but the case could still be the other way around, that the filter blurred out the low signal from the small spheres. For the visual evaluation, it is also worth noting that the noise characteristics between the two systems have in previous studies been shown to vary for a fixed β value, and a more equal evaluation could have been achieved if the clinical β value for each system had been used, which in SUS Lund is 550 for DMI and 400 for OMNI.

5.4.1 Visual evaluation in relation to peak-to-peak

To put the P2P values in relation to something visual, the mean values in Table 6 must be compared to what can be seen in the left panel of Figure 6a-6c. For the smaller phantom, the OSEM reconstructions ("mi osem-tof" and "om osem") show a somewhat higher median P2P for the DMI. Comparing the BSREM reconstructions ("mi bsrem-tof", "om bsrem" and "om bsrem-m") the DMI median is higher than OMNI median without AI filter, while it is approximately the same as the OMNI median with the filter. As said before, the mean score values for the Jaszczak phantom are very similar between the scanners, but the higher P2P for "mi bsrem-tof" is not shown in the visual score. Also, the higher P2P when applying the MPDL filter does not correspond to what was visually seen, since the filter images received a mean score of 4.7, while the images without images got 5.1. The fact that the P2P increases for AI filtered images could be explained by a more blurred background, hence decreasing the background peak value.

In cases of the XL phantoms, the spread in data points is much larger than for the Jaszczak phantom, making it hard to see a difference in P2P between the "om bsrem" and the reconstructions with applied AI filters. This makes the visual evaluation, in some cases, seem to correspond rather well to the P2P results. No clear difference is seen the BSREM methods on OMNI and a somewhat lower P2P is seen for the OMNI OSEM, which is seen in the visual evaluation for XL - water but not for XL - activity. The DMI OSEM-TOF is seen to be slightly lower than the DMI BSREM-TOF, in both P2P value and mean value of the visual estimate, for both phantoms. If comparing the OSEM between the systems, a clear winner can be found in both P2P and visually, which is OMNI. If comparing the BSREM reconstructions, the P2P again says that OMNI is the better choice, especially for the XL - water phantom. The visual evaluation does not show that big of a difference, but a trend could still be seen that the OMNI gave a higher sphere score for the BSREM reconstructions.

5.5 Variation in repeated measurements

Something that was observed in the P2P values, was the variation in the values between the different time frames. In the box plots, this is seen from the large difference in the minimum and maximum values and the eventual outliers. The distribution in data values varied between reconstruction methods and the trends were necessarily not correlated to the different phantoms. For the Jaszczak phantom, the DMI reconstructions showed a larger distribution in P2P values,

while for the XL water phantom, the OMNI reconstructions showed a somewhat larger distribution. In the case of the XL activity phantom, the distribution in P2P is large for both scanners, in fact larger than in both the other phantoms. The large spread in data values makes it hard to compare the different scanners and reconstructions, since for some data values one scanner gives higher P2P, but if other data values are compared, the other scanner might give higher P2P. The fact that the different time frames give such different P2P values is problematic, since this would mean that the detectability differs between scans of the exact same length. This observation was also seen in the visual evaluation, where for example 4 spheres were detected in some frames while in some other frames, for the same reconstruction method, only 2 spheres were detected. Which frames that gave higher P2P or visual score were completely random. Hence, there seems to be a random effect in the 90 second scans, affecting the detectability of the spheres. If this effect is seen in this phantom study, it most likely must be the same for patient scans, which makes this observation more concerning. A patient PET scan one time of the day would be assumed to give the same results as if the scan were performed under the same conditions but later in that day. With the random effect, this would possibly not always be the case and a small lesion could therefore hypothetically be detected in one scan, but not in another of the exact same length.

5.6 Limitations and future aspects

This work was limited to BSREM reconstructions with β value of 500. This value is known to give different noise characteristics for the two PET/CT systems, due to TOF and non-TOF, which therefore creates a limitation when comparing the noise between the systems. Hence, it would be of interest to add reconstructions with a few different β values to the comparison. It would also be interesting to acquire data over shorter and longer times than 90 seconds, since the acquisition time is an important factor seen from a clinical perspective, for such as patient movements. In this study, only one SBR was used, although various SBRs were desired to be used. It would be interesting to test the systems under conditions closer to those seen in the liver, which would be an SBR of approximately 3:1 or 2:1. A limitation that occurs in the visual evaluation is that the data was not blinded, which makes the reader aware of which reconstruction method that is viewed. A bias is also created for the same time frames of each measurement, which consist of the same data and come after one another when viewed. The images were for simplicity organised with all the DMI images before the OMNI images. It is possible that the reader's perspective of detectability changed over time, even though criteria on the fore hand were set, and hence favouring one of the systems. The possible change over time in a reader's perspective also becomes more pronounced when the number of readers is low, which in this study were limited to three. A methodological limitation worth noting is that it still remains unknown how the effect of the AI filters work on phantom data, since the AI models have been trained on patient data with and without TOF, not on phantom data. It would therefore be of interest to examine the AI filters' effect on a more human-like phantom geometry.

6 Conclusion

From this work, it can be stated that the ability to detect small lesions decrease with phantom size, but also that not just the total volume around the lesions, but the radioactive volume, in fact affect the detectability. The results from the P2P values showed that the benefit of true TOF was greater for the smaller phantom than the larger, which contradicts the theory. The evaluation of whether P2P can be considered a good metric of detectability was supported by the visual evaluation. For the larger phantom a higher number of spheres were detected with the OMNI scanner, while for the smaller phantom the scanners showed no big difference. The AI

filters did not show a significant increase in P2P compared to the corresponding reconstruction without filter applied, and through the visual evaluation, the AI filtered images did not show an increase in detectability compared to the non filtered images. The evaluation of the recovery curves did not always show correspondence to the P2P values, neither with the visual evaluation, and the recovery might therefore not be a reliable metric when searching for detectability.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors; **Sofia Kvernby** and **Anna Stenvall**. Thank you for the endless support and encouragement that you have given me through this project. It has been so fun to work with you and I have learned so much.

I would also like to thank **David Minarik**; for helping with the data analysis and design of the project, and **Jenny Oddstig**; for your curiosity around the project and for the willingness to exchange thoughts of the results.

A big thank you to 3D-centrum in Lund for bouncing ideas in designing the XL phantom, and for 3D printing the beautiful phantom, which I hope will come to use in later research projects.

Lastly, I want to thank my family for your constant support, and a big thank you to my classmates for these last three years together.

References

- [1] Meysam Dadgar et al. “Assessing the deep learning based image quality enhancements for the BGO based GE omni legend PET/CT”. In: *EJNMMI Physics* 11.1 (Oct. 2024). DOI: <https://doi.org/10.1186/s40658-024-00688-2>.
- [2] Magdy M. Khalil (Ed.) *Basic Sciences of Nuclear Medicine*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. ISBN: 9783540859611. DOI: <https://doi.org/10.1007/978-3-540-85962-8>.
- [3] Emilie Roncali and Simon R Cherry. “Application of silicon photomultipliers to positron emission tomography”. en. In: *Ann. Biomed. Eng.* 39.4 (Apr. 2011), pp. 1358–1377.
- [4] Michael Ljungberg (Ed.) *Handbook of Nuclear Medicine and Molecular Imaging for Physicists*. 2022. URL: <https://www.taylorfrancis.com/reader/download/63f02eb7-e658-4a0a-800b-384bbf053e60/book/pdf?context=ubx>.
- [5] Floris P Jansen. *The dBGO detector of Omni Legend – a new approach to PET detection [White paper]*. Tech. rep. GE HealthCare, 2023.
- [6] Dennis R Schaart. “Physics and technology of time-of-flight PET detectors”. In: *Physics in Medicine Biology* 66.9 (Apr. 2021), 09TR01. DOI: <https://doi.org/10.1088/1361-6560/abee56>.
- [7] Miles N Wernick and John N Aarsvold. *Emission tomography : the fundamentals of PET and SPECT*. Amsterdam ; Boston: Elsevier Academic Press, 2004. Chap. 10. ISBN: 9780127444826.
- [8] Abolfazl Mehranian et al. “Deep learning–based time-of-flight (ToF) image enhancement of non-ToF PET scans”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 49.11 (May 2022), pp. 3740–3749. DOI: <https://doi.org/10.1007/s00259-022-05824-7>.

- [9] Thomas F. Budinger. “Time-of-Flight Positron Emission Tomography: Status Relative to Conventional PET”. In: *Journal of Nuclear Medicine* (1983). URL: <https://jnm.snmjournals.org/content/jnumed/24/1/73.full.pdf>.
- [10] Scott Wollenweber Floris Jansen Abolfazl Mehranian. *Omni Legend™ with Precision DL — bringing the power of artificial intelligence to PET [White paper]*. Tech. rep. GE HealthCare, 2024.
- [11] Suleman Surti and Joel S. Karp. “Application of a Generalized Scan Statistic Model to Evaluate TOF PET Images”. In: *IEEE Transactions on Nuclear Science* 58.1 (Feb. 2011), pp. 99–104. DOI: <https://doi.org/10.1109/tns.2010.2072791>. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3234526/?utm_source=chatgpt.com.
- [12] Hui Liu et al. “PET Image Denoising Using a Deep-Learning Method for Extremely Obese Patients”. In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 6.7 (Sept. 2022), pp. 766–770. DOI: <https://doi.org/10.1109/trpms.2021.3131999>. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10241407/?utm_source=chatgpt.com.
- [13] Mitsuaki Tatsumi et al. “Impact of body habitus on quantitative and qualitative image quality in whole-body FDG-PET”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 30.1 (Jan. 2003), pp. 40–45. DOI: <https://doi.org/10.1007/s00259-002-0980-5>.
- [14] Suleman Surti and Joel S. Karp. “Advances in time-of-flight PET”. In: *Physica Medica* 32.1 (2016), pp. 12–22. ISSN: 1120-1797. DOI: <https://doi.org/10.1016/j.ejmp.2015.12.007>. URL: <https://www.sciencedirect.com/science/article/pii/S1120179715010492>.
- [15] S. Surti. “Update on Time-of-Flight PET Imaging”. In: *Journal of Nuclear Medicine* 56.1 (Dec. 2014), pp. 98–105. DOI: <https://doi.org/10.2967/jnumed.114.145029>. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287223/>.
- [16] K Lange and R Carson. “EM reconstruction algorithms for emission and transmission tomography”. In: *Journal of computer assisted tomography* 8.2 (Apr. 1984), pp. 306–16. URL: <https://pubmed.ncbi.nlm.nih.gov/6608535/>.
- [17] L. A. Shepp and Y. Vardi. “Maximum Likelihood Reconstruction for Emission Tomography”. In: *IEEE Transactions on Medical Imaging* 1.2 (Oct. 1982), pp. 113–122. DOI: <https://doi.org/10.1109/tmi.1982.4307558>.
- [18] A. Teymurazyan et al. “Properties of Noise in Positron Emission Tomography Images Reconstructed with Filtered-Backprojection and Row-Action Maximum Likelihood Algorithm”. In: *Journal of Digital Imaging* 26.3 (Aug. 2012), pp. 447–456. DOI: <https://doi.org/10.1007/s10278-012-9511-5>. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3649042/>.
- [19] Sangtae Ahn et al. “Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET”. In: *Physics in Medicine and Biology* 60.15 (July 2015), pp. 5733–5751. DOI: <https://doi.org/10.1088/0031-9155/60/15/5733>.
- [20] Fatemeh Sadeghi et al. “Phantom and clinical evaluation of Block Sequential Regularized Expectation Maximization (BSREM) reconstruction algorithm in 68Ga-PSMA PET-CT studies”. In: *Physical and Engineering Sciences in Medicine* 46.3 (July 2023), pp. 1297–1308. DOI: <https://doi.org/10.1007/s13246-023-01299-4>.
- [21] Steve Ross. *Q.Clear [White paper]*. Tech. rep. GE HealthCare. URL: https://www.gehealthcare.com/-/media/739d885baa59485aaef5ac0e0eeb44a4.pdf?srsltid=AfmB0op1RPk-o_qco0CFzxaeMwBczmHUD0q3kWeqyDjeX9rFI1_EKYM.

- [22] Judit Lantos et al. “Standard OSEM vs. regularized PET image reconstruction: qualitative and quantitative comparison using phantom data and various clinical radiopharmaceuticals”. In: *American journal of nuclear medicine and molecular imaging* 8.2 (2018), pp. 110–118. URL: <https://pubmed.ncbi.nlm.nih.gov/29755844/>.
- [23] Kenta Miwa et al. “Impact of factor in the penalty function of Bayesian penalized likelihood reconstruction (Q.Clear) to achieve high-resolution PET images”. In: *EJNMMI Physics* 10.1 (Jan. 2023), pp. 4–4. DOI: <https://doi.org/10.1186/s40658-023-00527-w>.
- [24] Läkemedelsverket. “Produktresumé”. In: (). URL: [https://docetp.mpa.se/LMF/Fludeoxyglucose%20\(18F\)%20SUS%20Lund,%20400%20MBq%20till%2090%20GBq,%20solution%20for%20injection%20SmPC_09001bee807a38f4.pdf](https://docetp.mpa.se/LMF/Fludeoxyglucose%20(18F)%20SUS%20Lund,%20400%20MBq%20till%2090%20GBq,%20solution%20for%20injection%20SmPC_09001bee807a38f4.pdf).
- [25] *Production and Quality Control of Fluorine-18 Labelled Radiopharmaceuticals*. Tech. rep. International Atomic Energy Agency, 2021. URL: <https://www-pub.iaea.org/MTCD/Publications/PDF/TE-1968web.pdf>.
- [26] *Cyclotron Produced Radionuclides: Principles and Practice*. Tech. rep. International Atomic Energy Agency, 2008. URL: https://www-pub.iaea.org/MTCD/Publications/PDF/trs465_web.pdf.
- [27] F. Hallouard et al. “Stability Study of [18F]Fludeoxyglucose and [18F]Fluorocholine in Different Medical Devices for Intravenous Administration”. In: *Current Radiopharmaceuticals* 11.2 (Aug. 2018), pp. 138–146. DOI: <https://doi.org/10.2174/1874471011666180703165639>.
- [28] Jenny Oddstig et al. “Head-to-head comparison of a Si-photomultiplier-based and a conventional photomultiplier-based PET-CT system”. In: *EJNMMI Physics* 8.1 (Feb. 2021). DOI: <https://doi.org/10.1186/s40658-021-00366-7>.
- [29] Ronald Boellaard et al. “Quality control for quantitative multicenter whole-body PET/MR studies: A NEMA image quality phantom study with three current PET/MR systems”. In: *Medical Physics* 42.10 (2015), pp. 5961–5969. DOI: <https://doi.org/10.1118/1.4930962>. eprint: <https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4930962>. URL: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4930962>.