Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

ACCEPTED MANUSCRIPT

Performance evaluation of digital breast tomosynthesis systems: physical methods and experimental data

To cite this article before publication: Nicholas W Marshall *et al* 2022 *Phys. Med. Biol.* in press https://doi.org/10.1088/1361-6560/ac9a35

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2022 Institute of Physics and Engineering in Medicine**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 134.58.253.56 on 10/11/2022 at 09:28

Performance evaluation of digital breast tomosynthesis systems: physical methods and experimental data

N.W. Marshall, H. Bosmans

UZ Gasthuisberg, Department of Radiology, Herestraat 49, 3000 Leuven, Belgium

Medical Imaging Research Center, Medical Physics and Quality Assessment , Katholieke Universiteit Leuven, 3000 Leuven, Belgium

Abstract

Digital breast tomosynthesis (DBT) has become a well-established breast imaging technique, whose performance has been investigated in many clinical studies, including a number of prospective clinical trials. Results from these studies generally point to non-inferiority in terms of microcalcification detection and superior mass-lesion detection for DBT imaging compared to digital mammography (DM). This modality has become an essential tool in the clinic for assessment and ad-hoc screening but is not yet implemented in most breast screening programmes at a state or national level. While evidence on the clinical utility of DBT has been accumulating, there has also been progress in the development of methods for technical performance assessment and quality control of these imaging systems. DBT is a relatively complicated "pseudo-3D" modality whose technical assessment poses a number of difficulties. This paper reviews methods for the technical performance assessment of DBT devices, starting at the component level in part one and leading up to discussion of system evaluation with physical test objects in part two. We provide some historical and basic theoretical perspective, often starting from methods developed for DM imaging. Data from a multi-vendor comparison are also included, acquired under the medical physics QC protocol developed by EUREF and currently being consolidated by a European Federation of Organisations for Medical Physics (EFOMP) working group. These data and associated methods can serve as a reference for the development of reference data and provide some context for clinical studies.

Accepted Manuscript

Performance evaluation of digital breast tomosynthesis systems: physical methods and experimental data

1. Introduction

Digital breast tomosynthesis (DBT) is an imaging modality in which projection views acquired over a limited angular range are reconstructed to produce cross-sectional planes through the breast (Niklason *et al.*, 1997). This modality offers a solution, at least in part, to the problems of tissue masking, whereby breast lesions are obscured by overlying anatomy, and of increased false positive findings from tissue superposition (Destounis, Morgan and Arieno, 2015; Gilbert, Tucker and Young, 2016). A DBT scan will often be complemented with a DM acquisition or more commonly a synthetic mammography image (SM), calculated from the DBT dataset (Durand, 2018). Numerous clinical studies have shown that DBT used in conjunction with digital mammography (DM) results in increased sensitivity and/or increased specificity in comparison to DM alone (Bernardi *et al.*, 2016; Zackrisson *et al.*, 2018; Skaane *et al.*, 2019). Despite increased sensitivity, the question remains open as to whether DBT detects more of the cancers that lead to increased morbidity (Yaffe, 2018; Mandoul *et al.*, 2019). Prospective studies have not revealed differences in interval breast cancer rates for women screened with DBT plus synthetic mammography (SM) compared to women screened with DM alone (Skaane *et al.*, 2018; Hovda *et al.*, 2020). Currently, a definitive answer to this question can only come from randomized controlled trials that deliver information on the distribution and types of interval cancers in the screened population. Comparison of DBT plus SM with DM alone should help to determine whether this modality improves detection of the small, dangerous cancers that lead to metastatic disease and increased morbidity or whether DBT largely improves detection of less clinically significant, small low grade, low stage cancers (Gilbert, Tucker and Young, 2016; Li, Marinovich and Houssami, 2018; Yaffe, 2018; Houssami *et al.*, 2021).

While clinical studies provide essential data on sensitivity via performance parameters such as small cancer detection performance of DBT compared to DM, they are often performed using an imaging device from one manufacturer (Mandoul *et al.*, 2019), even using one particular generation of a model. This can lead to difficulties in generalizing results from a given study to obtain some idea of 'DBT imaging' performance in general, where there can be substantial differences in acquisition regimes, dose operating point, reconstruction algorithms and associated image quality between devices from different manufacturers (Sechopoulos, 2013; Vedantham *et al.*, 2015; Yaffe, 2018). This has parallels with the Digital Mammography Imaging Screening Trial (DMIST) (Bloomquist, Yaffe, Pisano, *et al.*, 2006), where systems based around a wide range of x-ray acquisition technologies including photoconductor (a-Se) or scintillator (CsI) based flat panel detectors, CsI+charge-coupled device (CDD), photostimulable phosphors and CsI+CCD+slot-scan technology, were grouped under the term "DM" for the comparison against the established technology in breast screening at the time i.e. screen-film (S/F) mammography. While the choice of x-ray detector used in DBT devices has currently narrowed to CsI or a-Se flat panel detectors, considerable differences remain in parameters such as angular range, tube focus motion and reconstruction algorithm.

Accepted Manuscript

These technical parameters will likely influence clinical performance and a comprehensive set of technical measurements should help with the interpretation of these data and to situate the clinical results. Technical parameters assessed at the component or system level lie firmly in Stage 1 on the scale describing imaging system efficacy defined by Fryback and Thornbury (Fryback and Thornbury, 1991; Barrett *et al.*, 2015). Stage 2 on this scale is defined as diagnostic accuracy, often assessed in clinical studies and quantified using specificity and sensitivity. The difficulty in combining technical measurements established for an imaging system to form an idea of diagnostic accuracy is formidable. Under a given set of assumptions, task performance can be estimated using Fourier methods to characterize signal and noise transfer through the imaging system along with task functions for which some performance metric is calculated (ICRU, 1996). An alternative is the use of in-silico methods or virtual clinical trials (VCT) to assess imaging system performance, a field which has seen rapid growth recently (Abadi *et al.*, 2020; Badano, 2021; Barufaldi, Maidment, *et al.*, 2021). Underpinning these methods is the availability of basic performance data, obtained using well-defined and commonly applicable methods.

This review focuses on methods for the technical performance evaluation of DBT systems and includes some historical or theoretical context, which is then extended to include recent developments. In the first part, quality control (QC) data acquired by the medical physics service at our centre under the EUREF DBT QC protocol (van Engen *et al.*, 2018) are used to illustrate these methods. Evaluation at a component level is essential for understanding basic imaging performance of a system and for establishing baselines to track a system over time. Parameters assessed include yield and energy of the x-ray source, detector noise levels and dynamic range, system sharpness, and the spatial resolution orthogonal to the reconstructed DBT planes. While it is possible to combine parameters describing signal and noise transfer, for example using Fourier models to obtain an idea of imaging performance (ICRU, 1996; Zhao and Zhao, 2008b; Monnin *et al.*, 2020), this is not commonplace, certainly in a QC context. Far more common is the use of test objects which effectively combine specific aspects of system performance to produce a metric of overall imaging performance. The second part reviews the use of test objects, starting with the evaluation of automatic exposure control and dose and then proceeds to describe the range of test objects that are either available commercially or are under development to measure what we have defined as technical image quality.

2. Current DBT systems

All currently available DBT systems have a similar architecture, one that is derived from DM systems. A rotating arm holds the x-ray tube and the x-ray detector assembly. Tube and detector rotate together so that the standard craniocaudal (CC) and mediolateral oblique (MLO) views can be performed for both DM and DBT modes. When switching to DBT mode, the detector is either held static or rotates over a small angle (~2°) while the x-ray tube is moved over the angular range prescribed for the system. This involves decoupling the x-ray tube from the detector such that the source rotates around a point at the detector input plane or some distance above (the centre of rotation (COR)). Other system geometries have been implemented, for example the slit-scanning Sectra MicroDose L30 was adapted to generate projection

Accepted Manuscript

images by pivoting the x-ray tube, collimator and detector around a point 20 cm below the breast support table (Schmitzberger *et al.*, 2011). An observer study using a version of this system reported superior diagnostic accuracy for two-view tomosynthesis compared to FFDM for readers with the least reading experience in the study (Wallis *et al.*, 2012), however this system did not become commercially available.

The systems listed in Table 1 share the same basic architecture although there are considerable differences in the actual implementations by the different manufacturers. Some vendors supply a DBT system but are not the original equipment manufacturer (OEM) of the device. For example, the OEM for the Canon Artemis is Planmed Oy (i.e; the Planmed Clarity) while the OEM of the Villa Melody IIID TS is Metaltronica SpA (the Metaltronica Helianthus). Perhaps the parameter that most clearly distinguishes between the systems is the angular range, where different systems cover 'narrow', 'medium' and 'wide' angular spans. The wide angle systems currently include the Inspiration and Revelation units from Siemens Healthineers plus the Fuji Innovality in HR mode, with a range of ~50°. The medium range group includes the GE Pristina (25°) and the IMS Giotto Class and Planmed Clarity (both 30°) while the Hologic Selenia Dimensions and 3Dimensions and the Fuji Innovality in ST mode have a narrow angle (15°). Angular range is an obvious candidate for optimization and debate is ongoing as to the merits of wide versus narrow angle schemes. As seen in Table 1, there is not a clear link between the number of projections utilized by the manufacturers and the angular range, with the sampling density ranging from 1° per projection (Hologic Dimensions and Fuji ST mode), through 2° for the Siemens units to approximately 2.7° for the remaining systems. For systems with a continuously moving tube during acquisition, the angular range, and thus the angular velocity, sets some limitations on the total time for a scan. Large angular velocities lead to high levels of geometric blurring, the degree to which this influences system sharpness depends on the projection image pulse length (Zhou, Zhao and Zhao, 2007; Marshall and Bosmans, 2012). Scan times range from ~3.7s for the narrow angle systems (Hologic Dimensions and 3Dimensions and Fuji Innovality in ST mode), through ~9s for the medium angle systems and up to 21s for the Siemens Inspiration and Revelation units.

Table 1 lists the current SM image associated with the systems. SM imaging has an important role to play, especially in a screening context, potentially obviating the need for additional DM ('combo') views and thus reducing dose for the total examination. These images also facilitate comparison with previous (DM) screening images and can reduce radiologist reading time (Ratanaprasatporn, Chikarmane and Giess, 2017). SM imaging was approved for use in combination with the DBT dataset by the U.S. Food and Drug Administration (FDA) in 2013 (Durand, 2018).

As with reconstruction, the SM algorithms are proprietary with the exact method of production not disclosed (Ratanaprasatporn, Chikarmane and Giess, 2017; Durand, 2018). It is possible that both projection images and information from the reconstructed stack will be used to generate the SM image, perhaps with variable weights applied to the different projections. Algorithms may be employed to detect and highlight objects of interest or regions containing calcification-like or lesion-like objects by applying additional weighting compared to regions judged to have characteristics consistent with normal adipose and

Accepted Manuscript

fibroglandular tissue (Ratanaprasatporn, Chikarmane and Giess, 2017). The upshot of this is that SM algorithms may specifically process some breast tissue structures for improved visualization, and technical image quality evaluation of these algorithms is not straightforward. Some studies have utilized phantoms with homogeneous backgrounds such the ACR phantom (Gammex: A Sun Nuclear Company, Middleton, WI, USA) and the CDMAM phantom (Artinis Medical Systems, Elst, The Netherlands) while others have used test objects with heterogeneous backgrounds such as CIRS BR3D, Leuven L1, the Leeds TORMAS and the FDA printed phantoms (Nelson *et al.*, 2016; Baldelli *et al.*, 2018; Petropoulos *et al.*, 2020; Vancoillie *et al.*, 2020, 2021; Barca *et al.*, 2021; Ikejimba *et al.*, 2021). While SM imaging clearly has an important role to play in the implementation of DBT in screening (Durand, 2018) no protocol exists for the technical assessment of these images and evaluation methods are still under development. Consequently, SM imaging is discussed at points in this review, but the focus is on DBT system evaluation.

Accepted Manuscript

Table 1 Specifications and geometry of digital breast tomosynthesis systems

DBT system	Fujifilm Amulet Innovality	GEHC SenoClaire	GEHC Pristina	Hologic Selenia Dimensions	Hologic 3Dimensions	IMS Giotto Class	Metaltronica Helianthus	Planmed Clarity	Siemens Mammomat Inspiration	Siemens Mammomat Revelation
Focus size (mm)										
Target	W	Mo / Rh	Mo / Rh	W	W	W	W	W	W	W
Filter	700 μm Al	30 μm Mo 25 μm Rh	30 μm Mo 30 μm Ag	700 μm Al	700 μm Al	50 μm Ag	500 μm Al	75 μm Ag 50 μm Rh	50 μm Rh	50 μm Rh
X-ray tube motion	Continuous	Step and shoot	Step and shoot	Continuous	Continuous	Continuous	Continuous	Continuous with "sync and shoot"	Continuous	Continuous
Source – detector distance (mm)	650	660	659	700	700	688.5	658	656	650	650
Source – table top distance (mm)							636			
Centre of rotation – detector distance (mm)	46	40	46	0	0	40	45		40	40
Antiscatter grid	No	Yes	Yes	No	No	No	No	No	No	No
X-ray sensitive material	a-Se	CsI	CsI	a-Se	a-Se	a-Se	a-Se	a-Se	a-Se	a-Se
Detector Element Spacing (μm)	50*	100	100	140 (70 binned 2/2)	70	85	85	83	85	85
Detector motion	Static	Static	Static	Rotating	Rotating	Static	Static	Rotating during acquisition	Static	Static
Pixel size in focal plane (μm)		100	100	100 or 85	100 or 85	?	85	95	85	85
Number of projections	15	9	9	15	15	11	11, 13, 19	15	25	25
Nominal angular range (°)	15 (ST) 40 (HR)	25	25	15	15	30	±7,5, ±12, ±25	30	50	50
Angle/projection (°)	1 or 2.7	2.8	2.8	1	1	2.7	1.5, 2, 2.7	2	2	2
Time for scan (s)	3.5 (ST) 9.4 (HR)	9	9	3.7	3.7	9.2	2.5, 4, 7.7	18	21	21
Zero Degree mode acquisitions possible ?	Yes	Yes	Yes	Yes	Yes		Yes		No	Yes
'For Processing'-type projection images ?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Accepted Manuscript

Synthetic Mammography	S-view	V-Preview	C-View	Intelligent C-view	G-VIEW	(Yes)	Clarity S2D	Insight 2D	Insight 2D

Accepted Manuscript

3. PART 1 - Physical characterization of DBT systems: technical parameters

This part discusses DBT system technology and evaluation at the component level. The sub-sections include some historical perspective along with measurements obtained with the EUREF DBT protocol v1.03 (van Engen *et al.*, 2018). This protocol will be superseded by the European Federation of Organisations for Medical Physics (EFOMP) DBT protocol that is currently under development, however with the exception of dosimetry, differences between the methods in the two protocols are likely to be small. The data presented here were acquired for DBT systems known to stable, with no reported operational issues, set up according to manufacturer protocols. QC data were acquired by an experienced medical physics QC team, using the methods described in the EUREF protocol v1.03 (van Engen *et al.*, 2018).

3.1 X-ray source

Most evaluations of x-ray imaging systems start with the x-ray source; effective and consistent imaging would not be possible without a stable source of photons with a well-defined energy. A standard measurement set for DBT systems includes short term reproducibility of x-ray tube output and tube voltage (van Engen *et al.*, 2018). To evaluate this, an appropriate dosimeter is required, calibrated for the energy range used by DBT systems. Guidance on the required precision, accuracy, energy dependence and linearity for dosimeters used for mammography is available in an American Association of Physicists in Medicine (AAPM) report (Wagner *et al.*, 1992). In recent years, many medical physics departments have switched to solid state dosimeters which are incorporated into multi-meters and give a single shot measurement of half value layer (HVL) (Brateman and Heintz, 2015; Bemelmans *et al.*, 2021). Calibrations are available for all anode/filter (A/F) combinations used in DBT systems, including W anodes and Al filters. A crucial consideration when measuring tube output and HVL for DBT systems is the angular response of dosimeters. For accurate dosimetry, the radiation sensitive volume should be oriented perpendicular to the incident x-ray beam axis (International Atomic Energy Agency (IAEA), 2007). Furthermore, solid state dosimeters have input windows that limit the angle over which the dosimeter is sensitive (Martin, 2007) and will underestimate the measured x-ray output as DBT system angular range increases (Bradley, Carton and Maidment, 2010). This problem has since been solved through the introduction of a "zero degree" mode, in which all clinical functionality is retained but without tube rotation i.e. the tube remains in the zero degree projection while exposing. Most systems now support such a mode (Table 1). A number of early DBT systems did not have such a mode, or the mode was difficult to access; output and HVL were often assessed in DM mode as identical A/F settings were used for DM and DBT modes. In terms of technical specifications, the dosimeter has to cope with a frame rate of between 1.2 frame/s and 4 frame/s. Minimum air kerma/image is in the range of 5 μGy/image while maximum entrance air kerma is of the order 15 mGy, delivered in ~3.7 s. This performance is easily achievable with the current generation of dosimeters that can cope with fluoroscopy imaging rates.

Accepted Manuscript

The energy of the x-ray spectrum is an important parameter for x-ray imaging systems, with implications for patient dose and image quality (Fahrig and Yaffe, 1994). As imaging technology has changed, so has the optimal x-ray spectrum and the associated A/F compositions. There has been a progression from screen/film, where preservation of (radiographic) subject contrast was paramount and molybdenum (Mo) commonly used for both anode and filter (Fahrig and Yaffe, 1994; Sandborg *et al.*, 2000). With the switch to digital detectors in DM systems, signal to noise ratio (SNR) became the important imaging parameter (Fahrig and Yaffe, 1994). While there have been numerous studies on optimal spectra vs breast thickness for DM systems (Bernhardt, Mertelmeier and Hoheisel, 2006; K. C. Young *et al.*, 2006; Toroi *et al.*, 2007; Williams *et al.*, 2008), there has been little work on this explicitly for DBT systems, since early developmental work performed on this topic (Ren *et al.*, 2005). Measurement-based optimization studies often cannot be performed because many DBT systems will only operate with the anode/filter combination designated for DBT mode, suggesting that anode/filter has already been optimized for DBT operation. Current DBT systems are developments of DM devices and thus use the same x-ray detectors, either a scintillation phosphor (typically caesium iodide (CsI)) or a photoconductor (typically amorphous selenium (a-Se)). The influence of the x-ray detector on the optimal working point is expected to be similar between DM and DBT operation. These systems share a common optimization constraint, namely that the time for acquisition should be as low as possible to limit blurring due to patient motion, i.e. exposure time for a DM projection and scan time for DBT systems. An additional constraint exists for DBT systems where continuous tube motion during acquisition of the projection images leads to geometric blurring introduced by the potentially extended source size for these systems – this is discussed in more detail later. Taken together, these constraints favour higher output x-ray sources, such that the required x-ray signal can be reached in a shorter time.

X-ray beam filtration for DBT devices was studied using a narrow beam simulation by Shrestha *et al* (Shrestha, Vedantham and Karellas, 2017) using the tabulated spectra of Boone *et al* (Boone, Fewell and Jennings, 1997). A target signal difference to noise ratio (SDNR) was set and filtering regimes employing Al and K-edge filters were compared, using the metrics of breast dose and exposure time/photon fluence. Three filter materials were examined (700 μm Al, 50 μm Rh and 50 μm Ag), covering the filtration thickness and composition commonly used in current DBT systems. Three regimes were considered: fixed Al thickness and tube voltage matched between the Al and K-edge regimes, leading to a variation in HVL; fixed Al thickness with matched HVL but with varying tube voltage; and finally varying the tube voltage and Al thickness to match the HVL. All three alternatives showed an increase in photon fluence for the Al filtration scheme compared to the use of the K-edge filters. This is consistent with the results of QC x-ray tube output measurements made according to the EUREF DBT protocol at 28 kV, with the compression plate in the x-ray beam (Table 2). Radiation output at 1m, normalized for tube load (i.e. current-time product (mAs)), is approximately a factor of 2 higher for the systems with the Al filter than the units with Rh or Ag filtration, at 28 kV.

Accepted Manuscript

An important observation noted for K-edge filters was the influence of the K-edge on photon fluence. For Rh (K-edge at 23.2 keV) and Ag (K-edge at 25.5 keV), the filter strongly reduces the number of photons with energies between the K-edge energy and the peak tube voltage, when the tube voltage is greater than the K-edge (Figure 1). This limits the utility of an increase in tube voltage in order to increase photon fluence. Shrestha *et al* (Shrestha, Vedantham and Karellas, 2017) also performed analytical calculations of SDNR to examine whether the change from K-edge to Al filtration had a detrimental effect on image quality. For all three schemes described earlier, use of Al filtration enabled the SDNR of the K-edge filtered case to be matched, but with reduced exposure duration and some reduction in mean glandular dose (MGD). A limitation of the study was the use of an analytical, narrow beam calculation of SDNR based only on x-ray photon statistics, excluding the influence of x-ray source and detector on the results. Whether the positive result in favour of Al filtration applies to real systems requires further work, using either measurements and/or more realistic simulations.

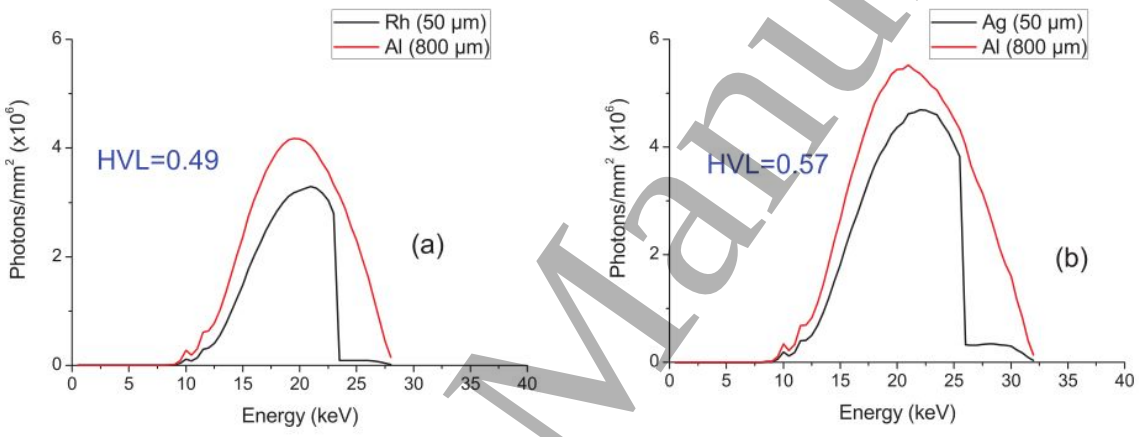


Figure 1. Comparison of spectra provided by 50 µm thick K-edge filters and Al filters of varying thickness at matched kV and HVL. (a) 28 kV Rh filter with 28 kV Al filter (800 m) (HVL = 0.49), (b) 32 kV Ag filter (50 µm) with 32 kV Al filter (800 µm) (HVL =0.57). (Figure taken from (Shrestha, Vedantham and Karellas, 2017)).

Table 2 Output and HVL measured with the compression paddle in the x-ray beam for the Mo/Mo and Rh/Ag settings. (Data measured using a calibrated RTI Piranha dosimeter (RTI, Mölndal, Sweden) as part of routine QC programme by UZ Leuven Medical Physics Department.)

System	Set tube voltage (kV)	Anode/filter (A/F)	Output (µGy/mAs at 1 m) with compression plate	HVL (mm Al)
Fuji Innovality	28	W/Al (700 µm)	30.12	0.45
Giotto Class	28	W/Ag (50 µm)	18.18	0.53
GE Pristina	26	Mo/Mo (30 µm)	30.09	0.33
GE Pristina	34	Rh/Ag (50 µm)	47.39	0.47
Hologic Selenia Dimensions	28	W/Al (700 µm)	27.76	0.41
Hologic 3Dimensions	28	W/Al (700 µm)	31.85	0.43
Siemens Inspiration	28	W/Rh (50 µm)	13.50	0.53
Siemens Revelation	28	W/Rh (50 µm)	11.77	0.55

Accepted Manuscript

Advances in source implementation include the development of multisource DBT systems in which a linear array of x-ray sources is distributed to cover some specific angular range (Qian *et al.*, 2012; Neculaes *et al.*, 2014; Calliste *et al.*, 2017; Becker *et al.*, 2020; Behling, 2021). Rapid sequential switching of the sources covers the angular range while the individual sources remain stationary during acquisition. This has two major benefits: eliminating motion blurring from a moving source and reducing total scan time, which in turn reduces blurring due to patient motion (Qian *et al.*, 2012; Calliste *et al.*, 2017). Given the short acquisition time, angular span can be increased without affecting total scan time and this should improve the in-depth (z-direction) resolution. Design constraints for these systems are the limited flux from the source and the x-ray detector readout speed.

The study of Calliste *et al* (Calliste *et al.*, 2017) develops the initial stationary source DBT system described by Qian *et al* (Qian *et al.*, 2012) that utilized carbon nanotube (CNT) x-ray sources. The second generation system utilized 21 CNT sources with increased flux, arranged linearly (i.e. not on an arc) to cover an angular span of 35°, compared to 30° in the first generation system. Inherent filtration was 0.5 mm Al and focus size was 1.1 mm. The x-ray detector was readout at the native 70 μm pixel size and also at 140 μm using binning. The second generation system showed increased in-plane resolution for 70 μm pixel readout; spatial frequency at the modulation transfer function (MTF) value of 10% increased from 6.7 mm^{-1} to 7.7 mm^{-1}. In-depth resolution was quantified using the artefact spread function (ASF) (see section 3.7); full width at half maximum (FWHM) fell from 4.30 mm for the earlier 28° angular range to 2.35 mm for the 35° span (Tucker, Lu and Zhou, 2013; Calliste *et al.*, 2017).

CNT tubes have stationary anodes and therefore limited x-ray flux compared to rotating anode tubes, restricting the reduction in total scan time that can be achieved. Recent work by Becker *et al* (Becker *et al.*, 2020) has described a multi-x-ray source array (MXA) using an 11-source rotating anode array with molybdenum anodes and a 1 mm thick Al tube window, shown in figure 2. Measured x-ray tube output at 1 m for a single source in the array ranged from 6 μGy/mAs at 20 kV to 170 μGy/mAs at 45 kV. HVL was found to vary from 0.5 mm Al at 30 kV to 0.8 mm Al at 45 kV. The x-ray pulse width achievable ranged from 20 ms to 110 ms, which covers the range seen in current DBT systems at typical automatic exposure control (AEC) settings (see Table 4). The MXA has a projected scan time for 15 projections of 2.0 s, assuming a detector such as the Varex Dexela CMOS 2923 with 75 μm dels, operated at 26 frame/s. This is an improvement over the value of ~3.1 s seen in the second generation CNT based system (Calliste *et al.*, 2017) and over the shortest total scan times of ~3.5 seen for current commercial systems (see Table 4).

Table 2 presents data for eight DBT units covering a range of A/F combinations. Here we see that output for the Mo/Mo and W/Al combinations is similar at ~30 μGy/mAs at 100 cm, while the output for W/Ag and W/Rh combinations is considerably lower, ranging between 12 and 18 μGy/mAs. This is partly a result of the K absorption edges of Rh and Ag, respectively lying at 23.22 keV and 25.51 keV, removing higher energy photons (see figure 1). Much higher output is seen for the Rh/Ag setting in the Pristina system,

Accepted Manuscript

mainly due to the higher tube voltage used for this combination. The higher output has some consequences in the measurement of the response function, discussed later.

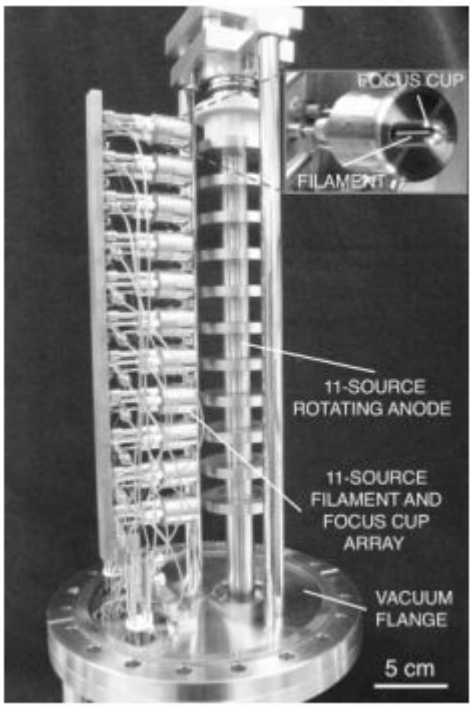


Figure 2. Photograph of the 11-source element DBT multi-x-ray source array (MXA) with a close-up of the filament and focus-cup-grid assembly of one source element (taken from Becker *et al* (Becker *et al.*, 2020))

3.2 Antiscatter grid

Of the systems listed in Table 1, only the GEHC SenoClaire and Pristina units have antiscatter grids. These are linear grids orientated such that the grid strips run parallel to the chest wall edge. They are both composed of 15 μm thick lead strips with a fibre interspace material, but with some differences in design. The SenoClaire has a grid ratio of 5:1, 104 lines/cm and graphite interspacing, while the grid in the Pristina has a ratio of 11:1, 67 lines/cm and fibre interspace material (Klausz *et al.*, 2018). The EUREF DBT protocol (van Engen *et al.*, 2018) does not describe tests to evaluate the antiscatter grid, and physicists would fall back on guidance for antiscatter grid evaluation in DM systems (van Engen *et al.*, 2006). This involves a measurement of the grid system factor, also known as the grid exposure factor (IEC, 2013) or Bucky factor, a parameter that quantifies the increase in exposure required to compensate for the radiation (scattered and primary) absorbed by the grid. The grid exposure factor describes the increase in patient exposure due to the grid but says nothing about the influence of the grid on image quality. Although important for film based imaging, where the absolute x-ray signal at the image receptor could influence the imaged contrast, the detectors used in DM and DBT systems have a large dynamic range and can compensate for a loss of contrast. For this type of detector, the effect of the grid in terms of SDNR transfer is the relevant parameter (Chan, Lam and Wu, 1990; Neitzel, 1992; Åslund *et al.*, 2006). This has been standardized by the IEC (IEC, 2013) with the definition of the image improvement factor (Q): for a constant entrance dose, the SDNR increases in proportion to the square root of the factor Q, when an antiscatter grid is applied. The grid makes a positive contribution to the imaging chain when Q is greater

Accepted Manuscript

than 1. The factor Q can be measured directly from SDNR for some object imaged with and without grid, however is defined in the IEC standard as:

$$Q = \frac{T_p^2}{T_t} \quad (1)$$

where T_p and T_t are the primary and total transmission of the antiscatter grid, respectively. The parameter T_t is the inverse of the grid exposure factor. These parameters can be measured under typical operating conditions, using a narrow-beam, low scatter geometry for T_p and a broad-beam geometry using a relevant x-ray scattering object for T_t (Monnin et al., 2017; Klausz et al., 2018). Poly(methyl) methacrylate (PMMA) blocks are typically used as breast simulating materials, positioned respectively at the x-ray tube and on the breast support table for the T_p and T_t image acquisitions. For the T_p measurement, Monnin *et al* (Monnin et al., 2017) used pinholes of decreasing diameter, x-ray beams filtered with Al to match the HVL of the PMMA blocks and a fitting function derived from the scattered radiation PSF to extrapolate the mean pixel value (PV) for zero pinhole diameter. An alternative is to use a 10 mm diameter aperture positioned at the exit of the PMMA, as described in the IEC standard (Aichinger et al., 2012; IEC, 2013). Klausz et al (Klausz et al., 2018) describe measurements for the two GEHC DBT grids, evaluated in DM mode with the x-ray tube orthogonal to the antiscatter grid. This is equivalent to evaluation in DBT stationary tube mode for the central (0°) projection. The Q factor for 40 mm PMMA for the Pristina grid was 1.06, rising to 1.34 for 70 mm, indicating that grid use is advantageous for objects thicker than ~40 mm PMMA equivalent thickness (Klausz et al., 2018) (assuming similar scattering properties). Grid performance data have not been published for other x-ray tube angulations. There are scant data on the influence of antiscatter grid on technical image quality measures for the grid used in the GEHC Pristina in DBT mode; the study of Monnin *et al* (Monnin et al., 2020) reported a Q factor of 1.20 for a phantom with a PMMA equivalent thickness of 50 mm. This can be compared to the Q value of 1.16 found measured by Klausz et al (Klausz et al., 2018).

Regarding the evaluation of the antiscatter grid for QC purposes, measurement of grid system factor alone is of limited utility, either for DBT systems with a grid or in fact for DM systems. Measurement of T_t together with T_p and calculation of the image improvement factor Q (IEC, 2013) will give a more useful characterization of the grid. Standardized assessment of the grid is made at 28 kV using a Mo/Mo A/F setting for a 50 mm thick PMMA phantom (IEC, 2013). The Standard describes the measurement of the scattered radiation transmission for a grid (T_s), which is a much more involved measurement than T_t. We would propose a characterization of T_t and T_p, and ultimately Q, over thicknesses from 20 mm to 70 mm PMMA in steps of 10 mm using x-ray factors selected by the clinically used AEC mode at each thickness. This would be performed at the Commissioning stage using zero degree mode. Phantom thickness will have a strong influence on T_t while only a weak dependence of T_p and T_t on energy is expected (Shen et al., 2006). Use of AEC derived x-ray factors is a practical point that that will not have a large influence on the measurement. Evaluation of grid efficiency at other angulations is definitely of interest, but difficult to

Accepted Manuscript

achieve practically, as the PMMA attenuation must be suspended at the tube x-ray port for the T_p measurement.

3.3 X-ray detectors: technology and characterization

DBT x-ray detectors are flat panel detectors used for DM and general radiography applications (Rowlands and Yorkston, 2000; Yaffe and Mainprize, 2004), but optimised to meet the requirements for DBT imaging. There are two main detector technologies employed, direct conversion using an amorphous selenium (a-Se) photoconductor and indirection conversion using a caesium iodide (CsI) scintillation phosphor. These layers are sensitive to x-ray photons and convert the x-ray energy to electron-hole pairs, in the case of a-Se detectors, and light photons for the CsI layer. Readout of these secondary signal carriers is performed by an array of detector elements or pixels that integrate the charge (a-Se) or the light photons (CsI). The sensitive pixel area is nominally square for most detectors with the exception of the detector in the Fuji Amulet Innovality, where the pixels are hexagonal and have an effective pitch of 68 μm. Pixel spacing and detector conversion layer information for common DBT devices is given in Table 1.

Manufacturers source the detectors from a range of suppliers. The detectors in the GEHC, Hologic and Fuji systems are not available to other manufacturers. The Planmed Clarity and Canon Artemis use a PaxScan 3024MX or modified version of this detector, produced by Varex (Varex Imaging, Salt Lake City, UT). The Siemens Inspiration and Revelation models and the IMS Giotto Class use a version of the Analogic axs-2430 a-Se detector. These detectors can switch between DM and DBT mode as required.

Cheung *et al* (Cheung *et al.*, 2005) presented a comprehensive evaluation of a version of the a-Se detector that was later used in the Hologic Selenia Dimensions DBT system, possibly after some further development. This was the DBT device used in a number of important breast screening trials and clinical studies (Ciatto *et al.*, 2013). The tests detailed by Cheung *et al* (Cheung *et al.*, 2005) cover many essential aspects of detector performance that will in turn influence the imaging performance of the DBT system, however some of the test conditions were not given in detail. Parameters assessed included the detector response function, MTF, noise power spectrum (NPS), detective quantum efficiency (DQE), detector ghosting and detector lag. Similar methods were applied by Zhao and Zhao (Zhao and Zhao, 2008a) for the a-Se x-ray detector implemented in Siemens Mammomat Novation^{TOMO} prototype DBT system. Following on from this prototype, Siemens introduced the Siemens Mammomat Inspiration and Revelation DBT systems; the Inspiration unit has been used in breast screening trials (Lång *et al.*, 2015). Again, parameters studied included the MTF, lag and ghosting and the DQE as a function of detector exposure level. Detailed characterization data for four DBT units, including the Selenia Dimensions and the Siemens Inspiration units, have been published by Mackenzie *et al* (Mackenzie *et al.*, 2017).

These are standard detector characterization methods but are generally too involved for implementation in QC and require specific detector readout modes, especially for lag and ghosting tests, that can only be achieved in a design, or system engineering setting. Some of these methods, however, have been

Accepted Manuscript

implemented in the EUREF protocol (van Engen *et al.*, 2018), as they give valuable insight into system performance and are a useful means of tracking changes in system performance for QC purposes.

3.3.1 Detector response function

The first step when evaluating detector performance is to measure the detector response function. The mean pixel value (MPV) at a defined position in the image is plotted against the x-ray signal, quantified using air kerma at the detector input plane (DAK). The reference ROI in the EUREF protocol (van Engen *et al.*, 2018) is 5 × 5 mm centred on a point 6 cm from the chest wall edge and centred between the left and right hand sides of the image. Zero degree mode is available for all current systems and should be used for the response function measurements (Table 1); this was not possible for some earlier system versions (Rodríguez-Ruiz *et al.*, 2016). The images are acquired with a 2 mm Al filter at the x-ray tube, using the tube voltage and A/F setting selected by the system for the standard block (40 mm PMMA). MPV is measured in the first projection to mitigate the influence of ghosting and lag on the pixel values. Lag effects are expected to be low as this is the first frame using manual exposures (no AEC test shot).

Table 3a presents response function curve fit coefficients for DM and DBT modes for 8 systems. All systems have a linear response of the form MPV = A+B×DAK with the exception of the Fuji Innovality where a logarithmic look up table (LUT) was used for export i.e. MPV = A+B×$\log_{10}$(DAK). Note that later software versions allow image export with a linear pixel intensity relationship. The model functions fit the response function data well, with R^2 = 1.000 for all units. A notable feature for all systems is that response function gradient is higher in DBT mode compared to DM mode. The ratio of the gradient for DBT to that for DM mode ranges from 2.5 to 7.4, with an average of 4.7. Rodriguez-Ruiz *et al* (Rodríguez-Ruiz *et al.*, 2016) also found an increase in gain for DBT mode for the Siemens Inspiration and Hologic Selenia Dimensions Systems, but by factors 7 and 15, respectively (Rodríguez-Ruiz *et al.*, 2016). Manufacturers increase system gain in DBT mode such that the signal intensity output from the detector at the lower DAK/projection fills the available range of output levels. It appears that the increase in amplification also moves the signal from the detector further from the noise floor, reducing the relative contribution of electronic noise to overall image noise. This is an important consideration in DBT system design, especially for systems with many projection images per scan. Given the constraints on the total MGD (van Engen *et al.*, 2018), DAK/projection must be kept low and can be ~10 μGy per projection for some systems. This is also a point of concern for manufacturers submitting data for a premarket approval (PMA) at the FDA. As part of this process, physical laboratory testing data are often supplied which show that the NPS follows the exposure level linearly and that quantum noise is the dominant noise source in the projections (*FujiFilm ASPIRE Cristalle SSED, FDA*, 2017). An alternative is to show that the DQE at low DAK levels remains within some percentage of the DQE for DM operation (*Siemens MAMMOMAT Inspiration SSED, FDA*, 2015).

Accepted Manuscript

Table 3a Detector response function fit coefficients for 2D and 3D imaging modes for eight DBT systems. Measured with a 2 mm thick Al sheet at the x-ray tube. (Units for the A coefficients are (arbitrary) digital units (DU); units for the B coefficient are (DU μGy^{-1})). Images acquired in zero degree mode, data taken from the 1st projection. * = Logarithmic response function. (Data shown are results from routine QC programme by UZ Leuven Medical Physics Department; DAK measured using a calibrated RTI Piranha dosimeter (RTI, Mölndal, Sweden.)

	2D mode					3D mode				
System	tube voltage (kV)	A/F	A	B	R^2	tube voltage (kV)	A/F	A	B	R^2
Fuji Innovality*	31	W/Al	69.0	449.2	1.000	31	W/Al	5256	1796	1.000
GE Pristina	26	Mo/Mo	3.06	6.08	1.000	26	Mo/Mo	-0.3	41.92	1.000
GE Pristina	34	Rh/Ag	18.9	9.41	1.000	34	Rh/Ag	-6.8	64.59	1.000
Giotto Class	27	W/Ag	-24.1	4.07	1.000	27	W/Ag	-69.0	18.94	1.000
Hologic Selenia Dimensions	28	W/Rh	52.2	2.51	1.000	30	W/Al	58.9	18.58	1.000
Hologic 3Dimensions	26	W/Al	52.1	2.45	1.000	30	W/Al	57.7	9.94	1.000
Planmed Clarity	30	W/Ag	-141.8	24.52	1.000	30	W/Ag	-182.5	61.50	1.000
Siemens Inspiration	28	W/Rh	57.4	2.91	1.000	28	W/Rh	57.8	8.72	1.000
Siemens Revelation	28	W/Rh	49.4	3.65	1.000	28	W/Rh	45.9	10.83	1.000

Table 3b Measured noise coefficients for DBT mode from a weighted fit to the linearized variance as a function of DAK for eight DBT systems. Two DAK levels are given that indicate where electronic and structure noise dominate the image noise. Also given is the fraction of quantum variance at the clinical DAK level. Measured with a 2 mm thick Al sheet at the x-ray tube. Images acquired in zero degree mode, data taken from the 1st projection. * = 4 mm Al filter used. (Data shown are results from routine QC programme by UZ Leuven Medical Physics Department.)

System	electronic noise coefficient	quantum noise coefficient	structure noise coefficient	electronic noise dominant below (μGy)	structure noise dominant above (μGy)	quantum fraction at clinical DAK
Fuji Innovality	0.000	5.968×10^{-3}	1.416×10^{-5}	-	422	91% at 40 μGy
GE Pristina (26 kV Mo/Mo)	2.750×10^{-2}	5.202×10^{-3}	8.537×10^{-5}	5.3	61	59% at 8.3 μGy
GE Pristina (34 kV Rh/Ag)	2.235×10^{-2}	5.093×10^{-3}	3.46×10^{-5}	4.4	147	~70% at 8.8 μGy
GE Pristina (34 kV Rh/Ag)*	5.732×10^{-3}	4.927×10^{-3}	2.258×10^{-5}	1.2	218	85% at 8.7 μGy
Giotto Class	0.000	1.717×10^{-2}	1.751×10^{-5}	-	980	97% at 19 μGy
Hologic Selenia Dimensions	0.000	1.072×10^{-2}	2.175×10^{-5}	-	493	93% at 36 μGy
Hologic 3Dimensions	1.023×10^{-3}	4.455×10^{-3}	4.181×10^{-5}	-	1065	96% at 45 μGy
Planmed Clarity	0.000	1.519×10^{-2}	1.425×10^{-5}	-	1066	96% at 25 μGy
Siemens Inspiration	4.994×10^{-2}	3.328×10^{-2}	0.000	1.5	-	86% at 10 μGy
Siemens Revelation	0.000	2.512×10^{-2}	3.083×10^{-5}	-	815	98% at 10 μGy

Accepted Manuscript

3.3.2 *Noise decomposition using polynomial model*

A measurement of variance can be made from the response function images, which can then be used for noise analysis (van Engen *et al.*, 2018). By applying a polynomial model to the variance as a function of DAK/projection, some basic information on the type and relative contribution to the total noise is obtained (K. C. Young *et al.*, 2006; Bouwman *et al.*, 2009; Monnin *et al.*, 2014). Three forms of noise are assumed: electronic noise, quantum noise and structured noise (Bouwman *et al.*, 2009; Monnin *et al.*, 2014). The use of data weighting during the curve fit is important as this increases the accuracy and robustness of the curve fit, as demonstrated for DM systems (Monnin *et al.*, 2014). Before performing the calculation, image pixel values are linearized to DAK using the inverse of the response function.

Results of this procedure applied to variance measured in the 1st projection are shown in Table 3b for eight DBT systems. As with the measurement of mean pixel value for the response function, lag and ghosting processes within the detector could influence the measured variance. While this is expected to be a small effect for current systems (Mackenzie *et al.*, 2017), use of the first projection should limit the influence of lag on this measurement. An approximate quantum limited range can be estimated from the fit coefficients to give the DAK/projection below which electronic noise variance forms more than 50% of the total noise and above which structure noise dominates. For five of the units, the electronic noise coefficient is zero, indicating no electronic noise (within the precision of the measurement) and that quantum noise will be the dominant noise at low DAK levels. In fact, electronic noise may be present but the method is not sufficiently sensitive to estimate the magnitude. Also shown is the fraction of variance due to the quantum noise in comparison to the total variance, evaluated at the clinical DAK/projection. This DAK level was taken from linearized MPV measured in an image acquired with a 40 mm PMMA block under AEC. All systems are quantum limited at the clinical exposure level at the detector, with the lowest fraction being ~60%. For most systems quantum variance comprises 85% or higher of the total image variance, consistent with the value of ~90% by Rodriguez-Ruiz *et al* (Rodríguez-Ruiz *et al.*, 2016). Of note are the data for the GEHC Pristina (34 kV, Rh/Ag) measured with 2 mm Al at the x-ray tube. The lowest DAK/projection on the response curve with 2 mm Al at the tube and minimum tube load setting (18 mAs) is ~20 μGy, while the clinical DAK/projection is ~9 μGy, meaning that the quantum variance fraction at clinical DAK has to be extrapolated (~70%). Switching to 4 mm Al for the response function gives a minimum DAK/projection of ~5 μGy and a value of 85% (without extrapolation) for quantum variance fraction at the clinical DAK/projection. If there is a trend for manufacturers to use higher energies for acquisition in DBT systems (Shrestha, Vedantham and Karellas, 2017) then the use of a thicker Al filter, of perhaps 3 mm or 4 mm for the response function may have to be considered.

3.3.3 Detector brightness uniformity

A basic requirement for most imaging systems is that images of a homogeneous test object, e.g. a 40 mm PMMA thick block that completely covers the breast support table, are free from artefacts and have uniform brightness. For FPD based imagers, this is achieved by the application of gain and offset corrections in a

Accepted Manuscript

procedure often referred to as 'flat fielding' (Moy and Bosset, 1999). Although there is some work on corrections for general x-ray imaging systems (Moy and Bosset, 1999; Kwan, Seibert and Boone, 2006; Schmidgunst, Ritter and Lang, 2007) and some literature available for DM detectors (Vedantham *et al.*, 2000; Konstantinidis, Olivo and Speller, 2011) there is no discussion of this for DBT devices. Although the principles will be broadly similar, vendors use proprietary algorithms that will differ to some extent. One of the difficulties of a implementing flat field correction for DBT systems is that DAK/projection is low, resulting in noisy gain correction images, and this noise can propagate into the corrected images (Konstantinidis, Olivo and Speller, 2011). Options include the use of an average of the gain correction images, or acquisition of the correction in 0° mode which is then applied to all projections, a procedure that results in images whose brightness uniformity changes with projection.

DBT projections are generally less uniform than those in DM detectors, where brightness uniformity (BNU) error, defined as the standard deviation of the ROI values divided by the mean value of the uniformity ROIs, is of the order of 2.5% (Marshall, Mackenzie and Honey, 2011). Figure 3 shows contour images of linearized PV for a Siemens Revelation system, calculated for the image of a 40 mm PMMA block (28 kV and W/Rh). The relative position of the source with respect to the x-ray detector varies with projection angle, resulting in BNU errors of 19.9%, 3.3% and 13.8%, for the -25°, 0° and 25° projections, respectively.

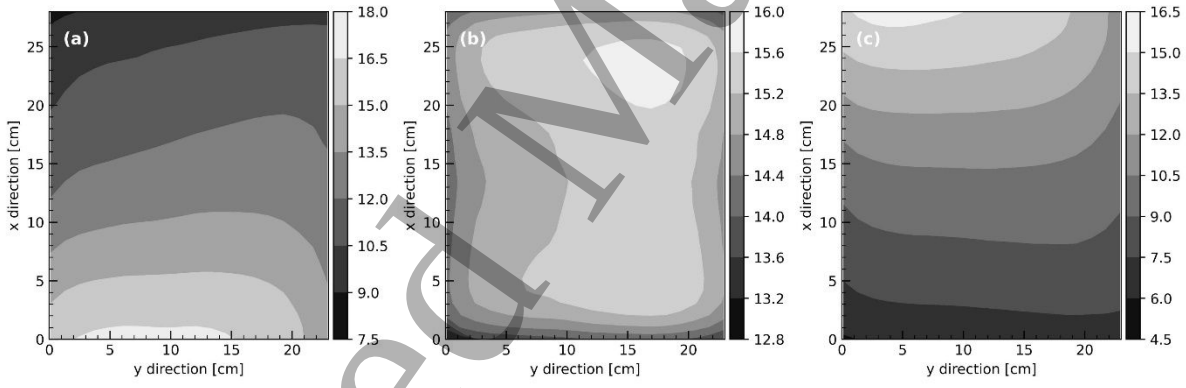


Figure 3. Contour images of linearized PV in homogeneous images for a Siemens Revelation system at +25°, 0° and -25°, for a 40 mm PMMA block imaged at 28 kV and W/Rh. (scale bar units are μGy)

Non-uniformity in the DBT projection images can therefore arise from two main sources: the varying source position for systems where the flat field correction is performed in (stationary) zero degree mode and from scattered radiation (I. Sechopoulos *et al.*, 2007), as most DBT systems do not have an antiscatter grid. Rather than assess non-uniformity in the projection images, the EUREF protocol specifies a test of homogeneity of the reconstructed planes (van Engen *et al.*, 2018) using images of the standard test block, which currently has a thickness of 45 mm and should be larger than the x-ray beam by 5 mm on all sides. Non-uniformity due to scattered radiation will depend on object thickness and a more comprehensive test may examine non-uniformity at different simulated breast thicknesses. A second consideration is that correction of non-uniformity at the reconstruction stage could involve a segmentation step which may

Accepted Manuscript

require a phantom with that simulates a breast outline in order to function correctly. Given these points, evaluation of uniformity error in the projections may be limited to applications such as system modelling, and should not be performed as part of routine QC.

3.3.4 Lag and Ghosting

Lag and ghosting are the results of temporal processes occurring within the x-ray detector (Siewerdsen and Jaffray, 1999), and can adversely affect the quality of DM (Bloomquist, Yaffe, Mawdsley, *et al.*, 2006) and DBT images (Zhao and Zhao, 2008a; Mainprize, Wang and Yaffe, 2009). Detailed modelling of lag in DBT systems can be found in Mainprize *et al* (Mainprize, Wang and Yaffe, 2009) and in Sanchez (Sánchez de la Rosa, 2019), where the analysis is extended to include contrast enhanced DBT imaging. Lag results from the release of signal over a time period longer than the frame integration time, which then appears in later images. One of the main sources of this lag is charge trapping, where electrons are retained in defects ("traps") within the detector material bulk and released over time. Also contributing to lag is scintillator afterglow or persistence and charge transfer, which occurs as the charge is read out (Sánchez de la Rosa, 2019). Ghosting also results from charge trapping and occurs when the sensitivity of the detector is changed due to previous exposures. If traps within the detector material remain filled after an x-ray exposure, this can lead to a higher signal at subsequent readout .

Cheung *et al* (Cheung *et al.*, 2005) measured lag and ghosting data for the early Hologic detector, reporting a ghosting level of <1% at 10s after an initial detector air kerma of 1.54 mGy to a 4.2 cm PMMA block that covered half the detector. The detector was then read out at 0.046 mGy/frame with the 4.2 cm PMMA fully covering the detector i.e. a differential air kerma of 1.49 mGy. Lag was measured using a 1 mm thick piece of Pb to block exposure the detector. Following an exposure of 1.43 mGy, lag was found to be <0.6% after 0.5s. Similar methods were applied for the a-Se x-ray detector implemented in Siemens Mammomat Novation^{TOMO} prototype DBT system (Zhao and Zhao, 2008a). First frame lag was ~4% for a frame readout interval of 0.8 s at full resolution, while ghosting was negligible.

Measurement of lag and ghosting in a QC setting is not trivial, as these effects are difficult to separate and depend on the exposure history of the detector in the preceding minutes or even hours. A measurement of ghosting is included in EUREF DM protocol (European Commission, 2013) in which the detector is half covered by the standard test block (eg 45 mm PMMA) and an image acquired at a typical patient exposure level. The detector is then fully covered by the block and a further image acquired with the test block plus a 0.1 mm thick Al sheet. The ghost factor is then estimated from MPVs measured in the regions that respectively had large and low detector exposure levels, related to the contrast of the Al sheet, and the result compared to a limiting value of 0.3. The extension of this to DBT QC testing is described by Mackenzie *et al* (Mackenzie *et al.*, 2020), where a number of options were examined. The DM version of the ghost test method was applied, but between successive DBT scans rather than DM projections. Next, a measure of the ghosting between DBT projections was obtained by moving the PMMA block from covering half the detector to covering the whole detector after the first projection. Finally, a method in which the 0.1 mm Al

Accepted Manuscript

sheet is imaged in conjunction with a 3 mm thick Al sheet (to simulate typical breast attenuation) was positioned in the compression plate, at a height of 100 mm above the breast support table. A number of difficulties were noted, including the fact that brightness uniformity can change as a function of projection, due to the flat field correction employed on some systems (Section 3.3.3). This can influence the measured ghost factor values. Quantification of ghosting between projections may be valuable at the Type Testing stage (European Commission, 2013), as ghost factor values between DBT projections were higher than in DM images (Mackenzie *et al.*, 2020). Further work on the need for a practical QC test for the assessment of ghosting is required, given that there is currently little evidence for the presence of ghosting (or lag) in studies reporting clinical data. One aspect to consider is the use of "Combo" imaging, a single high dose DM can be acquired directly before the DBT scan, which could induce some ghosting. It may be sufficient to perform the standard DM ghost test to assess this.

3.4 Sharpness: detector and system

A sharp image is one in which spatial detail travels largely unimpeded from the original scene to the final image. The sharpness influences the (spatial) resolution of a detector or system and its ability to separate or resolve adjacent features in the image. Sharpness is generally assessed under noise-free or close to noise-free conditions, while the ability of an imaging system to resolve adjacent details also depends on the level of noise present. This leads on to the use of signal to noise ratio (SNR) when describing imaging performance which is discussed later; this section focuses on sharpness evaluation, without considering noise for the moment. The most common means of assessing the sharpness of a medical imaging system or component is through the MTF (Rossmann, 1969; Metz and Doi, 1979). Some basic MTF theory is first presented and followed by a description of MTF used to quantify detector sharpness and DBT system sharpness from projection images i.e. excluding the reconstruction algorithm. Evaluation of DBT system sharpness in the reconstructed planes is then presented, before discussing some alternative methods of sharpness assessment.

3.4.1 Basic MTF theory for DBT projection images

Excellent references are available on the theory of MTF applied in radiological imaging (Rossmann, 1969; Metz and Doi, 1979; Cunningham, 1992). Assuming that the components or sub-systems exhibit linearity (are linearizeable, or have small signal linearity), and are shift invariant to some extent, then the total system MTF is the product of the MTFs of the sub-components (Metz and Doi, 1979). For the simplified case where the x-rays are normally incident on the sensitive x-ray layer, the MTF of a DBT system can be written as (Giger and Doi, 1984):

$$MTF_d(f_x,f_y) = MTF_{geom}(f_x,f_y)MTF_{conv}(f_x,f_y)MTF_{aperture}(f_x,f_y)\Sigma_{k_x=-\infty}^{\infty}\delta\left(f_x-\frac{f_x}{\Delta x}\right)\Sigma_{k_y=-\infty}^{\infty}\delta\left(f_y-\frac{f_y}{\Delta y}\right) \quad (2)$$

where $MTF_{geom}(f_x,f_y)$, $MTF_{conv}(f_x,f_y)$ and $MTF_{aperture}(f_x,f_y)$ are MTFs due to source, converter and pixel aperture, respectively. The sum of delta functions represents the sampling action of the pixel matrix. The product of the source, detector and pixel aperture is defined as the presampling MTF ($MTF_{pre}(f_x,f_y)$):

Accepted Manuscript

$$MTF_{pre}(f_x,f_y) = MTF_{geom}(f_x,f_y)MTF_{conv}(f_x,f_y)MTF_{aperture}(f_x,f_y) \quad (3)$$

The term $MTF_{geom}(f_x,f_y)$ can be written in terms of the blurring due to x-ray focus size (g_{fs}) with associated $MTF_{fs}(f_x,f_y)$, and x-ray focus motion (g_{fm}) with associated $MTF_{fm}(f_x,f_y)$ (Metz and Doi, 1979; Zhou, Zhao and Zhao, 2007):

$$MTF_{geom}(f_x,f_y) = MTF_{fs}(f_x,f_y)\, MTF_{fm}(f_x,f_y) \quad (4)$$

$$MTF_{pre}(f_x,f_y) = MTF_{fs}(f_x,f_y)\, MTF_{fm}(f_x,f_y)\, MTF_{conv}(f_x,f_y)MTF_{aperture}(f_x,f_y) \quad (5)$$

The product of the detector converter MTF with the pixel aperture MTF is the standard x-ray detector presampling MTF i.e. $MTF_d(f_x,f_y)$, measured for example as part of QC for DM systems. The presampling MTF for a DBT system is then the product of the two geometric MTFs and the detector MTF:

$$MTF_{pre}(f_x,f_y) = MTF_{fs}(f_x,f_y)\, MTF_{fm}(f_x,f_y)\, MTF_d(f_x,f_y) \quad (6)$$

The projected size of the focus due to motion, for example, at a distance d_1 from focus, and height d_2 above the detector is given by :

$$a_{fm} = \frac{d_2}{d_1} g_{fm} \quad (7)$$

The size of focus due to motion can be calculated from knowledge of the total distance travelled by the focus during the DBT scan, and the exposure time for a projection image expressed as a fraction of the total exposure time. It is clear that x-ray tube output can play an important role in limiting geometric blurring due to focus motion, as discussed by Shrestha *et al* (Shrestha, Vedantham and Karellas, 2017). Considering just the 1D case and assuming a rectilinear shape for the focus, sinc functions can be used to model the focus MTFs (Marshall and Bosmans, 2012):

$$MTF_{geom}(f_x) = \frac{|\sin(\pi f_x a_{fs})|}{\pi f_x a_{fs}} \frac{|\sin(\pi f_x a_{fm})|}{\pi f_x a_{fm}} \quad (8)$$

The above is a simplified estimation of source blur from a moving tube, with the assumption of a static focus incident on the anode; a more detailed calculation of source PSF is available in Zheng *et al* (Zheng, Fessler and Chan, 2019).

3.4.2 X-ray detector presampling MTF using projection images

For an explicit evaluation of detector presampling MTF in DBT mode, the geometric blurring should be minimized by setting MTF_{geom} ~1.0. To do this, the edge should be positioned on the breast support table, close to the x-ray detector and images acquired in 0° mode i.e. without focus motion (IEC, 2020a). Practically, a radio-opaque edge test object is most often used for MTF measurement in projection images (Samei, Flynn and Reimann, 1998; Buhr, Günther-Kohfahl and Neitzel, 2003). The paper by Carton *et al* (Carton *et al.*, 2005) discusses aspects specific to mammography. The EUREF protocol (van Engen *et al.*, 2013) specifies a 0.8 mm thick 120×60 mm Fe edge for MTF measurement for DM systems however an explicit measurement of detector presampling MTF is not specified in the EUREF DBT protocol (van

Accepted Manuscript

Engen *et al.*, 2018). Note that the DBT QC protocol would be applied in conjunction with the DM protocol and thus a measurement of detector MTF would already have been performed; a potential difference could be that binning is applied in DBT mode for some systems. Extensive analysis of longitudinal MTF data measured for 55 DM systems (16 different models) as part of a QC programme is available in the study of Wigati *et al* (Wigati *et al.*, 2021).

3.4.3 DBT system presampling MTF using projection images

In order to assess the system MTF, equation (6) is used and the geometrical blurring MTF is included by positioning the edge at clinically relevant heights above the breast support table and using a moving x-ray tube. The exposure time for a projection image can have a large influence on geometric blurring and may depend on the x-ray factors and therefore on breast thickness. The relevant x-ray factors to set for this measurement must be established by assessing the AEC as a function of breast equivalent thickness. Exposure time can be measured directly for a given thickness (i.e. tube voltage and load (mAs)) using a dosimeter that can record exposure time, or from the DICOM header (once verified as accurate). For systems where the source is stationary during the acquisition of projection images, MTF_{fm} is taken to be 1.0.

Figure 4 plots DBT system MTF curves measured with an edge test object positioned at 40 mm above the breast support table, for the tube scan direction and orthoscan directions, following the EUREF protocol. These curves were measured in the 0° projection images of a moving tube scan, using 2 mm Al at the x-ray tube and x-ray factors set manually to match the factors selected by the system AEC for 40 mm PMMA. For longitudinal testing, assessing the 0° projection is probably sufficient to identify changes in blurring within the system. Additional system MTF measurements could be performed as part of the Commissioning tests for non-zero projection angles in order to characterize blurring due to oblique entry. To provide some context with respect previous technology, presampling MTFs curves for DM imaging using an Anrad a-Se based system and the averaged MTF curve for powder CR detectors (Marshall *et al.*, 2011) are shown. The MTF curves for the DBT systems generally lie in the range between those for powder CR and a-Se photoconductor technology. Additional scan information is given in Table 4, along with the spatial frequencies for the points where MTF reaches 50% ($MTF_{50\%}$) and 10% ($MTF_{10\%}$), respectively.

Accepted Manuscript

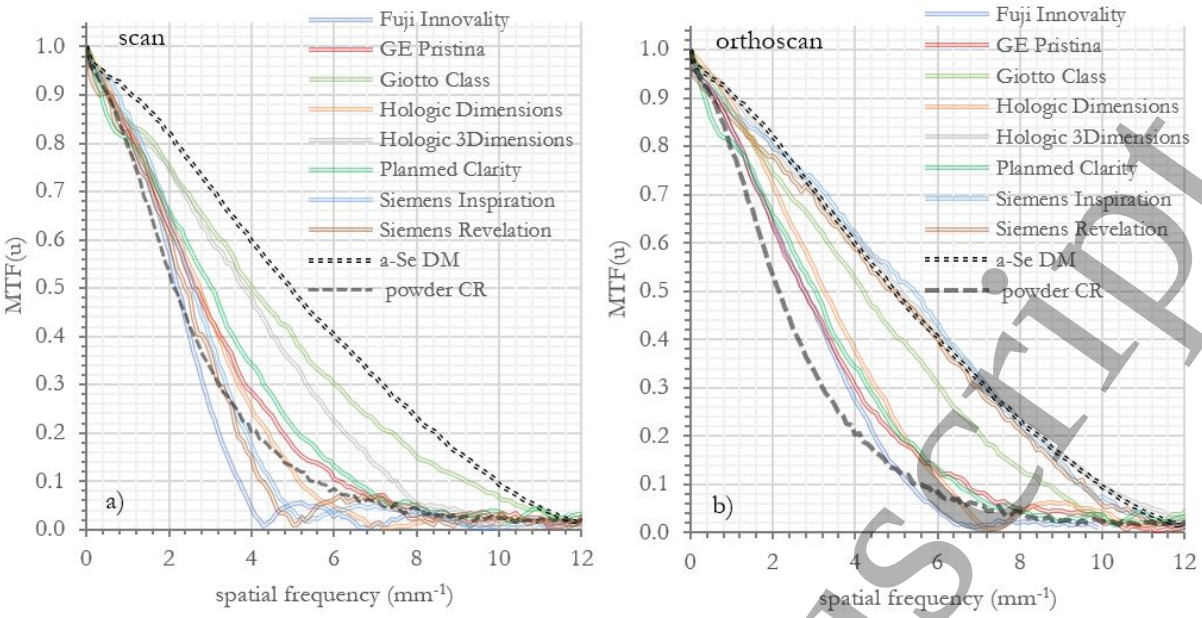


Figure 4. DBT system presampling MTF curves measured in the 0° projection for the scan (a) and orthoscan (b) directions with an edge test object positioned at 40 mm above the breast support. Acquired with tube voltage, A/F and load used by the system for 40 mm PMMA under AEC control. Moving tube scan. (Results taken from routine QC programme by UZ Leuven Medical Physics Department.)

Table 4 Tube load (mAs), total scan time, exposure time per projection, focus size due to tube motion, and spatial frequency at $MTF_{50\%}$ and $MTF_{10\%}$) for the orthoscan and scan directions, measured using an edge test object positioned 40 mm above the breast support table.

	mAs	Total scan time (s)	Exposure time (ms)	Focus motion size (mm)	$MTF_{50\%}$ (mm^{-1})		$MTF_{10\%}$ (mm^{-1})	
System					Scan	orthoscan	scan	orthoscan
Fuji Innovality ST	37.2	3.50	45	1.54	2.20	2.70	3.70	5.30
Fuji Innovality HR	60.4	9.39	56	1.92	-	-	-	-
GE Pristina	25.4	8.90	73	0.00	2.60	2.70	6.10	5.30
Giotto Class	119.0	9.20	88	0.00	4.00	4.20	9.10	8.50
Hologic Selenia	53.0	3.55	16	0.83	2.70	3.20	5.10	6.10
Hologic 3Dimensions	50.0	3.75	16	0.79	3.80	5.10	7.50	9.90
Planmed Clarity	119.0	18.00	152	2.67	3.00	3.00	6.40	6.40
Siemens Inspiration	152.0	21.20	96	2.40	2.50	5.40	4.40	9.40
Siemens Revelation	141.0	21.70	84	2.05	2.40	4.90	4.20	9.50

The GEHC Pristina and Giotto Class are step and shoot systems, and MTF curves for the scan and orthoscan directions are similar, as expected for isotropic blurring. MTF is higher for the Giotto unit, a result of the higher intrinsic sharpness of the a-Se layer compared to the CsI phosphor (Hajdok, Battista and Cunningham, 2008). For the continuous focus motion systems, presampling MTF in the orthoscan direction is not affected by focus motion blurring. The a-Se based DBT systems that use unbinned projections have a presampling MTF that remains close to that of a typical a-Se detector. As expected, MTF is reduced for a-Se based systems that use binned projections (Hologic Dimensions and Fuji Innovality); in fact these lie close to the MTF curve found for the GEHC Pristina, which uses a CsI based FPD. In the

Accepted Manuscript

scan direction, there is a general reduction in DBT MTF for the continuous focus motion systems, due to the additional geometric blurring from the extended source size. Of these system types, the Hologic Selenia 3Dimensions has the highest MTF curve, because of the unbinned projections, a short pulse length of 16 ms per projection and corresponding focus motion size of ~0.8 mm. The standard Hologic Selenia Dimensions uses ~16 ms for a 40 mm PMMA thickness, however detector binning reduces the MTF, which is similar to that for the Pristina and the Siemens Revelation and Inspiration systems. The Planmed Clarity uses an acquisition method called Sync-and-Shoot, in which the x-ray tube moves continuously during the acquisition but while a projection image is being acquired, the table, detector and breast tissue rotate to follow the x-ray source as x-rays are emitted. At the end of each exposure, the table returns to the starting point, ready for the next exposure (Jousi *et al.*, 2019). This results in an MTF curve that lies slightly above that of the GEHC Pristina in the scan direction. As predicted by Zheng *et al* (Zheng, Fessler and Chan, 2019), rotating the detector in synchrony with the source during exposure does not ameliorate focus motion blurring.

In the initial stages of DBT system development, a number of studies used the MTF to study how the different acquisition steps influence the final MTF. Ren *et al* (Ren *et al.*, 2005) showed that both 2×2 detector binning and motion of the source during acquisition cause a degradation of MTF for a prototype version of the Hologic Dimensions system. In this early system, total exposure time was 800 ms delivered over 11 projections i.e. 73 ms per projection, in which the source moved ~1.0 mm. In the study of Zhao *et al* (Zhao *et al.*, 2009), a prototype Siemens unit had values for focus motion per projection of 0.65 mm and 1.15 mm for un-binned and 2×1 binned operation, respectively. While earlier versions of the Siemens Inspiration used a pulse length of ~150 ms per projection (Marshall and Bosmans, 2012), both the Siemens Inspiration and the newer Revelation units now keep the pulse length at ~90 ms for most breast thicknesses. For the continuous tube motion systems, the extended focus is shortest for the Hologic systems at ~0.8 mm, followed by the Fuji Innovality in ST mode, at 1.54 mm. For the remaining continuous motion systems, focus motion length is 2.0 to 2.5 mm, with a pulse length in the region of 90 ms.

It is instructive to examine the relative contributions of the x-ray detector, focus size, focus motion and detector binning to the total system MTF (Ren *et al.*, 2005; Zhao and Zhao, 2008b). Figure 5 shows MTF curves for a Hologic Selenia Dimensions system where sinc functions (e.g. equation (8)) have been used to model the resultant MTF for focus motion size (21 ms pulse length, 1.04 mm focus travel), and binning (2×2 binning of 70 μm pixels to 140 μm), for an object positioned 40 mm above the breast support table. Also shown are the focus size MTF (0.42 mm focus) and the expected oblique entry MTF calculated according to Que and Rowlands (Que and Rowlands, 1995) for an a-Se detector for an angle of 7.5°. These data show that binning and focus motion have a strong influence on the final system MTF. Note that the Hologic 3Dimensions system has a newer detector that is operated without binning.

Accepted Manuscript

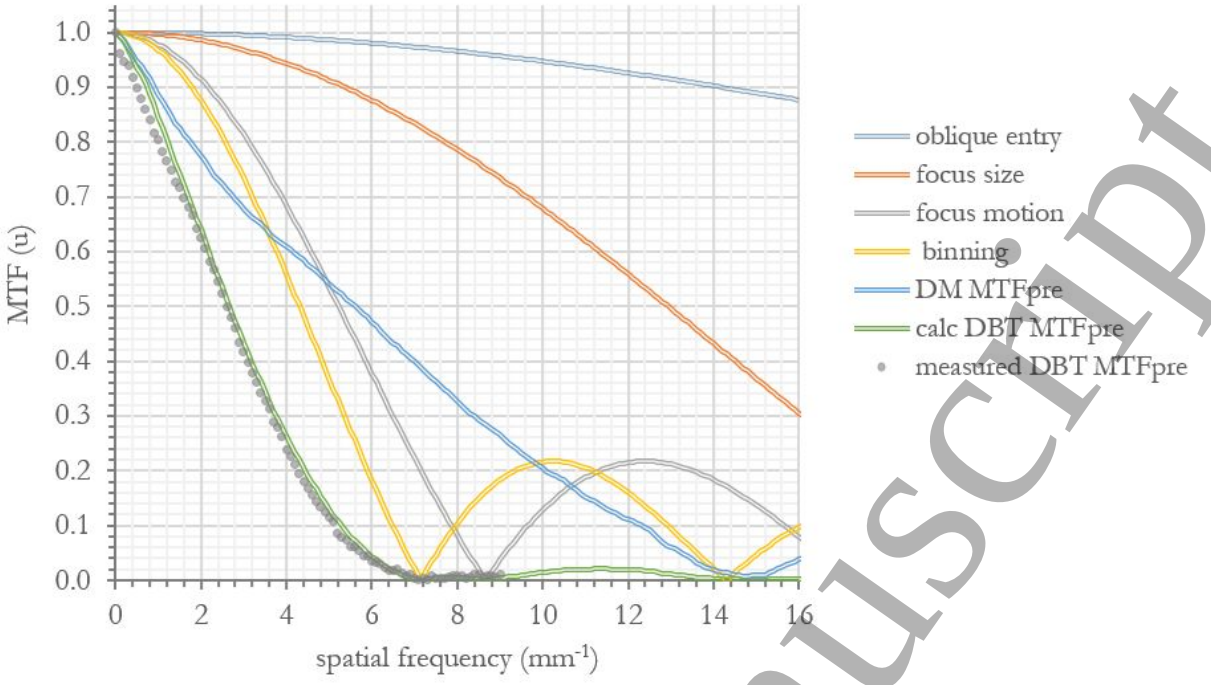


Figure 5 Theoretical MTF curves for the various components of a Hologic Selenia Dimensions DBT system, together with the final calculated and measured system MTF measured at tube projection angle of -7.5° using an edge positioned 40 mm above the breast support table.

The MTF and focus length data in Figure 4 and Table 4 are consistent with detailed results presented by Mackenzie *et al* (Mackenzie *et al.*, 2017), who measured MTF and NPS for the purpose of DBT system modelling. In addition to MTF measurements made for the central projection, Mackenzie *et al* also determined MTF in the first projection and compared this with central projection data. $MTF_{50\%}$ was found to be lower by 18% and 28% for the Fuji Innovality HR and Siemens Inspiration units, respectively, while very little difference was found for the 15° scan systems.

We can further examine this for a wide and narrow angle system in Table 4 – the Siemens Revelation and the Hologic Dimensions. Four scans were acquired on each system, with the edge supported 4 cm above the breast support platform and re-positioned between scans such that the edge produced low-to-high and high-to-low signal transitions in the tube motion direction. MTF curves were calculated in a region ~6 cm from the chest wall edge, averaged from four scans and the $MTF_{50\%}$ and $MTF_{10\%}$ spatial frequencies established. These values are plotted in Figure 6a while Figure 6b plots $MTF_{50\%}$ and $MTF_{10\%}$ relative to the value for the 0° projection for the wide angle system. The magnitude of $MTF_{50\%}$ and $MTF_{10\%}$ for the 0° projection are consistent with the data in Table 4 for the two systems. For both systems, the MTF stays within ~3% of the value for the 0° projection over the 'narrow angle' range of ±7.5°. In fact, for the wide angle system, $MTF_{50\%}$ and $MTF_{10\%}$ remain with 10% of the value at the central projection for an angular range of 30° (±15°). Only in the outer projections, covering -15° to -25° and 15° to 25° is there a greater reduction in MTF, reaching ~20% to 25%, a figure consistent with the data of Mackenzie *et al* (Mackenzie *et al.*, 2017).

Accepted Manuscript

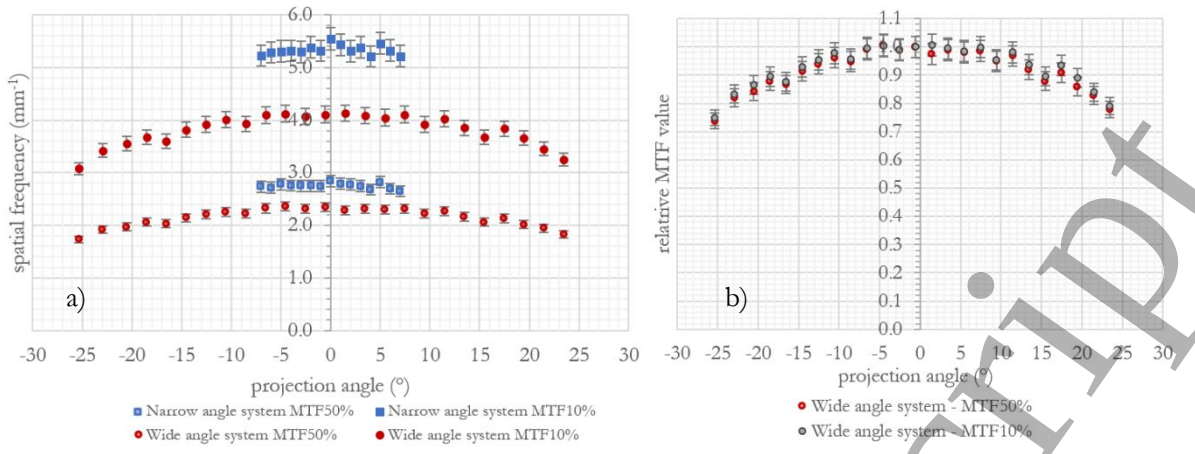


Figure 6 a) $MTF_{50\%}$ and $MTF_{10\%}$ plotted as a function of projection angle for systems with wide angle and narrow geometries b) $MTF_{50\%}$ and $MTF1_{0\%}$ relative to the value for the 0° projection, for the wide angle system. Error bars represent standard deviation over four independent measurements.

This reduction at the wider angles is due to oblique entry of the x-ray beam into the x-ray sensitive layer of the detector at the wider angles and has been studied by a number of authors, analytically (Que and Rowlands, 1995; Mainprize *et al.*, 2006; Freed, Park and Badano, 2010; Acciavatti and Maidment, 2011; Badano and Freed, 2011), via Monte Carlo methods (Badano *et al.*, 2007) and by measurement (Mainprize *et al.*, 2006; Freed *et al.*, 2009). The reduction in MTF due to oblique entry is spatial frequency dependent. The typical spatial frequency for $MTF_{50\%}$ lies between 2.5 mm^{-1} and 5 mm^{-1} for the Siemens and Fuji HR systems, depending on scan direction. Mainprize *et al* (Mainprize *et al.*, 2006) measured reductions in MTF of approximately 5% and 15% for angular incidences of 20° and 30°, respectively for a CsI detector at 4 mm^{-1}. Similar figures were found by Zhao *et al* (Zhao *et al.*, 2017) for a CMOS active pixel sensor with a 50 μm pixel pitch (DynAMITe) coupled to a CsI scintillator.

The principal physical effect leading to sharpness loss in CsI and a-Se layers from oblique entry is the geometry of the path along which the primary x-ray photons and secondary particles deposit energy (Que and Rowlands, 1995; Badano and Freed, 2011), which leads to an increase in the size of the point spread function. Also of importance for a-Se detectors is the size of the electron cloud created by the primary photoelectrons (Que and Rowlands, 1995). From analytical studies of turbid granular phosphors (Acciavatti and Maidment, 2011) and from Monte Carlo simulation of a-Se detectors (Hajdok and Cunningham, 2004), only a small influence on the NPS is expected from oblique entry. However, the reduction in MTF will reduce DQE which in turn will influence the SNR of images, and thus task performance. The consequences for imaging performance are difficult to predict as many other factors also influence the final image; focus motion, for example, causes a far larger reduction in $MTF_{50\%}$ compared to oblique entry, even for the current wide angle systems. As with detector binning, the physical effect on signal transfer associated with oblique entry occurs at the x-ray detector, independent of object position relative to the detector while focus motion blurring is a geometric effect whose severity varies with object height (Zhou, Zhao and Zhao, 2007).

Accepted Manuscript

In the EUREF protocol, system MTF data are acquired with the edge on the breast support and at 4 cm and 7 cm above the breast support to characterise system sharpness in the projection images as breast thickness changes. X-ray acquisition factors used for these measurements are the AEC settings for 2 cm, 4 cm and 7 cm PMMA.

3.4.4 DBT system MTF in the reconstructed planes

The MTF can be assessed at two points in the imaging chain. Access to 'For Processing' type projection images allows a measurement of presampling MTF as discussed previously, which gives information on the blurring present in the DBT projection images prior to reconstruction. MTF can also be measured in the reconstructed image stack giving a total system MTF which also includes the influence of image reconstruction on image sharpness. A number of methods have been used to do this. Zhao *et al* (Zhao *et al.*, 2009) used a 0.2 mm Al sheet supported at 40 mm above the breast support table for the prototype Siemens system. No additional filtration was at the x-ray tube and the edge transition between the Al sheet and air was used to generate the edge spread function (ESF) from which the MTF was calculated. In this method, a semi-transparent edge is imaged which produces limited contrast in the projection images such that the logarithmic transformation applied to the projection images before reconstruction can be regarded as linear (Zhao *et al.*, 2009). A number of other authors have since applied the semi-transparent edge method using 0.2 mm Al together with a 15 mm or 20 mm PMMA sheet (Marshall and Bosmans, 2012; Maldera *et al.*, 2017). Monnin *et al* (Monnin *et al.*, 2020) have extended this method to use a circular Al disc of diameter 50 mm and thickness 0.2 mm, which is imaged with 40 mm PMMA plus an additional 0.5 mm Al sheet, at a height of 20 mm above the breast support (see Figure 7). A total of 180 radial ESFs are generated by plotting PV as a function of distance from the centre of the disc over a 4° angular aperture and overlapping by 2° to reduce noise. The 2D angular MTF is formed by interpolating the angular 1D MTFs onto a grid, from which the radially averaged MTF is calculated. Monnin *et al* (Monnin *et al.*, 2020) present 2D and 1D radially averaged system MTF curves for six different DBT systems, and the method demonstrated an increase in system MTF for the latest Fuji IST iterative reconstruction over the older FBP-based algorithm. The Selenia 3Dimensions had improved MTF compared to the standard Dimensions unit.

Accepted Manuscript

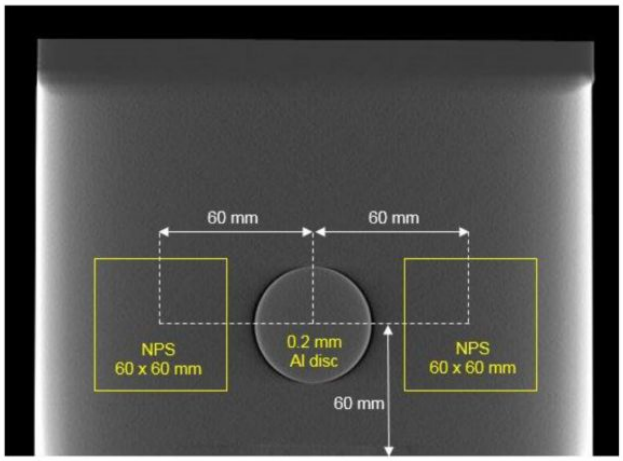


Figure 7 In plane image of the non-prewhitening model observer with eye filter (NPWE) phantom with the calculation areas. (Taken from Monnin *et al* (Monnin *et al.*, 2020))

An alternative to the semi-transparent edge is to use a thin radio-opaque wire (Cunningham, 1992). Marshall and Bosmans (Marshall and Bosmans, 2012) describe practical aspects of using a 25 μm diameter W wire to assess Siemens Inspiration and Hologic Selenia Dimensions systems. Some later studies have applied the wire method to a wider range of DBT units (Rodríguez-Ruiz *et al.*, 2016; Sundell *et al.*, 2019). Rodríguez-Ruiz *et al* explored the use of W wires of diameter 25, 50 and 100 μm, imaged at 45 mm above the breast support and found that the 25 μm wire gave consistently higher MTF results. Sundell *et al* (Sundell *et al.*, 2019) imaged a 25 μm wire with 50 mm PMMA, with the wire positioned at 40 mm from the breast support. The EUREF protocol includes the in-plane MTF measured with a 25 μm diameter W wire as an optional test.

Some practical aspects apply to both MTF types i.e. measurement in the projections and in the reconstructed planes. The edge must be supported at some relevant height above the table and held stationary and vibration free during the scan. Given the low contrast of the edge, image noise can prove a problem and standard means of controlling this include some local smoothing or conditioning of the ESF (Samei, Flynn and Reimann, 1998; Maidment and Albert, 2003). These methods must applied with caution, as the ESF or LSF may contain features such as strong undershoot or overshoot, characteristic of some reconstruction filters. For the wire method, the LSF is generated directly (without differentiation) and non-uniformity in the background containing the wire must be controlled as this biases the MTF at low spatial frequencies (Schafer, Wiegert and Bertram, 2007). A problem associated with thin test objects such as a wire or edge is that these objects will in general not be positioned physically at the exact centre between two reconstructed planes (Zhao *et al.*, 2009), leading to some potential phase (position) dependence in the results. Furthermore, if the edge does not remain parallel to the reconstructed planes (and to the detector plane, typically) then the MTF will also exhibit some position dependence. For example, there is a ~2° gradient between the breast support table and detector in the orthoscan direction for the Siemens Inspiration and Revelation systems. If a wire running in the orthoscan direction is parallel to the breast support then this wire will not remain in a single plane and this could influence the measured in-plane MTF.

Accepted Manuscript

The wire and associated test object must have sufficient contrast to survive any thresholding or segmentation step that may be applied by the reconstruction algorithm, as this can remove region containing the wire (Marshall and Bosmans, 2012)

DBT systems produce what is essentially a volumetric (3D) dataset that has non-isotropic resolution, with small reconstructed pixel size in-plane and typically a 1 mm spacing between planes. Thus the system MTF is 3D in nature and techniques adapted from CT or CBCT system evaluation have been applied. Thornton and Flynn (Thornton and Flynn, 2006) described the use of a sphere to measure the MTF of multi-slice CT scanners. The method used a Teflon sphere of diameter e.g. 25 mm, embedded in curable silicon. The interface of the sphere and the silicon background approximate a step function in a given direction. The centre of mass of the sphere is found and used to assemble surface spread functions by binning voxels that lie within a certain angle, for a given direction. Respective MTF curves are then calculated from the differentiated surface spread functions. This method offers a number of advantages over using a thin wire: there are no phase effects because of the oversampling and the method is relatively insensitive to position, a large number of voxels are used giving a good SNR and MTF data can be produced for all directions from a single acquisition. Samei *et al* (Samei, Murphy and Richard, 2013) applied this method to a Siemens Inspiration DBT system using both simulation and measurement via a 19.05 mm diameter sphere suspended in a vessel containing vegetable oil. Compared to CT imaging, DBT datasets are reconstructed from a limited number of projections acquired over a limited angle and as a result, DBT images can suffer from noise and long tails (artefact spreading) due to the large cone angle. A number of authors have used iterative methods to overcome these difficulties. Chen and Ning (Chen and Ning, 2004) imaged a 6 mm diameter plastic sphere from which edge profiles were derived. A Gaussian model was assumed for the PSF and iteratively applied to a sharp rectangular profile derived from the sphere such that the resultant profile matched that measured on the system. The paper of Lee and Baek (Lee and Baek, 2015) extended this method to measure spatially varying 3-D MTFs by modelling FDK reconstruction locally, calculating the plane integrals of an ideal sphere and comparing these to images of (real) reconstructed sphere phantoms. Directional plane spread functions were estimated using a Richardson-Lucy (RL) deconvolution method. To cope with the anisotropic resolution of DBT systems when using the sphere method to estimate in-plane MTF, Lee and Baek (Lee and Baek, 2017) took the Fourier transform of the reconstructed sphere phantom and divided this by the Fourier transform of the ideal sphere phantom, and then performed the plane integral along the direction orthogonal to reconstructed planes (f_z direction). Figure 8 shows the sphere phantom used with the experimental CBCT system while figure 9 shows the volumetric image of the reconstructed sphere along with the x-y, x-z and y-z planes through the centre of the sphere (Lee and Baek, 2017). Lee and Baek describe an updated method with more robust Richardson-Lucy deconvolution, that estimates the full, spatially varying 3D MTF (Lee, Song and Baek, 2020). These methods derived from CT and CBCT evaluations have yet to be implemented in DBT evaluation and QC programmes.

Accepted Manuscript

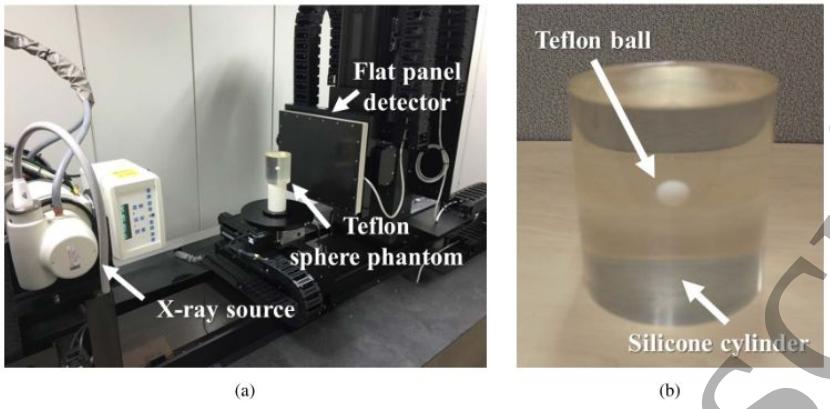


Figure 8 a) The bench-top experimental system. B) Teflon sphere phantom. (Taken from Lee and Baek (Lee and Baek, 2017))

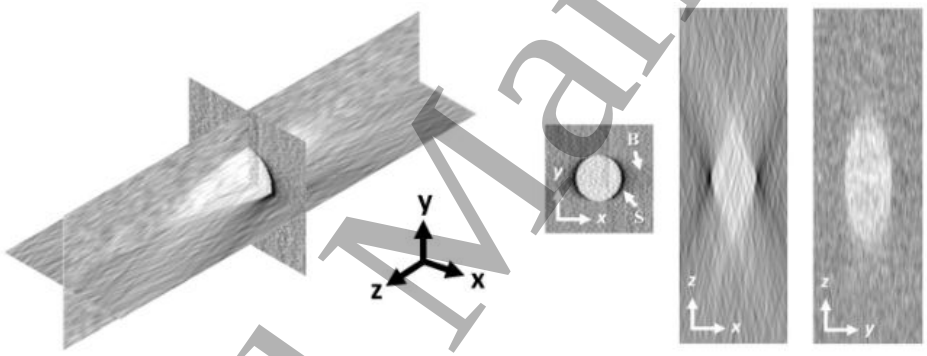


Figure 9 Detrended reconstruction volume image, x-y image, x-z image, and y-z image of the Teflon sphere phantom without the ST filter using 15 mA source current. Unit of the reconstructed values is cm^{-1}, and the display window is [-0.2 0.1]. Shadow artifact and blurring shown in in-plane images are denoted by 'S' and 'B', respectively. (Taken from Lee and Baek (Lee and Baek, 2017))

Common to many of the MTF methods described above is the use of a slanted test object, designed to eliminate the presence of aliasing by increasing the sampling rate for the ESF or LSF. Any modulation present in the presampling MTF curve above the Nyquist frequency indicates the presence of aliasing but also indicates the potential for super-resolution in DBT systems (Acciavatti and Maidment, 2012). Super-resolution arises as a result of the oblique x-ray incidence projecting the image of an object at different positions with respect to the array of detector elements, as the projection angle is changed, resulting in an increased sampling rate. Consequently, frequencies aligned parallel to the direction of source motion will be subject to super-resolution. Reconstruction using in-plane pixel spacing smaller than the detector native pixel spacing is required in order to produce DBT planes that can support super-resolution. Vent *et al* (Vent, Acciavatti and Maidment, 2021) have proposed the Fourier distortion metric (FSD) for tomosynthesis systems, calculated as a means of quantifying the degree of aliasing present. The FSD is calculated from

Accepted Manuscript

one quadrant of the image of a star resolution pattern: tangential profiles are extracted at decreasing radii, starting at the radius (lowest frequency, 1.5 lp/mm) until the centre is reached (highest frequency, 16 lp/mm). To generate the FSD graph, the normalized 1D FFTs of the profiles are plotted as a function of star pattern frequency (lp/mm). The FSD is explicitly designed to quantify the harmonics, spectral leakage, and aliasing present in DBT planes, factors to which the MTF method is insensitive (Vent, Acciavatti and Maidment, 2021).

Measurement of the standard, in-plane reconstructed MTF is an option in the EUREF protocol where the wire method is described. A 25 μm W wire held within a 15 mm PMMA sheet is supported such that the wire is 40 mm above the breast support and imaged using the x-ray factors selected by the system under AEC control for a 40 mm PMMA block. Acquisitions are made with the wire running parallel and perpendicular to the scan direction, and the standard clinical reconstruction is used. Projection image MTF curves are normalized to the zero frequency bin as this has the highest value, however this is generally not the case for DBT reconstruction images and it is common to normalize to maximum of the curve (Zhao *et al.*, 2009).

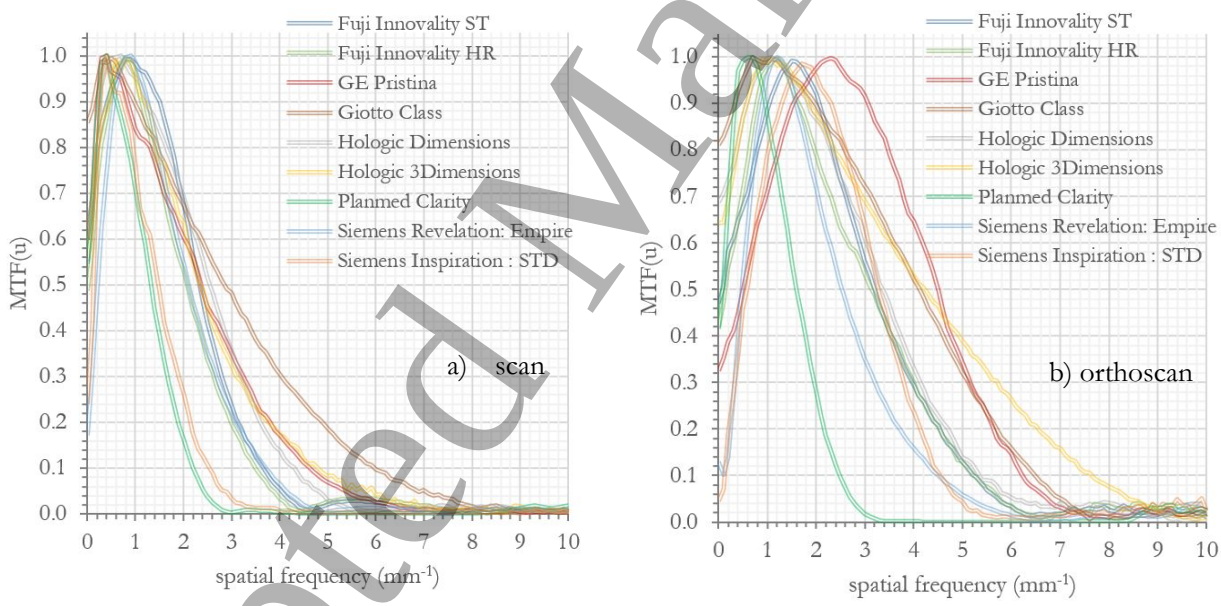


Figure 10. DBT MTF measured in the scan a) and orthoscan b) directions using a 25 μm diameter W wire test object positioned at 40 mm above the breast support. Acquired with a moving tube scan, tube voltage, A/F and mAs used by the system for 40 mm PMMA under AEC control, using the standard clinical reconstruction.

Figure 10 shows DBT system MTFs measured in the scan and orthoscan of eight systems, normalized to the curve peak. The $MTF_{50\%}$ and $MTF_{10\%}$ points taken from these curves are listed in table 5. The system DBT MTF is the product of the system projection MTF with the reconstruction filters followed by applied clinical image processing; this is explicitly studied in the cascaded systems analysis of filtered back projection (FBP) reconstruction for DBT systems by Zhao and Zhao (Zhao and Zhao, 2008b) and by Tward and Siewerdsen (Tward and Siewerdsen, 2008) for CBCT systems. A basic analysis taken from these studies can

Accepted Manuscript

be used to illustrate or explain some aspects of DBT reconstruction and is given in Appendix 1. If greater detail is required then the above papers should be consulted.

The main factors influencing DBT system MTF, measured at some height above the breast support table, are a) the presampling system MTF i.e. the detector and geometric MTF (with focus motion for continuous motion systems), b) the reconstruction filters and c) the angular range. Figure 10 clearly shows the influence of the presampling system MTF on the full DBT MTF as superior MTF curves are seen for the orthoscan direction compared to the scan direction, which is subject to geometric blurring for the continuous motion systems. The Giotto Class has the highest MTF in the scan direction, resulting from using a step and shoot scheme with a sharp a-Se detector. The difference in the use of detector binning can be clearly seen in the results for the Selenia Dimensions (binned) and the 3Dimensions units (unbinned), which has the highest MTF in the orthoscan direction of all the systems.

Different filtering schemes are applied in the scan and orthoscan directions for all systems, with stronger filtering applied in the scan direction. This results in differences between scan and orthoscan MTF curves, even for the step and shoot systems (GE Pristina and Giotto Class) where the presampling MTF is ~isotropic, however the reconstruction results in an anisotropic system MTF. In fact, for the Fuji and GE systems, the $MTF_{50\%}$ point for the system MTF measured in the planes is higher than $MTF_{50\%}$ for the presampling MTF, consistent with a high degree of filtering during the reconstruction. The effect of the ramp filter is clearly seen in the low spatial frequency region of these figures i.e. below ~1.0 mm^{-1} and this is especially prominent in the curves of the GE Pristina and Siemens units. From these curves, we might expect somewhat reduced large area contrast for these systems, which can result in the 'grey' appearance of reconstructed images. The data show improved in-plane MTF in the scan direction for the Siemens Revelation compared to the earlier Inspiration system. MTF is approximately isotropic for the Planmed Clarity system, and lower than for the other systems. These results are consistent with data from Sundell *et al* (Sundell *et al.*, 2019), for MTF measured at 4 cm above the breast support table using a 25 μm W with total PMMA thickness of 5 cm.

Table 5 DBT system MTF results from the wire method showing the spatial frequency at which the MTF has fallen to 50% ($MTF_{50\%}$) and 10% ($MTF_{10\%}$) for the orthoscan and scan directions. The test object was positioned 40 mm above the breast support table. Moving tube scan and clinical reconstruction mode were used. * indicates step and shoot system.

	$MTF_{50\%}$ (mm^{-1})		$MTF_{10\%}$ (mm^{-1})	
System	scan	orthoscan	scan	orthoscan
Fuji Innovality ST	2.3	3.1	3.6	5.2
Fuji Innovality HR	2.0	3.1	3.4	5.2
GE Pristina*	2.3	4.4	4.6	6.2
Giotto Class*	2.8	4.1	5.9	6.4
Hologic Selenia Dimensions	2.5	3.3	3.1	5.3
Hologic 3Dimensions	2.3	4.2	4.7	7.6
Planmed Clarity	1.4	1.5	2.2	2.4
Siemens Inspiration: STD	1.8	3.2	3.2	4.5
Siemens Revelation: EMPIRE	2.1	2.5	3.7	4.5

Accepted Manuscript

Given the well characterized sources of blurring in DBT systems, modelling and inclusion of these sources has the potential to improve the reconstructed images via resolution recovery methods. Michielsen *et al* (Michielsen *et al.*, 2013) developed a position dependent resolution model for the Siemens Inspiration moving focus DBT system. The blurring due to motion was incorporated as a Gaussian smoothing kernel in an iterative reconstruction algorithm, increasing the accuracy of the forward projection in the reconstruction. The results were evaluated via contrast to noise ratio (CNR) measurements and two observer studies performed in a power law background, the first a free search task for clusters of spherical simulated microcalcifications, and the second, a task to classify whether microcalcifications are smooth or irregular. Improved CNR and microcalcification visualization was demonstrated in both observer studies.

In related work, Zheng *et al* (Zheng, Fessler and Chan, 2019) developed a detailed model of the blurring due to the x-ray source for continuous focus motion DBT systems. Formulae were derived for the source PSF as a function of source size and location and effective source sizes were calculated for the Hologic Selenia Dimensions, Siemens Mammomat Inspiration and the Fujifilm Aspire Cristalle. An example calculation of the source PSF at 0° and 30° is shown in Figure 11. Technical data for the AEC performance were only available for the Hologic system, and therefore tube current and exposure time for the projection images were estimated from tube power considerations (~5kW for both tubes) for the Siemens and Fujifilm systems. Exposure time per projection was estimated to be ~21.5 ms for a 40 mm breast thickness, for the Hologic Dimensions, which is comparable to the measured value of 16 ms (Table 4). In the same paper, estimated values are lower for the Siemens Inspiration and Fujifilm systems for 40 mm or 45 mm breast thickness are 46.8 ms and 24.5 ms, compared to measured values of 90 ms and 45 ms. As a result, Zheng *et al* (Zheng, Fessler and Chan, 2019) estimated smaller source motion values than those in Table 4, at 0.7 mm, 0.4 mm and 0.7 mm for the Hologic, Siemens and Fujifilm units. Measured values (Table 4) were 0.79 mm, 2.40 mm and 1.54 mm.

Zheng *et al* (Zheng, Fessler and Chan, 2019) used CatSim (De Man *et al.*, 2007; Carvalho, 2014) to simulate line pair and bead test objects in order to quantify blurring at different points across the image field. A detector that rotated in synchrony with the source was also considered. They found that the size of a stationary x-ray size source (0.3 mm in this example) had little influence on measured resolution. As expected, for a moving source, blurring occurred parallel to source motion and increased as objects were moved towards the source i.e. blurring from the source was strongly non-stationary. Moving the detector in synchrony with the source during exposure did not substantially mitigate blurring from the source. This is consistent with measured MTF data for the Planmed Clarity (Figure 4), a system which adopts this strategy.

Accepted Manuscript



Figure 11. The source blur PSFs for a source of nominal size of 0.3 mm for pinholes at $z_{pinhole} = -(d_{so,GEN2}cos\theta - d_{so,GEN2})/2$ with different projection angles: (a) θ = 0°; (b) θ = 30°. We enlarged each projected focal spot shape by a factor of 20 while fixing its center at the original projected location to facilitate visualization. (taken from Zheng *et al* (Zheng, Fessler and Chan, 2019)).

3.5 Noise evaluation using Noise Power Spectrum

3.5.1 Detector NPS measured in the projection images

As with MTF evaluation, noise analysis can be applied in the projection images or in the reconstructed planes. Noise evaluation of the x-ray detector in the EUREF document is primarily performed using the optional test of assessing variance as a function of DAK, which is then decomposed using the polynomial model (section 3.3.2). This simple method provides useful information on detector noise performance but has limitations: variance evaluates the magnitude of the fluctuations, e.g. within a region of the image, but does not provide information on correlation present in the fluctuations. The evaluation typically includes the influence of system gain on the variance along with the pixel aperture. A more comprehensive evaluation of image (or detector) noise properties can be made using the NPS (Dainty and Shaw, 1974; Cunningham, 2000), assuming that the imaging process is linear and stationary. The NPS provides information on the correlation of the noise and puts the fluctuations on an absolute scale. The NPS is defined as the variance per unit spatial frequency bin of a stochastic signal in the spatial domain and is calculated via Fourier Transform (FT) of ROIs extracted from the image (Dainty and Shaw, 1974). As with the MTF, the NPS can be calculated from the projection images but also from the stack of DBT planes, resulting in a volumetric NPS. While NPS is included in the EUREF DM update as an optional test (van Engen *et al.*, 2013), NPS measured in the DBT projections is not included in the EUREF DBT protocol (van Engen *et al.*, 2018). Although optional in the DM protocol, NPS is used as an additional means of detecting artefacts, changes in noise level and blurring in the x-ray detector. It is also assumed in the EUREF protocol that the imaging system is first assessed in DM mode (including image receptor assessment) before

Accepted Manuscript

proceeding to the DBT protocol. Problems specific to the detector in DBT mode are expected to be found via the response function and variance decomposition, while homogeneity faults or artefacts e.g. from incorrect detector calibration will be found with the system homogeneity test. Recent papers reporting technical QC measurements for DBT systems do not report NPS data calculated in the projection images (Rodríguez-Ruiz *et al.*, 2016; Maldera *et al.*, 2017).

A standard equation for calculating NPS is given by Dobbins *et al* (James T. Dobbins *et al.*, 2006) and forms the definition used in the IEC standard (IEC, 2007):

$$NPS(f_x,f_y) = \frac{lim \quad lim \quad \Delta x\Delta y}{N_xN_y\to\infty M\to\infty \; N_xN_y}\sum_{m=1}^{M}\left|\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\{I(x_i,y_j) - \bar{I}\}e^{-2\pi(f_x x_{ix} + f_y x_j)}\right|^2 \quad (9)$$

where $I(x_i,y_j)$ is the pixel value at position (x_i,y_j) in the image. $\bar{I}$ can be the mean of the image region from which the ROIs are extracted or can be a 2nd order polynomial fitted to this region (Williams, Mangiafico and Simoni, 1999; IEC, 2007). The spatial coordinates x and y correspond to frequencies u and v in the spatial frequency domain, Δx and Δy are the pixel spacings and N_x and N_y are the number of pixels in the ROI, respectively in the horizontal and vertical directions.

The NPS given in (9) is the absolute NPS as used in the IEC standard (IEC, 2007), calculated from images converted to number of photons using the conversion function (also termed the detector response function). The NPS is not normalized and increases with increasing exposure per image. It is also common to work in terms of a relative or normalized NPS (NNPS) whose magnitude is inversely proportional to detector exposure (Nishikawa *et al.*, 1987; Dobbins III *et al.*, 1995). This is calculated by dividing by the square of the mean signal of the regions from which the NPS ROIs were extracted (MPV_{lin}):

$$NPS(f_x,f_y) = \frac{1}{MPV_{lin}^2}\frac{lim \quad lim \quad N_xN_y\Delta x\Delta y}{N_xN_y\to\infty M\to\infty \quad M}\sum_{m=1}^{M}\left\langle\left|FT_{nk}\{I(x,y) - \bar{I}\}\right|^2\right\rangle \quad (10)$$

Here, the FT is normalized by $\frac{1}{N_xN_y}$. There are many factors influencing a measurement of the NPS, and these are covered in the general literature on this topic that also applies to DBT devices, for example (Dobbins III *et al.*, 1995; Williams, Mangiafico and Simoni, 1999; J. T. Dobbins *et al.*, 2006; Zhou *et al.*, 2011).

The NPS measured in the projection images was used in a study by Mackenzie *et al* (Mackenzie *et al.*, 2017) to quantify noise from different physical processes occurring within the system, with the aim of improving image simulation for virtual clinical studies. Noise associated with lag, ghosting, x-ray quanta and from readout electronics (dark noise) was measured from a series of DBT acquisitions acquired with or without a lead shutter to block the x-ray beam at specific points during the sequence. Projections acquired with no beam block contain noise from all four of these sources while those acquired with the beam blocked for the entire sequence contain just dark noise. If the sequence is blocked until the central projection then the first unblocked projection contains quantum and dark noise but no lag or ghosting noise, while introducing the lead shutter one projection before the central projection results in images containing only lag and dark

Accepted Manuscript

noise. The NPS in the central projection was calculated for these four acquisition types and suitable subtraction of the different spectra enabled the NPS for these different sources to be isolated. For the NPS calculation, ROIs of 64 × 64 pixels were used and the final NPS radially averaged (Mackenzie *et al.*, 2017). Results showed that the noise from extra signal due to lag and ghosting was negligible compared to the quantum noise level, as seen in Figure 12. Note that the signal resulting from lag and ghosting processes is not negligible (Zhao and Zhao, 2008a; Mainprize, Wang and Yaffe, 2009; Marshall *et al.*, 2010; Mackenzie *et al.*, 2017) and may have to be included in system modelling.

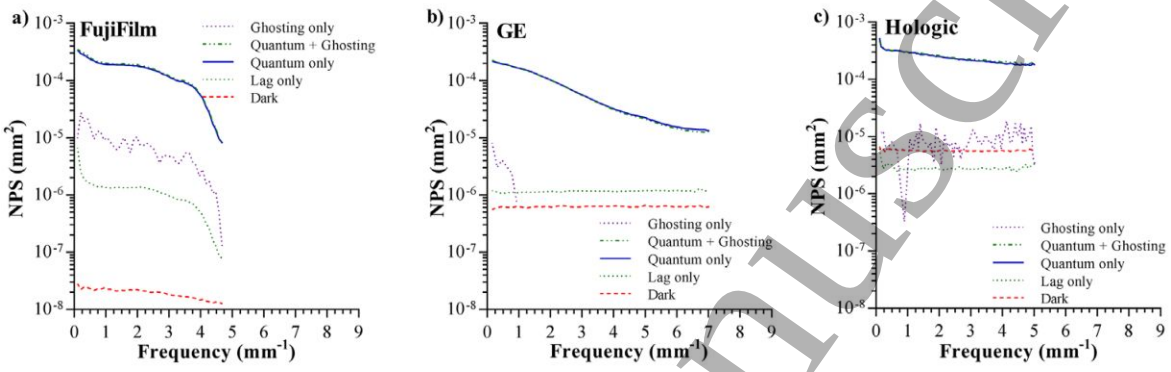


Figure 12. NPS associated with lag (W_L) and ghosting (W_G) compared the quantum noise with (W_{QG}) and without (W_Q) ghosting and the dark noise (W_D) at high dose: (a) FujiFilm (6.93 GeV mm^{-2}), (b) GE (6.78 GeV mm^{-2}), (c) Hologic (9.21 GeV mm^{-2}). Figure taken from (Mackenzie *et al.*, 2017).

3.5.2 NPS measured in the reconstructed planes

Outside of the measurements accompanying the detailed analysis of Zhao *et al* (Zhao and Zhao, 2008b; Zhao *et al.*, 2009), there are relatively few studies giving NPS in the reconstructed planes, especially for commercially available systems. For example, Chen *et al* (Chen, Lo and Dobbins III, 2006) provide NPS results for three different reconstruction methods applied to images acquired on an early version of a Siemens DBT device. They found that the NPS was sensitive to differences between the algorithms, showing for example that a matrix inversion method had improved noise performance in mid-range frequencies while a filtered back projection (FBP) method had lower high frequency noise due to the apodization filter. In a recent study, Sundell *et al* (Sundell *et al.*, 2019) measured NPS normalized to the peak value for four different DBT systems, with the Fujifilm system evaluated in HR and ST modes. Clear differences in NPS shape were seen between the systems, resulting from differences in the projection NPS and from the reconstruction algorithms. The lack of information on NPS absolute magnitude adds to the difficulty when trying to use these differences to understand clinical imaging performance.

These studies highlight two difficulties in using a metric such as the system NPS for DBT systems. First, there is no fixed output scale applied to NPS calculated in the reconstructed planes, in contrast to CT systems where the Hounsfield scale quantitatively relates PV to object linear attenuation coefficient, using water as a reference. Hounsfield calibration is difficult to achieve in DBT systems due to the large fraction of scattered radiation in the images. For the projection NPS measured in 'For Processing' images, the gain

Accepted Manuscript

is fixed and can be measured, allowing the image fluctuations to be placed on a known scale. For DBT systems the scaling applied to pixel data can vary greatly depending on the image content, and the scaling will certainly be different between different vendors. Gain or scaling must be accounted to obtain a valid comparison between systems. Second, although NPS can give deep insight into system performance when detailed information is available on the reconstruction filters (Tward and Siewerdsen, 2008; Zhao and Zhao, 2008b), NPS alone is not sufficient to predict system imaging performance. ICRU Report 54 (ICRU, 1996) gives methods to assess image quality of medical imaging systems and this includes task based assessment using Fourier-based metrics i.e. the NPS and MTF. A large body of literature exists on this topic and these methods are now widely used in medical image assessment; some discussion of these methods is given later.

Practically, 3D NPS in the reconstructed planes can be assessed from acquisitions of a homogeneous PMMA block (Zhao *et al.*, 2009; Monnin *et al.*, 2020), typically 4 or 5 cm in thickness. The 3D NPS is calculated from volumes of interest (VOIs) using:

$$NPS(f_x,f_y,f_z) = \frac{1\,\Delta x\Delta y\Delta z}{MN_xN_yN_z}\sum_{m=1}^{M}\langle|FT\{I(x,y,z) - \bar{I}(x,y,z)\}|^2\rangle \qquad (11)$$

where Δx, Δy, and Δz and N_x, N_y and N_z are respectively the voxel dimensions and the VOI length in the x, y and z directions. Note that there is no normalizing factor (~$1/(N_x\,N_y\,N_z)$) for the FT used (11). In Zhao *et al* (Zhao *et al.*, 2009), VOIs of 128 × 128 ×40 were extracted from a larger VOI of 512 ×512 × 40 and $\bar{I}(x,y,z)$ was the mean of this volume. In the study of Monnin *et al* (Monnin *et al.*, 2020), non-overlapping VOIs of 60 × 60× 30 were used but no de-trending correction was applied. The 2D in-plane NPS for a particular direction (plane) is calculated by integrating the 3D NPS in a direction orthogonal to this plane, thus for the x-y plane the 2D NPS is (Zhao *et al.*, 2009):

$$NPS_{x,y}(f_x,f_y) = \int_{-f_{z,\,Nyq}}^{f_{z,\,Nyq}} NPS(f_x,f_y,f_y)df_z \qquad (12)$$

This gives the 2D synthesized NPS as defined by Siewerdsen *et al* (Siewerdsen, Cunningham and Jaffray, 2002). In-plane 2D NPS was calculated for five DBT systems by Sundell *et al* (Sundell *et al.*, 2019), but using the method referred to as the extraction method by Siewerdsen *et al* (Siewerdsen, Cunningham and Jaffray, 2002), in which the NPS is calculated from a single DBT plane. The extraction method can be prone to error as correlations orthogonal to x-y plane are not accounted for (Siewerdsen, Cunningham and Jaffray, 2002).

Figure 13 presents 2D NPS taken from Monnin *et al* (Monnin *et al.*, 2020) calculated for the x-y plane for the Fuji Innovality in ST mode and Siemens Inspiration systems, normalized to the number of photons (scattered and primary) at the x-ray detector. The data show the influence of the ramp and apodization filters in the x (scan) direction, while the NPS along the y-axis (in the orthoscan direction) is zero. These figures show that both the magnitude and the bandwidth of the NPS is reduced for the newer reconstruction algorithms. This reduction in noise was achieved without compromising signal transfer, as system efficiency and contrast detail performance was improved for both systems (Monnin *et al.*, 2020).

Accepted Manuscript

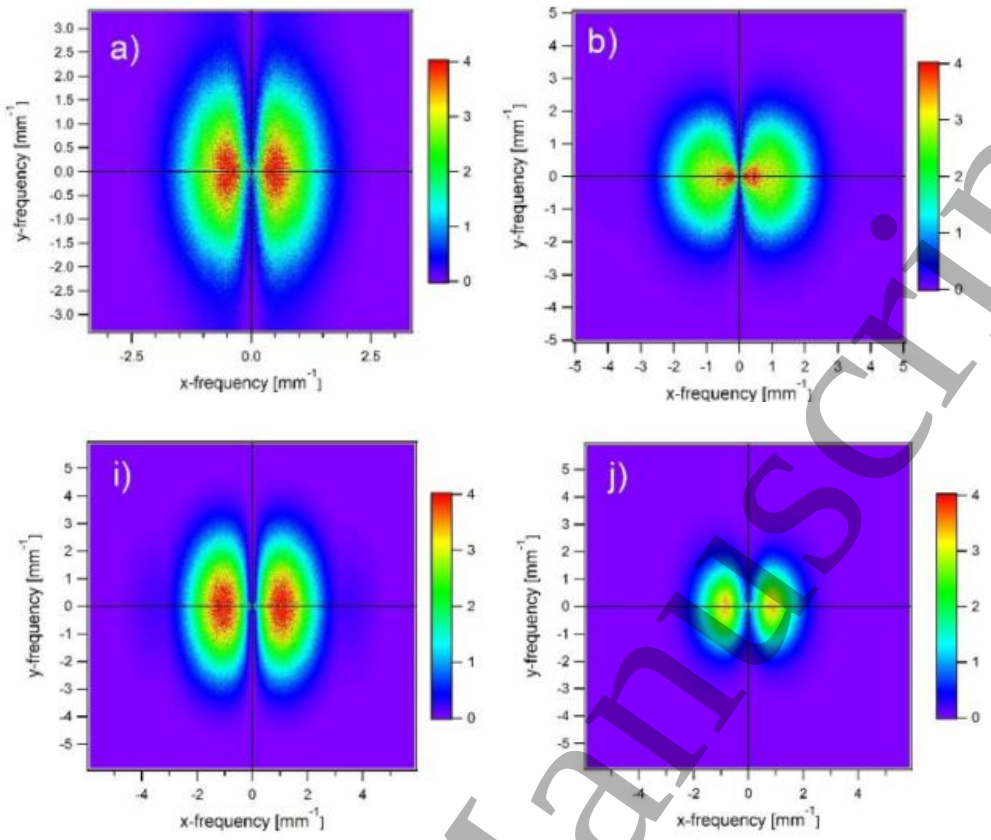


Figure 13. 2D in-plane NPS normalized for number of photons at the detector for a) FujiFilm Innovality in ST mode and b) FujiFilm Innovality in ST mode with the new ISR algorithm i) Siemens Inspiration with STD reconstruction j) Siemens Inspiration with the newer EMPIRE reconstruction. Excerpt of figure from (Monnin *et al.*, 2020)

3.6 Detective Quantum Efficiency

3.6.1 DQE calculated for the projection images

One of the principal uses of NPS is in image receptor characterization, where this parameter quantifies the magnitude and spatial frequency distribution of the image noise, within the constraints of detector stationarity (Cunningham, 2000). The presampling detector MTF (MTF_d) is used to quantify signal transfer in the spatial frequency domain, and together with the NPS, these parameters are used to estimate the squared output signal to noise ratio (SNR_{out}) of the image receptor (Shaw, 1978). This squared output SNR can be interpreted a measure of the effective number of x-ray quanta per unit area that contributes to image formation, under the assumption that all image noise arises from photon counting statistics, and is defined as the noise equivalent quanta (NEQ) (Shaw, 1978). Image receptor efficiency is calculated by comparing the NEQ measured for the real device to an ideal photon counting detector imaging Q_0 photons mm^{-2} i.e. the detective quantum efficiency (DQE) (Cunningham, 2000; IEC, 2007):

$$DQE(f_x,f_y) = \frac{NEQ(f_x,f_y)}{Q_0} = \frac{Q_0\, MTF_d^2\,(f_x,f_y)}{NPS(f_x,f_y)} \quad (13a)$$

Accepted Manuscript

or in terms of the NNPS:

$$DQE(f_x,f_y) = \frac{MTF_d^2(f_x,f_y)}{Q_0\, NNPS(f_x,f_y)} \quad (13b)$$

Some DQE results are available from early studies on DBT systems, however data on this topic remain scare. Cheung *et al* (Cheung *et al.*, 2005) give results for a version of the x-ray detector used in the Hologic Selenia Dimensions systems. DQE was measured at 28 kV Mo/Rh and 4 cm PMMA and found to be ~0.4 at a detector air kerma of 4.2 μGy, a reduction of ~10% compared to DQE at 8.0 μGy. A binning scheme of 2x2 pixels (70 μm binned to 140 μm) was applied and this is also used in the Hologic Selenia Dimensions. Hologic have recently introduced a new a-Se detector which operates without binning in DBT mode, i.e. with 70 μm pixel pitch, however published DQE results are not available. Zhao and Zhao (Zhao and Zhao, 2008a) measured DQE for a detector operated in binned and non-binned mode at a range of DAK levels for a prototype Siemens DBT system. At ~8 μGy and above, DQE was largely independent of DAK, while DQE at 3.5 μGy was within 20% of the peak value at higher air kerma levels indicating low electronic noise levels for the system (Zhao and Zhao, 2008a). Marshall *et al* (Marshall *et al.*, 2010) presented DQE data for an early Siemens Inspiration DBT system consistent with this result, where DQE (at ~1.0 mm^{-1}) at 24 μGy was a factor of 1.8 higher in DBT mode than in DM mode, due to the reduced influence of electronic noise. The zero degree acquisition mode was not available on the early generation Inspiration systems and therefore there was some influence of focus motion on the DQE data in this study. The DQE data of Olgar *et al* (Olgar, Kahn and Gosch, 2013) for a Selenia Dimensions operated in binned zero degree mode showed that DQE was constant for DAK levels above approximately ~13.5 μGy. Given the essentially flat NPS curves seen for a-Se based DBT systems, DQE(f) has a shape that follows $MTF^2(f)$ and therefore DQE(f) falls rapidly in binned mode compared to non-binned operation (Zhao and Zhao, 2008a; Olgar, Kahn and Gosch, 2013).

Zhao *et al* (Zhao *et al.*, 2015) developed a version of cascaded linear system analysis for the low and high gain modes of a CsI-CMOS x-ray detector (Dexela 2923 MAM) with emphasis on the detector electronic noise performance. Calculated results were compared to NNPS and DQE data acquired in two previous studies for this detector using broadband x-ray spectra at 28 kV (Patel *et al.*, 2012) and using monochromatic synchrotron radiation of 17 keV (Konstantinidis *et al.*, 2013). Both low and high gain modes were found to maintain DQE down to DAK levels of ~7 μGy, while the high gain mode had negligible electronic noise even at ~1.7 μGy. These are consistent with the variance decomposition results for other x-ray detectors in section 3.3.2.

3.6.2 System DQE calculated in the reconstructed planes

In projection imaging, the concept of DQE for a detector is well established (Shaw, 1978) and can be conceptualized as the departure from an ideal detector. The concept of DQE in CBCT and therefore DBT imaging needs careful definition where the logarithmic amplification and the back projection stages complicate the definition of SNR and therefore NEQ. Readers are referred to the discussion of DQE for

Accepted Manuscript

CBCT systems on this point by Tward and Siewerdsen (Tward and Siewerdsen, 2008). Both Tward and Siewerdsen (Tward and Siewerdsen, 2008) and Zhao and Zhao (Zhao and Zhao, 2008b) use the concept of 'deterministic NPS', as described by Siewerdsen and Jaffray (Siewerdsen and Jaffray, 2003), to define NEQ. The sampling stage 'freezes-in' or fixes the effect of the reconstruction filtering (ramp, apodization and interpolation) on the NPS (Appendix, equation (a10)). As a result, the NEQ and therefore the DQE can depend on the reconstruction filtering (Siewerdsen and Jaffray, 2003). The 3D DQE is written as the ratio of 3D NEQ to the total incident x-ray fluence (Shaw, 1978; Tward and Siewerdsen, 2008; Zhao and Zhao, 2008b).

There are few experimental estimates of system DQE for DBT devices in the literature. While the simulations of system MTF and NPS calculated using cascaded systems analysis for the Siemens Mammomat Novation^{TOMO} unit (Zhao and Zhao, 2008b) were experimentally validated (Zhao *et al.*, 2009) experimental, measured DQE results were not given. Chumin Zhao *et al* (Zhao *et al.*, 2017) measured the MTF, NPS and DQE as a function of oblique angle incidence for a CMOS active pixel sensor detector (DynAMITe) with 50 μm pixels and combined these data with the cascaded systems theory of (Zhao and Zhao, 2008b) to model the impact of angular range, dose and focal spot size for a theoretical step and shoot system. They found that oblique entry of ±30° did not greatly influence in-plane DQE but did improve the low frequency sampling in Fourier space and thus improved DQE below 2mm^{-1}. Experimental results for the 3D DQE were not presented and the study did not examine the influence of reconstruction filters on the in-plane DQE.

The paper by Monnin *et al* (Monnin *et al.*, 2020) develops a framework for system NEQ and DQE for DBT systems, calculated from the reconstructed planes. In-plane system DQE is written as the ratio of the NEQ in the image to the NEQ at the system input, in the presence of scattered radiation:

$$DQE_{sys}(f_x,f_y) = \frac{IRF_d^2(f_x,f_y)}{\left(\frac{\Delta\bar{q}}{\bar{q}}\right)^2 (1-SF_{in})^2 \bar{q}_{in}\, NPS_d(f_x,f_y)} \tag{14}$$

where $IRF_d(f_x,f_y)$ and $NPS_d(f_x,f_y)$ are respectively the in-plane impulse response function and noise power spectrum, expressed in image pixel values without linearization or normalization, and measured from images of the phantom shown earlier containing the Al disc of thickness 0.2 mm and diameter 5 cm (Section 3.4.4, Figure 6). The radiant contrast of this disc $\left(\frac{\Delta\bar{q}}{\bar{q}}\right)$ was measured using an ionization chamber, while SF_{in} and $\bar{q}_{in}$ are respectively, the scatter fraction and number of photons/mm² (primary and scattered) measured at the breast support table (i.e. before the antiscatter grid, if present). A key point in this analysis is that the NEQ is shown to be self-normalizing and thus measurement of the zero-frequency signal for MTF normalization (problematic for DBT systems, as discussed earlier) is not needed. Instead, the IRF and NPS are left un-linearized and un-normalized, but have to be measured from the same image, with the assumption that system gain is the same for the image regions used for IRF and NPS calculation. The study presents radially averages in-plane DQE_{sys} results for 5 DBT systems (Figure 14).

Accepted Manuscript

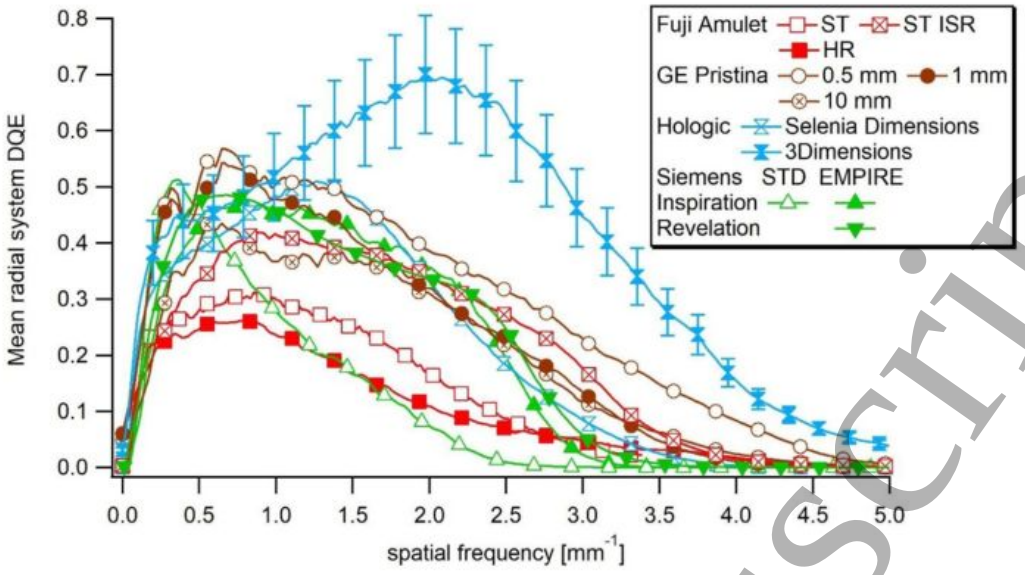


Figure 14. Mean radial DQE_{sys} (Figure taken from (Monnin *et al.*, 2020))

The results showed that the new iterative algorithm for the Fuji Amulet Innovality (ISR) increased peak DQEsys from 30% to 42% for ST mode at ~1.0 mm^{-1} and resulted in higher DQEsys up to 3.5 mm^{-1}. A similar effect was seen for the Siemens Inspiration unit, where peak DQEsys at ~0.5 mm^{-1} was similar for the STD and EMPIRE reconstructions, however, there was a strong reduction for the STD algorithm while the iterative EMPIRE algorithm maintained a high DQEsys, between 0.5mm^{-1} and 2.5 mm^{-1}. This confirms the importance of the reconstruction algorithm for overall system efficiency (Siewerdsen and Jaffray, 2003).

While equation (14) includes the influence of scattered radiation on system efficiency, imaging efficiency written this way remains 'task generic' i.e. is not weighted by any particular imaging task. Tapiovaara and Wagner (Tapiovaara and Wagner, 1985) show that the concept of imaging efficiency cannot be separated from the imaging task. In equation (14), the imaging system is compared to an ideal photon counter in the presence of scattered radiation. There are a number of ways in which the concept of imaging efficiency could be extended; for simplicity, this is done for the case without scattered radiation. One means of incorporating imaging task could be to use a Fourier-based model observer to define a detectability index (d') and use d' for the ideal detector in a task-based DQE definition (Persson, Rajbhandary and Pelc, 2018):

$$DQE^{task}(f) = \frac{d'^2(f)}{d'^2_{ideal}(f)} \tag{15}$$

As an example, assume a Fisher-Hotelling observer (ICRU, 1996; Richard and Siewerdsen, 2008):

$$d'^2(f) = \int \frac{MTF^2(f,E)\Delta S(f,E)}{NNPS(f,E)}\, dE = \int |\Delta S(f,E)|^2\, NEQ(f,E) dE \tag{16}$$

Accepted Manuscript

Here, E is the energy, f is the radial spatial frequency and $\Delta S(f, E)$ is a spatial frequency domain description of the imaging task function, given by the Fourier transform of the difference between two hypotheses (ICRU, 1996):

$$\Delta S(f,E) = \left| F[h_1(x,y) - h_2(x,y)\,] \right| \tag{17}$$

For the ideal detector we can write (Persson, Rajbhandary and Pelc, 2018):

$$d'_{ideal}{}^2(f) = \int |\Delta S(f)|^2\, \overline{q_0}(E)\, dE \tag{18}$$

where $\overline{q_0}$ is the number of x-ray photons per unit area at energy E. This leads to a measure of task weighted DQE at spatial frequency u:

$$DQE^{task}(f) = \frac{\int |\Delta S(f,E)|^2\, NEQ(f,E) dE}{\int |\Delta S(f)|^2\, \overline{q_0}(E)\, dE} \tag{19}$$

in which the efficiency of the imaging system would depend on the specified task. Persson *et al* integrate over energy and leave the frequency dependency in the DQE term, as in the standard definition of DQE.

Another possibility is the inclusion of anatomical noise in the definition of DQE, as described by Richard *et al* (Richard *et al.*, 2005) when studying dual energy chest imaging. The standard definition of DQE is used as a starting point:

$$DQE(f) = \frac{MTF^2(f)}{\overline{q_0}\, NNPS_{sys}(f)} \tag{20}$$

where $\mathrm{NNPS}_{sys}(f)$ is the normalized noise power spectrum, including the quantum and electronic noise contributions from the system and $\overline{q_0}$ is the mean photon fluence at the x-ray detector (Richard *et al.*, 2005). Rather than quantum or system electronic noise, anatomical noise can be the most important factor in limiting object detectability, especially for masses (Bochud *et al.*, 1999; Burgess, Jacobson and Judy, 2001). Anatomical background noise is included as a separate NPS term whose magnitude varies as a 1/f power law (Burgess, Jacobson and Judy, 2001; Richard *et al.*, 2005):

$$NNPS_{bkg}(f) = \frac{\mathrm{K}}{f^{\beta}} \tag{21}$$

where K and ß respectively characterize the magnitude and correlation of background structures. The conventional DQE is written with the quantum ($NNPS_{quantum}(f)$) and additive electronic noise ($NNPS_{add}(f)$) terms separated:

$$DQE(f) = \frac{MTF^2(f)}{\overline{q_0}\,(NNPS_{quantum}(f) + NNPS_{add}(f))} \tag{22}$$

and a generalized DQE (GDQE) defined with the anatomical background noise ($NNPS_{bkg}(f)$):

Accepted Manuscript

$$GDQE(f) = \frac{MTF^2(f)}{q_0\,(NNPS_{quantum}(f) + NNPS_{bkg}(f)MTF^2(f) + NNPS_{add}(f))} \quad (23)$$

The MTF has been separated from $NNPS_{bkg}(f)$ by Richard *et al* (Richard *et al.*, 2005) so that this term explicitly describes the spatial frequencies of image structures that act as a noise source, independent of the imaging system. Means of estimating $NNPS_{bkg}(f)$, both using analytical models and by measurement, are described (Richard *et al.*, 2005). Note that the derivation does not account for the influence of scattered radiation on DQE but could perhaps be combined with the approach of Monnin *et al* (Monnin *et al.*, 2020). While developed for dual energy chest imaging, this work has yet to be taken up in DBT imaging as a means of quantifying the degree of anatomical noise erasure effected by a system and the influence of residual anatomical background noise on overall imaging efficiency.

We must reiterate that MTF, NPS and DQE are Fourier-based metrics that assume a system is linear and shift-invariant. As discussed, the reconstruction will contain non-linear steps such as the logarithmic function and small signal linearity must be assumed when applying Fourier metrics. Non-linear iterative reconstruction methods are well established in DBT imaging (Wu *et al.*, 2004) and already applied by some manufacturers, for example in the Fuji Innovality and in the GEHC Pristina reconstruction algorithms. With the introduction of artificial intelligence (AI) and deep learning methods to medical imaging applications such as image reconstruction, the use of non-linear or locally variant methods is likely to increase. Using these techniques, transfer of signal and noise through the system can be separated to some extent and sharpness of structures in processed images can depend on both the dose level and the contrast of local structures. As a consequence, one must interpret the MTF and DQE curves in Figures 10 and 14 with caution.

Considering MTF or NPS in isolation is potentially misleading, as illustrated in the study by Urbanczyk et al (Urbanczyk *et al.*, 2012) for DM imaging, where the GEHC FineView algorithm was examined using a low contrast edge. The algorithm was found to increase MTF and NPS as a function of spatial frequency, an effect that was dose dependent. When MTF and NPS calculated from FineView processed images were combined to form the DQE, the filter effects cancelled giving a DQE that was unaffected by the processing. This indicated signal and noise were not de-correlated and that the signal to noise ratio transfer of the system was not affected by the processing and remained in some sense meaningful. A measurement of threshold contrast detectability for the processed and un-processed images was consistent with the DQE results. Similar results were found in the study by Monnin *et al* (Monnin *et al.*, 2020) where MTF and NPS measured with a low contrast edge were used in an NPWE model and the calculated detectability followed detectability results measured with a contrast-detail phantom. While the MTF is useful for detecting changes in aspects of system performance that affect signal transfer and sharpness, these studies highlight that a task based evaluation of SNR, either built using Fourier metrics or measured directly using test objects, is more robust than considering just MTF or NPS in isolation.

Accepted Manuscript

3.7 Artefact spread and resolution in the z-direction

The reduced angular range used in current DBT systems (≤50°) results in reduced spatial resolution in the direction perpendicular (z direction) to the reconstructed planes (x-y direction). This can be explained in terms of Fourier sampling of frequency space by the source in the DBT scan as follows. If the distance from the source to the object is infinite, then a parallel beam geometry applies. When this is the case, the central slice theorem states that the 2D Fourier transform of a transmission image (i.e. a DBT projection image) is equal to a plane through the origin of the 3D Fourier transform of the object (Kak and Slaney, 1988), as illustrated in Figure 15 (adapted from Wu *et al* (Wu *et al.*, 2004)). The effect of the limited angular range can be seen as two regions in frequency space, above and below the origin, that are not sampled (or very sparsely sampled). As a result, at low spatial frequencies along the x-direction, the spatial frequencies along the z-spatial frequency axis are not sampled, leading to very low resolution in the z-direction. Consequently, larger objects such as mass lesions will be poorly resolved in the z-direction while smaller objects such as microcalcifications, which contain a larger fraction of higher spatial frequencies, will have higher sampling of the spatial frequency content in the z-direction. As a result, one might expect small objects such as microcalcifications in a cluster to be separated between planes.

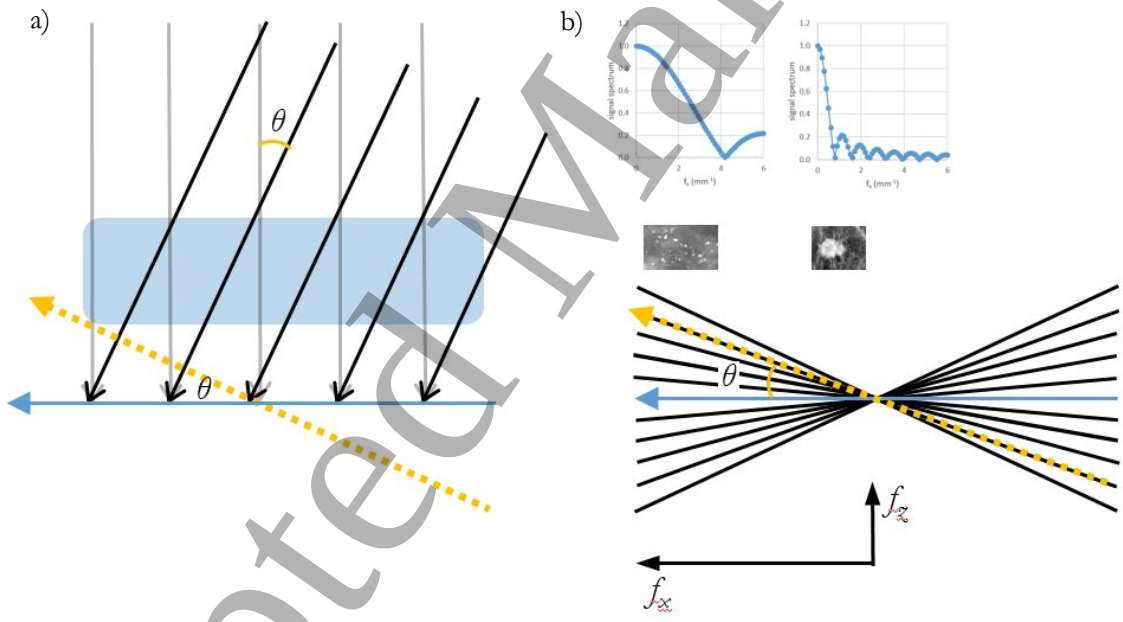


Figure 15 a) Schematic illustrations of two tomosynthesis projections acquired at 0° and at projection angle θ b) approximate tomosynthesis sampling in the Fourier domain under parallel-beam approximation for the two projections. The sampling figure shows low sampling of spatial frequencies in the z-direction at low spatial frequencies in the x-direction (~corresponding to mass lesions). Sampling of z-frequency content is higher at higher spatial frequencies i.e. for objects ~corresponding to microcalcifications. The small inset figures show a schematic representation of the spatial frequency content of large and small diameter circular details.(Figure adapted from Wu *et al* (Wu *et al.*, 2004)).

Accepted Manuscript

A further consequence of the limited sampling in the z-direction is the generation of artefacts of a real feature in a given plane that then appear in adjacent planes (Wu *et al.*, 2004). The artefacts have the appearance of the real object and the strength of the artefacts weakens as the distance from the plane containing the object increases. This was termed the artefact spread function (ASF) (Wu *et al.*, 2004) and defined for some object as:

$$ASF(z) = \frac{\overline{PV}_{artefact}(z) - \overline{PV}_{bkg}(z)}{\overline{PV}_{object}(z_0) - \overline{PV}_{bkg}(z_0)} \quad (24)$$

where z_0 is the plane containing the real feature, z is the location of a plane containing the artefact of the feature, $\overline{PV}_{object}(z_0)$ and $\overline{PV}_{bkg}(z_0)$ are the mean pixel values measured respectively in the object and background in plane z_0, and $\overline{PV}_{artefact}(z)$ and $\overline{PV}_{bkg}(z)$ are the mean pixel values in the artefact of the object and in the background, measured in adjacent planes z. The ASF quantifies the intensity of the artefact with respect to the signal from the object causing the artefact and is related to resolution of the DBT system in the z-direction.

The ASF has been implemented as a technical parameter for the characterization of DBT system resolution in the direction perpendicular to the reconstructed planes (z-direction), for assessing DBT artefacts, and as a QC tracking measure (Maki, Mainprize and Yaffe, 2016; van Engen *et al.*, 2018). The FWHM is used to quantify the width of the profile (van Engen *et al.*, 2018). The influence of the geometry and reconstruction coordinate system on the ASF is examined in the study of Maki *et al* (Maki, Mainprize and Yaffe, 2016). Two approaches are generally adopted by the vendors. If a cone beam geometry is used for reconstruction but a Cartesian coordinate system used to define coordinates in the reconstructed volume, then the FWHM of the ASF will depend on the position of the object relative to the on-axis point i.e. the point where the x-ray beam is orthogonal to the detector (Maki, Mainprize and Yaffe, 2016). As the object is moved off axis, the ASF will be orientated towards the focus. An alternative is used by the Hologic Selenia Dimensions system, in which the reconstructed volume is defined using a cone-beam coordinate system. For this system, all points lying on a straight line from the focus have the same x and y coordinate, independent of the z position. In order to transform from cone beam to Cartesian coordinates, the (x,y) coordinates must be multiplied by the geometric magnification factor *m* for height z above the x-ray detector, defined as $m = \frac{SDD}{(SDD - z)}$ (Maki, Mainprize and Yaffe, 2016). An equivalent means of performing this transformation is to specify an in-plane (x-y) pixel spacing that depends on the magnification of the plane. This is stored within the DICOM header for the Hologic systems. The ASF for systems using cone beam coordinates remains perpendicular to the reconstructed planes.

These points are illustrated in Figure 16 (taken from Maki *et al* (Maki, Mainprize and Yaffe, 2016)) where the x-z plane is shown for a phantom containing 0.79 mm diameter Al spheres arranged in a line parallel to the chest wall edge, imaged on GE, Hologic and Siemens DBT systems. For the GE and Siemens systems, which use Cartesian coordinates, the angle of ASF *wrt* the z-axis changes as the sphere moves further off-

Accepted Manuscript

axis, while for the Hologic the ASF remains perpendicular to the reconstructed planes. The upshot of this is some position dependence of the FWHM; Maki *et al* (Maki, Mainprize and Yaffe, 2016) show that transforming a reconstructed volume from Cartesian to cone beam coordinates improved the reproducibility of longitudinal measurements of FWHM made for QC and reduced the position dependence of the FWHM. Resampling the volume to use cone beam coordinates increases the FWHM as the worst direction of the ASF is aligned with the z-axis (4.9 mm compared to 4.0 mm, when using Cartesian coordinates for the volume) (Maki, Mainprize and Yaffe, 2016). The influence of vendor coordinate system on QC measurements of ASF can be minimized by ensuring that the object (typically a sphere) has the same position from test to test and is imaged on axis. Where an array of spheres is imaged to examine the ASF at different positions throughout the reconstructed volume (Strudley *et al.*, 2015), then this effect must be accounted for. Strudley *et al* (Strudley *et al.*, 2015) used the maximum value measured in-plane when measuring the mean pixel value for the out of plane artefact, which accounts for the tilt of the ASF as the object is moved off axis for Cartesian coordinate systems.

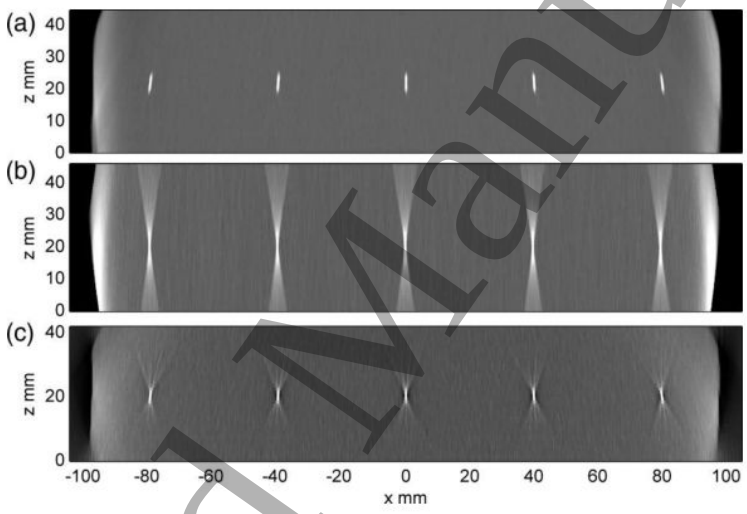


Figure 16. Slices in the x-z plane through reconstructed volumes of a phantom containing a row of spheres a) GE b) Hologic (without slice specific voxel spacing) c) Siemens. (taken from Maki *et al* (Maki, Mainprize and Yaffe, 2016))

Another means of quantifying the signal transfer between planes is an angled wire, first proposed by Eberhard *et al* (Eberhard *et al.*, 2006) for DBT characterization and also applied to chest tomosynthesis system evaluation (Li *et al.*, 2007). This method uses a wire positioned orthogonally to the scan direction on the breast support table, and tilted at some angle β *wrt* to the table, as described by Hu *et al*(Hu, Zhao and Zhao, 2008). The angle is calculated by forming a vector of y-position points that have the maximum intensity for each reconstructed plane and then using the equation:

$$\beta = \arctan\left(\frac{\Delta z}{\Delta y}\right) \tag{25}$$

where Δz is the total thickness in z of the region analysed containing the wire and Δy is the difference in maximum and minimum y positions of the points with maximum intensity. Following Hu *et al* (Hu, Zhao

Accepted Manuscript

and Zhao, 2008) this method yields the PSF in the z direction, calculated from a DBT plane containing the image of the angled wire. The first step is to establish the point with maximum intensity; this is the in-focus point (defined as the origin) i.e. the point where the wire has the same z-position as the plane being analysed (Hu, Zhao and Zhao, 2008). A point along the wire in the y direction at some distance p from the origin is at some distance q in the z-direction from the plane i.e.

$$q = p\tan(\beta) \tag{26}$$

The sampling distance in the z direction (d_z) is given by $d_z = d_x\tan(\beta)$ where d_x is the in-plane (x-y) pixel spacing, allowing the in-plane intensity distribution along the wire to be converted into the PSF in the z-direction. Hu *et al* (Hu, Zhao and Zhao, 2008) measured the z-direction PSF using a 70 μm diameter W wire for a prototype Siemens Novation^{TOMO} DBT system and found good agreement between the measurement and a cascaded linear system model. The z-direction PSF was explored as a function of angular range, from ±20° down to ±5°, and reducing angular range degraded resolution in the z-direction, smearing image details in a direction perpendicular to the planes. In their study, Hu *et al* (Hu, Zhao and Zhao, 2008) also measured and simulated the ASF expected for a 0.4 mm steel sphere and found that a wider angular range reduced the intensity of out of plane artefacts, as characterised by the ASF.

More recent studies have used both of these techniques to study the ASF and z-direction PSF for commercially available DBT systems. Marshall and Bosmans (Marshall and Bosmans, 2012) used a W wire of 0.22 mm diameter embedded in the 45 mm PMMA thick Agatha phantom (Jacobs *et al.*, 2010). Rodríguez-Ruiz *et al* (Rodríguez-Ruiz *et al.*, 2016) also used a 45 mm thick PMMA block but with W wires of diameter 25, 50 and 100 μm and calculated the z-direction PSF from a plane at 30 mm above the breast support table. The study found a slight dependency on wire diameter, with FWHM values that were between 5% and 19% smaller for the 25 μm wire compared to the 100 μm wire. The low contrast of the 25 μm diameter wire made computation difficult. FWHM results were consistently smaller than those reported by Marshall and Bosmans (Marshall and Bosmans, 2012), and this is probably due to the large diameter of the W wire (0.22 mm) in the Agatha phantom. Rodriguez-Ruiz *et al* (Rodríguez-Ruiz *et al.*, 2016) also measured ASF using an array of 1 mm Al spheres but did not take the coordinate system into account during analysis and found reduced reproducibility for off-axis spheres, consistent with Maki *et al* (Maki, Mainprize and Yaffe, 2016). In their implementation of ASF, a correction was made for the noise (standard deviation) in the planes under analysis (Rodríguez-Ruiz *et al.*, 2016):

$$ASF(z) = \frac{\overline{PV}_{artefact}(z) - \overline{PV}_{bkg}(z)}{\overline{PV}_{object}(z_0) - \overline{PV}_{bkg}(z_0)} \times \frac{\sigma_{bkg}(z_0)}{\sigma_{bkg}(z)} \tag{27}$$

The study of Maldera *et al* (Maldera *et al.*, 2017) used the Agatha phantom for evaluation of z-direction PSF while ASF was calculated from spheres of composition 30%/70% fatty to glandular equivalent material that were 3 mm and 5 mm in diameter, effectively simulating small mass lesions rather than microcalcifications. Strong differences in ASF and the associated sphere images were found between

Accepted Manuscript

systems, with for example the artefact disappearing completely at some point for the GE system. This is consistent with the image seen in Figure 16a, where the artefact spread in the z-direction is strongly suppressed by the reconstruction algorithm. Maldera *et al* (Maldera *et al.*, 2017) found FWHM values for the z-PSF that were 20% smaller compared to Marshall and Bosmans, despite using the same test object. Differences in reconstruction algorithm most likely account for this difference, for example in Marshall and Bosmans (Marshall and Bosmans, 2012), the 'Phantom_Contrast' and 'Flatfield QC' reconstruction was used, for the Siemens and Hologic systems, respectively.

The EUREF protocol specifies the use of an array of 1.0 mm diameter Al spheres embedded in a 5 mm thick PMMA sheet that can be imaged at different heights about the breast table when supported by different PMMA thicknesses. Table 6 shows the FWHM of the ASF for eight different DBT systems for such a sphere positioned centrally left-right (i.e. on the z-axis), 20 mm from the chest wall edge and imaged at a height of 20 mm above the table. Also shown are the ASF FWHM results for the 0.5 mm diameter Al sphere in the Agatha phantom (Jacobs *et al.*, 2010), which is embedded in a 45 mm PMMA block, at a height of 30 mm above the breast support table. The data were not corrected for the potential use of Cartesian coordinates in the reconstruction (Maki, Mainprize and Yaffe, 2016), however the on-axis spheres were used for analysis and the resulting difference would be small. The related ASF curves are plotted in Figure 17. As might be expected, there is a large range in ASF between systems measured with both sphere diameters, almost by a factor of 2, and this to some extent reflects the difference in angular range. The results for the 1.0 mm sphere are largely consistent with the data presented by a number of studies (Strudley *et al.*, 2015; Rodríguez-Ruiz *et al.*, 2016; Sundell *et al.*, 2019). It is interesting to note strong differences in the shape of the ASF between the different systems in Figure 17, with a number of systems having a flat top that persists for 3 or 4 planes for the 1.0 mm sphere. It is also clear that the FWHM for the 0.5 mm sphere is generally smaller than for the 1.0 mm, especially for the Fuji and Hologic units. On the other hand, there is not a large influence of sphere diameter on ASF for the Siemens units, however it can be seen that the reconstruction algorithm has a strong influence on ASF shape for these systems. Both the GEHC Pristina and the Siemens have an approximately top hat appearance, which is distinct from the Gaussian function that is sometimes used as a curve fit model for the ASF (Maki, Mainprize and Yaffe, 2016).

A small decrease in FWHM has been found for some systems as the sphere height above the breast support table is increased (Strudley *et al.*, 2015; Rodríguez-Ruiz *et al.*, 2016), although this was not seen for the Fuji Innovality or Planmed systems (Sundell *et al.*, 2019). As the height of the sphere is increased above the detector, this increases the apparent angular range at the position of the sphere, leading to a reduction in FWHM (Strudley *et al.*, 2015). One might expect the ASF to be sensitive to any mechanical changes or tolerances that may occur within the DBT unit during an acquisition. For 6 measurements taken over a period of 20 months, (Maki, Mainprize and Yaffe, 2016) found average COVs of 1.5 % and 4.3 %, for FWHM evaluated using cone beam and Cartesian coordinates, respectively, for the GE, Hologic and Siemens units. The QC data from the TOMMY trial presented by Strudley *et al* (Strudley *et al.*, 2015) found

Accepted Manuscript

no significant variation in FWHM for the Hologic Dimensions units in the trial, again over the course of approximately 20 months.

These data suggest that the degree of signal spread between planes, as quantified via the ASF, strongly depends on the magnitude of the initial in-plane stimulus, and thus on the size and composition of the object used. Furthermore, to obtain reproducible results, ASF should be measured on the z-axis, with the sphere positioned at a consistent height, unless correction is made (as required) for systems using a Cartesian coordinate system. The ASF should be thought of as characterizing the degree of artefact spreading rather than resolution in the z-direction; to do this an oversampled method such as the angled wire is required (Hu, Zhao and Zhao, 2008), although there may still be some dependence of the resolution on wire diameter (Rodríguez-Ruiz *et al.*, 2016). While ASF remains an interesting parameter that should relate to the ability of a DBT system to supress the spread of signals between planes, work on making this link is currently lacking in the literature. One of the challenges lies in incorporating the ASF - or more likely the resolution in the z-direction (z-PSF) – into models describing in-plane detectability for small and large objects in a range of clinically relevant backgrounds, for example using the methods described by Gang *et al* (Gang *et al.*, 2011).

Table 6 Artefact spread function measured using 0.5 mm and 1.0 mm diameter aluminium spheres.

System	ASF FWHM (mm)		ratio of FWHM (1.0 mm) to FWHM (0.5 mm)
	0.5 mm sphere	1.0 mm sphere	
Fuji Innovality ST	5.9	-	-
Fuji Innovality HR	3.3	-	-
GEHC Pristina	4.6	6.1	1.3
Giotto Class	2.7	4.4	1.7
Hologic Selenia Dimensions	4.6	10.7	2.3
Hologic 3Dimensions	4.6	11.7	2.5
Planmed Clarity	1.9	3.1	1.6
Siemens Inspiration: STD	4.6	4.9	1.1
Siemens Revelation: EMPIRE	3.1	5.5	1.8

Accepted Manuscript

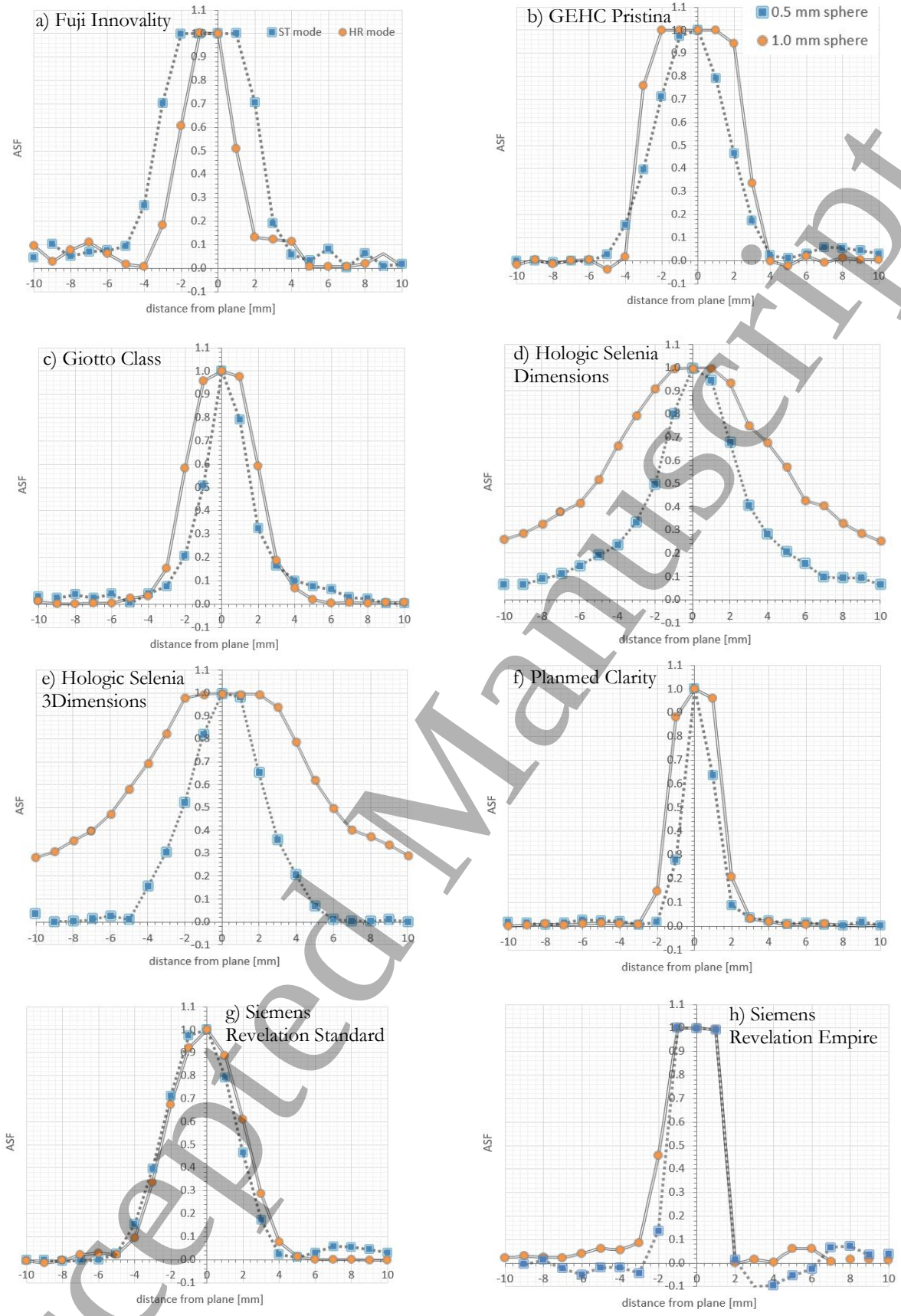


Figure 17 Artefact spread function measured using 0.5 mm and 1.0 mm diameter aluminium spheres for eight DBT systems. For the Fuji Innovality, the figure shows data obtained with the 0.5 mm sphere for ST and HR modes.

Accepted Manuscript

3.8 Summary of physical characterization of DBT systems: technical parameters

The parameters described in the preceding text are evaluated by manufacturers and by medical physicists for a number of reasons: to characterize component performance, establish the contribution of a component to overall imaging performance or to detect a change in a component or system performance. When designing a QC protocol, one must consider the time taken to set up the different tests, acquire and transfer the images and perform the analysis. The protocol must be practical, meaning that the tests are easy to set up while software tools are essential when processing the results efficiently (Marshall, 2009; Donini *et al.*, 2014; Zhang *et al.*, 2017; Samei *et al.*, 2018; Porzio and Konstantinidis, 2021). A further key point is that a test protocol should ideally be based on risk assessment, with limiting values that ensure the safe use of the system. An attempt to introduce an evidence and risk based approach has been described by a working group of the Institute of Physics and Engineering in Medicine (IPEM), for general x-ray tube and generator equipment (Honey *et al.*, 2018) and for fluoroscopy devices (Worrall *et al.*, 2020). The aim was to assess the effectiveness of various QA tests in terms of patient and staff safety, in the light of rates of failure and the consequence of a failure. Such data are not available for mammography or DBT devices, although Stuffins *et al* (Stuffins *et al.*, 2022) have used the fault database for DM units of the UK National Health Service breast screening programme (NHSBSP), collated by the National Centre for National Coordinating Centre for the Physics of Mammography (NCCPM). These are faults reported by the radiography staff that prevented the use of a mammography system, rather than those found by medical physics QC testing.

There is guidance available from the IEC on the evaluation and routine testing of mammography systems used in DBT mode in the form of IEC Standard 61223-3-6:2020 (IEC, 2020b). The Standard describes tests of the alignment, AEC performance, image receptor response function, MTF in the projection image, z-resolution, geometric distortion and detector element failure, using the same or similar methods and test objects employed in the EUREF protocol v1.03. For example, the AEC is evaluated using square PMMA blocks of thickness 10 mm and a 0.2 mm thick 10 × 10 mm Al square for an SDNR measurement. Z-resolution is assessed using an ASF measured with a phantom containing a 1 mm diameter sphere, while the projection MTF is assessed with an MTF positioned 40 mm above the breast support platform. The dosimetry method employed is that of Dance *et al* (Dance, Young and van Engen, 2011). The IEC specify an overall (visual) assessment of image quality in reconstructed planes using a phantom with a suggested thickness between 40mm and 60 mm, containing microcalcification, mass and architectural distortion simulating lesions, in structured background that simulates breast parenchymal tissue. Ideally, these lesions should match linear attenuation coefficients of real lesions and should be sufficiently challenging to allow differentiation of performance. IEC acknowledge current work from a number of groups in the literature and state that such a phantom will be included in the Standard, when validated and available. Progress in the development of phantom for DBT testing is discussed in part 2 of this work.

Accepted Manuscript

It is important to note that DBT tests are generally not performed in isolation but are preceded by QC test of the system in DM mode, in which x-ray tube, collimation, x-ray detector, AEC system and overall image quality will be assessed (van Engen *et al.*, 2013). Of course, the x-ray tube A/F selection, x-ray detector and AEC modes *etc* are generally different in DBT mode, hence the need for additional QC tests to assess DBT performance. At the moment, the extent to which technical problems in DBT mode differ from those found in DM mode is not known. Given the tube motion/rotation in DBT one might expect more faults related mechanical or stability faults. Relating fault conditions or gradually changing performance in a component to overall performance is useful and some studies have examined the use of parameters such as MTF, NNPS and DQE to do this for DM systems (Marshall, 2006b, 2007; Wigati *et al.*, 2021). The work of Wigati *et al* (Wigati *et al.*, 2021) established the stability of MTF measurements in longitudinal testing and examined the relation between changes in MTF and technical image quality, quantified using a threshold contrast-detail test object. We are not currently in a position to do this for DBT systems. One problem is the lack of a technical image quality test that is applied routinely, which would allow changes in a technical parameter, or the case where a parameter exceeds a remedial level, to be related to the technical IQ delivered by the DBT system. Outside of the context of QC, the use of virtual imaging studies enables physicists to study the impact of system parameters on imaging performance (Barufaldi *et al.*, 2020) at the system design stage, discussed in the companion paper.

Following examination of the tests in the EUREF DBT protocol and related literature, some amendments or additions to the methods used within the protocol can be suggested (see table 7). The protocol is, however, already quite involved and we do not wish to further add to the testing burden without good reason. One option could be to increase the testing performed at the Type Testing stage (if performed) or at the Commissioning stage, where more extensive tests can be performed without increasing the testing required for routine QC.

Accepted Manuscript

Table 7. Technical parameters or components and suggested amendments to the tests or test methods.

Test	Comment
X-ray source	No update. Multisource x-ray tubes may be introduced and dosimeters should be able to assess these sources.
Antiscatter grid	Evaluate grid primary and grid total transmission in zero degree mode as a function of PMMA thickness; calculate the image improvement factor. Evaluation of grid efficiency at oblique angles would be useful but is difficult to implement practically – consider this at Type Testing stage.
Sharpness/spatial resolution assessment in projection images	Measure system MTF as a function of projection angle to establish impact of beam obliquity for outer projections at Type Test and perhaps at Commissioning.
Sharpness/spatial resolution assessment in reconstructed planes	Routine measurement of MTF in the planes at 4 cm above the breast support. Consider the need for techniques that are not super-sampled and can therefore evaluate the influence of sampling rate on spatial resolution for new generation DBT systems that may employ super-resolution techniques (Vent *et al.*, 2022).
Brightness uniformity error	Only measure in the projections at Type Testing stage or if detailed system characterization is needed. Consider measurement of brightness uniformity in reconstructed planes for 2 cm and 6 cm thick PMMA blocks at Commissioning.
Lag/ghosting	Further investigate means of assessing lag and ghosting as this may be needed if scan times are reduced using multisource x-ray tubes.
Detector response function	Consider a switch to a thicker Al filter such that clinically relevant DAK/image levels can be achieved for the response function. This will also be needed when measuring the response function for the high energy spectrum in contrast enhanced dual energy imaging mode.
Noise decomposition	Routinely apply noise decomposition via a weighted polynomial curve.
System DQE	Measure system DQE at Type Testing stage. This will require a measurement of in-plane MTF, in-plane NNPS and scatter fraction.
Resolution in the z-direction	Consider the use of a thin angled wire, positioned orthogonal to the scan direction and angled from the breast support.

Accepted Manuscript

4. PART 2 - System performance assessment using test objects

Assessment of technical parameters gives a detailed picture of individual component or sub-system performance. For example, the system MTF measured in the projection images depends to some extent on x-ray tube output, physical focus size, pulse-length used for a projection, the angular velocity of the x-tube, x-ray detector MTF and the stability of the gantry. We would expect that tracking this parameter as part of a QC protocol would reveal changes in one or more of these sub-system components. However the link between such parameters and the performance achieved by the imaging system for the clinical tasks undertaken is far from obvious (Barrett et al., 2015). One option is to combine parameters characterizing system signal and noise transfer using an observer model of object detection (ICRU, 1996; Monnin et al., 2020). An alternative, and by far the most common method, is to image a test object containing details that are relevant to the imaging task and to evaluate the resulting object detectability. The test object score combines or integrates many aspects of system performance into a figure of merit that reflects system design choices and the chosen system dose operating point. Limits to modelling and test object realism i.e. the task present in the test object, limit the degree to which clinical performance is estimated, however, more realistic and comprehensive observer models and test objects should lead to evaluations that approach clinical task performance. This section focuses on system evaluation using test objects, and begins with a description of the simpler but nonetheless important job of assessing DBT automatic exposure control performance, which in turn leads on to dosimetry. Developments in the use of test objects to assess image quality are then reviewed.

4.1 Automatic exposure control (AEC)

The AEC system is a crucial element of a mammography imaging device, governing the balance between dose and image quality as breast thickness and breast composition changes (Åslund *et al.*, 2005; Klausz and Shramchenko, 2005). The AEC system has to compensate for changes in breast thickness and transmission. This is done using a certain set of acquisition parameters or 'operating point' in the AEC parameter space, such that the resulting images consistently meet some predefined quality level (Klausz and Shramchenko, 2005). As the technology of mammography imaging systems has advanced over the years, so has AEC design and implementation. Before discussing the evaluation of AEC performance in DBT systems, some background is provided on AEC development and on the phantom materials used for AEC assessment in mammography.

For many years, prior to the introduction of DM and DBT, screen/film (S/F) mammography systems were often used at a single energy and A/F combination; for example, 28 kV and Mo/Mo was used for most breast thicknesses in the UK NHSBSP until the mid-1990s (Young, Ramsdale and Rust, 1996). If higher penetration were needed to image thicker breasts then the tube voltage could be increased to 30 kV. There was evidence that a switch to other tube voltages and A/F combinations for S/F imaging could lead to a dose reduction while largely preserving image quality (Desponds *et al.*, 1991). Towards the end of the 1990s AECs were implemented that selected tube voltage and A/F according to compressed breast thickness

Accepted Manuscript

and/or transmitted radiation measured by the AEC sensor (Young, Ramsdale and Rust, 1996; Thilander-Klang *et al.*, 1997). As a result, dose to thicker breasts could be reduced by between 30% and 75%, depending on A/F combination and breast thickness (Young, Ramsdale and Rust, 1996; Thilander-Klang *et al.*, 1997). The AECs of all S/F systems were set up such that the system reached the target optical density for the clinically relevant compressed breast thickness range, as this influenced target detectability in test objects (Robson, Kotre and Faulkner, 1995) and in patient populations (Kotre, Robson and Faulkner, 1994; Young, Wallis and Ramsdale, 1994).

A major change came with the introduction of DM with integrated FPDs, where the signal from the image receptor itself was used to control the exposure, rather than an AEC sensor positioned behind the S/F cassette (Berns and Hendrick, 2002; Shramchenko and Klausz, 2004). The AEC in an early DM model, the GEHC 2000D, can be used as an example. A pre-exposure is performed and a region extending 14 cm along the chest wall edge and 15 cm out from the chest wall edge is readout in large ROIs of 1×1 cm. The ROI with the lowest mean PV is selected as this generally corresponds to the most attenuating (i.e. densest) area within the AEC sensing region. Using this region to control the main exposure avoids under exposure and thus regions with higher quantum noise. X-ray factors are set such that the densest region receives the required target signal level. AEC implementations and specifics vary between vendors however most DM systems use this type of AEC and this now forms the basis of AEC function in the current generation of DBT systems. There are differences in implementation, for example test shot can be performed in the 0° position (Siemens Revelation), or once the x-ray tube has moved to the largest angle (Hologic Selenia 3Dimensions). There are also differences in the definition of the sensing region and in the size and permitted position of the ROIs used in the pre-exposure. Unfortunately, there is little in the literature describing the design or operation of AEC systems in DBT.

Evaluation of the AEC performance in the EUREF protocol (van Engen *et al.*, 2018) has three main aspects: short term reproducibility of mAs and SNR (measured in 'For Processing' projection images), the performance of the AEC as a function of attenuator thickness and the response of the AEC to changes in local and global breast density. The x-ray factors selected by the AEC versus thickness are then used to make an estimate of mean glandular dose (MGD). The attenuator used to assess AEC should therefore be a reasonable match in terms of attenuation and scattered radiation exiting the phantom, for the range of breasts expected in the population. Numerous breast simulating materials have been explored. Most of these studies define "glandularity" as the percentage by mass of glandular tissue in the breast, excluding the skin layer (in fact many studies use the term 'weight' instead of mass). One can convert between from glandular fraction by mass (f_g) to glandular volume fraction (v_g) using the relationship given by Boone (Boone, 1999):

$$v_g = \left[\frac{(1-f_g)\rho_g}{f_g \quad \rho_a} + 1\right]^{-1} \tag{28a}$$

Accepted Manuscript

where ρ_g is the density of 100% glandular tissue (ρ_g = 1.04 g/cm³ (Hammerstein *et al.*, 1979)) and ρ_a is the density of 100% adipose tissue (ρ_a = 0.92 g/cm³ (Berger *et al.*, 1998)) For a given glandular volume fraction, the glandular fraction by mass is given by:

$$f_g = \left[\frac{(1-v_g)\rho_a}{v_g \quad \rho_g} + 1\right]^{-1} \tag{28b}$$

Stanton *et al* (Stanton *et al.*, 1978) used PMMA containers filled with water and polyethylene (PE) sheets to simulate adipose tissue. The phantom could simulate breast thicknesses of 47 mm, 59 mm and 72 mm, with an average attenuation coefficient within ± 5% of assumed breast composition, in the energy range 15 to 45 keV. Furthermore, three breast types could be simulated: "fatty", "average" and "dense", corresponding to composition of 25/75, 50/50 and 75/25 percentages of water/fat by volume, respectively. White *et al* (White *et al.*, 1977) developed solid phantom mixtures for use as tissue substitutes. The material named 'BR12' simulated a breast with a homogeneous 50/50 mixture of glandular and adipose simulating material by mass, then considered to represent the average breast (Hammerstein *et al.*, 1979). In addition to the 50/50 model, Skubic and Fatouros (Skubic and Fatouros, 1989) considered phantoms with glandular/adipose compositions of 70/30 and 30/70 by mass, manufactured by CIRS Inc.(Norfolk, VA, USA). As might be expected, higher glandular content was found to increase absorbed dose and reduce the contrast of mammographic features.

The study of Argo *et al* (Argo, Hintenlang and Hintenlang, 2004) used a mixture of epoxy resin and associated hardener, PE powder, phenolic microspheres and magnesium powder to produce a solid breast tissue-equivalent series (BRTES) of phantoms that matched the glandularity by mass vs compressed breast thickness distribution found in the work of Geise and Palchevsky (Geise and Palchevsky, 1996). Slabs ranged from thickness of 1.86 cm with a glandularity of 70% up to a series with a composite thickness of 6.96 cm with 20% glandularity.

Despite the accuracy that can be achieved using these materials, many current QC protocols use PMMA as a material for testing AEC devices. PMMA is affordable, easily obtainable and a reasonable surrogate for the breast model currently used (Dance, 1990). For these reasons, PMMA has found widespread use in many mammography QC and dosimetry protocols, although care must be taken that the PMMA slabs are accurately machined to the desired thickness (International Atomic Energy Agency (IAEA), 2007; Reis *et al.*, 2013). The conversion factor, (i.e. the *p* factor in the early Dance model) that converts air kerma incident on a PMMA phantom to that for the standard breast (50/50 composition of glandular to adipose mixture with 5 mm outer shield of adipose tissue) was shown to range between 1.06 and 1.10, depending on HVL (Dance, 1990). As a result, PMMA thickness required to simulate some average breast (model) thickness is between 3 mm and 5 mm less than the breast model thickness. In fact, a number of studies have shown that the glandularity of typical breasts, as a function of breast thickness, deviates from the 50% value initially used in breast dosimetry modelling (Geise and Palchevsky, 1996; Young, Ramsdale and Rust, 1996; Klein *et al.*, 1997; Beckett and Kotre, 2000; Dance, Skinner, *et al.*, 2000; Yaffe *et al.*, 2009). For example Geise and

Accepted Manuscript

Palchevsky (Geise and Palchevsky, 1996) showed for all patients combined in their study, the average thickness was 5.2 cm and average glandularity by mass was 34%. This is in close agreement with the data of Dance *et al* (Dance, Skinner, *et al.*, 2000) where glandularities by mass of 41% and 33% were found in the screening population for compressed breast thicknesses of 4.5 cm and 5.0 cm, respectively. Both of these studies used adipose and glandular tissue simulating blocks together with the delivered tube load (i.e. mAs) to estimate glandularity.

Evaluation of the AEC in the EUREF DBT protocol is performed using sheets of PMMA, which ensures continuity with QC protocols used for DM AEC testing and phantom dosimetry (van Engen *et al.*, 2006). From MC simulations (Dance, Skinner, *et al.*, 2000), PMMA slabs of thickness 20 to 70 mm can be used to simulate breasts in the thickness range 20 to 90 mm. For AECs that rely solely on mechanical (compressed) thickness to set the tube voltage and/or A/F combination, physical spacers must be used such that the AEC selects the factors relevant to the simulated compressed breast thickness. This can be achieved using physical spacers that are radio-translucent or are positioned outside of the AEC sensing region. Bouwman *et al* (Bouwman, 2013) proposed the combined use of PMMA and PE slabs to make a phantom with the relevant thickness and glandularity according to the Dance model (Dance, Skinner, *et al.*, 2000). This avoids the need for spacers, provided the mechanical thickness indicated by the system is accurate (Mawdsley *et al.*, 2009) and the phantom in combination with the compression plate produces the relevant mechanical thickness. The PMMA/PE combined slabs improve the thickness-attenuation match and keep the maximum discrepancy between mean energy absorbed per unit area for the phantom and that for modelled breast to at most 10%, and typically within 6% (Bouwman, 2013). In a study comparing patient dose with dose estimated by PMMA and the PMMA/PE phantoms, Bouwman *et al* (Bouwman *et al.*, 2015) found equally good performance for the two phantoms. Averaged over all breast thicknesses, the ratio of patient to phantom MGD was 1.14 and 1.15 for the PMMA and PMMA/PE phantoms, respectively in DM mode and 1.00 and 1.02 in DBT mode.

A fundamental requirement of an AEC system is the ability to consistently produce the target signal level in a series of images. Table 8 shows results of the short term reproducibility test i.e. maximum deviation from the mean for the mAs, MPV and SNR; all the systems comfortably pass the EUREF protocol limiting values of <5% and <10% for mAs and SNR, respectively. Also shown are results of the local dense area test, in which 2 mm thick plates of PMMA (20 x 10 mm) are used to test whether the AEC can locate and respond to simulated higher density (i.e. higher attenuation) regions. This is a test originally introduced for DM systems (Bouwman *et al.*, 2016), and has now been carried over to DBT systems. The protocol expects signal level within the higher attenuation region to be maintained at the pre-defined level set by the manufacturer for the AEC at a particular thickness; low signal levels could suffer from relatively higher levels of electronic noise (see noise decomposition in section 3.3.2). This is quantified using the SNR measured within a region of the small plates. Table 8 shows that all the AEC systems manage to hold the SNR within the 20% limiting value (van Engen *et al.*, 2018), in line with data presented for DM systems

Accepted Manuscript

(Bouwman *et al.*, 2016). Particularly low variation in SNR is seen for the two Siemens units. Although the current protocol suggests that SNR should be held within limiting values, there could be a change in tube voltage and/or A/F setting rather than just tube load (mAs) as the PMMA blocks are added, and this would lead to a change in SDNR. A future revision of this test may examine the use of SDNR as the parameter to be measured.

Table 8 Short term reproducibility of the AEC measured using the standard test block and the response of the AEC to simulated local dense area. (Results taken from routine QC programme by UZ Leuven Medical Physics Department.)

	Short term reproducibility			local dense area
	Maximum deviation on the mean			Maximum deviation on the mean
System	mAs (%)	MPV(%)	SNR(%)	SNR(%)
Fuji Innovality ST	0.2	0.2	0.9	18.0
GE Pristina	1.3	0.7	0.5	12.0
Giotto Class	1.3	1.1	2.2	16.1
Hologic Selenia Dimensions	0.0	0.9	1.5	18.2
Hologic 3Dimensions	0.0	0.8	1.0	17.6
Planmed Clarity	0.6	0.5	1.7	18.0
Siemens Inspiration: STD	0.2	0.3	1.3	1.7
Siemens Revelation: EMPIRE	0.1	0.3	1.9	1.8

Table 9 lists results of the AEC evaluation of eight DBT systems as described in the EUREF protocol using PMMA slabs in conjunction with spacers positioned outside of the AEC sensing region. The tables show the MPV and standard deviation measured in the background near the Al square using a 5 × 5 mm ROI, along with MPV measured within the 0.2 mm Al square. These values have been linearized to an approximate DAK using the inverse of the measured response function (section 3.3.1) (Mackenzie, 2008). Also shown are the contrast of the 0.2 mm Al square along with the SDNR calculated from the linearized MPV and standard deviation values as a function of PMMA thickness. The SDNR values for each system are presented normalized to the SDNR for the 50 mm PMMA block, as SDNR is not directly comparable between different DBT systems, given the differences in image processing, pixel apertures and presampling blurring that can be present (Burgess, 1999). The thickness of 50 mm was chosen as this represents a 60 mm breast thickness (Dance, Skinner, *et al.*, 2000), which is within the range of average breast thicknesses found in surveys for DM imaging (Robinson, 2008; Hendrick *et al.*, 2010; Kelaranta *et al.*, 2015).

Accepted Manuscript

Table 9 AEC performance data for DBT systems. Linearized MPV and standard deviation (σ) measured in the image background, and MPV measured within the 0.2 mm thick Al detail. Contrast for the Al detail and SDNR relative to SDNR for a 50 mm PMMA block. Also listed is the mean glandular dose (MGD) for DBT mode and along with the ratio of MGD for DBT to MGD for DM mode. (Results taken from routine QC programme by UZ Leuven Medical Physics Department.)

PMMA (mm)	A/F	kV	mAs	linearized MPV Al (~μGy)	linearized MPV bkg (~μGy)	linearized σ (~μGy)	contrast (%)	relative SDNR	MGD DBT (mGy)	MGD DBT : MGD DM
Fuji Innovality ST										
20	W/Al	27	28.1	17.8	20.1	0.33	11.65	1.74	0.98	1.93
30	W/Al	29	29.1	19.3	21.2	0.35	8.92	1.31	1.18	1.50
40	W/Al	31	37.2	26.6	28.6	0.43	7.14	1.15	1.59	1.55
45	W/Al	32	44.8	33.2	35.5	0.48	6.52	1.16	1.86	1.54
50	W/Al	33	50.4	38.6	40.9	0.56	5.65	1.00	2.12	1.55
60	W/Al	36	56.7	55.1	57.7	0.70	4.58	0.92	2.70	1.28
70	W/Al	37	70.2	59.8	62.0	0.70	3.60	0.78	3.06	1.05
Fuji Innovality HR										
20	W/Al	27	33.5	19.6	22.2	0.47	11.57	1.57	1.13	2.24
30	W/Al	29	43.8	26.9	29.3	0.56	8.48	1.27	1.44	1.82
40	W/Al	31	60.4	39.8	42.6	0.72	6.55	1.11	2.05	2.00
45	W/Al	32	72.5	49.3	52.3	0.80	5.70	1.06	2.55	2.12
50	W/Al	33	79.6	55.8	58.8	0.86	5.12	1.00	2.94	2.15
60	W/Al	35	85	62.7	65.3	0.95	4.07	0.80	3.43	1.62
70	W/Al	37	85.3	65.7	67.9	0.99	3.16	0.62	3.68	1.26
GEHC Pristina										
20	Mo/Mo	26	18.1	6.7	8.4	0.30	20.43	1.39	0.51	1.11
30	Mo/Mo	26	38.9	6.7	8.3	0.29	19.28	1.32	0.84	0.98
40	Rh/Ag	34	25.4	9.6	11.0	0.27	12.05	1.16	1.09	0.99
45	Rh/Ag	34	29.5	8.6	9.7	0.26	11.70	1.04	1.24	1.02
50	Rh/Ag	34	35.9	8.1	9.1	0.25	11.47	1.00	1.38	0.99
60	Rh/Ag	34	53.7	7.4	8.2	0.24	10.58	0.86	1.85	1.01
70	Rh/Ag	34	84.6	7.0	7.8	0.24	10.08	0.80	2.59	0.97
Giotto Class										
20	W/Ag	25	68.5	15.6	17.5	0.55	10.41	1.60	0.89	1.34
30	W/Ag	27	68	26.4	28.8	0.72	8.26	1.59	0.95	1.35
40	W/Ag	28	101.7	22.5	24.0	0.67	6.37	1.10	1.37	1.33
45	W/Ag	30	108.8	29.5	31.3	0.77	5.77	1.13	1.80	1.31
50	W/Ag	31	120.2	30.5	32.1	0.80	5.16	1.00	2.10	1.30
60	W/Ag	32	178	35.2	36.7	0.86	4.15	0.86	3.12	1.28
70	W/Ag	34	219.4	40.8	42.1	0.96	2.98	0.63	4.22	1.25

Accepted Manuscript

Table 9 cont. AEC performance data for DBT systems. Linearized MPV and standard deviation (σ) measured in the image background, and MPV measured within the 0.2 mm thick Al detail. Contrast for the Al detail and SDNR relative to SDNR for a 50 mm PMMA block. Also listed is the mean glandular dose (MGD) for DBT mode and along with the ratio of MGD for DBT to MGD for DM mode.

PMMA (mm)	A/F	kV	mAs	linearized PV Al (~μGy)	linearized PV bkg (~μGy)	linearized σ (~μGy)	contrast (%)	relative SDNR	MGD DBT (mGy)	MGD DBT : MGD DM
Hologic Selenia Dimensions										
20	W/Al	26	42	11.2	12.8	0.30	12.75	1.73	0.95	1.48
30	W/Al	28	44	13.2	14.6	0.36	9.58	1.23	1.00	1.16
40	W/Al	30	52	18.2	19.7	0.43	7.45	1.10	1.24	1.06
45	W/Al	31	65	25.3	27.0	0.50	6.29	1.09	1.61	1.08
50	W/Al	33	67	34.4	36.3	0.59	5.11	1.00	1.93	1.10
60	W/Al	36	84	60.8	62.8	0.85	3.13	0.74	2.91	0.97
70	W/Al	42	78	111.5	113.1	1.22	1.48	0.44	4.01	1.01
Hologic 3Dimensions										
20	W/Al	26	40	12.3	14.1	0.63	12.37	1.50	1.06	1.53
30	W/Al	27	62	17.0	18.8	0.74	9.89	1.37	1.45	1.71
40	W/Al	30	53	22.1	23.8	0.90	7.06	1.03	1.47	1.23
45	W/Al	32	59	33.1	35.1	1.17	5.95	0.98	1.90	1.19
50	W/Al	33	71	43.0	45.4	1.32	5.31	1.00	2.36	1.14
60	W/Al	36	81	65.9	68.7	1.73	4.07	0.89	3.26	1.29
70	W/Al	42	70	112.1	115.3	2.47	2.77	0.71	4.05	1.19
Planmed Clarity										
20	W/Ag	29	57	29.0	32.1	0.66	9.60	1.71	1.09	1.94
30	W/Ag	29	83	26.9	29.3	0.61	8.26	1.45	1.32	1.60
40	W/Ag	30	119	28.4	30.4	0.65	6.75	1.17	1.85	1.58
45	W/Ag	31	133	29.0	30.9	0.67	6.12	1.04	2.17	1.55
50	W/Ag	32	154	30.6	32.4	0.68	5.68	1.00	2.67	2.17
60	W/Ag	33	183	27.2	28.5	0.65	4.77	0.77	3.14	1.95
70	W/Ag	34	219	24.6	25.6	0.61	3.80	0.58	3.68	1.67
Siemens Inspiration										
20	W/Rh	26	63	8.8	10.5	0.51	15.49	1.47	0.84	1.26
30	W/Rh	27	91	8.8	10.2	0.51	14.14	1.30	1.05	2.32
40	W/Rh	28	132	8.7	9.9	0.52	12.59	1.10	1.38	2.10
45	W/Rh	29	147	8.7	9.9	0.51	12.01	1.07	1.56	1.68
50	W/Rh	30	163	8.7	9.8	0.51	11.34	1.00	1.72	1.53
60	W/Rh	31	223	8.8	9.8	0.53	10.21	0.86	2.26	1.26
70	W/Rh	32	297	9.0	9.9	0.56	9.11	0.74	2.85	1.29
Siemens Revelation										
20	W/Rh	26	71	10.8	12.5	0.59	14.29	1.45	1.06	2.47
30	W/Rh	27	103	10.7	12.2	0.59	12.81	1.28	1.36	2.32
40	W/Rh	28	150	10.7	12.1	0.60	11.65	1.12	1.82	2.10
45	W/Rh	29	166	10.6	11.9	0.58	11.13	1.08	2.16	2.26
50	W/Rh	30	184	10.7	11.9	0.58	10.27	1.00	2.42	1.98
60	W/Rh	31	255	11.0	12.1	0.63	9.04	0.83	3.23	1.67
70	W/Rh	32	334	11.1	12.1	0.64	8.27	0.75	4.04	1.55

Accepted Manuscript

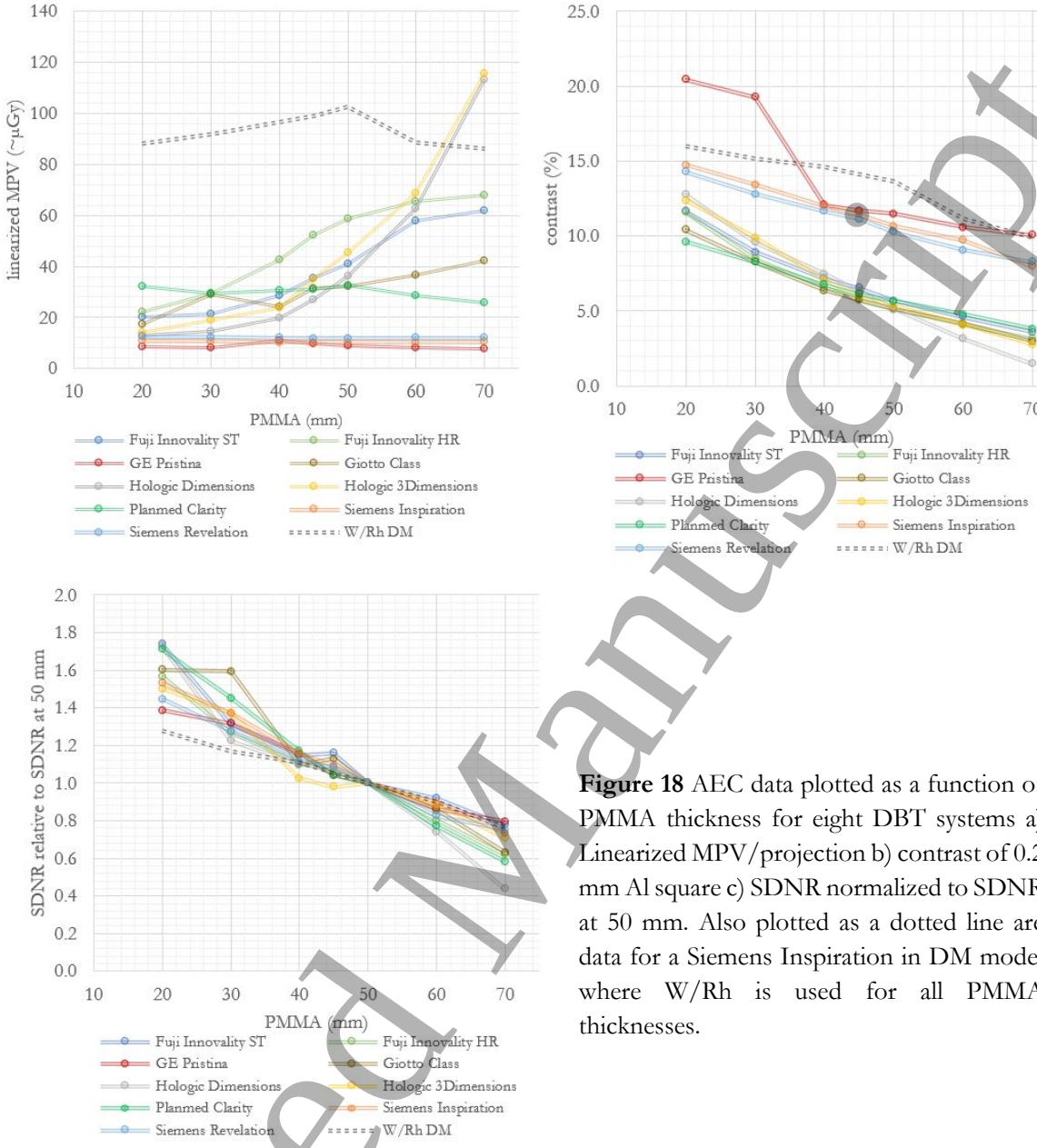


Figure 18 AEC data plotted as a function of PMMA thickness for eight DBT systems a) Linearized MPV/projection b) contrast of 0.2 mm Al square c) SDNR normalized to SDNR at 50 mm. Also plotted as a dotted line are data for a Siemens Inspiration in DM mode, where W/Rh is used for all PMMA thicknesses.

The data in Figure 18a show large differences in the way linearized MPV is modulated in the projection images as PMMA thickness is changed. The Siemens Inspiration and Revelation, the GEHC Pristina and the Planmed Clarity all hold linearized MPV approximately constant as PMMA thickness increases, with the linearized MPV for the Planmed also constant but consistently higher by a factor of ~3. One reason for the lower linearized MPV/projection for the Siemens units is the greater number of projections, compared to the other systems. Linearized MPV/projection is low for the GEHC due to the low total scan dose for this unit. The other units generally have a higher MPV and the MPV is programmed to increase as PMMA thickness increases, especially notable for the Hologic Dimensions and 3Dimensions systems. This modulation of MPV/projection as object thickness increases is consistent with the findings of Salvagnini *et al* (Salvagnini *et al.*, 2016) for DM systems, where dose/image was increased for thicker breasts,

Accepted Manuscript

increasing lesion detectability (primarily for microcalcifications). Also shown in this figure are linearized MPV data for a Siemens Inspiration in DM mode, where MPV is held approximately constant with PMMA thickness, at ~90 μGy/image.

Figure 18b shows results for measured contrast of the Al square for the different systems. While it is recognized that contrast is of less direct importance for imaging performance of DM and DBT systems than for F/S imaging (Dance, Thilander Klang, *et al.*, 2000; Åslund *et al.*, 2005), the signal difference generated by the structures of interest will influence task performance, e.g. detectability, regardless of imaging modality (Prakash *et al.*, 2011). The contrast gives some basic insight into the spectral differences between systems. Contrast of the Al square is highest for the GEHC Pristina. This system uses an antiscatter grid, and in conjunction with the Mo/Mo A/F setting, this results in high contrast for the 20 mm and 30 mm PMMA blocks. There is a step loss in contrast as the Rh/Ag A/F setting is engaged, but the grid helps to maintain contrast at the high PMMA thicknesses. For the Siemens systems, the W/Rh setting produces systematically higher contrast than that for the group of DBT systems including the Hologic Dimensions, Hologic 3Dimensions, Giotto Class, Planmed Clarity and Fuji Innovality, where W/Ag and W/Al A/F settings are used. For systems with lower contrast, the higher MPV/projection seen in Figure 18, especially as breast equivalent thickness increases, helps to reduce noise in the projections and allows the imaging of these smaller signal differences (Salvagnini *et al.*, 2016). Figure 18b also plots contrast for the Siemens Inspiration operated in DM mode, where the W/Rh A/F setting gives consistently higher contrasts due to the use of the antiscatter grid.

Figure 18c shows that similar relative SDNR performance is found for the DBT systems, regardless of the results for the MPV/image or the contrast of the Al square. There is a consistent reduction in SDNR as PMMA thickness increases, due to the reduction in signal difference from scattered radiation and increased mean energy of the x-ray spectrum. The relative SDNR reduction is steeper than that typically found in DM mode, shown by the example result for the Siemens Inspiration DM using W/Rh in Figure 18c. A similar result was found (Rodríguez-Ruiz *et al.*, 2016) for the Siemens Inspiration, Hologic Dimensions and Fuji Innovality systems.

4.2 Dosimetry

In the EUREF protocol (van Engen *et al.*, 2018), the AEC data are used to calculate doses expected for typical breasts as simulated by PMMA using the dosimetry model of Dance *et al* (Dance, Skinner, *et al.*, 2000), modified to account for the influence of tube rotation around the breast (Dance, Young and van Engen, 2011). The review of dosimetry in x-ray based breast imaging by Dance and Sechopoulos (Dance and Sechopoulos, 2016) describes the methods of both Sechopoulos (Ioannis Sechopoulos *et al.*, 2007; Sechopoulos *et al.*, 2014) and Dance (Dance, Young and van Engen, 2011) focuses on DM imaging but can be extrapolated to DBT imaging. The obvious difference in DBT is source rotation, to a smaller or larger degree, around the breast, influencing the calculated dose conversion coefficients. Both studies use factors that relate dose at some projection angle to the dose for the 0° projection, e.g. the relative glandular dose

Accepted Manuscript

coefficient of Sechopoulos *et al* (Sechopoulos and D'Orsi, 2008). Dance *et al* (Dance, Young and van Engen, 2011) introduced the tomosynthesis factor, $t(\theta)$, calculated from the mean glandular dose at projection angle θ divided by the dose for the 0° projection:

$$t(\theta) = \frac{D(\theta)}{D(0°)} \tag{29}$$

The mean glandular dose for a complete scan (D_T) is then given by

$$D_T = K_T gcsT \qquad \text{where} \quad T = \frac{1}{N}\sum_i t(\theta_i) \tag{30}$$

for the case where the dose per projection is held constant (currently the case for all systems). Dance *et al* (Dance, Young and van Engen, 2011) included tabulations of $t(\theta)$ such that MGD could be calculated for individual projections and also of T for the angular ranges currently used by the vendors. Additional g, c and s factors were also calculated to cover the higher energies used in DBT.

In terms of dosimetry modelling and calculation, there is continued focus on the spatial distribution of the glandular tissues within the breast model and on the skin thickness used. Sarno *et al* (Sarno et al., 2017) used GEANT4 MC simulations to investigate the influence of a number of factors on the magnitude of normalized glandular dose (DgN) factors for DM. Inclusion of bremsstrahlung radiation and compression paddles of thickness up to 5 mm had only a small effect on DgN values, within the range of 1 – 2%. Sarno *et al* (Sarno et al., 2017) used the skin thickness of 1.45 mm originally found from breast CT data by Huang *et al* (Huang *et al*., 2008); this produced an increase of up to 19% in DgN compared to the 4 mm skin thickness used in the Boone model (Nosratieh *et al*., 2015).

The INFN group have recently generated a set of normalized glandular dose coefficients for DBT (DgN_{DBT}) using a homogenous breast model (Sarno *et al*., 2021) with a 1.45 mm thick skin envelope. These coefficients take as input the air kerma measured at the breast surface for a 0° projection along with the tube load (mAs) for the complete scan. Instead of using a tomosynthesis correction factor applied to glandular dose coefficients calculated for DM (Ioannis Sechopoulos *et al*., 2007; Dance, Young and van Engen, 2011), the DgN_{DBT} coefficients are calculated directly as a function of projection angle using data specific to one of six commercially available systems (Fujifilm Innovality, GEHC SenoClaire, GEHC Pristina, Hologic Selenia Dimensions, Siemens Inspiration and IMS Giotto Class). This includes setting the source-to-isocentre and air gap distances, compression plate thickness (2.0 – 2.7 mm) and composition (typically polycarbonate). DgN_{DBT} factors are presented as [mGy mGy^{-1}] as a function of compressed breast thickness (and glandular fraction by mass) for the six systems along with the factor *gcsT* [mGy mGy^{-1}] from the Dance method as implemented in the EUREF protocol. Maximum relative difference from the Dance results ranged from -17.6% to +6.2%, with the larger deviations occurring at the smaller breast thicknesses. These deviations arose from differences in breast models and the scanner specific calculation. An Excel workbook is provided in the supplementary material that interpolates DgN_{DBT} for a given input (DBT model, compressed breast thickness, etc).

Accepted Manuscript

In 2012, Sechopoulos *et al* (Sechopoulos *et al.*, 2012) used a dataset of 20 breast CT images to study the influence of the actual, heterogeneous distribution of glandular tissue on the calculated normalized glandular dose. This was done for a breast CT geometry and, by simulating the compression process, for projection mammography. To estimate normalized glandular dose for the homogeneous case, the heterogeneous distribution was replaced by a homogeneous mixture of glandular and adipose tissues such that the glandular fraction corresponded to the volumetric breast glandular density of a given breast. The homogenous tissue mixture approximation was found to overestimate the normalized glandular dose on average by 27% for the particular dataset of 20 breasts used. This finding was consistent with an earlier initial study made by Dance *et al* (Dance *et al.*, 2005) that highlighted the importance of knowing the position of the glandular for accurate breast dosimetry. Hernandez *et al* (Hernandez, Seibert and Boone, 2015) subsequently confirmed this for a larger dataset of 219 breast CT images, where the overestimate arising from using the homogenous mixture assumption was found to be ~30%. To do this, Hernandez *et al* (Hernandez, Seibert and Boone, 2015) introduced distributions of glandular tissue derived from the breast CT datasets that had a Gaussian shape and were centred in the coronal plane.

These improvements in breast modelling accuracy are currently being consolidated by a joint EFOMP and AAPM Task Group ("Development of a new universal breast dosimetry method (TG282)") that is developing a new model and dosimetry method covering DM, DBT, contrast enhanced mammography and DBT. The new model utilizes improved knowledge of the spatial position of the glandular parenchyma within the breast (Fedon et al., 2021) and covers DM, DBT, contrast enhanced mammography and breast computed tomography (BCT) modalities. A new set of breast heterogeneous digital breast phantoms incorporating using these results have been developed for breast dosimetry in DM and DBT (Caballo et al., 2022).

It is important to note that even with the improved modelling that is underway in TG282, this will not result in patient specific doses. Glandular dose to a specific breast can only be estimated when there is detailed knowledge of actual glandular tissue distribution (Dance *et al.*, 2005), the skin and adipose layer thickness and true compressed breast thickness and this information is unlikely to be available for projection imaging or DBT, where depth resolution is severely limited. Recent work by Teuwen *et al* (Teuwen *et al.*, 2021) describes a deep learning based reconstruction technique, using neural networks trained using virtual phantoms from two sources, that improved the accuracy of the breast density estimation which then enabled improved accuracy in the glandular dose estimate. Density and dose estimates within respectively ±3% and ±20% were achieved without bias.

If the dose is to be used in a risk/benefit calculation, then a point is reached when the combined influence of other factors begins to limit the accuracy of calculation, including the free atom assumption used in the MC transport calculation (Dance and Sechopoulos, 2016), the equation used to estimate risk and the choice of the dose and dose rate effectiveness factor (DDREF) (BEIR, 2006; ICRP, 2007; Warren, Dance and Young, 2016). Absolute accuracy of glandular tissue distributions is less important when comparing systems

Accepted Manuscript

or modalities; the models should reflect the differences in dose due to irradiation geometry and the energy used; this is the case for the currently used models (Dance, Young and van Engen, 2011; Sechopoulos *et al.*, 2014).The T factors are generally >0.93 for current systems, for breast thickness up to 110 mm.

There are studies presenting doses from DBT systems, but perhaps not as many as might be expected. The study by Svahn *et al*(Svahn *et al.*, 2015) reviewed patient dose data for DBT systems used in early clinical studies. Two view DM was considered as the reference modality and doses from one view DBT, two view DBT and DBT used as an adjunct to DM were compared by calculating the dose ratio D_{DBT}/D_{DM}. Data from clinical studies published between 2008 and 2014 were included, and all of the systems except the Hologic Dimensions were investigational units. It must be noted that many of the systems in the study are no longer available or under development (Siemens Novation, GEHC Senograph DS, Xcounter Xmamo-3T, and the Sectra (Philips) MicroDose). Only the Sectra MicroDose and Hologic Dimensions utilized an AEC, x-ray factors for the remaining units were set manually. Dosimetry data were taken from two sources. Option one was to make acquisition of a test block which represents a 45 mm thick breast of 50%/50% glandular to adipose tissue composition, as defined by the American College of Radiology (ACR), from which MGD could be estimated (Boone, 1999; Sechopoulos *et al.*, 2014). Option two was to use the MGD displayed by the imaging system, calculated from the patient acquisition factors using a specified method (Wu, Barnes and Tucker, 1991; Wu *et al.*, 1994; Dance, Skinner, *et al.*, 2000; Dance, Young and van Engen, 2011). Svahn *et al* (Svahn *et al.*, 2015) found dose ratios (DBT to DM) in the range 0.34 -1.0 for one-view DBT used stand-alone, while for two view stand-alone DBT this ratio was 0.68 – 1.17. When combined with DM, the dose ratio for one view DBT was between 1.03 – 1.5 and for two view DBT this ratio was 2.0 – 2.23, compared to two view standalone DM.

The Hologic Selenia Dimensions system has been used in a number of trials and consequently more dose results are available for this system in the literature. In the study by Gennaro *et al* (Gennaro, Bernardi and Houssami, 2018), acquisition data for a subset of patients from the STORM-2 prospective trial were analysed. Images were acquired using a Hologic Selenia Dimensions system under AEC control in COMBO mode, enabling a direct comparison between DM and DBT exposures. For the CC view, MGD for DM was 1.37 mGy compared to 1.86 mGy for DBT, indicating an increase of 38% over DM, when averaged over all CCs. The Selenia Dimensions system was also used in the Oslo trial where the average MGD (assumed averaged over CC and MLO views) for DM and DBT was 1.58 mGy and 1.95 mGy, respectively i.e. an increase of 23% (Skaane *et al.*, 2013). From the STORM-2 data, the MGD for the CC and MLO averaged together was respectively 1.37 mGy and 1.88 mGy for DM and DBT, giving an increase of 37%. Thus, the doses for the Oslo trial were slightly higher, with a smaller difference between DM and DBT. A recent study by Ritlumbert *et al* (Ritlumlert *et al.*, 2020) reported dose results calculated following IAEA Report 457 (International Atomic Energy Agency (IAEA), 2007) for a group of 50 patients, again imaged using a Hologic Selenia Dimensions. Averaged MGD was 1.36 mGy and 1.63 mGy, for DM and DBT modes respectively, indicating an increase of 20%. Finally, Østerås *et al* (Østerås *et al.*, 2018) returned to the

Accepted Manuscript

data from the Oslo trial to perform a detailed analysis of dose, including the influence of breast density on dose quantified using QUANTRA™ breast density software, as did the study of Gennaro *et al* (Gennaro, Bernardi and Houssami, 2018) where Volpara breast density estimation software was used. The volumetric breast density measured by QUANTRA™ was converted to glandularity and then used to select the appropriate c-factor for the MGD calculation (equation (30)) (Dance, Young and van Engen, 2011). DICOM meta-data were used along with measured radiation output, HVL and breast thickness to calculate MGD from the patient data (van Engen *et al.*, 2018). Disregarding breast type, the ratio of DBT/DM dose per view was 1.24, close to the figure of 23% originally reported for this study (Skaane *et al.*, 2013). Breast doses were also analysed in terms of 'fatty' and 'dense' classes, as determined by QUANTRA™. When imaging dense breasts, MGD was higher for DM acquisitions and similar for DBT, leading to lower DBT/DM dose ratios for dense breasts (1.08) compared to fatty breasts (1.33).

The Siemens Inspiration system used in the Malmö breast screening trial (Lång *et al.*, 2015), where a single MLO view was acquired for the DBT part of the study, gave an MGD of 1.2 mGy and 1.6 mGy for DM and DBT. Note that this was measured using a 45 mm PMMA block, i.e. an equivalent breast thickness of ~53 mm breast, following the EUREF protocol (van Engen *et al.*, 2018). This is an increase of 33% for DBT mode over DM, per view.

Looking further at test object QC data, Figures 15a and 15b plot MGD measured using PMMA blocks together with spacers positioned outside the AEC sensing regions to set the relevant breast equivalent thickness for 7 systems in DM and DBT modes. These data show quite a variation in MGD between systems in DBT mode, for all PMMA thicknesses. The GEHC Pristina is clearly operated at a lower MGD than the other systems. Acquisition factors are the same for DM and DBT for the Pristina, leading to the ~same dose for the two modes, apart from the influence of the T factor. The DM data show that doses are generally lower than DBT doses with less spread between systems for PMMA thicknesses up to about 50 mm, while there is greater spread above this thickness in DBT. Some DM systems, for example the Hologic units, increase the MPV above 60mm PMMA equivalent, similar to that seen for DBT in Figure 19a and this leads to slightly higher MGD. Figure 19c shows the ratio of MGD in DBT mode to that in DM mode averaged over PMMA thickness, with the spread in values quantified using the standard deviation (shown by the error bar). Averaging over all thickness and systems, we could say that MGD per view is a factor of ~1.5 higher for DBT than for DM, however this is too simplistic as some systems have no dose increase (GEHC Pristina) or a somewhat smaller increase (~1.1 to 1.3). The dose increase for the Hologic units is smaller due to the fact that dose in DM mode for these systems is typically higher than for the other systems. The standard deviation in Figure 19c is low for the Pristina and Giotto Class, showing that dose in DBT is the same as in DM (Pristina) or is a constant factor higher, independent of thickness (Giotto Class). At most, DBT dose per view is a factor of ~2 higher for DBT compared to DM, for example for the Fuji Innovality HR and the Siemens units. These results are similar to those found by Maldera *et al* (Maldera *et al.*, 2017) for the Hologic, Siemens, Fuji and GEHC Essential systems.

Accepted Manuscript

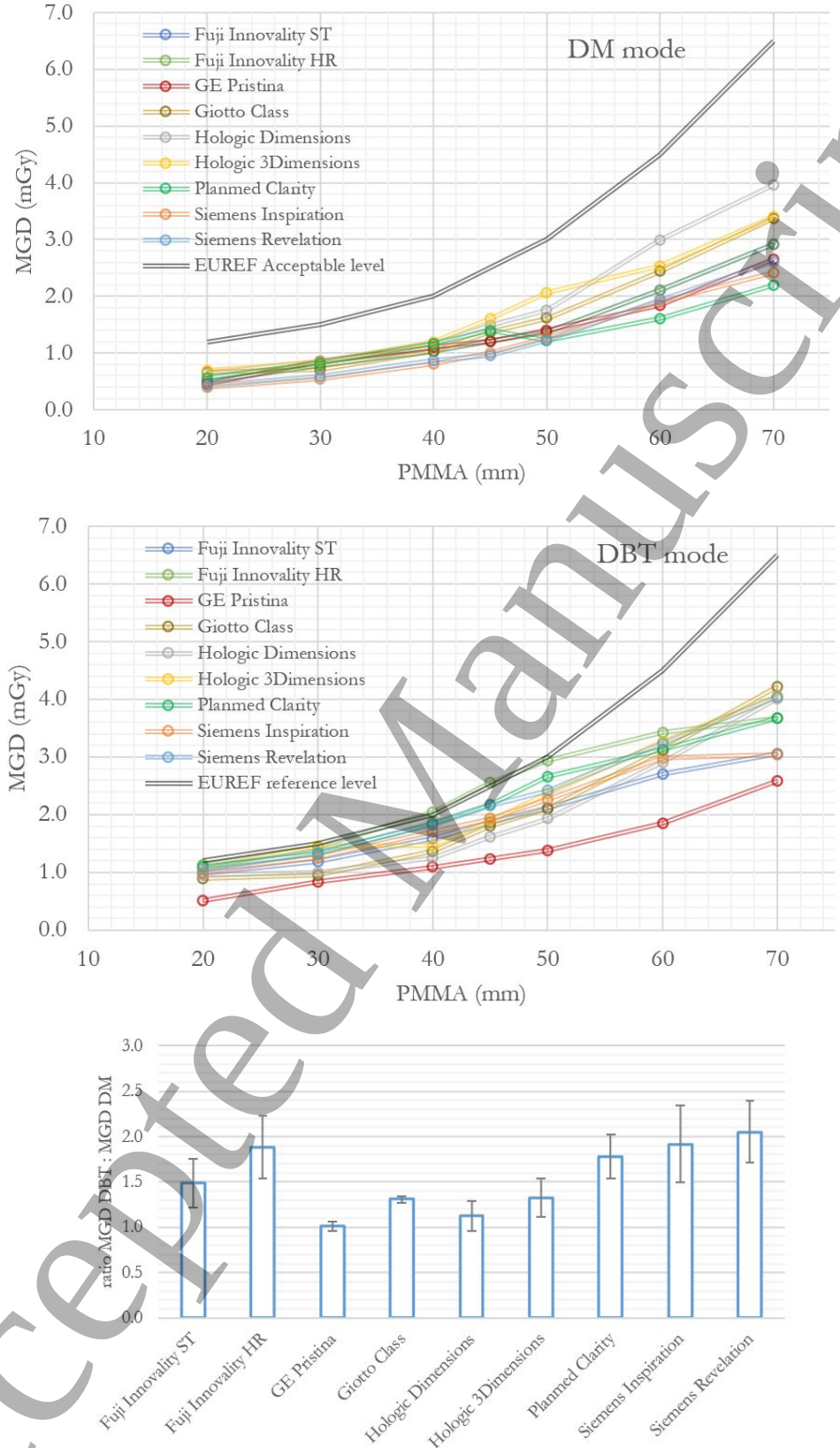


Figure 19 MGD assessed using PMMA blocks and spacers for eight DBT systems operated in a) DM mode b) DBT mode; c) ratio of MGD in DBT mode to MGD in DM mode, averaged for PMMA thicknesses 2 to 7 cm (error bar shows the standard deviation of the ratio).

Accepted Manuscript

4.3 Technical image quality assessment with test objects

This section discusses the evaluation of imaging performance using test objects, defined here as a measurement of *technical image quality* in order to distinguish this from assessment of *clinical image quality*, which relates to the task performance of an imaging modality with radiologists reading real patient data, under clinical or close-to-clinical working conditions. Obviously clinical IQ is crucial and is one of the determining factors of system performance in controlled clinical studies and even large trials. Over the years there has been strong interest in establishing whether DBT offers an improvement in terms of cancer detection rate and interval cancer detection when used in population breast screening, in comparison to the current standard of DM imaging (Houssami *et al.*, 2021). Detection tasks therefore play a vital role in determining system performance, and in these studies, systems are used to image a wide ranging sample of the lesions present in the screening population, set against a varying range of anatomical backgrounds and breast types. This gives a robust estimate of the detection performance of the system for the relevant tasks. It is currently not possible to simulate these conditions using physical test objects, however measurements of technical image quality using test objects are valuable and have their place in a quality assurance programme, provided the limitations of the methods are recognised (Yaffe *et al.*, 2006). Examining the implementation of technical image quality measurements for DM imaging throws some light on what might be expected or required for these evaluations in DBT imaging. We distinguish between test objects with targets in a homogeneous background and those with structured backgrounds designed to simulate or mimic breast structure. We further consider two separate ways in which test objects are used: i) testing system performance against some established reference level and ii) tracking system performance over time.

4.3.1 Homogeneous test objects

Minimum performance level – testing against a standard

We will first discuss comparison against a reference level using homogeneous test objects. In the USA, mammography facilities have to meet the Mammography Quality Standards Act (MQSA) which includes an evaluation of image quality using the American College of Radiology (ACR) phantom. This phantom is imaged under AEC control, and one image that has been reviewed and approved by the radiologist is submitted to the ACR. For the original accreditation phantom used with S/F systems and with DM systems in the DMIST trial, four fibres, three speck groups and three masses must be seen by a trained human reader for the system to pass. For the new DM version of the accreditation phantom, two fibres, three speck groups and two masses must be seen. This standard applies to DM systems used for both screening and diagnostic examinations. In Europe, many countries adopt the European Protocol (van Engen *et al.*, 2006) which specifies image quality criteria in terms of threshold gold thickness for circular gold discs ranging in diameter from 0.1 mm to 2.0 mm, measured using the CDMAM contrast-detail (c-d) test object. These minimum quality criteria were set such that the c-d performance of DM systems was the same or superior to S/F systems used in breast screening programmes at the time (K. Young *et al.*, 2006). Given

Accepted Manuscript

that S/F and DM are both projection-based imaging modalities using similar x-ray energies, it was reasonable to assume that this means of specifying image quality is valid for both modalities. In doing this, the assumption is made that small detail detectability as measured using CDMAM (i.e. disc targets against a homogeneous background, without clinical image processing applied) sufficiently characterizes or predicts the performance of DM and S/F systems when performing typical mammography imaging tasks, such as detection of microcalcifications and mass lesions by radiologists. The influence of the projected overlying anatomical information that reduces sensitivity in 2D mammography, is not quantified using CDMAM but is assumed to be the same for both S/F and DM imaging. Only later were studies published that linked c-d, measured using CDMAM, with radiologist microcalcification detection, evaluated using a JAFROC detection study of simulated microcalcifications, for a number of different DM technologies (Warren *et al.*, 2012; Mackenzie *et al.*, 2016; Salvagnini *et al.*, 2016). The focus of CDMAM testing is generally on small detail detectability, with the assumption that microcalcification detection plays an important role in determining overall performance of system in a screening programme.

When trying to establish the technical image quality performance of DBT systems, comparison against 2D projection imaging in the form of DM is useful as this is the current standard for mammography x-ray imaging. If the previous EUREF approach is adopted, in which DM imaging had to achieve or surpass the quality criteria of S/F, then one would require DBT to provide the same or superior performance using a test object that can be imaged on both DM and DBT devices. However this requires a suitable test object containing relevant anatomy simulating structures. Phantoms without breast simulating structures, such as CDMAM, would bias the result towards DM as no overlying anatomical structures are present, and acquiring all of the projection information in a single, high dose acquisition made at 0° would give the best test object score (normalized for dose) i.e. DM has a grid to control scattered radiation, very little electronic noise, 'default' or the minimum oblique entry blurring, no focus motion blurring, no interpolation in the reconstruction step. The improved detectability expected from DBT systems due to reduced anatomical clutter would not be tested.

In contrast to the EUREF DM protocol, no minimum image quality standards are given in EUREF DBT protocol (van Engen *et al.*, 2018), mainly due to the lack of a validated, commercially available test object with targets and structured background. This is echoed by the IEC DBT Standard (IEC, 2020b), where a test object requiring a breast simulating structured background is specified. Instead, the EUREF protocol describes a test of the image quality stability in the x-y plane, to be done routinely using a test object such as the CDMAM or Leeds TORMAM (Objects, 2021). The focus is therefore on longitudinal testing with the aim of detecting changes in system performance.

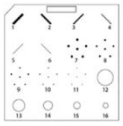
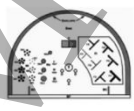
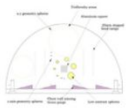
Tracking changes – longitudinal QC testing

Most test objects used to test against a standard can also be used to monitor DBT system imaging stability over time. Sage *et al* (Sage *et al.*, 2019) have reviewed seven test objects for QC testing, although only three were explicitly developed to assess DBT devices. Five of these phantoms are manufactured by CIRS

Accepted Manuscript

(Norfolk, VA, USA) (CIRS, 2021), the TOMOPHAN is produced by The Phantom Laboratory (Salem, NY, USA) (*The Phantom Laboratory DBT phantom*, 2021) while Quart (Zorneding, DE) (QUART-Gmbh, 2021) produces the Mam Digi/EPQC test object. Detailed information is available on the design and construction of the TOMOMAM test object, including software for the readout of various metrics (Fitton *et al.*, 2022). Note that other phantoms are available such as the Agatha phantom, produced by Leeds Test Objects (Boroughbridge, UK) and in fact CDMAM or TORMAM could be used, as described in the EUREF protocol (van Engen *et al.*, 2018).

Table 10 Principal specifications of the seven phantoms included in the study of Sage *et al* (Sage *et al.*, 2019)

Manufacturer	CIRS	CIRS	CIRS	CIRS	CIRS	PHANTOM LABORATORY	QUART
Model	ACR Mammographic Accreditation Phantom Model 015	BR3D Breast Imaging Phantom Model 020	DBT QC Model 021 (prototype version)	MTM 100	TOMOMAM*	TOMOPHAN* TSP004	Mam Digi/EPQC
Shape	Rectangular	Semi-circular	Semi-circular	Semi-circular	Semi-circular	Semi-circular	Semi-circular
Dimensions	108 mm × 102 mm	100 mm × 180 mm	100 mm × 180 mm	180 mm × 130 mm	ϕ:190 mm	100 mm × 180 mm	180 mm × 240 mm
Thickness (mm)	44	60 6 plates of 10 mm	60 6 plates of 10 mm	45	60	42 or 56[c]	45
Background	Homogeneous - PMMA	Homogeneous – Epoxy resin[a]	Homogeneous	Homogeneous – Epoxy resin[b]	Homogeneous – Epoxy resin[a]	Homogeneous	Homogeneous
Microcalcifications (groups)	5 (Al_2O_3) From 0.16 to 0.54 mm	6 ($CaCO_3$) From 0.13 to 0.40 mm	6 NC	7 From 0.09 to 0.30 mm	6 (glass) From 0.14 to 0.33 mm	/	12[d] From 0.18 to 0.60 mm
Masses	5 From 0.25 to 2 mm	6 From 1.8 to 6.3 mm	6 NC	7 From 1.8 to 6.3 mm	6 From 1.8 to 5.5 mm	8 (acrylic) From 1 to 10 mm	/
Fibres	6 (nylon) From 0.4 to 1.56 mm	7 From 0.15 to 0.60 mm	/	7 (nylon) From 0.30 to 1.25 mm	6 (nylon) From 0.15 to 0.41 mm	/	/
Drawings							
Other parameters available	No	No	Yes	No	Yes	Yes	Yes
Designed for DBT	No	No	Yes	No	Yes	Yes	Yes

a Two tissue-equivalent materials mimicking 100% adipose and 100% gland tissues "swirled" together in an approximate 50/50 ratio.
b Equivalent to 50% adipose and 50% gland tissues.
c Thickness used in this study.
d Each group is composed of six Landolt objects

Table 10 shows a number of similarities between the test objects in that six of the seven have a semi-circular shape and all but one have a homogenous background (Sage *et al.*, 2019). The semi-circular shape may be important if the reconstruction algorithm makes assumptions about breast shape or if a square shaped phantom generates strong artefacts that interfere with the phantom score. The phantoms are composed of either PMMA or 50%adipose/50%glandular simulating materials. Thickness is either fixed at 44 or 45 mm to simulate an approximately median breast thickness or the phantoms have 6 x 10 mm plates so that thickness can be varied to assess AEC function.

Small specks are used to simulate microcalcifications in five of the phantoms, composed either of Al_2O_3, $CaCO_3$ or glass (~$Si0_2$), with diameters ranging from 0.09 mm up to 0.54 mm across the different phantoms. The Phantom Laboratory TOMOPHAN and QUART "mam digi/EPQC" phantoms do not have

Accepted Manuscript

microcalcification simulating specks, but respectively include 0.18 mm diameter tungsten carbide beads for MTF measurement and Landolt C objects composed of 12.5 μm thick titanium, with gaps ranging from 50 μm to 160 μm for an image sharpness assessment. The tungsten carbide beads are designed for evaluation of MTF and ASF, rather than subjective scoring by an observer. Mass simulating objects are included in all phantoms except the QUART "mam digi/EPQC". These range in diameter from 1.8 mm up to 8 mm; the composition depends on the manufacturer and includes the CIRS material 'Breast Carcinoma' for the CIRS phantoms or acrylic (TOMOPHAN). Note that masses for the ACR and CIRS model 021 are specified as a thickness rather than a diameter. Five of the phantoms also include fibres, whose composition is not always specified but is typically Nylon, with diameters ranging from 0.4 mm to 1.25 mm.

Some of these test objects also include targets that enable the calculation of technical parameters, such as tungsten wires of diameter 25 μm or 50 μm and 0.18 mm diameter tungsten carbide spheres for MTF calculation. ASF and 2D and 3D accuracy is assessed in the CIRS model 021 phantom using 0.8 mm diameter ceramic spheres, although these are positioned off the z-axis, while the 0.18 mm diameter tungsten carbide spheres in the TOMOPHAN are also used for ASF measurement. A 0.4 mm steel sphere, positioned centrally (x-y) at height of 30 mm (z), is used for ASF evaluation in TOMOMAM (Fitton *et al.*, 2022). Finally, the CIRS model 021 includes an Al square of thickness 0.2 mm for SDNR assessment; an Al square is also in the TOMOPHAN test object but thickness is not specified.

The study by Sage *et al* (Sage *et al.*, 2019) required a visual evaluation of the test object to produce a score for tracking system performance in time. Images were generated under AEC control on three DBT systems: a GEHC SenoClaire (2 units), a Hologic Selenia Dimensions (2 units) and a Siemens Inspiration (1 unit), and the in-focus plane was scored by three human observers. The ASF was calculated and the images were also evaluated for the presence of artefacts generated by high contrast structures in the images. Sage *et al* (Sage *et al.*, 2019) found that artefacts affected the observer scoring in four of the test objects. Some of the targets were always visible at the AEC dose level used and are therefore not sensitive to changes in system performance. Step size, in terms of contrast or target diameter between successive details in the test object, should be balanced between sensitivity and reproducibility i.e. should be sufficiently sensitive such that changes in imaging performance can be detected (Yaffe *et al.*, 2006) yet not too small so that reader reproducibility error becomes a problem. There were also inconsistencies in the ASF measured for the systems when calculated from the different phantoms, a consequence of the different sizes, positions and compositions used in the test objects (Section 3.9). There has been no attempt to develop a minimum image quality level using any of the test objects described in the study of Sage *et al* (Sage *et al.*, 2019) in the EUREF protocol.

Turning to the use of CDMAM for longitudinal testing, Strudley *et al* (Strudley *et al.*, 2015) performed CDMAM measurements at 6 monthly intervals on the seven Hologic Selenia used in the TOMMY trial (Gilbert *et al.*, 2015). The CDCOM automated readout method (Karssemeijer and Thijssen, 1996) was used to score the plane containing the in-focus CDMAM image, although some image processing was required

Accepted Manuscript

to remove the low frequency non-uniformity in the image (Strudley *et al.*, 2015). They found that the change in threshold gold thickness remained within measurement error and that the plane of best focus did not change over the duration of the trial. The best focus plane could differ between systems by ± 1 plane, probably due to differences in indicated thickness calibration. Slightly higher threshold gold thickness was found for DBT compared to DM mode for the Selenia Dimensions, however threshold gold thickness was on or slightly superior to the Achievable level, consistent with good in-plane small detail detectability (Strudley *et al.*, 2015). It should be noted that the EUREF protocol states that DBT threshold gold thickness must not be compared against the Acceptable and Achievable levels that have been established for DM 2D imaging. The levels were developed in 2D projection imaging, without overlying projected anatomy. Projected structure noise is ~constant for the different DM systems, while for DBT images the projected in-plane structure is likely to vary between the different DBT units, depending on the angular range and thus the PSF in the z-direction (Gang *et al.*, 2010; Ikejimba *et al.*, 2021).

Automated readout

When visually scoring test objects, this must be possible within a reasonable amount of time, or an alternative automated method should be provided. An example of this is the readout of CDMAM for DM images, where 6 CDMAM images are required in the EUREF DM protocol, and typically 3 observers should each read 2 images. At approximately 15 minutes per image, the total reading time of 90 minutes per system represents quite a burden on medical physics resources. This motivated the development of computer readout of CDMAM images via the CDCOM module (Karssemeijer and Thiijssen, 1996); Young *et al* (K. Young *et al.*, 2006) later compared a number of additional processing steps for automatic readout against human readout for 20 DM systems. There have been a number of papers describing automatic readout for DM images of other phantoms including the ACR phantom (Brooks *et al.*, 1997; Chakraborty, 1997; Asahara and Kodera, 2012) and the Leeds TORMAS phantom (Gennaro *et al.*, 2007), primarily using template matching techniques in the spatial and Fourier domains. Test object alignment and target segmentation or location remains a challenge and should prompt the inclusion of alignment methods into the test object design, such as the grid used in CDMAM.

Automated readout greatly benefits both types of test object use, *i.e.* testing against a standard and longitudinal testing, by reducing within-observer and between-observer error (Green and Swets, 1966; Cohen, McDaniel and Wagner, 1984). Both of these factors introduce a bias to test object scores that complicate interpretation of a result or could even lead to an incorrect conclusion. Ultimately, test object scoring of DBT image quality phantoms will be performed using computational readout methods also referred to as model observers (Barrett *et al.*, 1993; Bouwman *et al.*, 2017; Petrov, Marshall, K. C. Young, *et al.*, 2019; Petrov, Marshall, K. Young, *et al.*, 2019; Zeng *et al.*, 2020).

Other methods

There is potential for the application of methods referred to as *no-reference* or "blind" quality assessment methods, where algorithms detect distortions within the image, using metrics such as the structural

Accepted Manuscript

similarity index (SSIM) (Wang *et al.*, 2004; Mittal, Moorthy and Bovik, 2012). Some of these methods include aspects of the human visual system such as the contrast sensitivity function and contrast masking, an example being the visible difference predictor (VDP) proposed by Daly (Daly, 1992). These methods remain relatively unexplored in test object image evaluation, where the current focus is firmly on model observer methods (Barrett *et al.*, 1993; Barrett and Myers, 2003), but may prove useful in mimicking human observer scoring or in detecting changes in a series of test object images when compared to a baseline reference image.

A number of studies have shown that object detectability in c-d test objects is related to the SNR transfer of the imaging system, which can be approximated using the well-known theory of Rose (Rose, 1948; Burgess, 1999) or Fourier-based models of SNR (Marshall, 2006a; Segui and Zhao, 2006; Monnin *et al.*, 2011; Liu *et al.*, 2014). The study of Monnin *et al* (Monnin *et al.*, 2020) has extended the link between c-d imaging tasks and the Fourier NPWE model to in-plane detectability in five DBT systems, to show that 'in-plane' SNR measured using CDMAM and the NPWE are consistent, for the DBT systems studied. A detectability index that encapsulates aspects of system performance, which in turn influence small detail detectability in a homogeneous background, offers an alternative QC method for tracking the imaging system over time compared to the human reader (visual) evaluation specks and spheres in test objects (Sage *et al.*, 2019). These methods are not based on the evaluation of a threshold stimulus and are not subject to inter- or intra-reader uncertainties.

4.3.2 Anthropomorphic test objects for DBT

If we wish to compare the detection performance for microcalcification and mass lesion simulating details in DBT imaging to that in DM images, then an object with a relevant structured background is required. Some of the latest developments in anthropomorphic phantoms for evaluation of breast x-ray imaging systems have been reviewed in two recent papers, first by Glick and Ikejimba (Glick and Ikejimba, 2018) and by Bliznakova (Bliznakova, 2020). Obviously, the aim of these phantoms is to simulate the attenuation properties and structures found in real breasts, which in turn depends on the availability of tissue simulating materials. McGarry *et al* (McGarry *et al.*, 2020) have reviewed advances in tissue simulating materials, including for mammography and BCT applications. Given that this topic is directly relevant to the evaluation of DBT imaging systems, we briefly describe the physical anthropomorphic phantoms that have potential for use in technical image quality assessment. Almost all of these phantoms are currently research efforts and are not currently commercially available.

CIRS BR3D 'swirl' phantom

This phantom is produced by CIRS (CIRS, 2021) by mixing together glandular and adipose simulating materials in the ratio of 50/50 to generate a phantom with a textured background. This generates a nominal structure, which when imaged in DM or DBT modes, does not resemble breast anatomy but some metric calculated from the images are similar to those found in breast images. Cockmartin *et al* (Cockmartin, Bosmans and Marshall, 2013) showed that the power spectrum magnitude (~magnitude of the image

Accepted Manuscript

structure, K) was within the range found for patients, while the ß exponent characterizing the correlation of structures in the image was somewhat lower than the patient range (see equation (21)). This test object is commercially available, and contains a 10 mm thick slab embedded with specks, fibres and spheroidal masses. The $CaCO_3$ specks range in diameter from 0.13 mm to 0.40 mm while the masses, composed of an epoxy resin that simulates breast carcinoma, have diameters from 1.8 mm to 6.3 mm.

The studies of Chan *et al* (Chan *et al.*, 2014) and Goodsitt *et al* (Goodsitt *et al.*, 2014) used the BR3D phantom to examine microcalcification cluster detection for the prototype step and shoot GEHC DBT system. Chan *et al* (Chan *et al.*, 2014) found narrower angle geometries for the prototype system gave higher detection sensitivity and conspicuity of subtle calcifications, however mass lesion was not considered. In the complementary study using mass-like simulating lesions by Goodsitt *et al* (Goodsitt *et al.*, 2014), wider angle geometries gave higher CNR, a result that generally correlated with human reader scores. More recently, Chen *et al* (Chan *et al.*, 2020) have imaged the BR3D phantom using a GEHC Pristina DBT system to study the effect of dose on microcalcification detection by radiologists. A dose increase from 1.34 to 2.07 mGy for a 5 cm thick breast simulating phantom increased microcalcification conspicuity, improved the detection sensitivity by radiologists, increased the confidence in identifying true clusters and reduced the likelihood of marking false positives.

A problem one will encounter when trying to implement an image quality standard using the BR3D test object is that the structured patterns differ from phantom to phantom. The position of the targets, especially the specks, with respect to the swirl pattern will have a strong influence on object detectability, and this will be specific to a particular instance of the BR3D phantom.

Duke physical phantom

The use of 3D printing methods (or more formally, additive manufacturing) to produce realistic phantoms that are based on human breast data, has been explored at Duke University and described in a series of papers (Kiarashi *et al.*, 2015; Ikejimba *et al.*, 2016; Rossman *et al.*, 2019). Patient data were taken from the XCAT family of virtual breast phantoms (Hsu *et al.*, 2013), which in turn were generated from a BCT dataset acquired at UC Davis (Lindfors *et al.*, 2008). In the initial studies, two phantom types were produced, called the Doublet and Singlet, however the methods were not able to reproduce the full dynamic range of glandular to adipose tissue found in a real breast population. Improvements to the methods and phantoms were described by Rossman *et al* (Rossman *et al.*, 2019). A cheaper alternative to adipose tissue was assessed (Jf Flexible material) however the lower breast density limit using printing materials remained at 36%. The previous limiting maximum breast density of 64% was increased using two methods. Effective linear attenuation coefficients (μ_{eff}) were assessed using a 28 kV, W/Rh spectrum with 4 cm of 50/50 adipose/glandular material at the tube exit port. First, the newly available printing material VeroPureWhite was found to have a volumetric equivalent density of 92% (μ_{eff} = 0.65 cm^{-1}). Second, adipose simulating material Jf Flexible was doped with either zinc acetate or tungsten to increase attenuation. Using 3% zinc acetate by weight gave a good match to glandular tissue (μ_{eff} = 0.68 cm^{-1}) but proved not to be stable.

Accepted Manuscript

Doping using tungsten nanopowder (1.6% by weight) raised μ_{eff} to 0.75 cm^{-1}, which is equivalent to 140% glandularity and therefore printing using Jf Flexible and tungsten doped Jf Flexible would produce phantoms with approximately the correct dynamic range (or contrast) between adipose and glandular tissue, yet the overall attenuation would be slightly higher than breast equivalent.

Earlier versions of the these phantoms were made in two halves and imaged with a sheet of paper inserted between the halves containing circular (effectively 2D) disks printed with iodine doped ink (Ikejimba *et al.*, 2016), but did not contain 3D lesion targets. Steps taken to include masses, iodinated masses and microcalcifications are described in Rossman *et al* (Rossman *et al.*, 2019). Inserts of thickness 0.5 cm were printed that contained just healthy anatomy and the healthy anatomy with a set of 64 masses; detailed information on how the masses were integrated into the background breast structure is not given. The micro-lobulated mass models used were those described by Kiarashi *et al* (Kiarashi *et al.*, 2016) and were printed from Jf Flexible material with and without tungsten, covering attenuation levels 100% to 110% relative to fibroglandular tissue. Figure 20 shows a simulated mammography image of the mass insert, along with DBT and DM images acquired of the phantom with the insert. Rossman *et al* (Rossman *et al.*, 2019) also produced an insert containing calcification, simulated using crushed and sieved eggshells (calcium carbonate, $CaCO_3$) with a length ranging from 53 μm to 710 μm and an average of 200 μm. The calcification test object was treated as a 2D insert, i.e. all the calcifications lying in a single plane. To complement the clinical targets, a homogeneous region along with a set of test objects suitable for measuring technical parameters such as the ASF and MTF was fabricated, designed to be imaged at the chest wall edge.

To date, no detailed human observer detection studies have been performed using this phantom. The prototype version was included in a survey of phantoms for use in synthetic mammography (SM) technical image quality assessment (Vancoillie *et al.*, 2020). Eventually, the group suggests that a set of phantoms could be produced from the XCAT catalogue such that a range of breast types can be used to evaluate DM and BT systems, using model observers for automatic readout of a set of lesions in a specific background type. For the moment, these phantoms remain in the research domain.

Accepted Manuscript

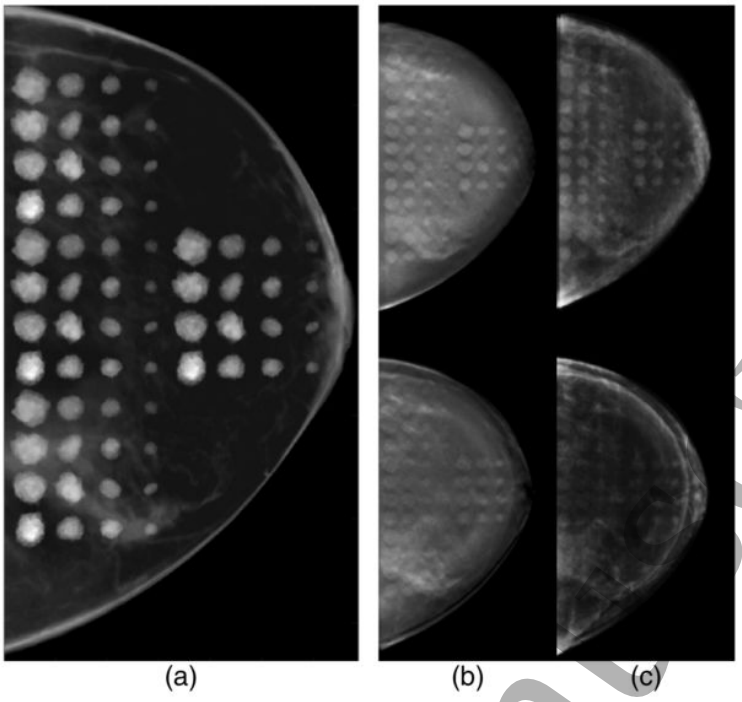


Figure 20 Breasts imaged with the mass contrast-detail insert. Tungsten doped breast on the top and VeroPureWhite breast on the bottom. (a) Simulated mammography image of the virtual phantom containing the lesion insert with exaggerated contrast to illustrate lesion location and details, (b) DBT reconstructed slice image of the phantoms, and (c) mammography image of the phantoms. (from Rossman *et al* (Rossman *et al.*, 2019).

FDA printed phantom

The phantom developed by a research group at the Center for Diagnostic and Radiological Health (CDRH)/FDA in the USA is produced by printing a virtual breast model onto pages of parchment paper (Ikejimba *et al.*, 2017) using an inkjet printer (Figure 21). Adding different quantities of iodine produced a radiopaque medium that generated contrast when imaged, a method that has been explored previously for c-d test objects to assess fluoroscopy systems (Theodorakou *et al.*, 2004). In fact, a study by Kotre and Porter (Kotre and Porter, 2005) showed that the iron oxide (Fe_3O_4) present in laser jet toner could generate sufficiently high contrast at mammographic energies so that c-d test objects could be printed without adding heavier elements.

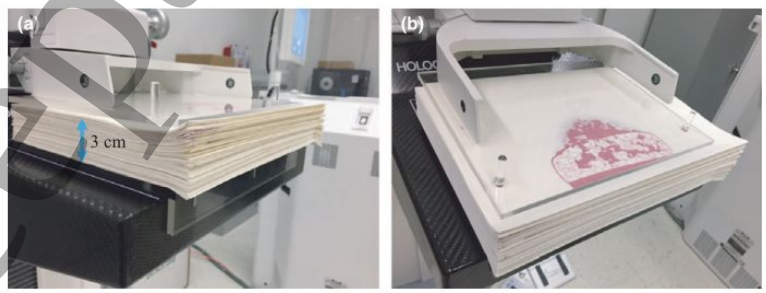


Figure 21. The completed stack of parchment paper. The complete physical phantom was imaged on clinical DM and DBT systems, shown here with the (a) side and (b) top views. (taken from Ikejimba *et al* (Ikejimba *et al.*, 2017)).

Accepted Manuscript

Ikejimba *et al* (Ikejimba *et al.*, 2017) measured effective linear attenuation coefficient (μ_{eff}) data on a Lorad Selenia DM system, with a Mo/Mo A/F combination at 28kV and 35 kV. From the measured pixel intensity, μ_{eff} was calculated as:

$$\mu_{eff} = \frac{1}{-d}\ln\left(\frac{I_1}{I_0}\right) \quad (31)$$

Ikejimba *et al* (Ikejimba *et al.*, 2017) also calculated the effective attenuation data for adipose and glandular tissue using the Boone spectral data (Boone, Fewell and Jennings, 1997) and attenuation data taken from the NIST database (Berger *et al.*, 1998):

$$\mu_{eff} = \frac{1}{-d}\ln\left(\frac{\int \varphi(E)E\, e^{-\mu(E)d}dE}{\int \varphi(E)E\, dE}\right) \quad (32)$$

where φ is the photon spectrum, E is the photon energy and μ is the linear attenuation coefficient.

Printing resolution was found to be largely isotropic, with a setting of 363 dots/inch giving a resolution of 70 μm. Two types of parchment paper were examined; physical thickness was ~70μm while the μ_{eff} values were close to that of adipose tissue (e.g. 0.61 ± 0.05 cm^{-1} compared to 0.57 ± 0.05 cm^{-1} for adipose tissue). Results for two levels of ink iodination were presented; printing with 75% ink plus 25% iohexol on parchment paper gave μ_{eff} = 0.89 ± 0.06 cm^{-1}, while parchment paper printed with 67% ink plus 33% iohexol resulted in a μ_{eff} = 0.94 ± 0.06 cm^{-1}. For comparison, μ_{eff} = 0.90 ± 0.03 cm^{-1} for glandular tissue. The method was used to print the virtual phantom developed by Graff (Graff, 2016). Generation of the virtual phantom with 70 μm voxel size took ~2 hours, while total production of the phantom including drying time and assembly required approximately 2 days. Overall, the method had good accuracy, a reproducibility of 3% when printing successive test patches and a variability of <1% across batches of ink. Figure 22 shows a plane of the phantom with a 5 mm diameter lesion, a DBT plane of the imaged phantom and a DM image. A limitation associated with the use of iodine as a dopant is the K-edge at 33 keV, which restricts use of the phantom to lower energies. Switching to dopants such as gadolinium and titanium dioxide nanoparticles would allow imaging at higher energies (Ikejimba *et al.*, 2017).

Accepted Manuscript

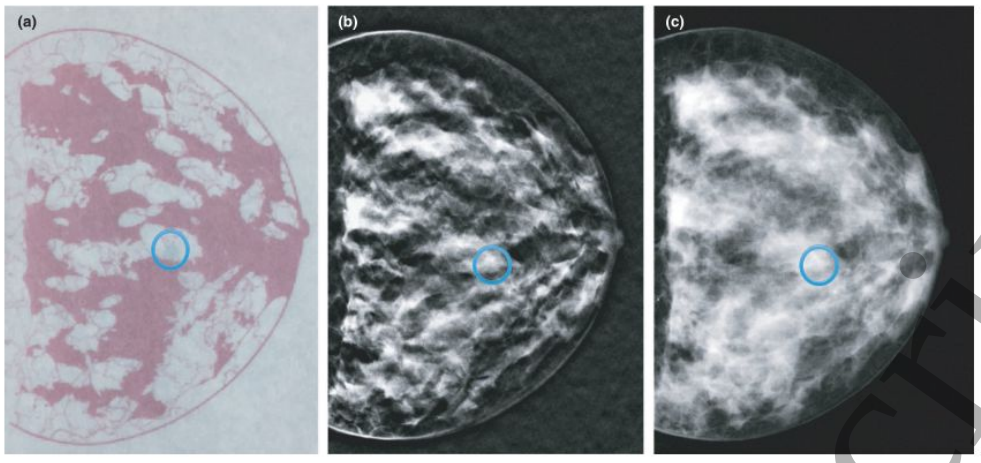
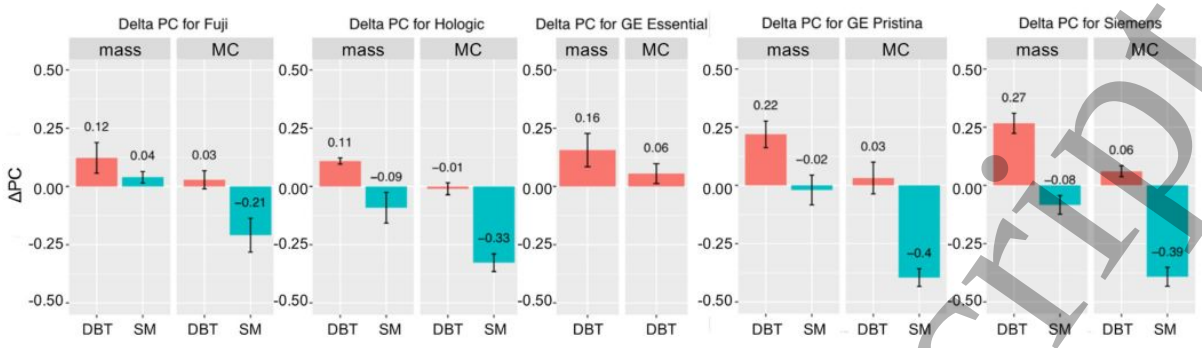


Figure 22. Phantom with 5-mm lesion. (a) A photograph of the slice containing the lesion is compared to (b) a 1-mm slice through the reconstructed volume for DBT, and (c) the 2D DM image. A circle specifies the location of the lesion. (taken from Ikejimba *et al* (Ikejimba *et al.*, 2017)).

In a recent study, Ikejimba *et al* (Ikejimba *et al.*, 2021) describe the use of this printed phantom to evaluate five commercially available DBT systems. Spiculated masses of diameter 5 mm generated using the method described by de Sisternes *et al* (de Sisternes *et al.*, 2015) were simulated into the virtual phantom (Graff, 2016). A set of 18 identical masses were printed with a spacing of 20 mm centre-to-centre. Four such inserts were made to give a total of 72 masses, with the set of masses shifted by 2 – 5 mm in the x-y plane. A new ink formulation based on potassium iodide was used for the mass printing to increase attenuation, and the print performed using two cartridges, one containing the previous fibroglandular simulating ink and one with the potassium iodide ink. Microcalcifications were fabricated from a mixture of calcium hydroxyapatite powder and polyvinylpyrrolidone, sieved to give specks with a size between 150 and 180 μm.

Accepted Manuscript

The phantom was imaged at AEC dose level on five DBT systems: the Fuji Innovality in ST mode (15°), the Hologic Selenia, GE Essential, GE Pristina and Siemens Inspiration. A 4-alternative forced choice



(AFC) study using seven medical physicists was conducted to evaluate the DM images, the DBT stack and the SM images. The change in percentage correction detection (ΔPC) from the DM imaging mode was calculated to show relative performance from DM images (Figure 23). Reduced microcalcification detection was found for SM images compared to DM and DBT images. DBT had the highest mass detection performance, and mass detection was found to improve as angular range increased. This study shows the utility of this test object, which can be produced at reasonable cost using open source methods for the virtual phantom. As demonstrated by Ikejimba *et al* (Ikejimba *et al.*, 2021), applications include technical image quality comparison between the modalities, allowing DBT to be compared against DM using relevant imaging tasks.

Figure 23. ΔPC for all systems. The ΔPC was computed relative to full-field digital mammography for digital breast tomosynthesis (DBT) and SM, with results given for masses and MCs. (taken from Ikejimba *et al* (Ikejimba *et al.*, 2021)

KU Leuven L1 phantom

There is a class of phantoms that apply a statistical method to the simulation of breast structure, using plastic or PMMA beads with a range of diameters to produce a self-similar structure that matches the first and second order statistical properties of real breasts (Gang *et al.*, 2010; Park *et al.*, 2010; Cockmartin *et al.*, 2017). The phantom described by Cockmartin *et al* (Cockmartin *et al.*, 2017), now named "L1", contains PMMA spheres ranging in diameter from 15.88 mm down to 1.58 mm, in a container that is 48 mm thick (~60 mm breast equivalent thickness). The PMMA spheres occupy 75% of volume while the remaining 25% volume is filled with water, leading to a linear attenuation coefficient of 0.71 cm^{-1} at 20 keV (Hubbell and Seltzer, 1995). Included in the phantom are microcalcification and mass lesion simulating objects, glued onto two PMMA plates (2 mm thick) that are supported at a height of 20 mm within the phantom. Although the 2mm PMMA is not visible in DBT planes, this plate could influence distribution of spheres in the neighbourhood of a lesion, leading to slightly different background statistics near a lesion. To limit this effect, the width of the PMMA plate is kept to 20 mm. The calcification specks are simulated using calcium carbonate (CaO_3) and are arranged in five groups of five, with average diameters of 90–100, 112–125, 140–160, 180–200 and 224–250 μm. A spiculated and non-spiculated mass lesion model were selected from those described by Shaheen *et al* (Shaheen *et al.*, 2014) and rescaled to generate five non-spiculated lesions

Accepted Manuscript

with average diameters from 1.6 to 6.2 mm and four spiculated masses with average diameters from 3.8 to 9.7 mm. Four printing materials were tested and Next was selected as this material had the best mechanical properties and water resistance of the four (see Figure 24). The linear attenuation coefficient at 20 keV of Next was measured (Fredenberg *et al.*, 2013) and found to be 0.83 ± 0.001 cm^{-1}, giving an attenuation difference between the Next masses and the L1 background of 0.12 cm^{-1}. This can be compared to the attenuation difference of ~0.22 cm^{-1} at 20 keV for breast cancerous tissue, assuming an attenuation coefficient of 0.84 cm^{-1} for cancerous tissue (Johns and Yaffe, 1987) against a breast background of 50% glandular/50%, where an attenuation coefficient of 0.62 cm^{-1} can be assumed (Hammerstein *et al.*, 1979). For a given diameter, the Next lesions therefore generate less contrast than expected for typical cancerous lesions in 50%/50% breast tissue. Cockmartin *et al* (Cockmartin *et al.*, 2017) looked at the use of polyethylene spheres to reduce the attenuation of the L1 background, however the spheres contained air bubbles which were obvious upon imaging and PMMA was used instead.

While this type of phantom shares some statistical features with real breast images, the visual realism is limited, as confirmed in the study of Marinov *et al* (Marinov *et al.*, 2020). Five radiologists were asked to score how closely four phantoms represented real breast structure using a five point rating scale. Two software phantoms, the Stochastic Solid Breast Texture (SSBT) phantom (Li *et al.*, 2016) and a power law noise texture (Burgess, Jacobson and Judy, 2001) and two physical phantoms, the L1 and CIRS BR3D phantoms, were evaluated. The SSBT phantom was rated as a good representation of real breast texture for both DM and DBT images, while the BR3D and L1 were considered the least visually realistic. Whether the backgrounds in these phantoms (physical and software) can achieve similar realism as breast tissue when used in detectability studies, i.e. in a task-based comparison (Badano, 2017), is still to be evaluated. An advantage of L1 is that shaking the phantom produces another background and this is a prerequisite when producing images for observer studies or training of a MO. Figure 25 shows a photograph of L1 along with DM and DBT images. An update of this phantom is underway, in which the water surrounding the spheres has been replaced by liquid paraffin oil (Salomon *et al.*, 2020).



Figure 24. Photograph of 3D printed models of a spiculated (left) and non-spiculated (right) mass using the printing material 'Next'.

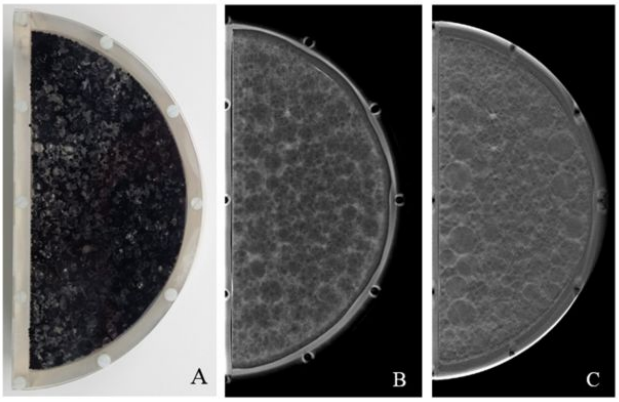
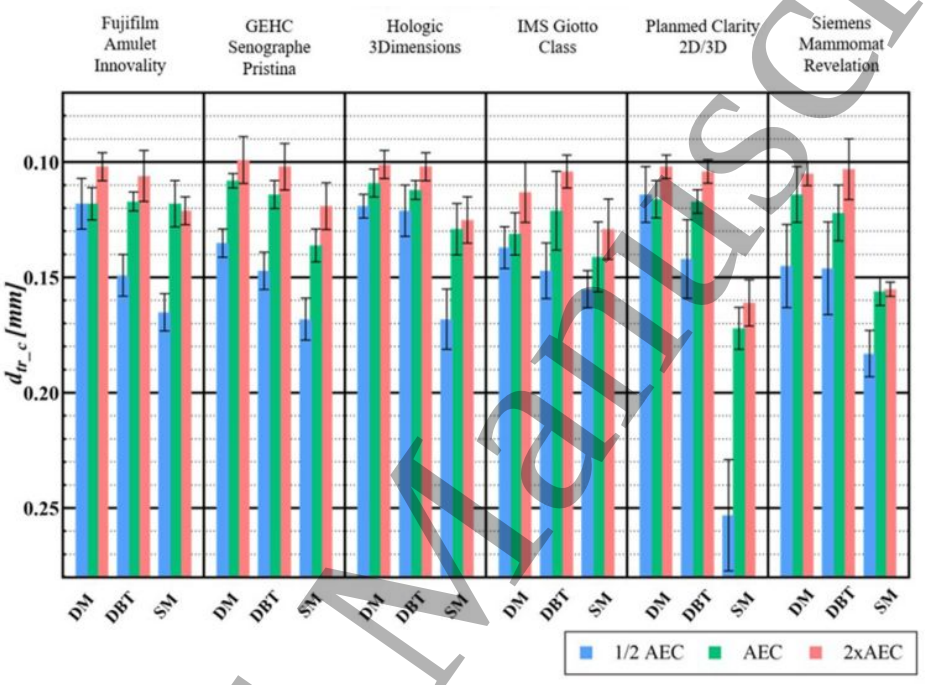


Figure 25. Photograph of the structured phantom with inserted target objects embedded in a background of PMMA spheres and water (A) its 2D mammographic image (B) and a

Accepted Manuscript

breast tomosynthesis reconstructed plane (C). (Figures taken from Cockmartin *et al* (Cockmartin *et al.*, 2017))

The L1 test object has been used in a multi-vendor study comparing detection of masses and calcification specks in DM, DBT and SM images (Vancoillie *et al.*, 2021). Devices from six vendors (Fujifilm Amulet Innovality in ST mode, GEHC Senographe Pristina, Hologic 3Dimensions, IMS Giotto Class, Planmed Clarity 2D/3D and Siemens Mammomat Revelation) were included and the three imaging modes were compared at three dose levels. Images were read by six medical physicists in a 4-AFC study, resulting in



threshold diameters for calcification specks and non-spiculated masses. As shown in Figure 26, no significant difference in microcalcification threshold diameter was found between DBT and DM images for any of the systems, however threshold diameter was always larger for SM images than for DBT and DM images, significantly so for three systems/SM algorithms. For both spiculated and non-spiculated masses (see Figure 26), DBT had smaller threshold diameters compared to SM and DM images (except the spiculated masses result for the Fujifilm system). A reduction in dose impacted calcification speck detectability, especially for SM images, but did not influence the detection of either mass type.

Figure 26 Threshold diameters [for microcalcification specks] with 95% confidence interval at the different dose levels for the three different imaging modes for six systems: Fujifilm Amulet Innovality ST (a), GEHC Senographe Pristina (b), Hologic 3Dimensions (c), IMS Giotto Class (d), Planmed Clarity 2D/3D (e), and Siemens Mammomat Revelation (f). Note that the y-axis is inverted such that smaller d_{tr} values (indicating better performance) have larger bars. (Figure taken from Vancoillie *et al* (Vancoillie *et al.*, 2021))

Accepted Manuscript

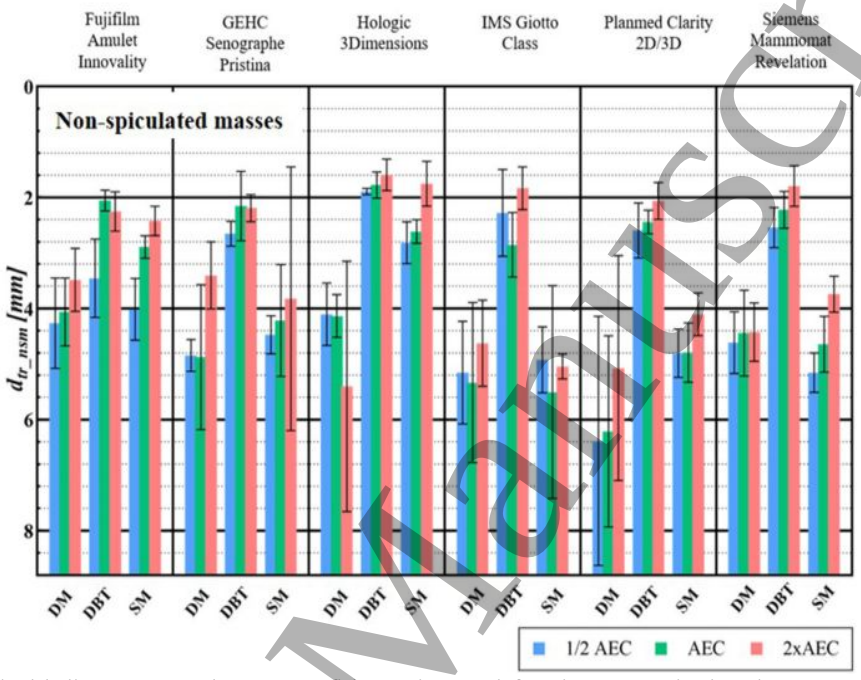


Figure 27. Threshold diameters and 95% confidence interval for the non-spiculated masses at three dose levels for DM, DBT, and SM images for six systems: Fujifilm Amulet Innovality ST, GEHC Senographe Pristina, Hologic 3Dimensions, IMS Giotto Class, Planmed Clarity 2D/3D, and Siemens Mammomat Revelation. Note that the y-axis is inverted such that smaller d_tr values (indicating better performance) have larger bars. (Figure taken from Vancoillie *et al* (Vancoillie *et al.*, 2021))

Computational read out has been developed for the L1 phantom (Petrov, Marshall, K. C. Young, *et al.*, 2019; Petrov, Marshall, K. Young, *et al.*, 2019; Petrov *et al.*, 2020). Recent work describes a deep learning method to generate the channels (DL-CHO) from pooled human reader data acquired on six DBT systems at three dose levels (Petrov *et al.*, 2020). In developing the channels, the DL-CHO was trained with the human reader responses where the percentage correct result in a given 4-AFC trial was greater than 50%. The resulting DL-CHO was able to predict human observer results for seven DBT systems.

Sunnybrook anthropomorphic phantom

The group at the Sunnybrook Research Institute in conjunction with GEHC has described the development and fabrication of an anthropomorphic phantom for 2D and 3D breast x-ray imaging (Mainprize *et al.*, 2020). The starting point was the SSBT model developed by Li *et al* (Li *et al.*, 2016) that generates a stochastic 3D texture based on medium and small-scale glandular and adipose tissue from BCT datasets in the UC Davis database (Lindfors *et al.*, 2008). The physical phantom was produced in four slabs using additive

Accepted Manuscript

manufacturing. The fibro-glandular part of the voxel model was deleted and the adipose part printed in polyamide-12 (a nylon polymer) using selective laser sintering (SLS). The pockets in the adipose section were filled with a resin doped with zinc oxide (ZnO) nanoparticles. The polymide-12 and the resin with ZnO were respectively shown to match adipose and glandular tissue from 15 keV to 45 keV, as verified by x-ray spectroscopy measurements (Mainprize *et al.*, 2020). Two measures were applied to assess whether selected statistical properties calculated for the phantom were within the range found for real breasts, the power spectrum β parameter (equation (21)) (Burgess, Jacobson and Judy, 2001) and the Laplacian fractional entropy (Abbey *et al.*, 2012). These parameters were found to lie within the patient range, although overall some increase in entropy in both the virtual and physical phantom at spatial frequencies up to 2 mm^{-1} would improve agreement with the patient dataset. Given the excellent matching both in terms of breast tissue attenuation and structure, this phantom has excellent potential for performance evaluation in DBT and contrast enhanced mammography. The software version of the SSBT scored well in the visual realism study of Marinov *et al* (Marinov *et al.*, 2020). Figure 28 shows DBT images of the virtual and real phantoms, showing a close match between the two.

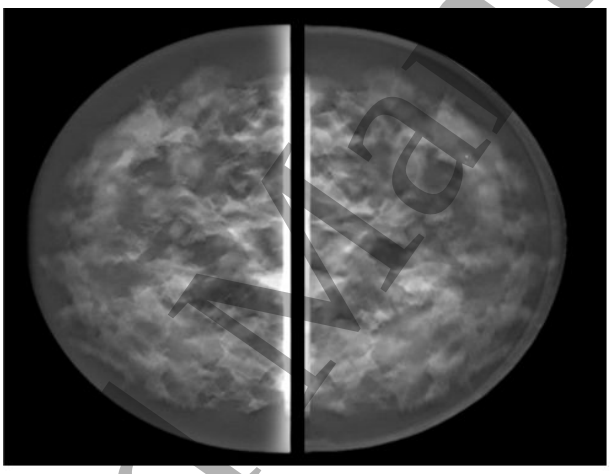


Figure 28. (a) DBT slices of the virtual (left) and physical (right) phantoms (mirrored) at 15 mm demonstrating good correspondence between structures in the phantoms. Window and level settings were manually adjusted to match contrast. (taken from (Mainprize *et al.*, 2020))

Other phantoms

A physical version of the digital breast phantom developed at the University of Pennsylvania (UPENN) (Bakic *et al.*, 2002) was built by Carton *et al* (Carton *et al.*, 2011). The method is effectively the inverse of that used for the Sunnybrook anthropomorphic phantom. Although steps have been taken to produce a commercial version by CIRS (Glick and Ikejimba, 2018; CIRS, 2021), this test object does not appear to be currently available.

There is class of phantom that uses variation in phantom height to vary the attenuation, which varies the fluence across the image plane when the 2D projection image is formed, provided the acquisition and geometry used to design the phantom match (Caldwell and Yaffe, 1990). This method takes as input a mammography image of a real breast, which should ideally be acquired on very high quality DM system (sharp, low noise, wide dynamic range), as the transfer properties are included in the image used to make

Accepted Manuscript

the phantom. Two such examples using this method are given in Badal *et al* (Badal, Clark and Ghammraoui, 2018) and in (Schopphoven *et al.*, 2019). These methods rely on a calculation of the required thickness of printing material at a given (x,y) point and need high resolution to print the fine detail visible in DM images. These methods can produce impressively realistic DM images, although the printing resolutions of ~300 μm and 200 μm were estimated for (Badal, Clark and Ghammraoui, 2018) and (Schopphoven *et al.*, 2019) result in some loss of detail when fine structures are viewed under magnification. As discussed, these methods do not generate 3D anatomy and therefore cannot be used for DBT system evaluation.

Jahnke *et al* (Jahnke *et al.*, 2017) describe the use of an Inkjet printer to print CT datasets using potassium iodide-doped ink (600 mg/mL) onto 80 mg/cm² plain paper. Samples were printed and scanned to establish a calibration curve between printer greyscale setting and the resulting Hounsfield unit. The utility of the method was demonstrated by printing a phantom from an abdomen CT scan. The method has been applied to generate a mammography phantom of thickness 3.6 cm from the Graff model (Graff, 2016), available as one piece or sectioned into 4 slabs each of 0.9 cm thickness (PhantomX-Gmbh, 2021) (Berlin, DE). Close matching of adipose and gland tissue phantoms is claimed, however no tasks are included and there is no evaluation of this phantom in the scientific literature yet.

Petropoulos *et al* (Petropoulos *et al.*, 2020) describe the use of TORMAM (Leeds Test Objects) to study microcalcification imaging performance for DM, SM and DBT images. A Hologic Selenia Dimensions was used to image TORMAM combined with PMMA to give a total phantom thickness ranging between 20 mm and 60 mm. Calcification clusters in both homogeneous and heterogeneous (tissue simulating) regions of the test object were segmented and the SDNR calculated and used as a metric of detectability. Results from the homogeneous region indicated superior performance for DM and DBT modes over SM imaging for the smallest calcification clusters (106 μm – 177 μm), for all phantom thicknesses. In the heterogeneous region, which covers a range of effective parenchymal densities, SDNR was similar in DM and SM images, while improved performance was seen for DBT images over DM and SM imaging in the denser regions. It must be noted that the TORMAM test object was designed for the evaluation of projection mammography imaging; the phantom is 15 mm thick and the thickness of the heterogeneous material is just 10 mm. The VOXMAM phantom is available from Leeds Test Objects and this was designed for use with DBT systems. The test object is 45 mm thick and contains 4 groups of microcalcification simulating particles randomly located in the x-y plane, at approximately 10mm intervals in the z plane. The breast simulating structure varies between phantoms and the specific positions of the calcification particles is not given. Cockmartin *et al* (Cockmartin, Bosmans and Marshall, 2013) found the β coefficient (equation (21)) to be close to that of patients but the magnitude of the anatomical noise (K) was notably higher than that measured for the patient population.

Finally, Cockmartin *et al* (Cockmartin *et al.*, 2015) describe a hybrid method in which physical, lesion simulating targets are immersed in a homogeneous, tissue simulating material (for example vegetable oil) and imaged on the system to be evaluated. The targets are segmented to form templates and multiplied into

Accepted Manuscript

projection images of a patient dataset, also acquired on the imaging system. Image processing/reconstruction is then performed and the resulting images can be evaluated in a reader study or by model observer methods. Advantages of this technique include the use of real patient backgrounds and systems sharpness, noise and scattered radiation are accurately encapsulated, however the method is time consuming to implement, requires access to the system reconstruction and needs physical lesion simulating targets, and is thus not suitable for QC testing.

4.4 Summary of system assessment using test objects

In summarizing the latest developments in anthropomorphic phantom development, it is useful to list the desired characteristics of a phantom for use with DBT systems and/or to compare DM and DBT systems (Table 11). The intended use of the phantom must be clearly defined, here it is assumed that the goal is to determine detection performance of model microcalcification and mass-like lesions in breast simulating structure, for an imaging system at a given clinical site. An alternative and probably easier option would be to define a test object without breast simulating structure for longitudinal testing against a baseline.

Table 11. Desired characteristics of breast phantom for performance assessment of DBT systems via lesion detection

Property of phantom
Sensitive to changes in system performance
Gives a result that can be compared against a relevant standard
Breast simulating thickness 20 to 90 mm, or median thickness if just one thickness
Match energy equivalence of adipose, fibro-glandular tissue, microcalcification, cancerous tissue, perhaps skin, over imaging system energy range (~16 keV to ~30 keV, currently)
Breast simulating structure is sufficiently realistic
Lesions are realistic and relevant
Variable glandularity / "difficulty" as thickness changes, or evaluate at median difficulty?
Can be used for DM and DBT (and contrast enhanced digital mammography (CEDM) and BCT?)
Can be produced reproducibly
Physically stable over time
Produces a reproducible metric, without needing too many images
Has computational readout
Inexpensive

Regarding phantom thickness there are some options available, depending on whether the AEC function and (technical) image quality at different thicknesses is to be assessed using the same phantom set. Currently, for DM imaging, CDMAM is combined with 40 mm PMMA sheets to give a PMMA equivalent thickness of 50 mm, and breast equivalent thickness of ~60 mm. The image quality is assessed at the median breast thickness, and SDNR measured during the AEC test is used to extrapolate threshold gold thickness to other breast equivalent thicknesses (Van Engen *et al.*, 2010; Salvagnini *et al.*, 2015). This is particularly important in DBT where pulse length can vary with breast/phantom thickness and x-ray factors must be relevant to a particular phantom thickness, e.g. via spacers. Extrapolation of anthropomorphic structures

Accepted Manuscript

is not straightforward, as thinner breasts tend to have higher glandular content and increased background complexity, while thicker breasts tend to be more adipose and have reduced complexity (Klausz and Shramchenko, 2005). However, one could question whether this is needed for QC, and that this should be reserved for initial optimization or Type Testing stages (Van Engen *et al.*, 2010). Nevertheless, there is the potential to do this using the phantom technologies described above. Simulated breasts, either stochastically modelled (Graff, 2016; Li *et al.*, 2016) or derived from breast databases (Hsu *et al.*, 2013; Rossman *et al.*, 2019) could be realized using additive manufacturing or doped-ink printing on paper with breast glandularity relevant to a given thickness. The alternative is to design a phantom representative of the median thickness (Cockmartin *et al.*, 2017), and just evaluate at that breast equivalent thickness.

The degree to which the energy response of materials used for structures in the phantoms is required to match those in real breast depends on the thickness range to be assessed and the modalities to be compared. PMMA is currently used extensively for DM system assessment and this has worked well for many years, as long as the increased density/attenuation is recognized (Dance, Skinner, *et al.*, 2000). The fact that gold is used in the CDMAM test object can influence the extrapolation of the SDNR results at different thickness, but this can be corrected (Van Engen *et al.*, 2010). The use of PMMA in test objects such as the Leuven L1 phantom therefore remains relevant when comparing evaluating and comparing DBT systems, however the linear attenuation coefficient of mass model material is only known at 20 keV (Cockmartin *et al.*, 2017). Extending the use to higher energies would require a characterization of the printing material at higher energies (Ivanov *et al.*, 2018) and selection of a relevant material for the breast structure. The doped-ink-on-paper method of Ikejimba *et al* (Ikejimba *et al.*, 2017) allowed a close match adipose and glandular tissue, however the use of iodine-based dopants limits the energy range to just below the iodine K-edge. If phantoms are required for use at higher energies, for example in CEDM or BCT, then alternative printing materials will be needed. Mainprize *et al* (Mainprize *et al.*, 2020) have shown that additive manufacturing can also closely match breast tissue attenuation through the use of printing materials doped with heavier elements. The recent work by Dahal *et al* (Dahal *et al.*, 2018) has described a method to adjust the attenuation of gelatine-based materials in the mammography energy range by the addition of micro-bubbles.

Cockmartin *et al* (Cockmartin *et al.*, 2017) found Next printing material to be a good match for cancerous breast lesion tissue, however this material is no longer available; VeroPureWhite or tungsten powder doped Jf Flexible may be suitable alternatives (Rossman *et al.*, 2019). A number of materials have been used to mimic microcalcifications, including calcium hydroxyapatite combined with a polyvinyl (Ikejimba *et al.*, 2021), calcium oxalate (Shaheen *et al.*, 2011) and calcium carbonate (Cockmartin *et al.*, 2017; Rossman *et al.*, 2019). These are reasonable choices, however care must be taken when selecting a material. Warren *et al* (Warren *et al.*, 2013) have investigated the properties of microcalcifications extracted from seven thinly sliced mastectomy samples, and imaged using a specimen cabinet. A total of 558 calcifications were included in the analysis; diameter of the single calcifications ranged from 15 – 491 μm, with a mean of 96 μm and a shape that was close to spherical. The equivalent aluminium thickness was found to be 0.85 times the

Accepted Manuscript

diameter of the calcification, suggesting that aluminium can be used to simulate calcification in computer simulation with a 0.85 factor applied to the attenuation coefficient. Calcium oxalate was found to have similar attenuation to aluminium and is thus a reasonable choice for simulation. Aluminium equivalence of calcium hydroxyapatite was 2.19 times the object thickness, and therefore the use of pure calcium hydroxyapatite in phantoms or computer simulation would lead to contrasts that are too high (Warren *et al.*, 2013). When placed in a physical phantom, calcifications can be grouped in fives, usually in a simple geometric pattern (Cockmartin *et al.*, 2017), but other choices are possible. Ikejimba *et al* (Ikejimba *et al.*, 2021) defined a 5 × 9 grid of 5 mm diameter circles, with 5 microcalcification specks randomly placed in each circle. Use of a grid increases the statistical power, in particular when computer read-out is employed.

The job of an anthropomorphic breast phantom is to act as a realistic substitute for a breast when imaged on an x-ray unit, however the degree of realism required of the breast simulating structures is somewhat open. Images of physical phantoms derived from patient datasets or those produced by models can look more realistic than images produced by sphere-based or texture swirl phantoms, especially in DBT images. There are a number of ways to assess how closely structures within a phantom x-ray image represent structures characteristic of average breast images. Feature or texture related metrics include the power spectrum (Burgess, Jacobson and Judy, 2001) and more recently Laplacian fractional entropy (Abbey *et al.*, 2012, 2019; Mainprize *et al.*, 2020; Barufaldi, Abbey, *et al.*, 2021). An alternative is rating scale experiments in which radiologists score the realism of simulated and real lesions. ROC analysis is then applied to determine whether the reader response differs significantly for the real and simulated datasets (Shaheen *et al.*, 2011; Elangovan *et al.*, 2017). In the discussion of realism, (Badano, 2017) defines these as "fool the radiologist" type studies (Badano, 2017). Ultimately, Badano (Badano, 2017) suggests that rather than just having the appearance of a breast image, the phantom and real images should present the same visual information to the observer. Simulation methods would be required to test whether this is the case. For example, the physical lesion simulating objects known to be present in a phantom could be simulated into the projections of a patient image dataset. Detection performance for the lesions in the phantom backgrounds could then be compared against thresholds measured in the real patients.

If benchmarking against DM performance is required, then obviously the phantom must be able to evaluate both DM and DBT systems. All phantoms discussed above should allow this, with the exception of the Rachel-type test objects (Badal, Clark and Ghammraoui, 2018; Schopphoven *et al.*, 2019). It is questionable whether any of these test objects can provide a meaningful evaluation of the SM images as they are produced with using algorithms that potentially make assumptions about the anatomy being imaged (Durand, 2018; Vancoillie *et al.*, 2020).

Some practical aspects need to be discussed. The phantom should be commercially available, or open access instructions should be available such that the phantom can be produced by a physics service. Currently, it appears that the CIRS 3BRD 'swirl' phantom and the Leeds VOXMAM (Leeds Test Objects) are the only commercially available structured/anthropomorphic phantoms. The phantoms and fabrication methods

Accepted Manuscript

described by Ikejimba *et al* (Ikejimba *et al.*, 2017, 2021) and by Cockmartin *et al* (Cockmartin *et al.*, 2017) are sufficiently detailed that physics services or research groups could make versions of these phantoms, although obviously the phantoms must be fabricated reproducibly. Phantoms must be stable over time, however there is little in the literature on long term stability of printed materials, for example when exposed to radiation or submerged in water.

Phantom readout must be reproducible, without needing too many images or DBT datasets to be acquired. Both the studies by Cockmartin *et al* (Cockmartin *et al.*, 2017) and Ikejimba *et al* (Ikejimba *et al.*, 2021) used a 4-AFC readout via human observers. Cockmartin *et al* (Cockmartin *et al.*, 2017) found that ten DBT scans of the L1 phantom gave reasonable reproducibility. In the work of Ikejimba *et al* (Ikejimba *et al.*, 2021), the four mass inserts were acquired. This was followed by the sheet containing the microcalcification specks, which was imaged multiple (undefined) times, shifted slightly between each acquisition. A single acquisition was used for the signal absent images, with VOIs extracted in a half-overlapping pattern. The fact that lesion is strongly influenced by local background structures presents a difficulty for fixed, 3D printed phantoms containing lesions, where ideally a number of realisations of the phantom with lesions shifted slightly against the background should be used to give an averaged detectability score.

In developing an automatic scoring method for the L1 phantom, (Petrov, Marshall, K. C. Young, *et al.*, 2019) showed that a robust, non-biased MO observer could be built for the non-spiculated mass models in the L1 for DBT images acquired on a Siemens Inspiration system, reconstructed using the EMPIRE algorithm. The MO was tuned to a set of human observer readings and a total of 24 scans were needed for this, 12 for training and 12 for the reading set. A two layer MO was developed to score the microcalcification groups in L1 (Petrov, Marshall, K. Young, *et al.*, 2019), again for the Siemens Inspiration system with EMPIRE reconstruction and tuned to human observer results. The study also used 12 scans of L1 for reading stage, and found a factor ~2 improvement in reproducibility for the MO compared to human readers. Automatic readout offers a solution to intra-reader bias that can confound comparison against an absolute standard, provided the reference values are generated using the sample computational algorithm. For targets in a homogeneous background, imaged on a quantum noise limited x-ray system, the test object score will scale monotonically with x-ray exposure, usually as the square root (Rose, 1948). When this is the case, image score for all object sizes can be changed/improved by changing dose. If the test object containing the two lesion types has a structured background then the use of dose as a remedial action will improve microcalcification detection to some extent but is not likely to improve mass lesion score (Ikejimba *et al.*, 2021; Vancoillie *et al.*, 2021).

Ideally, the phantom should not be expensive to buy or produce, if the test objects are to be widely used by medical physics departments. Both the studies of Ikejimba *et al* (Ikejimba *et al.*, 2021) and Cockmartin *et al* (Cockmartin *et al.*, 2017) describe phantoms that do not require expensive materials or methods for their production. This is in contrast to the current generation of additive manufacturing phantoms, where

Accepted Manuscript

fabrication cost is generally high, although quantitative information on this is difficult to find as many papers do not list these costs (Filippou and Tsoumpas, 2018; Bliznakova, 2020).

5. Conclusion

The current range of methods used for the technical evaluation of DBT systems provides a detailed picture of performance via a number of metrics, measured at the component and system levels, with clear applications in QC testing. There has been considerable progress in the development of (detection) task-based test objects that combine individual aspects of system performance, producing metrics that characterize overall technical imaging performance. With regard to technical performance testing ('setting the standard'), research is still needed to bring these testing methods closer to the clinical reality, however the preparatory work described here provides a solid foundation upon which to build.

Acknowledgements

We would like to acknowledge the QC data and images provided by Joke Binst, Annelies Jacobs, Kim Lemmens, Liesbeth Vancoillie and Hannelore Verhoeven. We would like to thank Pascal Monnin for helpful discussions regarding the technical image quality metrics presented in this work.

Accepted Manuscript

Appendix 1

The application of cascaded linear system theory to DBT and CBCT systems (Tward and Siewerdsen, 2008; Zhao and Zhao, 2008b) is a Fourier based description of the propagation of signal and noise through the different stages of the imaging system, from x-ray interaction in the photoconductor or scintillation phosphor through to the reconstructed DBT planes. Tward and Siewerdsen (Tward and Siewerdsen, 2008) focus solely on the noise transfer for CBCT systems while Zhao and Zhao include the additional stage of focal spot blur that is needed when modelling flying focus DBT systems (Zhao and Zhao, 2008b). Chumin Zhao et al (Zhao *et al.*, 2017) included the impact of oblique blurring in the x-ray detector. Such analysis can be complex and readers are referred to these papers for an in-depth description of this topic; an overview is given here as this offers some insight into the interpretation of MTF and NPS in the reconstructed planes. A flowchart detailing the final stages of the cascaded model is given in figure a1.

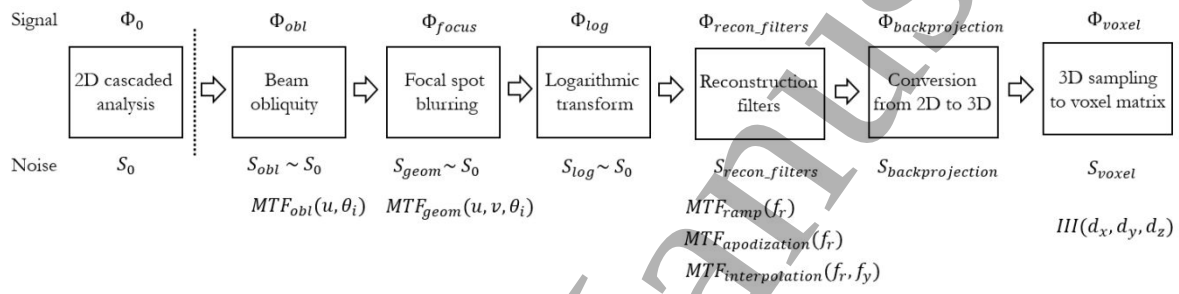


Figure a1. Flowchart showing the final stages of the cascaded model description of signal (Φ) and noise (S) transfer from projection images to the 3D voxel volume (adapted from (Zhao and Zhao, 2008b; Zhao *et al.*, 2017)).

The starting point is the signal and noise spectra from the end point fo the cascaded description of the x-ray detector. The next stage describes the influence of oblique entry on the signal spectrum, which is a small effect compared to the reconstruction filters (Hu, Scaduto and Zhao, 2012), while the noise is not greatly affected (Mainprize *et al.*, 2006). Blurring due to the focus (size and motion related size) follows. Both the oblique entry and focus blurring stages influence the signal spectrum but not the noise spectrum and therefore degrade the DQE (Zhao and Zhao, 2008b). In the log-normalization stage, detector signal is first converted to attenuation values by dividing by the mean image of an air or 'blank' scan, followed by a logarithmic transformation applied to each projection. This step can be considered to be a linear stage for small signals or fluctuations (noise) (Tward and Siewerdsen, 2008) with a gain equal to 1/(KB) where K and B are respectively the x-ray exposure and gain of the x-ray detector, i.e. 1/(mean signal level) (Zhao and Zhao, 2008b). In the cascaded analysis, the signal spectrum and the NPS are multiplied by 1/(KB) and $1/(KB)^2$, respectively. For the noise, this is recognized as a normalization step, after which the noise magnitude in the reconstructed images is inversely proportional to exposure. There is no noise associated with this gain stage and therefore DQE is not affected (Zhao and Zhao, 2008b).

Reconstruction filters are then applied with the ramp filter applied first to counteract the blurring from the backprojection stage. This filter increases linearly as a function of spatial frequency ($\propto f$) and many

Accepted Manuscript

implementations will have a small non-zero value at zero spatial frequency as this avoids images with zero DC value. The ramp filter amplifies high frequency noise and this is controlled by applying a smoothing filter, for example a window function, which removes or suppresses parts of the signal and noise (also called apodization). In applying the reconstruction filters, there is a change of coordinates from the detector plane (f_x, f_y) to a tilted plane at angle θ_i with coordinates (f_r, f_y, θ_i), described in Zhao *et al* (Zhao *et al.*, 2017). The reconstruction filters are then applied, giving these expressions for the signal (a1) and noise (a2) at the reconstruction stage:

$$\Phi_{recon_{filters}}(f_x, f_y) = \Phi_{log}(f_r, f_y, \theta_i) MTF_{ramp}(f_r) MTF_{apodize}(f_r) MTF_{interp}(f_r, f_y) \quad \text{(a1)}$$

$$S_{recon_filters}(f_x, f_y) = S_{log}(f_r, f_y, \theta_i) \left(MTF_{ramp}(f_r)\, MTF_{apodize}(f_r) MTF_{interp}(f_r, f_y)\right)^2 \quad \text{(a2)}$$

Transfer functions for the ramp filter, apodization and interpolation filters are respectively (Zhao and Zhao, 2008b):

$$MTF_{ramp}(f_r) = 2\tan\theta \frac{|f_r|}{f_{r,Nyq}} \quad \text{(a3)}$$

$$MTF_{apodize}(f_r) = 0.5\left[1 + cos\left(\frac{\pi f_r}{A f_{r,Nyq}}\right)\right] \quad \text{(a4)}$$

$$MTF_{interp}(f_r, f_y) = sinc^2(a_x f_r)\, sinc^2(a_y f_y) \quad \text{(a5)}$$

The next step is the conversion to 3D using back-projection and the transfer function for this stage is inversely proportional to spatial frequency ($\propto 1/f$). This exactly cancels out the ramp filter and results in an MTF that is 1.0 at zero spatial frequency (Tward and Siewerdsen, 2008; Zhao and Zhao, 2008b). Following backprojection, the 3D signal spectrum can be written as:

$$\Phi_{backproj}(f_r, f_y, \theta_i) = \frac{N}{\theta f_r} \Phi_{recon_filt}(f_r, f_y, \theta_i) \quad \text{(a6)}$$

$$S_{backproj}(f_r, f_y, \theta_i) = \frac{N}{\theta f_r} S_{recon_filt}(f_r, f_y, \theta_i) \quad \text{(a7)}$$

The term $N/(\theta f_r)$ is a normalization factor required when converting from polar coordinates to Cartesian coordinates, sometimes referred to as the 'spoke density' i.e. lines along the angular coordinate θ_i. The spectra are then transformed from the tilted planes to an interpolated reconstructed volume (Zhao *et al.*, 2017):

$$\Phi_{backproj}(f_x, f_y, f_y) = \Phi_{backproj}(f_r, f_y, \theta_i) \quad \text{(a8)}$$

$$S_{backproj}(f_x, f_y, f_y) = S_{backproj}(f_r, f_y, \theta_i) \quad \text{(a9)}$$

Accepted Manuscript

where $f_x = f_r \cos\theta_i$ and $f_y = f_r \sin\theta_i$. The final stages is 3D sampling using a finite voxel spacing and is modelled a convolution with a 3D comb function; the 3D NPS is replicated at multiples of the Nyquist frequency and summed:

$$S_{voxel}(f_x,f_y,f_y) = S_{backproj}(f_x,f_y,f_y) *** III(f_x,f_y,f_y)$$

$$= \sum_{i,j,k} S_{backproj}(f_x,f_y,f_y)\delta\left(f_x - \frac{i}{d_x}, f_y - \frac{j}{d_y}, f_z - \frac{k}{d_z}\right) \quad (a10)$$

Post reconstruction binning can also be applied (Tward and Siewerdsen, 2008). The 3D sampling step introduces aliasing of the high frequency noise which appears at lower spatial frequencies. In the Siemens system modelled by (Zhao and Zhao, 2008b), voxel spacing was 0.085 mm in the x and y directions i.e. at the native pixel spacing of the image receptor. A plane spacing of 1 mm was used in the z-direction, which is a common distance used by many vendors although the GEHC systems can generate a 0.5 mm plane spacing. This large sampling distance results in a replication of the NPS in the z-direction and aliasing of the noise at high spatial frequencies. Various filtering schemes are discussed to alleviate aliasing (Zhao and Zhao, 2008b), including the application of a filter in the z-direction that attenuates higher frequencies and reduces spectral overlap.

References

Abadi, E. *et al.* (2020) 'Virtual clinical trials in medical imaging : a review', *Journal of Medical Imaging*, 7(4), pp. 042805-1–40.

Abbey, C. K. *et al.* (2012) 'Non-Gaussian statistical properties of breast images.', *Medical physics*, 39(11), pp. 7121–30. doi: 10.1118/1.4761869.

Abbey, C. K. *et al.* (2019) 'Evaluation of non-Gaussian statistical properties in virtual breast phantoms', *Journal of Medical Imaging*, 6(02), p. 1. doi: 10.1117/1.jmi.6.2.025502.

Acciavatti, R. J. and Maidment, A. D. A. (2011) 'Optimization of phosphor-based detector design for oblique x-ray incidence in digital breast tomosynthesis', 38(11), pp. 6188–6202. doi: 10.1118/1.3639999.

Acciavatti, R. J. and Maidment, A. D. A. (2012) 'Observation of super-resolution in digital breast tomosynthesis', *Medical Physics*, 39(12), pp. 7518–7539.

Aichinger, H. *et al.* (2012) *Radiation Exposure and Image Quality in X-Ray Diagnostic Radiology*. 2nd Editio. Berlin, Germany: Springer-Verlag Berlin Heidelberg.

Argo, W. P., Hintenlang, K. and Hintenlang, D. E. (2004) 'A tissue-equivalent phantom series for mammography dosimetry.', *Journal of applied clinical medical physics / American College of Medical Physics*, 5(4), pp. 112–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15738925.

Asahara, M. and Kodera, Y. (2012) 'Computerized scheme for evaluating mammographic phantom images', *Medical Physics*, 39(3), pp. 1609–1617. doi: 10.1118/1.3687159.

Åslund, M. *et al.* (2005) 'AEC for scanning digital mammography based on variation of scan velocity', *Medical Physics*, 32(11), p. 3367.

Accepted Manuscript

Åslund, M. *et al.* (2006) 'Scatter rejection in multislit digital mammography', *Medical Physics*, 33(4), pp. 933–940. doi: 10.1118/1.2179122.

Badal, A., Clark, M. and Ghammraoui, B. (2018) 'Reproducing two-dimensional mammograms with three-dimensional printed phantoms', *Journal of Medical Imaging*, 5(03), pp. 033501-1–9. doi: 10.1117/1.jmi.5.3.033501.

Badano, A. *et al.* (2007) 'Anisotropic imaging performance in breast tomosynthesis', *Medical Physics*, 34(11), p. 4076. doi: 10.1118/1.2779943.

Badano, A. (2017) '"How much realism is needed?" - the wrong question in silico imagers have been asking', *Medical physics*, 44(5), pp. 1607–1609. doi: 10.1002/mp.12187.

Badano, A. (2021) 'In silico imaging clinical trials: cheaper, faster, better, safer, and more scalable', *Trials*. Trials, 22(1), pp. 1–7. doi: 10.1186/s13063-020-05002-w.

Badano, A. and Freed, M. (2011) 'Oblique incidence effects in direct x-ray detectors : A first-order approximation using a physics-based analytical model', pp. 2095–2098. doi: 10.1118/1.3567497.

Bakic, P. R. *et al.* (2002) 'Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation', *Medical Physics*, 29(9), p. 2131. doi: 10.1118/1.1501143.

Baldelli, P. *et al.* (2018) 'A comparative study of physical image quality in digital and synthetic mammography from commercially available mammography systems', *Physics in Medicine and Biology*. IOP Publishing, 63(16), p. 165020 (19pp). doi: 10.1088/1361-6560/aad106.

Barca, P. *et al.* (2021) 'Technical evaluation of image quality in synthetic mammograms obtained from 15° and 40° digital breast tomosynthesis in a commercial system: a quantitative comparison', *Physical and Engineering Sciences in Medicine*. Springer International Publishing, 44(1), pp. 23–35. doi: 10.1007/s13246-020-00948-2.

Barrett, H. H. *et al.* (1993) 'Model observers for assessment of image quality.', *Proceedings of the National Academy of Sciences of the United States of America*, 90(21), pp. 9758–65.

Barrett, H. H. *et al.* (2015) 'Task-based measures of image quality and their relation to radiation dose and patient risk', *Physics in Medicine and Biology*. IOP Publishing, 60(2), pp. R1–R75. doi: 10.1088/0031-9155/60/2/R1.

Barrett, H. H. and Myers, K. J. (2003) *Foundations of Image Science*. Wiley.

Barufaldi, B. *et al.* (2020) 'MRMC ROC analysis of calcification detection in tomosynthesis using computed super resolution and virtual clinical trials', 1151313(May 2020), p. 58. doi: 10.1117/12.2564118.

Barufaldi, B., Abbey, C. K., *et al.* (2021) 'Computational Breast Anatomy Simulation Using Multi-scale Perlin Noise', *IEEE Transactions on Medical Imaging*, 40(12), pp. 3436–3445. doi: 10.1109/TMI.2021.3087958.

Barufaldi, B., Maidment, A. D. A., *et al.* (2021) 'Virtual Clinical Trials in Medical Imaging System Evaluation and Optimisation', *Radiation Protection Dosimetry*, (July), pp. 1–9. doi: 10.1093/rpd/ncab080.

Becker, A. E. *et al.* (2020) 'A prototype Multi-X-ray-source array (MXA) for digital breast tomosynthesis', *Physics in Medicine and Biology*, 65(23). doi: 10.1088/1361-6560/abc305.

Beckett, J. R. and Kotre, C. J. (2000) 'Dosimetric implications of age related glandular changes in screening mammography Dosimetric implications of age related glandular changes in screening mammography', *Physics in Medicine and Biology*, 45, pp. 801–813.

Behling, R. (2021) 'On a new multi-source X-ray tube concept for minimizing imaging time in digital breast tomosynthesis', *Physica Medica*, 88(June), pp. 20–22. doi: 10.1016/j.ejmp.2021.06.013.

Accepted Manuscript

BEIR (2006) *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII— Phase 2 2006*. doi: DOI: 10.17226/11340.

Bemelmans, F. *et al.* (2021) 'Investigation of single-shot beam quality measurements using state of the art solid-state dosimeters for routine quality assurance applications in mammography', *Physica Medica*, 88(July), pp. 242–249. doi: 10.1016/j.ejmp.2021.07.006.

Berger, M. *et al.* (1998) 'XCOM: photon cross sections database. NIST Standard Reference Database 8', pp. 3587–3597.

Bernardi, D. *et al.* (2016) 'Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study', *The Lancet Oncology*, 17(8), pp. 1105–1113. doi: 10.1016/S1470-2045(16)30101-2.

Bernhardt, P., Mertelmeier, T. and Hoheisel, M. (2006) 'X-ray spectrum optimization of full-field digital mammography: Simulation and phantom study', *Medical Physics*, 33(11), pp. 4337–4349.

Berns, E. A. and Hendrick, R. E. (2002) 'Performance comparison of full-field digital mammography to screen – film mammography in clinical practice', pp. 830–834. doi: 10.1118/1.1472497.

Bliznakova, K. (2020) 'The advent of anthropomorphic three-dimensional breast phantoms for X-ray imaging', *Physica Medica*. Elsevier Ltd, 79(December), pp. 145–161. doi: 10.1016/j.ejmp.2020.11.025.

Bloomquist, A. K., Yaffe, M. J., Mawdsley, G. E., *et al.* (2006) 'Lag and ghosting in a clinical flat-panel selenium digital mammography system', *Medical Physics*, 33(8), p. 2998. doi: 10.1118/1.2218315.

Bloomquist, A. K., Yaffe, M. J., Pisano, E. D., *et al.* (2006) 'Quality control for digital mammography in the ACRIN DMIST trial: Part I', *Medical Physics*, 33(3), p. 719. doi: 10.1118/1.2163407.

Bochud, F. O. *et al.* (1999) 'Estimation of the noisy component of anatomical backgrounds.', *Medical physics*, 26(7), pp. 1365–70.

Boone, J. M. (1999) 'Medical Physics Glandular Breast Dose for Monoenergetic and High-Energy X-ray Beams : Monte Carlo Assessment', *Radiology*, 213, pp. 23–37.

Boone, J. M., Fewell, T. R. and Jennings, R. J. (1997) 'Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography.', *Medical physics*, 24(12), pp. 1863–74.

Bouwman, R. W. *et al.* (2009) 'An alternative method for noise analysis using pixel variance as part of quality control procedures on digital mammography systems.', *Physics in Medicine and Biologyiology*, 54(22), pp. 6809–22. doi: 10.1088/0031-9155/54/22/004.

Bouwman, R. W. (2013) 'Phantoms for quality control procedures in digital breast tomosynthesis : dose assessment', 4423. doi: 10.1088/0031-9155/58/13/4423.

Bouwman, R. W. *et al.* (2015) 'Average glandular dose in digital mammography and digital breast tomosynthesis: Comparison of phantom and patient data', *Physics in Medicine and Biology*. IOP Publishing, 60(20), pp. 7893–7907. doi: 10.1088/0031-9155/60/20/7893.

Bouwman, R. W. *et al.* (2016) 'Simulating local dense areas using PMMA to assess automatic exposure control in digital mammography', *Radiation Protection Dosimetry*, 169(1), pp. 143–150. doi: 10.1093/rpd/ncw032.

Bouwman, R. W. *et al.* (2017) 'Toward image quality assessment in mammography using model observers: Detection of a calcification-like object', *Medical Physics*. doi: 10.1002/mp.12532.

Bradley, L. R., Carton, A.-K. and Maidment, A. D. A. (2010) 'Angular dependence of mammographic dosimeters in digital breast tomosynthesis', in *Proceedings of SPIE*, pp. 76225L-1–10. doi: 10.1117/12.845369.

Accepted Manuscript

Brateman, L. F. and Heintz, P. H. (2015) 'Solid-state dosimeters: a new approach for mammography measurements', *Med Phys*, 42(2), pp. 542–557. doi: 10.1118/1.4903263.

Brooks, K. W. *et al.* (1997) 'Automated analysis of the American College of Radiology mammographic accreditation phantom images', *Medical Physics*, 24(5), pp. 709–723. doi: 10.1118/1.597992.

Buhr, E., Günther-Kohfahl, S. and Neitzel, U. (2003) 'Accuracy of a simple method for deriving the presampled modulation transfer function of a digital radiographic system from an edge image', *Medical Physics*, 30(9), p. 2323. doi: 10.1118/1.1598673.

Burgess, a E. (1999) 'The Rose model, revisited.', *Journal of the Optical Society of America. A, Optics, image science, and vision*, 16(3), pp. 633–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10069050.

Burgess, A. E., Jacobson, F. L. and Judy, P. F. (2001) 'Human observer detection experiments with mammograms and power-law noise', *Medical Physics*, 28(4), p. 419. doi: 10.1118/1.1355308.

Caldwell, C. B. and Yaffe, M. J. (1990) 'Development of an anthropomorphic breast phantom', *Medical Physics*, 17, pp. 273–280.

Calliste, J. *et al.* (2017) 'Second generation stationary digital breast tomosynthesis system with faster scan time and wider angular span', *Medical Physics*, 44(9), pp. 4482–4495. doi: 10.1002/mp.12393.

Carton, A.-K. *et al.* (2005) 'Validation of MTF measurement for digital mammography quality control', *Medical Physics*, 32(6), p. 1684. doi: 10.1118/1.1921667.

Carton, A.-K. *et al.* (2011) 'Development of a physical 3D anthropomorphic breast phantom', *Medical Physics*, 38(2), pp. 891–896. doi: 10.1118/1.3533896.

Carvalho, P. M. De (2014) *Low-Dose 3D Quantitative Vascular X-ray Imaging of the Breast*. UNIVERSITE PARIS-SUD.

Chakraborty, D. P. (1997) 'Computer analysis of mammography phantom images (CAMPI): An application to the measurement of microcalcification image quality of directly acquired digital images', *Medical Physics*, 24(8), pp. 1269–1277. doi: 10.1118/1.598149.

Chan, H.-P. *et al.* (2014) 'Digital Breast Tomosynthesis: Observer Performance of Clustered Microcalcification Detection on Breast Phantom Images Acquired with an Experimental System Using Variable Scan Angles, Angular Increments, and Number of Projection Views', *Radiology*, 273(3), p. 132722. doi: 10.1148/radiol.14132722.

Chan, H.-P., Lam, K. L. and Wu, Y. (1990) 'Studies of performance of antiscatter grids in digital radiography: Effect on signal-to-noise ratio', *Medical Physics*, 17(4), pp. 655–664.

Chan, H. P. *et al.* (2020) 'Effect of Dose Level on Radiologists' Detection of Microcalcifications in Digital Breast Tomosynthesis: An Observer Study with Breast Phantoms', *Academic Radiology*. Elsevier Inc., pp. 1–8. doi: 10.1016/j.acra.2020.07.038.

Chen, Y., Lo, J. Y. and Dobbins III, J. T. (2006) 'Noise power spectrum analysis for several digital breast tomosynthesis reconstruction algorithms', *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. Spie, 6142, pp. 614259-614259–8. doi: 10.1117/12.652282.

Chen, Z. and Ning, R. (2004) 'Three-dimensional point spread function measurement of cone-beam computed tomography system by iterative edge-blurring algorithm.', *Physics in Medicine and Biology*, 49, pp. 1865–1880. doi: 10.1088/0031-9155/49/10/003.

Cheung, L. *et al.* (2005) 'Image performance of a new amorphous selenium flat panel x-ray detector designed for digital breast tomosynthesis', in *Proc. of SPIE: Medical Imaging Vol 5745*. SPIE, Bellingham, WA, pp. 1282–1290.

Accepted Manuscript

Ciatto, S. *et al.* (2013) 'Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): A prospective comparison study', *The Lancet Oncology*. Elsevier Ltd, 14(7), pp. 583–589. doi: 10.1016/S1470-2045(13)70134-7.

CIRS (2021) *CIRS mammography phantoms*. Available at: https://www.cirsinc.com/products/ (Accessed: 16 June 2021).

Cockmartin, L. C. *et al.* (2015) 'Comparison of digital breast tomosynthesis and 2D digital mammography using a hybrid performance test', *Physics in Medicine and Biology*, 60(10). doi: 10.1088/0031-9155/60/10/3939.

Cockmartin, L. C. *et al.* (2017) 'Design and application of a structured phantom for detection performance comparison between breast tomosynthesis and digital mammography', *Physics in Medicine and Biology*, 62(3). doi: 10.1088/1361-6560/aa5407.

Cockmartin, L. C., Bosmans, H. and Marshall, N. W. (2013) 'Comparative power law analysis of structured breast phantom and patient images in digital mammography and breast tomosynthesis', *Medical Physics*, 40(8), pp. 081920-1–17. doi: 10.1118/1.4816309.

Cohen, G., McDaniel, D. L. and Wagner, L. K. (1984) 'Analysis of variations in contrast-detail experiments', *Medical Physics*, 11(4), pp. 469–473.

Cunningham, I. A. (1992) 'Signal and Noise in Modulation Transfer Functiom Determinations using the Slit, Wire, and Edge Techniques', *Medical Physics*, 19(4), pp. 1037–1044.

Cunningham, I. A. (2000) 'Applied Linear-Systems Theory', in Jacob Beutel (ed.) *Medical Imaging Volume 1. Physics and Psychophysics*. Bellingham: SPIE - The International Society for Optical Engineering, pp. 79–156.

Dahal, E. *et al.* (2018) 'Stable gelatin-based phantom materials with tunable x-ray attenuation properties and 3D printability for x-ray imaging', *Physics in Medicine and Biology*. IOP Publishing, 63, p. 09NT01 (6pp). doi: 10.1088/1361-6560/aabd1f.

Dainty, J. C. and Shaw, R. (1974) *Image Science: Principles, Analysis and Evaluation of Photographic-type Imaging Processes*. Academic Press.

Daly, S. J. (1992) 'The visible differences predictor: An algorithm for the assessment of image fidelity', in *Proceedings of SPIE - The International Society for Optical Engineering,*. SPIE, Bellingham, WA, pp. 2–15.

Dance, D. R. (1990) 'Monte Carlo calculation of conversion factors for the estimation of mean glandular breast dose.', *Physics in Medicine and Biology*, 35(9), pp. 1211–9.

Dance, D. R., Skinner, C. L., *et al.* (2000) 'Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol', *Physics in Medicine and Biology*, 45(11), pp. 3225–3240. doi: 10.1088/0031-9155/45/11/308.

Dance, D. R., Thilander Klang, A., *et al.* (2000) 'Influence of anode/filter material and tube potential on contrast, signal-to-noise ratio and average absorbed dose in mammography: A Monte Carlo study', *British Journal of Radiology*, 73(874), pp. 1056–1067. doi: 10.1259/bjr.73.874.11271898.

Dance, D. R. *et al.* (2005) 'Breast dosimetry using high-resolution voxel phantoms', *Radiation Protection Dosimetry*, 114(1–3), pp. 359–363. doi: 10.1093/rpd/nch510.

Dance, D. R. and Sechopoulos, I. (2016) 'Dosimetry in x-ray-based breast imaging', *Physics in Medicine and Biology*. IOP Publishing, 61(19), pp. R271–R304. doi: 10.1088/0031-9155/61/19/R271.

Dance, D. R., Young, K. C. and van Engen, R. E. (2011) 'Estimation of mean glandular dose for breast tomosynthesis : factors for use with the UK , European and IAEA breast dosimetry protocols', *Phys. Med. Biol.*, 56, pp. 453–471. doi: 10.1088/0031-9155/56/2/011.

Desponds, L. *et al.* (1991) 'Influence of anode and filter material on image quality and glandular dose for

Accepted Manuscript

screen-film mammography.', *Physics in Medicine and Biology*, 36(9), pp. 1165–1182.

Destounis, S. V., Morgan, R. and Arieno, A. (2015) 'Screening for dense breasts: Digital breast tomosynthesis', *American Journal of Roentgenology*, 204(2), pp. 261–264. doi: 10.2214/AJR.14.13554.

Dobbins III, J. T. *et al.* (1995) 'DQE(f) of four generations of computed radiography acquisition devices', *Medical Physics*, 22(10), pp. 1581–1593.

Dobbins, J. T. *et al.* (2006) 'Intercomparison of methods for image quality characterization. II. Noise power spectrum', *Medical Physics*, 33(5), pp. 1466–1475.

Dobbins, James T. *et al.* (2006) 'Intercomparison of methods for image quality characterization. II. Noise power spectruma)', *Medical Physics*, 33(5), p. 1454. doi: 10.1118/1.2188819.

Donini, B. *et al.* (2014) 'Free software for performing physical analysis of systems for digital radiography and mammography', *Medical Physics*, 41(5), p. 051903. doi: 10.1118/1.4870955.

Durand, M. (2018) 'Synthesized Mammography: Clinical Evidence, Appearance, and Implementation', *Diagnostics*, 8(2), p. 22. doi: 10.3390/diagnostics8020022.

Eberhard, J. W. *et al.* (2006) 'High Speed , Large Angle Mammography Tomosynthesis System', in *Medical Imaging 2006: Physics of Medical Imaging*. Bethesda: SPIE, Bellingham, WA, pp. 61420C1-11. doi: 10.1117/12.653298.

Elangovan, P. *et al.* (2017) 'Design and validation of realistic breast models for use in multiple alternative forced choice virtual clinical trials', *Physics in Medicine and Biology*. IOP Publishing, 62(7), pp. 2778–2794. doi: 10.1088/1361-6560/aa622c.

van Engen, R. E. *et al.* (2013) *European guidelines for quality assurance in breast cancer screening and diagnosis*. 4th ed. European Union.

van Engen, R. E. *et al.* (2018) 'Protocol for the Quality Control of the Physical and Technical Aspects of Digital Breast Tomosynthesis Systems v1.03'.

Van Engen, R. E. *et al.* (2010) 'A supplement to the European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis 2 Adaptations Compared to the Fourth Edition of the European', *Proceeding of International Workshop Digital Mammography*, 6136(Perry 2006), pp. 643–650. doi: 10.1007/978-3-642-13666-5_87.

van Engen, R. *et al.* (2006) 'European protocol for the quality control of the physical and technical aspects of mammography screening. Part B: Digital mammography'. Luxembourg: European Commission.

European Commission (2013) *European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition supplement. Digital mammography update*, *Office for Official Publications of the European Communities*.

Fahrig, R. and Yaffe, M. (1994) 'A model for optimization of spectral shape in digital mammography', *Medical Physics*, 21(9), pp. 1463–1471. doi: 10.1118/1.597406.

Filippou, V. and Tsoumpas, C. (2018) 'Recent advances on the development of phantoms using 3D printing for imaging with CT, MRI, PET, SPECT, and ultrasound', *Medical Physics*, 45(9), pp. e740–e760. doi: 10.1002/mp.13058.

Fitton, I. *et al.* (2022) 'Design and initial evaluation of a novel physical breast phantom to monitor image quality in digital breast tomosynthesis', *Medical Physics*, 49(4), pp. 2355–2365. doi: 10.1002/mp.15498.

Fredenberg, E. *et al.* (2013) 'Measurement of breast-tissue x-ray attenuation by spectral mammography: First results on cyst fluid', *Physics in Medicine and Biology*, 58(24), pp. 8609–8620. doi: 10.1088/0031-9155/58/24/8609.

Accepted Manuscript

Freed, M. *et al.* (2009) 'Experimental validation of Monte Carlo (MANTIS) simulated x-ray response of columnar CsI scintillator screens', *Medical Physics*, 36(11), p. 4944. doi: 10.1118/1.3233683.

Freed, M., Park, S. and Badano, A. (2010) 'A fast, angle-dependent, analytical model of CsI detector response for optimization of 3D x-ray breast imaging systems', *Medical Physics*, 37(6), p. 2593. doi: 10.1118/1.3397462.

Fryback, D. G. and Thornbury, J. R. (1991) 'The Efficacy of Diagnostic Imaging', *Medical Decision Making*, 11, pp. 88–94.

FujiFilm ASPIRE Cristalle SSED, FDA (2017). Available at: https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160031B.pdf (Accessed: 4 January 2022).

Gang, G. J. *et al.* (2010) 'Anatomical background and generalized detectability in tomosynthesis and cone-beam CT', *Medical Physics*, 37(5), p. 1948. doi: 10.1118/1.3352586.

Gang, G. J. *et al.* (2011) 'Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance', *Medical Physics*, 38(4), pp. 1754–1768. doi: 10.1118/1.3560428.

Geise, A. and Palchevsky, A. (1996) 'Composition of mammographic phantom materials', *Radiology*, 196, pp. 347–350.

Gennaro, G. *et al.* (2007) 'Automated analysis of phantom images for the evaluation of long-term reproducibility in digital mammography.', *Physics in medicine and biology*, 52(5), pp. 1387–407. doi: 10.1088/0031-9155/52/5/012.

Gennaro, G., Bernardi, D. and Houssami, N. (2018) 'Radiation dose with digital breast tomosynthesis compared to digital mammography : per-view analysis', *Eur. Radiol.* European Radiology, 28, pp. 573–581. doi: 10.1007/s00330-017-5024-4.

Giger, M. L. and Doi, K. (1984) 'Investigation of basic imaging properties in digital radiography. I. Modulation transfer function', *Medical Physics*, 11(3), pp. 287–295.

Gilbert, F. J. *et al.* (2015) 'The TOMMY trial : a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme – a multicentre retrospective', *Health Technology Assessment*, 19(4). doi: 10.3310/hta19040.

Gilbert, F. J., Tucker, L. and Young, K. C. (2016) 'Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool', *Clinical Radiology*. The Royal College of Radiologists, 71(2), pp. 141–150. doi: 10.1016/j.crad.2015.11.008.

Glick, S. J. and Ikejimba, L. C. (2018) 'Advances in digital and physical anthropomorphic breast phantoms for x-ray imaging', *Medical Physics*, 45(10), pp. e870–e885. doi: 10.1002/mp.13110.

Goodsitt, M. M. *et al.* (2014) 'Digital breast tomosynthesis: Studies of the effects of acquisition geometry on contrast-to-noise ratio and observer preference of low-contrast objects in breast phantom images', *Physics in Medicine and Biology*, 59(19), pp. 5883–5902. doi: 10.1088/0031-9155/59/19/5883.

Graff, C. G. (2016) 'A new, open-source, multi-modality digital breast phantom', in *Medical Imaging 2016: Physics of Medical Imaging*, p. 978309. doi: 10.1117/12.2216312.

Green, D. M. and Swets, J. A. (1966) *Signal detection theory and psychophysics*. John Wiley.

Hajdok, G., Battista, J. J. and Cunningham, I. a. (2008) 'Fundamental x-ray interaction limits in diagnostic imaging detectors: Spatial resolution', *Medical Physics*, 35(7), p. 3180. doi: 10.1118/1.2924219.

Hajdok, G. and Cunningham, I. A. (2004) 'Penalty on the detective quantum efficiency from off-axis incident x rays', in *Proceedings of SPIE 5368 - The International Society for Optical Engineering*. SPIE, Bellingham,

Accepted Manuscript

WA, pp. 109–118.

Hammerstein, G. *et al.* (1979) 'Absorbed Radiation Dose in Mammography', *Radiology*, 130, pp. 485–491.

Hendrick, R. E. *et al.* (2010) 'Comparison of acquisition parameters and breast dose in digital mammography and screen-film mammography in the American College of Radiology imaging network digital mammographic imaging screening trial', *American Journal of Roentgenology*, 194(2), pp. 362–369. doi: 10.2214/AJR.08.2114.

Hernandez, A. M., Seibert, J. A. and Boone, J. M. (2015) 'Breast dose in mammography is about 30% lower when realistic heterogeneous glandular distributions are considered', *Medical Physics Med. Phys*, 42(101), pp. 6337–6348. doi: 10.1118/1.4931966.

Honey, I. *et al.* (2018) 'IPEM Topical Report: an evidence and risk assessment based analysis of the efficacy of tube and generator quality assurance tests on general x-ray units', *Physics in Medicine and Biology*, 63, p. 245011 (12pp). doi: https://doi.org/10.1088/1361-6560/aaf112.

Houssami, N. *et al.* (2021) 'Meta-analysis of prospective studies evaluating breast cancer detection and interval cancer rates for digital breast tomosynthesis versus mammography population screening', *European Journal of Cancer*, 148, pp. 14–23. doi: 10.1016/j.ejca.2021.01.035.

Hovda, T. *et al.* (2020) 'Interval and consecutive round breast cancer after digital breast tomosynthesis and synthetic 2D mammography versus standard 2D digital mammography in BreastScreen Norway', *Radiology*, 294(2), pp. 256–264. doi: 10.1148/radiol.2019191337.

Hsu, C. M. L. *et al.* (2013) 'Generation of a suite of 3D computer-generated breast phantoms from a limited set of human subject data', *Medical Physics*, 40(4), pp. 043703-1–11. doi: 10.1118/1.4794924.

Hu, Y.-H., Zhao, B. and Zhao, W. (2008) 'Image artifacts in digital breast tomosynthesis: Investigation of the effects of system geometry and reconstruction parameters using a linear system approach', *Medical Physics*, 35(12), p. 5242. doi: 10.1118/1.2996110.

Hu, Y., Scaduto, D. A. and Zhao, W. (2012) 'The Effect of Amorphous Selenium Thickness on Imaging Performance of Contrast Enhanced Digital Breast Tomosynthesis', pp. 9–16.

Huang, S. *et al.* (2008) 'The effect of skin thickness determined using breast CT on mammographic dosimetry', pp. 1199–1206. doi: 10.1118/1.2841938.

Hubbell, J. and Seltzer, S. (1995) *Hubbell, J. Seltzer, S., Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients 1 keV to 20 MeV for Elements Z = 1 to 92 and 48 Additional Substances of Dosimetric Interest*. Available at: http://physics.nist.gov/PhysRefData/XrayMassCoef/cover.html,.

ICRP (2007) *ICRP Publication 103. The 2007 Recommendations of the International Commission on Radiological Protection; 2007*.

ICRU (1996) *ICRU Rep. No. 54 - Medical Imaging The Assessment of Image Quality*. Bethesda, MD.

IEC (2007) 'IEC 62220-1-2:2007 Characteristics of digital X-ray imaging devices - Part 1-2: Determination of the detective quantum efficiency - Detectors used in mammography'. Geneva, Switzerland. Available at: https://webstore.iec.ch/publication/6598.

IEC (2013) 'IEC 60627:2013 Characteristics of General Purpose and Mammographic Anti-Scatter Grids'. Geneva, Switzerland: International Electrotechnical Commission (IEC). Available at: https://webstore.iec.ch/publication/2740.

IEC (2020a) 'IEC 60336:2020 X-ray tube assemblies for medical diagnosis - Focal spot dimensions and related characteristics'. Geneva, Switzerland: International Electrotechnical Commission. Available at: https://webstore.iec.ch/publication/62504.

Accepted Manuscript

IEC (2020b) 'IEC 61223-3-6:2020 Part 3-6: Acceptance and constancy tests – Imaging performance of mammographic X-ray equipment used in a mammographic tomosynthesis mode of operation'. IEC.

Ikejimba, L. C. *et al.* (2016) 'Assessing task performance in FFDM, DBT, and synthetic mammography using uniform and anthropomorphic physical phantoms', *Medical Physics*, 43(10), pp. 5593–5602. doi: 10.1118/1.4962475.

Ikejimba, L. C. *et al.* (2017) 'A novel physical anthropomorphic breast phantom for 2D and 3D x-ray imaging', *Medical Physics*, 44(2), pp. 407–416. doi: 10.1002/mp.12062.

Ikejimba, L. C. *et al.* (2021) 'Assessment of task-based performance from five clinical DBT systems using an anthropomorphic breast phantom', *Medical Physics*, 48(3), pp. 1026–1038. doi: 10.1002/mp.14568.

International Atomic Energy Agency (IAEA) (2007) *TECHNICAL REPORTS SERIES No. 457. DOSIMETRY IN DIAGNOSTIC RADIOLOGY: AN INTERNATIONAL CODE OF PRACTICE*. Vienna (Austria).

Ivanov, D. *et al.* (2018) 'Suitability of low density materials for 3D printing of physical breast phantoms', *Physics in Medicine and Biology*. IOP Publishing, 63(17), p. 175020 (14pp). doi: 10.1088/1361-6560/aad315.

Jacobs, J. *et al.* (2010) 'Towards an international consensus strategy for periodic quality control of digital breast tomosynthesis systems', in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. Bellingham: SPIE, Bellingham, WA, p. 76220G. doi: 10.1117/12.844228.

Jahnke, P. *et al.* (2017) 'Radiopaque three-dimensional printing: A method to create realistic CT phantoms', *Radiology*, 282(2), pp. 569–575. doi: 10.1148/radiol.2016152710.

Johns, P. C. and Yaffe, M. J. (1987) 'X-ray characterisation of normal and neoplastic breast tissues.', *Physics in medicine and biology*, 32(6), pp. 675–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3039542.

Jousi, M. O. *et al.* (2019) 'A new breast tomosynthesis imaging method : Continuous Sync-and-Shoot – technical feasibility and initial experience', *Acta Radiologica Open*, 8(3), pp. 1–9. doi: 10.1177/2058460119836255.

Kak, A. and Slaney, M. (1988) *Principles of Computerized Tomographic Imaging*. New York: IEEE.

Karssemeijer, N. and Thiijssen, M. A. O. (1996) 'Determination of contrast-detail curves of mammography systems by automated image analysis', in *Digital Mammography*. Amsterdam: Elsevier, pp. 155–160.

Kelaranta, A. *et al.* (2015) 'Conformance of mean glandular dose from phantom and patient data in mammography', *Radiation Protection Dosimetry*, 164(3), pp. 342–353.

Kiarashi, N. *et al.* (2015) 'Development of realistic physical breast phantoms matched to virtual breast phantoms based on human subject data', *Medical Physics*, 42(7), pp. 4116–4126. doi: 10.1118/1.4919771.

Kiarashi, N. *et al.* (2016) 'Impact of breast structure on lesion detection in breast tomosynthesis, a simulation study', *Journal of Medical Imaging*, 3(3), pp. 035504-1–13. doi: 10.1117/1.JMI.3.3.035504.

Klausz, R. *et al.* (2018) 'Properties of a new high-ratio anti-scatter grid in digital mammography', in *ECR EPOS*. European Society of Radiology, pp. 1–21. doi: 10.1594/ecr2018/C-2647.

Klausz, R. and Shramchenko, N. (2005) 'Dose to population as a metric in the design of optimised exposure control in digital mammography', *Radiation Protection Dosimetry*, 114(1–3), pp. 369–374. doi: 10.1093/rpd/nch579.

Klein, R. *et al.* (1997) 'Determination of average glandular dose with modern mammography units for two large groups of patients', *Physics in Medicine and Biology*, 42, pp. 651–671.

Konstantinidis, A. C. *et al.* (2013) 'X-ray Performance Evaluation of the Dexela CMOS APS X-ray Detector

Accepted Manuscript

Using Monochromatic Synchrotron Radiation in the Mammographic Energy Range', *IEEE Transactions on Nuclear Science*, 60(5), pp. 3969–3980.

Konstantinidis, A. C., Olivo, A. and Speller, R. D. (2011) 'Modification of the standard gain correction algorithm to compensate for the number of used reference flat frames in detector performance studies', *Medical Physics*, 38(12), pp. 6683–6687. doi: 10.1118/1.3664003.

Kotre, C. J. and Porter, D. J. T. (2005) 'A printed image quality test phantom for mammography', *British Journal of Radiology*, 78, pp. 746–748. doi: 10.1259/bjr/11930472.

Kotre, C. J., Robson, K. J. and Faulkner, K. (1994) 'Measurement of the frequency distribution of optical density in screening mammography', *British Journal of Radiology*, 67, pp. 856–859.

Kwan, A. L. C., Seibert, J. A. and Boone, J. M. (2006) 'An improved method for flat-field correction of flat panel x-ray detector', *Medical Physics*, 33(2), p. 391. doi: 10.1118/1.2163388.

Lång, K. *et al.* (2015) 'Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study', *European Radiology*. doi: 10.1007/s00330-015-3803-3.

Lee, C. and Baek, J. (2015) 'A New Method to Measure Directional Modulation Transfer Function Using Sphere Phantoms in a Cone Beam Computed Tomography System', *IEEE transactions on medical imaging*, 34(4), pp. 902–910.

Lee, C. and Baek, J. (2017) 'Inverse filtering approach to measure directional in-plane modulation transfer function using a sphere phantom for a digital tomosynthesis system', *Optics Express*, 25(15), p. 17280. doi: 10.1364/oe.25.017280.

Lee, C., Song, H. and Baek, J. (2020) '3D MTF estimation using sphere phantoms for cone-beam computed tomography systems', *Medical Physics*, 47(7), pp. 2838–2851.

Li, B. *et al.* (2007) 'Optimization of slice sensitivity profile for radiographic tomosynthesis', *Medical Physics*, 34(7), pp. 2907–2916. doi: 10.1118/1.2742499.

Li, T., Marinovich, M. L. and Houssami, N. (2018) 'Expert Review of Anticancer Therapy Digital breast tomosynthesis (3D mammography) for breast cancer screening and for assessment of screen-recalled findings : review of the evidence', *Expert Review of Anticancer Therapy*. Taylor & Francis, 18(8), pp. 785–791. doi: 10.1080/14737140.2018.1483243.

Li, Z. *et al.* (2016) 'A Novel 3D Stochastic Solid Breast Texture Model for X-Ray Breast Imaging', in *LNCS 9699*, pp. 660–667. doi: 10.1007/978-3-319-41546-8_82.

Lindfors, K. K. *et al.* (2008) 'Dedicated Breast CT : Initial Clinical Experience', *Radiology*, 246(3), pp. 725–733.

Liu, H. *et al.* (2014) 'Evaluation of clinical full field digital mammography with the task specific system-model-based Fourier Hotelling observer (SMFHO) SNR', *Medical Physics*, 41(5), pp. 051907-1–8. doi: 10.1118/1.4870377.

Mackenzie, A. (2008) 'Validation of correction methods for the non-linear response of digital radiography systems.', *The British Journal of Radiology*, 81(964), pp. 341–5.

Mackenzie, A. *et al.* (2016) 'The relationship between cancer detection in mammography and image quality measurements', *Physica Medica*. Associazione Italiana di Fisica Medica, 32(4), pp. 568–574. doi: 10.1016/j.ejmp.2016.03.004.

Mackenzie, A. *et al.* (2017) 'Characterisation of noise and sharpness of images from four digital breast tomosynthesis systems for simulation of images for virtual clinical trials', *Physics in Medicine and Biology*. IOP Publishing, 62(6), pp. 2376–2397. doi: 10.1088/1361-6560/aa5dd9.

Accepted Manuscript

Mackenzie, A. *et al.* (2020) 'Methodology for undertaking quality control testing of ghosting in digital breast tomosynthesis systems', in *15th International Workshop on Breast Imaging (IWBI2020)*. SPIE, Bellingham, WA, p. 115132I.

Maidment, A. D. A. and Albert, M. (2003) 'Conditioning data for calculation of the modulation transfer function', *Medical Physics*, 30(2), pp. 248–253. doi: 10.1118/1.1534111.

Mainprize, J. G. *et al.* (2006) 'Resolution at oblique incidence angles of a flat panel imager for breast tomosynthesis', *Medical Physics*, 33(9), p. 3159. doi: 10.1118/1.2241994.

Mainprize, J. G. *et al.* (2020) 'Full-size anthropomorphic phantom for 2D and 3D breast x-ray imaging', in *Proceedings of SPIE - The International Society for Optical Engineering Vol 11513*, p. 1151302. doi: 10.1117/12.2560358.

Mainprize, J. G., Wang, X. and Yaffe, M. J. (2009) 'The effect of lag on image quality for a digital breast tomosynthesis system', in *Proceedings of SPIE - The International Society for Optical Engineering*. Spie, pp. 72580R-72580R–12. doi: 10.1117/12.811889.

Maki, A. K., Mainprize, J. G. and Yaffe, M. J. (2016) 'Considerations for robust measurement of the slice-sensitivity profile in tomosynthesis mammography for quality control', *Medical Physics*, 43(8), pp. 4803–4807. doi: 10.1118/1.4959539.

Maldera, A. *et al.* (2017) 'Digital breast tomosynthesis: Dose and image quality assessment', *Physica Medica*. Associazione Italiana di Fisica Medica, 33, pp. 56–67. doi: 10.1016/j.ejmp.2016.12.004.

De Man, B. *et al.* (2007) 'CatSim: a new computer assisted tomography simulation environment', in *Medical Imaging 2007: Physics of Medical Imaging*, p. 65102G. doi: 10.1117/12.710713.

Mandoul, C. *et al.* (2019) 'Breast tomosynthesis: What do we know and where do we stand?', *Diagnostic and Interventional Imaging*, 100(10), pp. 537–551. doi: 10.1016/j.diii.2019.07.012.

Marinov, S. *et al.* (2020) 'Evaluation of the visual realism of breast texture phantoms in digital mammography', in *Proceedings Volume 11513, 15th International Workshop on Breast Imaging (IWBI2020);* SPIE, Bellingham, WA, p. 1151314. doi: https://doi.org/10.1117/12.2564124.

Marshall, N. W. (2006a) 'A comparison between objective and subjective image quality measurements for a full field digital mammography system', *Physics in Medicine and Biology*, 51, pp. 2441–2463 PHYSICS. doi: 10.1088/0031-9155/51/10/006.

Marshall, N. W. (2006b) 'Retrospective analysis of a detector fault for a full field digital mammography system', *Physics in Medicine and Biology*, 51(21), pp. 5655–5673. doi: 10.1088/0031-9155/51/21/018.

Marshall, N. W. (2007) 'Early experience in the use of quantitative image quality measurements for the quality assurance of full field digital mammography x-ray systems', *Physics in Medicine and Biology*, 52(18), pp. 5545–5568. doi: 10.1088/0031-9155/52/18/006.

Marshall, N. W. (2009) *Calculation of quantitative image quality parameters—notes describing the use of OBJ_IQ_reduced.* Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/442798/nhsbsp-equipment-report-0902.pdf.

Marshall, N. W. *et al.* (2010) 'Technical evaluation of a digital breast tomosynthesis system', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 350–356. doi: 10.1007/978-3-642-13666-5_47.

Marshall, N. W. *et al.* (2011) 'Image quality assessment in digital mammography: Part I. Technical characterization of the systems', *Physics in Medicine and Biology*, 56, pp. 4201–4220. doi: 10.1088/0031-9155/56/14/002.

Accepted Manuscript

Marshall, N. W. and Bosmans, H. (2012) 'Measurements of system sharpness for two digital breast tomosynthesis systems', *Physics in Medicine and Biology*, 57, pp. 7629–7650. doi: 10.1088/0031-9155/57/22/7629.

Marshall, N. W., Mackenzie, A. and Honey, I. D. (2011) 'Quality control measurements for digital x-ray detectors', *Physics in Medicine and Biology*, 56(4), pp. 979–999. doi: 10.1088/0031-9155/56/4/007.

Martin, C. J. (2007) 'An evaluation of semiconductor and ionization chamber detectors for diagnostic x-ray dosimetry measurements.', *Physics in medicine and biology*, 52(15), pp. 4465–80.

Mawdsley, G. E. *et al.* (2009) 'Accurate estimation of compressed breast thickness in mammography', *Medical Physics*, 36(2), p. 577. doi: 10.1118/1.3065068.

McGarry, C. K. *et al.* (2020) 'Tissue mimicking materials for imaging and therapy phantoms: A review', *Physics in Medicine and Biology*, 65, p. 23TR01. doi: 10.1088/1361-6560/abbd17.

Metz, C. E. and Doi, K. (1979) 'Transfer function analysis of radiographic imaging systems.', *Physics in Medicine and Biology*, 24(6), pp. 1079–1106.

Michielsen, K. *et al.* (2013) 'Patchwork reconstruction with resolution modeling for digital breast tomosynthesis', *Medical Physics*, 40(3), pp. 031105-1–10. doi: 10.1118/1.4789591.

Mittal, A., Moorthy, A. K. and Bovik, A. C. (2012) 'No-reference image quality assessment in the spatial domain', *IEEE Transactions on Image Processing*. IEEE, 21(12), pp. 4695–4708. doi: 10.1109/TIP.2012.2214050.

Monnin, P. *et al.* (2011) 'Image quality assessment in digital mammography: Part II. NPWE as a validated alternative for contrast detail analysis', *Physics in Medicine and Biology*, 56(14). doi: 10.1088/0031-9155/56/14/003.

Monnin, P. *et al.* (2014) 'Comparison of the polynomial model against explicit measurements of noise components for different mammography systems', *Physics in Medicine and Biology*, 59(19), pp. 5741–5761. doi: 10.1088/0031-9155/59/19/5741.

Monnin, P. *et al.* (2017) 'A comprehensive model for x-ray projection imaging system efficiency and image quality characterization in the presence of scattered radiation', *Physics in Medicine and Biology*, 62(14), pp. 5691–5722. doi: 10.1088/1361-6560/aa75bc.

Monnin, P. *et al.* (2020) 'In-plane image quality and NPWE detectability index in digital breast tomosynthesis', *Physics in Medicine and Biology*, 65(9). doi: 10.1088/1361-6560/ab8147.

Moy, J. and Bosset, B. (1999) 'How does real offset and gain correction affect the DQE in images from x-ray flat detectors?', in *SPIE Physics of Medical Imaging*. SPIE, Bellingham, WA, pp. 90–97.

Neculaes, V. B. *et al.* (2014) 'Multisource X-Ray and CT: Lessons learned and future outlook', *IEEE Access*. IEEE, 2, pp. 1568–1585. doi: 10.1109/ACCESS.2014.2363949.

Neitzel, U. (1992) 'Grids or air gaps for scatter reduction in digital radiography: A model calculation', *Medical Physics*, 19(2), pp. 475–481.

Nelson, J. S. *et al.* (2016) 'How does C-VIEW image quality compare with conventional 2D FFDM?', *Medical Physics*, 43(5), pp. 2538–2547. doi: 10.1118/1.4947293.

Niklason, L. T. *et al.* (1997) 'Digital tomosynthesis in breast imaging.', *Radiology*, 205(2), pp. 399–406. doi: 10.1259/bjr/80279516.

Nishikawa, R. M. *et al.* (1987) 'Scanned-projection digital mammography', *Medical Physics*, 14(5), pp. 717–727.

Accepted Manuscript

Nosratieh, A. *et al.* (2015) 'Mean glandular dose coefficients (DgN) for x-ray spectra used in contemporary breast imaging systems', *Physics in Medicine and Biology*. IOP Publishing, 60(18), pp. 7179–7190. doi: 10.1088/0031-9155/60/18/7179.

Objects, L. T. (2021) *Leeds Test Objects*, *Leeds Test Objects*. Available at: https://www.leedstestobjects.com/#phantoms-heading.

Olgar, T., Kahn, T. and Gosch, D. (2013) 'Quantitative Image Quality Measurements of a Digital Breast Tomosynthesis System.', *Fortschr Röntgenstr*. doi: 10.1055/s-0033-1350106.

Østerås, B. H. *et al.* (2018) 'Average glandular dose in paired digital mammography and digital breast tomosynthesis acquisitions in a population based screening program : effects of measuring breast density , air kerma and beam quality', *Physics in Medicine Biology*. IOP Publishing, 63, p. 035006 (14pp).

Park, S. *et al.* (2010) 'A statistical , task-based evaluation method for three-dimensional x-ray breast imaging systems using variable-background phantoms', *Medical Physics*, 37(12), pp. 6253–6270. doi: 10.1118/1.3488910.

Patel, T. *et al.* (2012) 'Detective Quantum Efficiency of a CsI-CMOS X-ray Detector for Breast Tomosynthesis Operating in High Dynamic Range and High Sensitivity Modes', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) Vol. 7361 IWDM 2012*, pp. 80–87.

Persson, M., Rajbhandary, P. L. and Pelc, N. J. (2018) 'A framework for performance characterization of energy-resolving photon-counting detectors', *Medical Physics*, 45(11), pp. 4897–4915.

Petropoulos, A. E. *et al.* (2020) 'Quantitative assessment of microcalcification cluster image quality in digital breast tomosynthesis, 2-dimensional and synthetic mammography', *Medical and Biological Engineering and Computing*. Medical & Biological Engineering & Computing, 58(1), pp. 187–209. doi: 10.1007/s11517-019-02072-0.

Petrov, D., Marshall, N. W., Young, K., *et al.* (2019) 'Model and human observer reproducibility for detection of microcalcification clusters in digital breast tomosynthesis images of three-dimensionally structured test object', *Journal of Medical Imaging*, 6(1), pp. 015503-1–8. doi: 10.1117/1.JMI.6.1.015503.

Petrov, D., Marshall, N. W., Young, K. C., *et al.* (2019) 'Systematic approach to a channelized Hotelling model observer implementation for a physical phantom containing mass-like lesions: Application to digital breast tomosynthesis', *Physica Medica*, 58, pp. 8–20. doi: 10.1016/j.ejmp.2018.12.033.

Petrov, D. *et al.* (2020) 'Deep learning channelized Hotelling observer for multi-vendor DBT system image quality evaluation', in *Proc. SPIE 11316, Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment*, p. 113160X. doi: 10.1117/12.2548998.

PhantomX-Gmbh (2021) *PhantomX*. Berlin. Available at: https://www.thephantomx.com/breast-phantoms/.

Porzio, M. and Konstantinidis, A. C. (2021) 'MAMMO_QC: Free software for quality control (QC) analysis in digital mammography and digital breast tomosynthesis compliant with the European guidelines and EUREF/EFOMP protocols', *Biomedical Physics and Engineering Express*. IOP Publishing, 7, pp. 067004-1–8. doi: 10.1088/2057-1976/ac2076.

Prakash, P. *et al.* (2011) 'Task-Based Modeling and Optimization of a Dedicated Cone-Beam CT Scanner for Musculoskeletal Imaging', *Medical Physics*, 38(10), pp. 5612–5628.

Qian, X. *et al.* (2012) 'High resolution stationary digital breast tomosynthesis using distributed carbon nanotube x-ray source array', *Medical Physics*, 39(4), pp. 2090–2099. doi: 10.1118/1.3694667.

QUART-Gmbh (2021) *QUART Gmbh*. Available at: https://quart.de/en/products/test-phantoms/mammography (Accessed: 16 June 2021).

Accepted Manuscript

Que, W. and Rowlands, J. A. (1995) 'X-ray imaging using amorphous selenium: Inherent spatial resolution', *Medical Physics*, 22(4), pp. 365–374.

Ratanaprasatporn, L., Chikarmane, S. A. and Giess, C. S. (2017) 'Strengths and Weaknesses of Synthetic Mammography in Screening', *RadioGraphics*, 37(7), pp. 1913–1927. doi: 10.1148/rg.2017170032.

Reis, C. *et al.* (2013) 'Quality assurance and quality control in mammography: A review of available guidance worldwide', *Insights into Imaging*, 4(5), pp. 539–553. doi: 10.1007/s13244-013-0269-1.

Ren, B. *et al.* (2005) 'Design and performance of the prototype full field breast tomosynthesis system with selenium based flat panel detector', in *Medical Imaging 2005: Physics of Medical Imaging SPIE Voll. 5745*, pp. 550–561.

Richard, S. *et al.* (2005) 'Generalized DQE analysis of radiographic and dual-energy imaging using flat-panel detectors', *Medical Physics*, 32(5), pp. 1397–1413. doi: 10.1118/1.1901203.

Richard, S. and Siewerdsen, J. H. (2008) 'Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images', *Medical Physics*, 35(11), pp. 5043–5053. doi: 10.1118/1.2988161.

Ritlumlert, N. *et al.* (2020) 'Assessment of average glandular dose received in full-field digital mammography and digital breast tomosynthesis', *Journal of Health Science and Medical Research*, 38(2), pp. 115–123. doi: 10.31584/jhsmr.2020730.

Robinson, M. (2008) 'Trends in compressed breast thickness and radiation dose in breast screening mammography', 81(March), pp. 214–218. doi: 10.1259/bjr/90916004.

Robson, K. J., Kotre, C. J. and Faulkner, K. (1995) 'The use of a contrast-detail test object in the optimization of optical density in mammography.', *The British journal of radiology*, 68, pp. 277–282.

Rodríguez-Ruiz, A. *et al.* (2016) 'Evaluation of the technical performance of three different commercial digital breast tomosynthesis systems in the clinical environment', *Physica Medica*, 32(6), pp. 767–777. doi: 10.1016/j.ejmp.2016.05.001.

Rose, A. (1948) 'The sensitivity performance of the human eye on an absolute scale', *J. Opt. Soc. Am.*, 38, pp. 196–208.

Rossman, A. H. *et al.* (2019) 'Three-dimensionally-printed anthropomorphic physical phantom for mammography and digital breast tomosynthesis with custom materials, lesions, and uniform quality control region', *Journal of Medical Imaging*, 6(02), pp. 021604-1–8. doi: 10.1117/1.jmi.6.2.021604.

Rossmann, K. (1969) 'Point Spread-Function, Line Spread Function, and Modulation Transfer Function', *Radiologic clinics of North America*, 93(August), pp. 257–272. doi: 10.1148/93.2.257.

Rowlands, J. and Yorkston, J. (2000) 'Flat panel detectors for digital radiology', in Van Metter, R., Beutel, J., and Kundel, H. (eds) *Handbook of Medical Imaging, Volume 1. Physics and Psychophysics*. Bellingham: SPIE, Bellingham, WA, pp. 223–328.

Sage, J. *et al.* (2019) 'Experimental evaluation of seven quality control phantoms for digital breast tomosynthesis', *Physica Medica*. Elsevier, 57(January), pp. 137–144. doi: 10.1016/j.ejmp.2018.12.031.

Salomon, E. *et al.* (2020) 'Equivalent breast thickness and dose sensitivity of a next iteration 3D structured breast phantom with lesion models', in *Medical Imaging 2020: Physics of Medical Imaging*. SPIE, Bellingham, WA, p. 113120K. doi: 10.1117/12.2548956.

Salvagnini, E. *et al.* (2015) 'Tailoring automatic exposure control toward constant detectability in digital mammography', *Medical Physics*, 42(7), pp. 3834–3847. doi: 10.1118/1.4921417.

Salvagnini, E. *et al.* (2016) 'Impact of compressed breast thickness and dose on lesion detectability in digital

Accepted Manuscript

mammography: FROC study with simulated lesions in real mammograms', *Medical Physics*, 43(9), pp. 5104–5116. doi: 10.1118/1.4960630.

Samei, E. *et al.* (2018) 'Report of AAPM Task Group 162: Software for planar image quality metrology: Software', *Medical Physics*, 45(2), pp. e32–e39. doi: 10.1002/mp.12718.

Samei, E., Flynn, M. J. and Reimann, D. A. (1998) 'A method for measuring the presampled MTF of digital radiographic systems using an edge test device.', *Medical Physics*, 25(1), pp. 102–113.

Samei, E., Murphy, S. and Richard, S. (2013) 'Assessment of multi-directional MTF for breast tomosynthesis', *Physics in Medicine and Biology*, 58, pp. 1649–1661. doi: 10.1088/0031-9155/58/5/1649.

Sánchez de la Rosa, R. (2019) *Simulations and virtual clinical trials for the assessment of the added clinical value of angio-tomosynthesis over angio-mammography*. l'Universit´e Paris-Saclay.

Sandborg, M. *et al.* (2000) 'a Voxel Phantom Based Monte Carlo Computer Program for Optimisation of Chest and Lumbar Spine X Ray Imaging Systems', *Radiation Protection Dosimetry*, 90, pp. 105–108.

Sarno, A. *et al.* (2017) 'A Monte Carlo study of monoenergetic and polyenergetic normalized glandular dose (DgN) coefficients in mammography', *Physics in Medicine and Biology*. IOP Publishing, 62(1), pp. 306–325. doi: 10.1088/1361-6560/62/1/306.

Sarno, A. *et al.* (2021) 'Normalized glandular dose coefficients for digital breast tomosynthesis systems with a homogeneous breast model Normalized glandular dose coef fi cients for digital breast tomosynthesis systems with a homogeneous breast model', *Physics in Medicine & Biology*. IOP Publishing, 66(6), p. 65024. doi: 10.1088/1361-6560/abe2e9.

Schafer, D., Wiegert, J. and Bertram, M. (2007) 'Method for the determination of the modulation transfer function (MTF) in 3D x-ray imaging systems with focus on correction for finite extent of test objects', *Proceedings of SPIE*. Spie, 6510, pp. 65104Q-65104Q–10. doi: 10.1117/12.708924.

Schmidgunst, C., Ritter, D. and Lang, E. (2007) 'Calibration model of a dual gain flat panel detector for 2D and 3D x-ray imaging', *Medical Physics*, 34(9), pp. 3649–3664. doi: 10.1118/1.2760024.

Schmitzberger, F. F. *et al.* (2011) 'Development of Low-Dose Photon-counting Contrast- enhanced Tomosynthesis with Spectral Imaging', *Radiology*, 259(2), pp. 558–564.

Schopphoven, S. *et al.* (2019) 'Breast phantoms for 2D digital mammography with realistic anatomical structures and attenuation characteristics based on clinical images using 3D printing', *Physics in Medicine and Biology*. IOP Publishing, 64, p. 215005 (12pp). doi: 10.1088/1361-6560/ab3f6a.

Sechopoulos, Ioannis *et al.* (2007) 'Computation of the glandular radiation dose in digital tomosynthesis of the breast', *Medical Physics*, 34(1), p. 221. doi: 10.1118/1.2400836.

Sechopoulos, I. *et al.* (2007) 'Scatter radiation in digital tomosynthesis of the breast', *Medical Physics*, 34(2), p. 564. doi: 10.1118/1.2428404.

Sechopoulos, I. *et al.* (2012) 'Characterization of the homogeneous tissue mixture approximation in breast imaging dosimetry', *Medical Physics*, 39(August), pp. 5050–5059.

Sechopoulos, I. (2013) 'A review of breast tomosynthesis . Part I . The image acquisition process', *Medical Physics*, 40(1), pp. 014301-1–12. doi: 10.1118/1.4770279.

Sechopoulos, I. *et al.* (2014) 'Radiation dosimetry in digital breast tomosynthesis : Report of AAPM Tomosynthesis Subcommittee Task Group 223', *Medical Physics*, 41(September), pp. 1–10.

Sechopoulos, I. and D'Orsi, C. J. (2008) 'Glandular radiation dose in tomosynthesis of the breast using tungsten targets', *Journal of Applied Clinical Medical Physics*, 9(4), pp. 161–171. doi: 10.1120/jacmp.v9i4.2887.

Accepted Manuscript

Segui, J. a. and Zhao, W. (2006) 'Amorphous selenium flat panel detectors for digital mammography: Validation of a NPWE model observer with CDMAM observer performance experiments', *Medical Physics*, 33(10), p. 3711. doi: 10.1118/1.2349689.

Shaheen, E. *et al.* (2011) 'The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis', *Medical Physics*, 38(12), pp. 6659–6671. doi: 10.1118/1.3662868.

Shaheen, E. *et al.* (2014) 'The simulation of 3D mass models in 2D digital mammography and breast tomosynthesis', *Medical Physics*, 41(8). doi: 10.1118/1.4890590.

Shaw, R. (1978) 'Evaluating the efficient of imaging processes', *Rep. Prog. Phys.*, 41, pp. 1103–1155.

Shen, S. Z. *et al.* (2006) 'Effect of scatter and an antiscatter grid on the performance of a slot-scanning digital mammography system', *Medical Physics*, 33(4), p. 1108. doi: 10.1118/1.2184445.

Shramchenko, N. and Klausz, R. (2004) 'Optimized exposure control in digital mammography', in *Proceedings of SPIE - The International Society for Optical Engineering*. Spie, pp. 445–456. doi: 10.1117/12.535463.

Shrestha, S., Vedantham, S. and Karellas, A. (2017) 'Towards standardization of x-ray beam filters in digital mammography and digital breast tomosynthesis: Monte Carlo simulations and analytical modelling', *Physics in Medicine and Biology*. IOP Publishing, 62(5), pp. 1969–1993. doi: 10.1088/1361-6560/aa58c8.

Siemens MAMMOMAT Inspiration SSED, FDA (2015).

Siewerdsen, J. H., Cunningham, I. a. and Jaffray, D. a. (2002) 'A framework for noise-power spectrum analysis of multidimensional images', *Medical Physics*, 29(11), p. 2655. doi: 10.1118/1.1513158.

Siewerdsen, J. H. and Jaffray, D. A. (1999) 'A ghost story: spatio-temporal response characteristics of an indirect-detection flat-panel imager.', *Medical physics*, 26(8), pp. 1624–1641.

Siewerdsen, J. H. and Jaffray, D. A. (2003) 'Three-dimensional NEQ transfer characteristics of volume CT using direct- and indirect-detection flat-panel imagers', *SPIE Medical Imaging 2003*, 5030, pp. 92–102. doi: 10.1117/12.479970.

de Sisternes, L. *et al.* (2015) 'A computational model to generate simulated three-dimensional breast masses', *Medical Physics*, 42(2), pp. 1098–1118. doi: 10.1118/1.4905232.

Skaane, P. *et al.* (2013) 'Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program', *Radiology*, 267(1), pp. 47–56. doi: 10.1148/radiol.12121373.

Skaane, P. *et al.* (2018) 'Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial', *Breast Cancer Research and Treatment*. Springer US, 169(3), pp. 489–496. doi: 10.1007/s10549-018-4705-2.

Skaane, P. *et al.* (2019) 'Digital mammography versus digital mammography plus tomosynthesis in breast cancer screening: The Oslo tomosynthesis screening trial', *Radiology*, 291(1), pp. 23–30. doi: 10.1148/radiol.2019182394.

Skubic, E. and Fatouros, P. (1989) 'The effect of breast composition on absorbed dose and contrast', *Medical Physics*, 16(4), pp. 544–552.

Stanton, L. *et al.* (1978) 'A breast phantom method for evaluating mammography technique', *Investigative Radiology*, 13(4), pp. 291–297.

Strudley, C. J. *et al.* (2015) 'Development and experience of quality control methods for digital breast tomosynthesis systems', *British Journal of Radiology*, 88(1056), pp. 1–11. doi: 10.1259/bjr.20150324.

Stuffins, M. *et al.* (2022) 'The relationship between age of digital mammmogrpahy systems and number of

Accepted Manuscript

reported faults and downtime', *Physica Medica*. Elsevier Ltd, 98(May), pp. 113–121. doi: 10.1016/j.ejmp.2022.04.015.

Sundell, V. *et al.* (2019) 'A phantom study comparing technical image quality of five breast tomosynthesis systems', *Physica Medica*. Elsevier, 63(January), pp. 122–130. doi: 10.1016/j.ejmp.2019.06.004.

Svahn, T. M. *et al.* (2015) 'Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography', *The Breast*. Elsevier Ltd, 24(2), pp. 93–99. doi: 10.1016/j.breast.2014.12.002.

Tapiovaara, M. J. and Wagner, R. F. (1985) 'SNR and DQE analysis of broad spectrum x-ray imaging', *Physics in Medicine and Biology*, 30(6), pp. 519–529.

Teuwen, J. *et al.* (2021) 'Deep learning reconstruction of digital breast tomosynthesis images for accurate breast density and patient-specific radiation dose estimation', *Medical Image Analysis*. Elsevier B.V., 71, p. 102061. doi: 10.1016/j.media.2021.102061.

The Phantom Laboratory DBT phantom (2021). Available at: https://www.phantomlab.com/tomophan-phantom (Accessed: 16 June 2021).

Theodorakou, C. *et al.* (2004) 'A novel method for producing x-ray test objects and phantoms', *Physics in Medicine and Biology*, 49, pp. 1423–1438. doi: 10.1088/0031-9155/49/8/004.

Thilander-Klang, A. C. *et al.* (1997) 'Influence of anode-filter combinations on image quality and radiation dose in 965 women undergoing mammography', *Radiology*, 203(2), pp. 348–354. doi: https://doi.org/10.1148/radiology.203.2.9114087.

Thornton, M. M. and Flynn, M. J. (2006) 'Measurement of the spatial resolution of a clinical volumetric computed tomography scanner using a sphere phantom', *Proceedings of SPIE*. Spie, 6142(7), pp. 61421Z-61421Z–10. doi: 10.1117/12.654969.

Toroi, P. *et al.* (2007) 'Experimental investigation on the choice of the tungsten/rhodium anode/filter combination for an amorphous selenium-based digital mammography system.', *European radiology*, 17(9), pp. 2368–75. doi: 10.1007/s00330-006-0574-x.

Tucker, A. W., Lu, J. and Zhou, O. (2013) 'Dependency of image quality on system configuration parameters in a stationary digital breast tomosynthesis system', *Medical Physics*, 40(3), pp. 031917-1–10. doi: 10.1118/1.4792296.

Tward, D. J. and Siewerdsen, J. H. (2008) 'Cascaded systems analysis of the 3D noise transfer characteristics of flat-panel cone-beam CT', *Medical Physics*, 35(12), p. 5510. doi: 10.1118/1.3002414.

Urbanczyk, H. *et al.* (2012) 'Characterization of the effects of the FineView algorithm for full field digital mammography', *Physics in Medicine and Biology*, 57(7). doi: 10.1088/0031-9155/57/7/1987.

Vancoillie, L. *et al.* (2020) 'Evaluation of possible phantoms for assessment of image quality in synthetic mammograms', in *Proc. SPIE Vol 11312*. Bellingham: SPIE, Bellingham, WA, pp. 113120J-1–8. doi: 10.1117/12.2548465.

Vancoillie, L. *et al.* (2021) 'The impact on lesion detection via a multi-vendor study: A phantom-based comparison of digital mammography, digital breast tomosynthesis, and synthetic mammography', *Medical Physics*, 48(10), pp. 6270–6292. doi: 10.1002/mp.15171.

Vedantham, S. *et al.* (2000) 'Full breast digital mammography with an amorphous silicon-based flat panel detector: physical characteristics of a clinical prototype.', *Medical physics*, 27(3), pp. 558–567. doi: 10.1118/1.598895.

Vedantham, S. *et al.* (2015) 'Digital Breast Tomosynthesis : State of the Art', *Radiology*, 277(3), pp. 663–684.

Accepted Manuscript

Vent, T. *et al.* (2022) 'Pre-clinical evaluation and optimization of image quality for a next generation tomosynthesis prototype', in *Proc. SPIE. 12031, Medical Imaging 2022: Physics of Medical Imaging*. SPIE, Bellingham, WA, p. 1203142. doi: https://doi.org/10.1117/12.2608811.

Vent, T. L., Acciavatti, R. J. and Maidment, A. D. A. (2021) 'Development and Evaluation of the Fourier Spectral Distortion Metric', *IEEE Transactions on Medical Imaging*, 40(3), pp. 1055–1064. doi: 10.1109/TMI.2020.3045325.

Wagner, L. K. *et al.* (1992) 'Recommendations on performance characteristics of diagnostic exposure meters Report of AAPM Diagnostic X-Ray Imaging Task Group No. 6', *Medical Physics*, 19(1), pp. 231–241.

Wallis, M. G. *et al.* (2012) 'Two-View and single-View Tomosynthesis versus Full-Field Digital Mammography: High-Resolution X-Ray Imaging Observer Study', *Radiology*, 262(3), pp. 788–796.

Wang, Z. *et al.* (2004) 'Image quality assessment: from error visibility to structural similarity.', *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 13(4), pp. 600–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15376593.

Warren, L. M. *et al.* (2012) 'Effect of image quality on calcification detection in digital mammography', *Medical Physics*, 39(6), pp. 3202–3213.

Warren, L. M. *et al.* (2013) 'Comparison of the x-ray attenuation properties of breast calcifications , aluminium , hydroxyapatite and calcium oxalate', *Physics in medicine and biology*, 58, pp. N103–N113. doi: 10.1088/0031-9155/58/7/N103.

Warren, L. M., Dance, D. R. and Young, K. C. (2016) 'Radiation risk of breast screening in England with digital mammography', *The British Journal of Radiology*, 89(1067), p. 20150897. doi: 10.1259/bjr.20150897.

White, D. R. *et al.* (1977) 'Epoxy-resin-based tissue substitutes.', *Medical physics*, 5(5), pp. 453–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/713980.

Wigati, K. T. *et al.* (2021) 'On the relevance of modulation transfer function measurements in digital mammography quality control', *Journal of Medical Imaging*, 8(02), pp. 023505-1–25. doi: 10.1117/1.jmi.8.2.023505.

Williams, M. B. *et al.* (2008) 'Optimization of exposure parameters in full field digital mammography', *Medical Physics*, 35(6), pp. 2414–2423.

Williams, M. B., Mangiafico, P. a and Simoni, P. U. (1999) 'Noise power spectra of images from digital mammography detectors.', *Medical physics*, 26(7), pp. 1279–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10435530.

Worrall, M. *et al.* (2020) 'IPEM Topical Report: An evidence and risk assessment based analysis of the efficacy of quality assurance tests on fluoroscopy units - Part II; Image quality', *Physics in Medicine and Biology*. IOP Publishing, 65(22), p. 195011. doi: 10.1088/1361-6560/abb92f.

Wu, T. *et al.* (2004) 'A comparison of reconstruction algorithms for breast tomosynthesis', *Medical Physics*, 31(9), p. 2636. doi: 10.1118/1.1786692.

Wu, X. *et al.* (1994) 'Normalized in Molybdenum and Rhodium', *Radiology*, 193, pp. 83–89.

Wu, X., Barnes, G. T. and Tucker, D. M. (1991) 'Spectral dependence of glandular tissue dose in screen-film mammography', *Radiology*, 179(1), pp. 143–148.

Yaffe, M. J. *et al.* (2006) 'Quality control for digital mammography: Part II recommendations from the ACRIN DMIST trial', *Medical Physics*, 33(3), p. 737. doi: 10.1118/1.2164067.

Yaffe, M. J. *et al.* (2009) 'The myth of the 50-50 breast', *Medical Physics*, 36(12), p. 5437. doi: 10.1118/1.3250863.

Accepted Manuscript

Yaffe, M. J. (2018) 'Towards improving accuracy, effectiveness, and efficiency in breast cancer screening', *The Lancet Oncology*. Elsevier Ltd, 19(11), pp. 1426–1427. doi: 10.1016/S1470-2045(18)30589-8.

Yaffe, M. J. and Mainprize, J. G. (2004) 'Detectors for digital mammography.', *Technology in cancer research & treatment*, 3(4), pp. 309–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19424702.

Young, K. *et al.* (2006) 'Comparison of software and human observers in reading images of the CDMAM test object to assess digital mammography systems', in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, pp. 614206-1–13. doi: 10.1117/12.653296.

Young, K. C. *et al.* (2006) 'Optimal beam quality selection in digital mammography.', *The British journal of radiology*, 79(948), pp. 981–90. doi: 10.1259/bjr/55334425.

Young, K. C., Ramsdale, M. L. and Rust, A. (1996) 'Dose and image quality in mammography with an automatic beam quality system', *British Journal of Radiology*, 69(822), pp. 555–562. doi: 10.1259/0007-1285-69-822-555.

Young, K. C., Wallis, M. G. and Ramsdale, M. L. (1994) 'Mammographic film density and detection of small breast cancers', *Clinical Radiology*, 49, pp. 461–465.

Zackrisson, S. *et al.* (2018) 'One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study', *The Lancet Oncology*, 19(11), pp. 1493–1503. doi: 10.1016/S1470-2045(18)30521-7.

Zeng, R. *et al.* (2020) 'Computational reader design and statistical performance evaluation of an in-silico imaging clinical trial comparing digital breast tomosynthesis with full-field digital mammography', *Journal of Medical Imaging*, 7(04), pp. 042802-1–18. doi: 10.1117/1.jmi.7.4.042802.

Zhang, G. *et al.* (2017) 'An adaptive toolkit for image quality evaluation in system performance test of digital breast tomosynthesis', in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. doi: 10.1117/12.2253871.

Zhao, B. *et al.* (2009) 'Experimental validation of a three-dimensional linear system model for breast tomosynthesis', *Medical Physics*, 36(1), p. 240. doi: 10.1118/1.3040178.

Zhao, B. and Zhao, W. (2008a) 'Imaging performance of an amorphous selenium digital mammography detector in a breast tomosynthesis system', *Medical Physics*, 35(5), p. 1978. doi: 10.1118/1.2903425.

Zhao, B. and Zhao, W. (2008b) 'Three-dimensional linear system analysis for breast tomosynthesis', *Medical Physics*, 35(12), p. 5219. doi: 10.1118/1.2996014.

Zhao, C. *et al.* (2015) 'Large area CMOS active pixel sensor x-ray imager for digital breast tomosynthesis: Analysis, modeling, and characterization', *Medical Physics*, 42(11), pp. 6294–6308. doi: 10.1118/1.4932368.

Zhao, C. *et al.* (2017) 'Three-dimensional cascaded system analysis of a 50 μ m pixel pitch wafer-scale CMOS active pixel sensor x-ray detector for digital breast tomosynthesis', *Physics in Medicine and Biology*. IOP Publishing, 62(5), pp. 1994–2017. doi: 10.1088/1361-6560/aa586c.

Zheng, J., Fessler, J. A. and Chan, H. P. (2019) 'Effect of source blur on digital breast tomosynthesis reconstruction', *Medical Physics*, 46(12), pp. 5572–5592. doi: 10.1002/mp.13801.

Zhou, J., Zhao, B. and Zhao, W. (2007) 'A computer simulation platform for the optimization of a breast tomosynthesis system', *Medical Physics*, 34(3), p. 1098. doi: 10.1118/1.2558160.

Zhou, Z. *et al.* (2011) 'Techniques to improve the accuracy of noise power spectrum measurements in digital x-ray imaging based on background trends removal', *Medical Physics*, 38(3), pp. 1600–1610. doi: 10.1118/1.3556566.

Accepted Manuscript