Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

ACCEPTED MANUSCRIPT

Performance evaluation of digital breast tomosynthesis systems: comparison of current virtual clinical trial methods

To cite this article before publication: Nicholas W Marshall *et al* 2022 *Phys. Med. Biol.* in press https://doi.org/10.1088/1361-6560/ac9a34

Performance evaluation of digital breast tomosynthesis systems: comparison of current virtual clinical trial methods

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

Virtual clinical trials (VCT) have been developed by a number of groups to study breast imaging applications, with the focus on digital breast tomosynthesis (DBT) imaging. In this review, the main components of these simulation platforms are compared, along with the validation steps, a number of practical applications and some of the limitations associated with this method. VCT platforms simulate, up to a certain level of detail, the main components of the imaging chain: the x-ray beam, system geometry including the antiscatter grid and the x-ray detector. In building VCT platforms, groups use a number of techniques, including x-ray spectrum modelling, Monte Carlo (MC) simulation for x-ray imaging and scatter estimation, ray tracing, breast phantom models and modelling of the detector. The incorporation of different anthropomorphic breast models is described, together with the lesions needed to simulate clinical studies and to study detection performance. A step by step comparison highlights the need for transparency when describing the simulation frameworks. Current simulation bottlenecks include resolution and memory constraints when generating high resolution breast phantoms, difficulties in accessing/applying relevant, vendor specific image processing and reconstruction methods, while the imaging tasks considered are generally detection tasks without search, evaluated by computational observers. A number of applications are described along with some future avenues for research.

1. Introduction

1.1 Background

In recent years, there have been significant advances in computer simulation applied to the evaluation of medical imaging systems in general (Frangi, Tsaftaris and Prince, 2018; Abadi *et al.*, 2020). The use of simulation to study medical imaging devices is a method that has become known as *in silico* imaging (Badano, 2011, 2021) or as a virtual clinical trial (VCT) (Abadi *et al.*, 2020; Barufaldi, Maidment, *et al.*, 2021). The scope can vary dramatically but these methods can be thought of lying somewhere between stages 1 and 2 on the scale expressing the efficacy of diagnostic imaging proposed by Fryback and Thornbury (Fryback and Thornbury, 1991; Barrett *et al.*, 2015), depending on degree of realism achieved in the simulations. We review recent progress in these methods, in the context of digital breast tomosynthesis (DBT) system performance evaluation.

In the companion paper, methods to perform a reasonably detailed, explicit characterization of DBT systems and the sub-components were described. This was followed by a description of current developments in the field of physical test objects used to evaluate DBT system technical image quality. These methods are practical and are designed to fulfil quality control (QC) and performance testing requirements for a particular imaging system at a specific clinical site (van Engen *et al.*, 2016). As we have seen, there are limitations to these methods, especially with regard to the task realism that can be achieved and the range of conditions that can be studied. The past two decades have seen rapid development in the ield of computational simulation of x-ray imaging systems. A range of applications have been developed, from use in supporting regulatory submissions, to focused technical studies (Hadjipanteli *et al.*, 2017), to a comparison of imaging modalities (Badano *et al.*, 2018; Barufaldi, Maidment, *et al.*, 2021), or in the development of technology within industry (Marchessoux, Kimpe and Bert, 2008). While applications vary considerably, VCTs have a common structure in which a virtual patient population is imaged using a virtual imaging system and the resulting images are interpreted by a virtual reader (Abadi *et al.*, 2020). The use of VCTs is a response to a number of difficulties associated with clinical trials including the complexity, expense, the length of time required to complete a trial and the lack of a ground truth (Frangi, Tsaftaris and Prince, 2018; Badano, 2021; Barufaldi, Maidment, *et al.*, 2021; Kopans, 2021). Further impetus comes from the increasing difficulty in obtaining approval for studies, with ethical committees and General Data Protection Regulation (GDPR) in the European Union being very strict. These methods obviously rely on the realism that can be achieved in the simulation; the degree of realism required for a VCT to accurately predict system or component performance in a real patient population is something of an open question, but will likely depend on the clinical question being addressed.

1.2 Selection of the VCT platforms for detailed study

The use of VCTs to evaluate x-ray breast imaging devices is an active field and a number of groups have made fundamental contributions to the conceptualization and application of VCTs in 2D digital mammography (DM) and DBT imaging. This review focuses on the various simulation platforms that have

been implemented, describes the physics models used for the x-ray components and the associated verification and validation steps, compares the breast and lesion models used and gives results from some of the studies that have been performed using these platforms. We have selected five simulation frameworks to examine in detail. These were chosen because detailed published documentation is available, a number of virtual studies have been performed and the platforms have been actively maintained and developed over a number of years. Data from the groups at the University of Pennsylvania (UPENN), the Federal Drug Administration (FDA), the OPTIMAM group at the University of Surrey/National Coordinating Centre for the Physics of Mammography (NCCPM), the group at KU Leuven and at the team at General Electric Healthcare (GEHC) are discussed. Detailed information of these five simulation platforms is given in table 1.

It must be noted that this is not an exclusive list, and a number of other platforms have been described in the literature. Some examples of breast imaging simulation platforms or studies not examined in depth are now briefly discussed. Petersson *et al* (Petersson *et al.*, 2016) used PENELOPE/PenEasy Imaging (Sempau, Badal and Brualla, 2011) to generate DBT projection images for a Siemens Inspiration system; realistic sharpness in the reconstructed planes could be achieved however levels of noise and object contrast evaluated in the projection images were higher than in the real images.

A group at Duke University used VCT methods to study the impact of breast structure on lesion detection in DBT (Kiarashi *et al.*, 2016). Twenty breast phantoms taken from the extended cardiac-torso (XCAT) family were generated and mass lesions derived from DBT data were embedded at 200 positions in the central plane of each breast phantom. Imaging characteristics of a Siemens MAMMOMAT Inspiration DBT system were applied and the resulting images evaluated using a composite hypothesis signal detection paradigm. Increasing breast density reduced detection performance. Background tissue density and heterogeneity also affected the test statistic differently under lesion absent and lesion present conditions, suggesting that background tissue variability must be considered and will influence VCT outcomes. While the group at Duke have made fundamental contributions to the development of virtual imaging methods (Abadi *et al.*, 2020), current work focuses on CT imaging (Abadi *et al.*, 2019).

An early simulation tool for radiographic imaging was described by Lazos *et al* (Lazos, Kolitsi and Pallikarakis, 2000; Lazos *et al.*, 2003) and this was later validated via simulations of a physical CIRS 011A phantom using synchrotron imaging (Bliznakova *et al.*, 2010). In later work, a software tool called BreastSimulator was developed (Bliznakova *et al.*, 2012) and validated by Mettivier *et al* (Mettivier *et al.*, 2017) for tomographic imaging, which led on to the Napoli-Varna-Davis project to perform virtual clinical trials in x-ray breast imaging (Mettivier *et al.*, 2019). The same group has since published a proof of concept for a simulation platform using the Geant4 Monte Carlo (MC) toolkit (di Franco *et al.*, 2020), referred to the Agata platform. This is still under development and has focused on breast dosimetry, while details such as x-ray photon interaction in the image receptor and lesion modelling are under development and not currently included in the simulation.

2. Components of the virtual chain

This section steps through the components of the virtual imaging chain. For each component, there is first a detailed comparison of the methods implemented in the five selected platforms. In some sections, where relevant, this is followed by a broader discussion of the related literature. A diagram illustrating the total simulation VCT platform developed by the OPTIMAM project is illustrated in Figure 1 (Elangovan *et al.*, 2014).

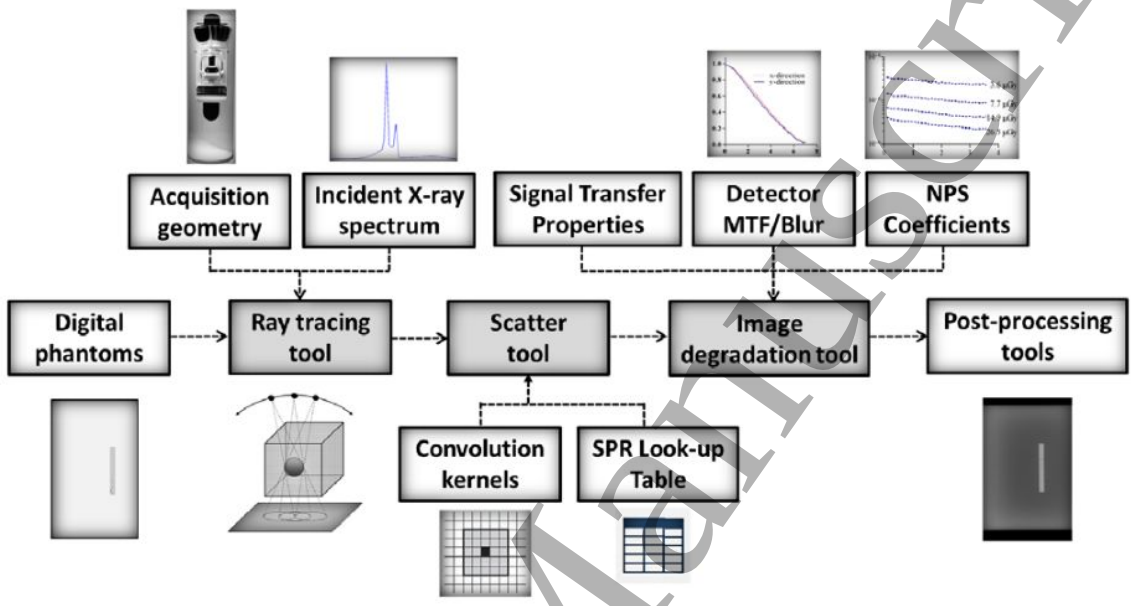


Figure 1. Overview of OPTIMAM platform used for total simulation of images. (From Elangovan *et al* (Elangovan *et al.*, 2014))

2.1 X-ray source

2.1.1 Spectral model

Table 1 lists the methods and data sources for the different platforms considered. System simulation begins at the x-ray source with the modelling of the energy and spatial distribution of the x-rays. With the exception of CatSim (Carvalho, 2014; Sánchez de la Rosa, 2019), most platforms use a poly-energetic source, generally implementing the spectral model developed by Boone *et al* (Boone, Fewell and Jennings, 1997). The simulation platform developed at UPENN (OpenVCT) utilizes a poly-energetic x-ray beam, although the spectral model is not identified.

The Boone model takes as input x-ray spectra measured at the Center for Devices and Radiological Health (CDRH) in the 1990s, to which interpolating cubic spline fits are applied. There have been a number of recent developments in spectral modelling, which may eventually lead to an update in the spectra used in simulation platforms. The maximum energy available in the data of Boone *et al* (Boone, Fewell and Jennings, 1997) is 42 kV, which will be a limitation at some point in the future, given the higher energies used in DBT, dual energy subtraction mammography and breast computed tomography (BCT). This prompted work by

Hernandez *et al* (Hernandez *et al*., 2017) in which the same cubic spline method was applied to spectra generated using the MC code MCNP6 for Mo, Rh and W anodes up to 49 kV and for use in DM and DBT. Data are also presented for W anodes up to 70 kV for the modelling of BCT spectra. In addition to this work, updated physics models for electron penetration in the target were introduced by Poludniowski (Poludniowski, 2007), and implemented in SpekCalc (Poludniowski *et al*., 2009). These models have been further developed in a series of papers by physicists in the Medical Radiation Physics and Nuclear Medicine group at Karolinska University Hospital in Sweden (Bujila, Omar and Poludniowski, 2020; Omar, Andreo and Poludniowski, 2020b, 2020a; Poludniowski *et al*., 2021). The validation study of Omar *et al* (Omar, Andreo and Poludniowski, 2020b) found improved agreement with MC simulations, especially for low tube voltage x-ray beams and should improve simulation framework accuracy, if incorporated. Figure 2 shows the agreement found between the modelled x-ray fluence compared to measurements made using a CdTe spectrometer for W/Rh, Mo/Rh and Mo/Mo anode/filter (A/F) combinations.

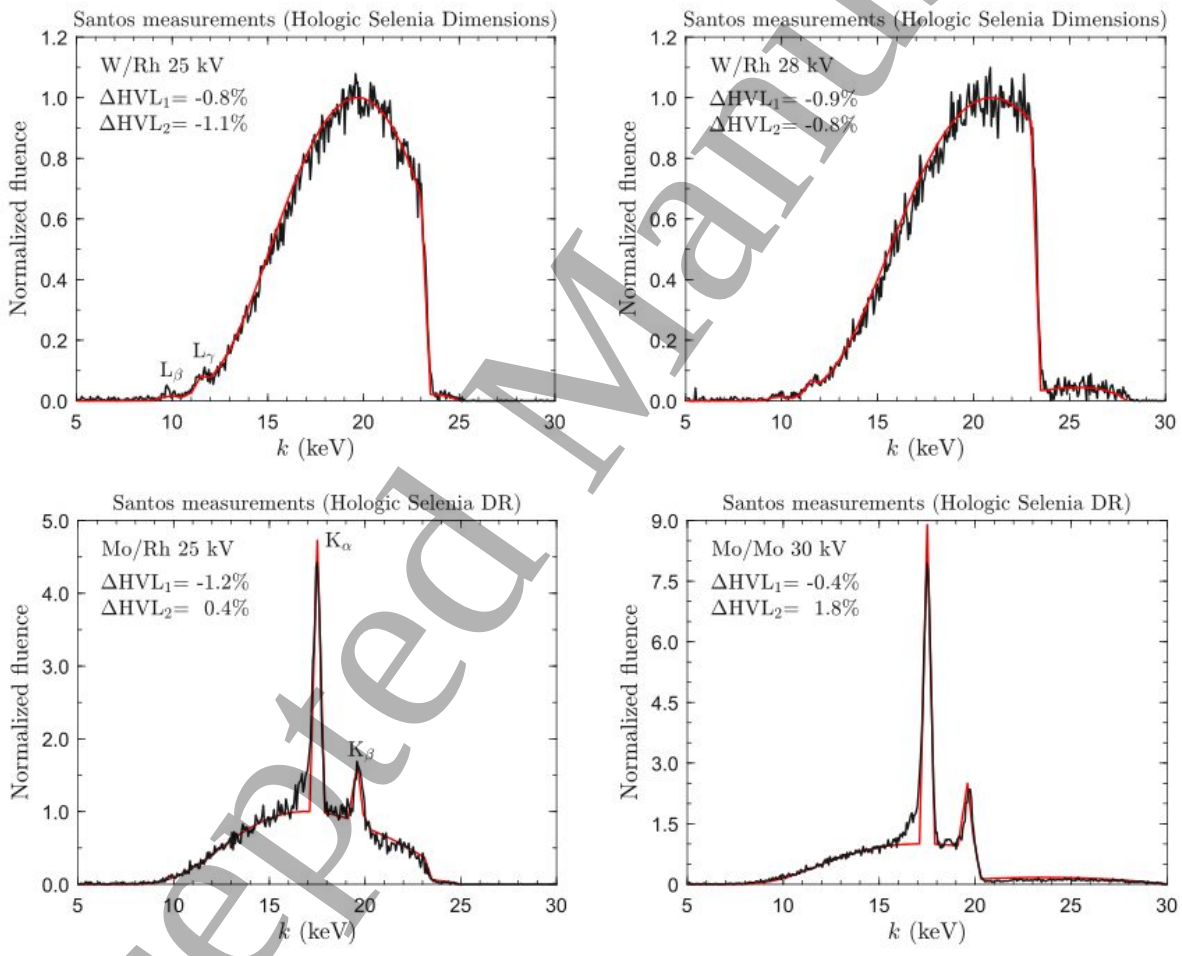


Figure 2. X-ray fluence differential in photon energy, k, calculated analytically (red lines) and measured by Santos et al. as described in Sections 2.C.2 and 2.C.3 (black lines). The results correspond to different target–filter combinations and tube potentials (for example, W/Rh 25 kV: tungsten/rhodium target–filter combination). The spectra are arbitrarily normalized to unity at their bremsstrahlung peaks. Also shown is the relative difference (%) in first and second aluminum half-value layer thickness (ΔHVL). (adapted from Omar *et al* (Omar, Andreo and Poludniowski, 2020b))

While including accurate poly-energetic source models will improve simulation realism, the use of a spectrum also increases computational complexity in the ray tracing part where an image is created for each energy bin. These mono-energetic images are then combined to give the total transmission probability for a given detector element for the spectrum considered. The approximation introduced by using a mono-energetic simulation will depend on the polychromaticity of the spectra; this has not been examined extensively in the simulation literature. Polychromatic simulations will include the effects of beam hardening and will more accurately model the x-ray energy distribution in glandular tissue as a function of depth within the phantom when estimating the mean glandular dose (MGD). While the use of poly-chromatic sources to model scattered will further increase computational complexity and workload, this has been accomplished using the MC-GPU package in VICTRE (Badal *et al.*, 2021) and in the work of Diaz *et al* (Diaz *et al.*, 2019). When using a simulation framework to investigate optimal anode-filter and spectral filtration for the source then factors such as the heat load rating of the x-ray tube, the maximum permissible cathode current and the specific radiation output of the tube must of be included (Fahrig, Rowlands and Yaffe, 1996; Shrestha, Vedantham and Karellas, 2017). These parameters will influence the exposure time per projection and the total time required to acquire all the projections, which in turn influence the geometric blurring present in the system and the degree of patient motion that may be present. The source model used in VICTRE (Virtual Imaging Clinical Trial for Regulatory Evaluation) (Badal *et al.*, 2021) includes the specific radiation output of the x–ray tube but all platforms do not explicitly model tube current or anode heat loads at the moment.

2.1.2 Spatial distribution of the x-ray source

A second important aspect of source modelling is to include the influence of the spatial distribution of the focus during x-ray emission. Regarding the physical size and shape of the focus, there are some variations, as both Gaussian (Badal *et al.*, 2021) and square/rectangular (Shaheen *et al.*, 2010; Elangovan *et al.*, 2014) approximations are used. The shape or size is not specified in the OpenVCT platform (Barufaldi, Bakic and Maidment, 2019) while the study into microcalcification detection using CatSim (Li *et al.*, 2018) used a point source, although an extended source can be used (Carvalho, 2014). Focus size simulated ranges from a point source, through 0.30 mm (Badal *et al.*, 2021) up to maximum of 0.33 × 0.54 mm² (Elangovan *et al.*, 2014). Focal spot data measured by Marshall and Bosmans (Marshall and Bosmans, 2012) found approximately square foci of size 0.43×0.53 mm² and 0.40×0.43 mm², for the Hologic Dimensions and Siemens Inspiration DBT systems, measured in DM mode.

For continuous motion DBT systems, the extended source due to motion during exposure may have to be modelled, and there are two approaches to this. First, during ray tracing, multiple rays can be cast from within the extended source region, whose size is calculated from the tube angular velocity and exposure time for a projection image (Michielsen *et al.*, 2013). Both the VICTRE and OPTIMAM platforms use this method (Elangovan *et al.*, 2014; Badal *et al.*, 2021), sampling randomly from within the source area. The magnitude of blurring experienced by structures in the breast model varies as a function of height above the

table and this method applies the correct blurring during the ray tracing stage. An alternative method can be used in partial simulation frameworks, where a small template containing the lesion is adapted to the sharpness and contrast properties of the imaging system. Here, a modulation transfer function (MTF) that characterizes the system blurring is estimated for height above the table at which the lesion is to be inserted into the voxel model or real breast image (Carton *et al.*, 2003; Shaheen *et al.*, 2011). This is the presampling system MTF that contains geometric blurring due to source size and motion, detector converter and pixel aperture. The corrected template is then multiplied into the image, typically followed by the reconstruction step and the application of clinical image processing. Use of a Fourier method assumes stationarity within the imaged volume; this is a reasonable approximation in the x-y direction (although will be limited due to oblique x-ray entry) but does not hold in the vertical (z) direction for systems with large sources of geometric blurring (Zheng, Fessler and Chan, 2019). The extended focus size in the scan direction used in the OPTIMAM simulation was 1.4 mm for the Hologic-like system (angular range ±7.5°) and 2.2 mm for the wider angle Siemens-like system (angular range ±25°). For comparison, values of 0.80 mm and ~2.2 mm for the focus motion size were found for the Hologic and Siemens systems.

2.1.3 Angular range

Regarding the angular range and number of projections of the systems simulated using these platforms, OpenVCT and OPTIMAM have implemented both typical narrow angle systems i.e. Hologic-like (±7.5°) and wide angle i.e. Siemens-like (±25°) (Elangovan *et al.*, 2014; Barufaldi, Bakic and Maidment, 2019; Hadjipanteli *et al.*, 2019). Note that the OpenVCT platform used just 15 projections to cover the ±25° range (3.3° per projection), while the Siemens Revelation uses 25 projections (2° per projection). This reduced angular sampling has the potential to introduce artefacts. The current VICTRE and Leuven platforms only simulate the Siemens device (Shaheen, Marshall and Bosmans, 2011; Vancoillie *et al.*, 2020; Badal *et al.*, 2021).

Table 1 Five platforms used for DM and DBT image simulation and VCTs (see main text for definition of acronyms). NS = not specified

Group	University of Pennsylvania	FDA	OPTIMAM	KU Leuven	GEHC
Platform Name	OpenVCT	VICTRE	OPTIMAM	Leuven Platform	CatSim
Simulation type	Total	Total	Total and partial	Partial	Total
Development language	GPU/XML/SQL/Java	C, NVIDIA/CUDA, python	NS	MATLAB	MATLAB, C++
DM/DBT systems simulated	Selenia-like, Inspiration-like	Inspiration-like	Hologic-like; Siemens-like	Siemens Inspiration/Revelation	GEHC Pristina
X-ray source					
- Model/spectra used	NS	(Boone, Fewell and Jennings, 1997)	(Boone, Fewell and Jennings, 1997)	(Boone, Fewell and Jennings, 1997)	GE SpeXim (Birch and Marshall, 1979; Cranley *et al.*, 1997) and (Boone, Fewell and Jennings, 1997)
- Mono/poly-chromatic?	Poly-energetic	Poly-energetic	Poly-energetic	Poly-energetic	Mono-energetic
- Anode material	W	W	W	W	Mo, Rh, and W
- Filter materials/thickness	0.05 mm Rh (DM); 0.7 mm Al (DBT)	1.0 mm Be window, 0.05 mm Rh (DM and DBT)	0.05 mm Rh (DM); 0.70 mm Al (DBT)	0.05 mm Rh (DM and DBT)	Mo, Rh, Be, Cu, Sn; 0.03 mm Rh and 0.03 mm Ag (DM and DBT)
- Focus size	NS	0.30 mm	0.33×0.54 mm² (DM), 0.46×0.37 mm² (DBT) (Elangovan *et al.*, 2014) 0.4×0.4 mm² (DM); 0.4×1.4 mm² (±7.5° DBT); 0.4×2.2 mm² (±25° DBT) (Hadjipanteli *et al.*, 2019)	0.45×0.45 mm² (DM) 0.4×2.05 mm² (DBT), corresponding to 90 ms exposure time	Point source
- Focus shape	NS	Gaussian	Square or rectangular	Square (sinc fn in Fourier plane)	N/A
- Focus motion/sampling	Continuous; step and shoot; details not given	Continuous; 0.18° arc, corresponding to 90 ms exposure time	Continuous; random sampling within the focal spot region	Continuous; Fourier based, applied via MTF to object (not background)	Step and shoot
- Angular range/number of projections	±7.5°, ±15°, ±25°/ 15 projections	±25°/ 25 projections	±7.5°/15 or ±25°/25 projections	±25°/25 projections	±12.5°/9 projections
- Source-to-image-distance	700 mm and 652 mm	652 mm	NS	NS	660 mm
AEC					
AEC function/factors	Measured AEC performance (Feng and Sechopoulos, 2012)	28 kV, 30 kV 85.1 mAs for 5.5-cm scattered glandularity breast model, 3.4×10^{11} histories simulated;	31 kV W/Rh (0.55 mm HVL) for DM 33 kV W/Al (0.59 mm HVL) for DBT (Hadjipanteli *et al.*, 2019)	AEC factors taken from image DICOM header	Taken from AOP AEC data

		50% more x-rays used for DBT acquisition (1.5 mGy) vs DM (1.0 mGy)			
Exposure time	NS	90 ms	NS	90 ms or as specified in DICOM header	NS
Ray tracing					
Algorithm	(Siddon, 1985; De Greef *et al.*, 2009)	MC-GPU / Penelope	(Siddon, 1985)	(Siddon, 1985)	GE tracing algorithm
Attenuation coefficient database	ICRU Report 44 (ICRU, 1989)	Rayleigh (Cullen, Hubbell and Kissel, 1997) Total cross section (Berger *et al.*, 2005)	(Berger *et al.*, 1998)	(Berger *et al.*, 2005)	NS
Breast Phantom					
- Model	UPENN Bakic 2002 plus later modifications (Bakic *et al.*, 2002; Pokrajac, Maidment and Bakic, 2012)	(Graff, 2016)	OPTIMAM (Elangovan *et al.*, 2017)	None implemented; Partial simulation into real patient images	Stochastic Solid Breast Texture (SSBT) model (Li *et al.*, 2016) using BCT dataset at UC Davis (Lindfors *et al.*, 2008); a voxelized structure
- Voxel size	100 μm³ ; 200 μm³	50 μm³	100 μm³		analytical, can generate voxel models at e.g. 100 μm³
- Modelled size	Full breast, 6.33 cm compressed thickness	Full breast; 3.5, 4.5, 5.5, and 6 cm compressed thickness	Full breast cropped into 30×30×30 mm³ VOIs; 6 cm compressed thickness	Thickness determined by patient selection	e.g. 50×50×50 mm³ VOIs extracted. 5 cm compressed breast thickness
- View	CC; MLO	CC; MLO	N/A		N/A
- Breast types	Adjustable volumetric density, 10% and 30% studied	Variable density: dense (0.548 glandular fraction), heterogeneously dense (0.339), scattered (0.143), and fatty (0.071)	Volumetric glandular density 17% to 19%	Real breast, determined by patient selection	Variable density e.g. average glandular density of 27%, (BI-RADS 2), average glandular density of 55% (BI-RADS 3) (Li *et al.*, 2018)
Scattered radiation					
- Included?/method	No/-	Full scatter distribution / MC-GPU + Penelope	Yes / scatter kernels to generate look up tables (Diaz *et al.*, 2014). Geant4 MC(Agostinelli *et al.*, 2003) to generate the scatter kernels	Yes / Contrast of lesion template is adjusted by scatter fraction (SF). SF data from (Boone *et al.*, 2000; Salvagnini *et al.*, 2012)	Hybrid analytic - MC method with MC engine developed by GEHC. Scatter kernels calculated from ESF measurements (Sánchez de la Rosa, 2019)
- Grid for DM?/method	No/-	Yes / analytic (Day and Dance, 1983) 31 lp/cm ratio 5:1; 65 μm Pb strips with polystyrene interspace	NS/NS	Yes / SF data measured for Hologic and Siemens grids (Salvagnini *et al.*, 2012)	Yes / Scatter kernel estimated for grid acquisitions (Sánchez de la Rosa, 2019)
X-ray detector					

- Material	a-Se	a-Se	a-Se	s-Se	CsI
- Thickness	-	200 μm	NS	NS	NS
- Pixel matrix	2048×1664 and 3584×2816	3000×1500	NS	2816×3584	1000×1000
- Detector element spacing	0.140 mm; 0.085 mm	0.085 mm	0.070 mm DM; 0.140 mm for both ±7.5° and ±25° DBT	0.085 mm	0.100 mm
- Oblique entry modelled?/method	No/-	Yes	No	Yes (only for template) (Que and Rowlands, 1995)	No
- Detector physics	Perfect absorber	MC-GPU / Penelope. Absorbed γ energy converted to charge. Effective gain of 50 keV/ehp, σ given by Swank factor 0.99, Poisson additive electronic noise of mean 5200 e- added to each pixel	Detector response fn, MTF and NPS characterized for the imaging system. NPS of electronic, quantum and structure noise sources included. Applied to the projection images after ray tracing in the Fourier domain.	Lesion signal is energy absorbed in detector material; MTF applied to template includes focus size, focus motion blurring and detector presampling MTF	Energy integrating detector; ray tracing to generate noise-free, blur-free images. Detector converter and photodiode (pixel) blurring applied in Fourier domain. Gaussian additive noise added to simulate electronic noise
- Temporal effects included? (e.g. lag, ghosting)	No	No	No	No	No (yes for contrast enhanced digital mammography CEDM))
Breast Lesions					
- Calcification model	Voxel polycubes; calcification clusters	Shape not defined	One calcification extracted from database of (Shaheen *et al.*, 2011) replicated/rotated 5 times to make clusters	Real calcification clusters imaged using micro-CT (Shaheen *et al.*, 2011)	Ellipsoids (max deviation of 5% of calc diameter allowed for each axis), Perlin noise used to create 20 μm irregularities on calcification surface (Li *et al.*, 2018)
- Calcification composition /attenuation	Calcium hydroxyapatite $Ca_5(PO_4)_3$, weighting factor applied to control contrast	Calcium oxalate (CaC_2O_4) with mass density scaled by 0.84 (1.78 g/cm²)	Calcium oxalate (CaC_2O_4) with mass density scaled by 0.84 (1.78 g/cm²)	Calcium oxalate (CaC_2O_4)	Al; μ_{Al} at 22 keV scaled by 20%, 40% and 60%, 2.73 kg/m³ mass density
- Calcification size	100 μm³ polycubes, in groups of 1 to 4; Individual calcifications are ~spherical with diameter between 300 μm and 600 μm	171, 179 and 195 μm, grouped as 5 calcifications in 5 mm² volume	Cluster with 5 calcifications in 2.5 mm³, microcalcification diameters in total range 110 μm to 275 μm, ~50 μm range in diameter within group eg 110 - 155 μm	Not defined	Average diameters 100μm, 200μm, 400μm and 600μm
- Supersampling for calcification insertion ?	No	NS	Yes	No	Yes
- Masses shape / size	Oblate spheroids; / 7 mm diameter and 0.5 – 2 mm thick or 9 mm diameter sphere. Concentric rings used	Spiculated mass (de Sisternes *et al.*, 2015) / 5 mm diameter	Mass with irregular border (A Rashidnasab *et al.*, 2013) / diameter 4.7 mm to 10.3 mm at 30 mm above the table (Hadjipanteli *et al.*, 2019)	Irregular (non-spiculated) mass lesions segmented from Magnetic Resonance Imaging (MRI); spiculated mass lesions following method of	Irregular and spiculated mass lesions (Carvalho, 2014), generated as analytic structure /

	to control blending with background			(Shaheen *et al.*, 2014; de Sisternes *et al.*, 2015) / 6.2 mm to 8.3 mm diameter	
- Mass composition/attenuation	$\mu_{mass} = 1.04 \times \mu_{gland}$	mass density 2% higher than normal glandular tissue $\mu_{mass} = 1.02 \times \mu_{gland}$	$\mu_{mass} = \mu_{gland}$	$\mu_{mass} = \mu_{gland}$	NS
Reconstruction					
- Algorithm	FBP (Briona (commercially available))	FBP (Fessler, 2018; Sengupta *et al.*, 2018)	FBP (Briona (commercially available))	Vendor supplied (currently Siemens EMPIRE)	GEHC
- In-plane pixel spacing	0.100 mm and 0.085 mm	0.085 mm	NS	0.085 mm	0.100 mm
- Plane spacing	1 mm	1 mm	NS	1 mm	0.5 mm, 1 mm and 10 mm
Display					
- Included?/method	Yes/MeVIC (Marchessoux, Kimpe and Bert, 2008)	No	No	No	No
Image interpretation					
- Study type	MRMC ROC	MRMC ROC	4-AFC	ROC, FROC	ROC
- Output metric	AUC, *d'*	AUC	Percentage correct (PC); threshold diameter (mm); detectability index d'	AUC	AUC
- Human/computer?	Computer	Computer	Human	Computer and Human	Computer and Human
- Computer algorithm	CHO with 15 LG channels, Gaussian spread=15, 20, 22, 25, 26, 31; 150 x 150 pixel region of interest (ROIs); 252 to 500 training set image pairs	CHO with 5 Convolutional LG channels; Masses: width 30 pixels for DM, 25 for DBT; Calcs: Gaussian spread = 1.5 pixels; 3D CHO for DBT; ROIs varied in size from 65 x 65 to 109 x 109 pixels; 100 training set image pairs			CHO, single slice applied to in-focus DBT plane, multi-slice CHO also applied to 5 DBT planes; Gaussian spread ranged from 100μm to 800μm, 1 to 29 channels considered. 200 image pairs used for training. (Li *et al.*, 2018)
Validation					
- Reference	NS	(Badal *et al.*, 2021)	(Elangovan *et al.*, 2014)	(Vancoillie *et al.*, 2020)	(Carvalho, 2014)
Miscellaneous					
- Time to generate DM + DBT	121.5 s for 200 μm³ phantom; 520.4 s for 100 μm³ phantom	~55 min to 450 min, depending on breast size and complexity	~15 min for a breast phantom which then generates many VOIs	90s for DM without image processing 220 s for DBT without reconstruction	NS
- Website for platform	https://sourceforge.net/projects/openvct/	https://github.com/DIDSR/VICTRE	Not publically available	Not publically available	Not publically available

2.2 Ray-tracing / x-ray transport

The next aspect considered is transport of x-rays from the source through the breast model and here there is a clear difference between the VICTRE platform (Badal *et al.*, 2021) and the other platforms. VICTRE uses MC-GPU MC simulation code to transport x-ray photons through a voxelized model, where each x-ray photon is tracked until either absorbed in the x-ray detector or it leaves the MC world volume. Each photon can either be scattered or undergo no scattering and be directly absorbed in the x-ray detector (i.e. primary), while in the other platforms the primary and scattering steps are treated in two separate stages (see Figure 1 for the OPTIMAM example). After adjusting for intrinsic x-ray source efficiency, the VICTRE approach transports a similar number of x-ray histories as used in a real system and therefore directly generates accurate quantum image noise (assuming accurate detector modelling), scattered x-ray spatial distribution, blurring from the extended focus, oblique entry in the x-ray detector and the relevant glandular tissue dose. This is achieved using the PENELOPE MC transport code (Salvat, Fernandez-Varea and Sempau, 2006).

The OpenVCT, OPTIMAM and Leuven platforms use the Siddon ray tracing algorithm (Siddon, 1985) or a related method (De Greef *et al.*, 2009), while the ray tracing algorithm is not specified for CatSim. The attenuation path length through the breast model is calculated in this step and combined with exponential attenuation to give the probability of x-ray transmission for a given pixel. This attenuation map must then be combined with the x-ray scatter signal to give the total signal at each detector pixel, although this step is not always clearly elaborated in the platforms. Attenuation data generally come from the NIST XCOM Photon Cross Section Database (Berger *et al.*, 1998, 2005) for the total cross-section, while the Rayleigh cross sections used in PENELOPE come from Cullen *et al* (Cullen, Hubbell and Kissel, 1997). Attenuation coefficients in OpenVCT are taken from ICRU Report 44 (ICRU, 1989; Barufaldi *et al.*, 2022), while the data source is not specified in CatSim.

2.3 Scattered radiation

There are even larger differences between the platforms in the methods used to include scattered radiation within the simulations. Accurate modelling of the spatial distribution and magnitude of scattered photons is important, especially if studies are performed comparing DBT against DM imaging performance, as the antiscatter grid influences the contrast of structures within the breast and the level of quantum noise (Chen *et al.*, 2015). At one extreme, the full MC-GPU simulation in VICTRE generates a scattered radiation distribution specific to a given breast model, while the OpenVCT platform does not simulate scattered radiation. For DM acquisitions in VICTRE, the exact grid composition was unknown and therefore a 1D focused grid was assumed with 65 μm Pb lamellae, 31 lines/cm, with a polystyrene interspace material and an aspect ratio of 5:1 (Badal *et al.*, 2021). Transport of x-ray photons within the grid is handled with an analytic model (Day and Dance, 1983). The OPTIMAM simulation uses a scatter kernel method that can estimate scatter radiation to within 10% across most of the breast area within ~1.5 h of computation per condition (Diaz *et al.*, 2014). Note that this is still a significant computational burden if trying to produce a large number of patient images for a VCT, leading to a time of ~37.5 h to generate scatter kernels for a

Siemens system with 25 projections. Scatter kernels for DBT (generally a grid out geometry) are described by Diaz *et al* (Diaz *et al.*, 2014) but details on scatter simulations for DM with grid-in are not given. Diaz *et al* (Diaz *et al.*, 2019) describe a fast method to estimate scatter distributions for DBT geometries for use in VCTs using normalized scatter maps generated from pre-calculated MC simulations of low resolution homogenous phantoms.

A scatter kernel method is also implemented in the CatSim platform for mammography imaging (Sánchez de la Rosa, 2019). Instead of using MC simulations to produce the scatter kernels, Sánchez de la Rosa (Sánchez de la Rosa, 2019) used an analytical model of the scatter point spread function (PSF) (Ducote and Molloi, 2010; Leon, Brateman and Wagner, 2014). Parameter values in the scatter PSF specific to the imaging system and object thickness were determined using an edge spread function (ESF) acquired with a radio-opaque edge positioned on top of poly(methyl)methacrylate (PMMA) plates (Chan and Doi, 1983; Cooper *et al.*, 2000). The scatter-to-primary ratio (SPR) and MTF derived from the ESF were used to generate a thickness dependent scatter kernel. The method was applied to grid-in and grid–out acquisition geometries. Application of the kernel to a DM image of 1000×1000 took approximately 3.4 minutes. The Leuven platform is somewhat different as only a small template containing the lesion (microcalcification cluster or mass) is multiplied into the projection image of a real breast, where scattered radiation is already present. The template contrast has to be modified using a scatter fraction (SF) relevant to the breast thickness, composition and energy; SF data measured for PMMA blocks using the beam stop and MTF-based methods (Salvagnini *et al.*, 2012) for grid in and grid-out geometries are used for this. Some limitations hold in that lesions are simulated towards the central region of the breast, where the SF values are assumed to be approximately constant (Sechopoulos *et al.*, 2007) and breast thickness is converted to PMMA thickness for the SF calculation using the relation given by Dance *et al* (D R Dance, Skinner, *et al.*, 2000).

2.4 Breast phantoms

In general, the group developing the framework has also implemented a breast phantom model for use in the framework (see Table 1). More detail on these models can be found in reviews by Bliznakova (Bliznakova, 2020) and by Glick and Ikejimba (Glick and Ikejimba, 2018). Here, we will briefly look at some of the practical points regarding implementation within a VCT framework.

2.4.1 Bakic phantom

OpenVCT uses the UPENN phantom developed from the Bakic phantom (Bakic *et al.*, 2002), which is voxelized at 100 μm^3 or 200 μm^3 and renders cranio-caudal (CC) or medio-lateral oblique (MLO) views of the full breast. In studies, the group has simulated breasts of volume 700 ml and compressed thickness 6.33 cm, with dense compartments constituting 10% and 30% of the breast volume (Bakic, Barufaldi, Higginbotham, *et al.*, 2018). In recent work, multiscale Perlin noise has been included in the phantom as a means of improving the anatomical realism of this phantom (Barufaldi, Abbey, *et al.*, 2021). This type of noise was developed by Perlin (Perlin, 1985) as a means of generating textures to increase realism in computer graphic scenes. To generate the noise, a grid is defined at some spacing and populated with

random, unit length vectors. The random vector set and the spacing defines the basic pattern and scale of the noise. Although the grid can be n-dimensional, 2D or 3D grids have been used for phantom applications (Dustler *et al.*, 2015; Barufaldi, Abbey, *et al.*, 2021). An interpolating function is applied at random points over the grid to produce smoothly varying structures with values in the range -1 to 1. The textures produced have a characteristic size, which can be tuned to the requirements of the scene or object by changing the values of the coefficients (Barufaldi, Abbey, *et al.*, 2021).

2.4.2 Graff phantom

VICTRE implements the open source phantom developed by Graff (Graff, 2016), which also produces full CC and MLO views of variable volumetric density, with a 50 μm³ voxel size. For the in-silico trial reported by Badano *et al* (Badano *et al.*, 2018), four density categories were studied, defined by the glandular volume fractions (GVF): extremely dense (0.548 GVF), heterogeneously dense (0.339 GVF), scattered (0.143 GVF), and fatty (0.071 GVF). These correspond to compressed thicknesses of 3.5, 4.5, 5.5, and 6 cm.

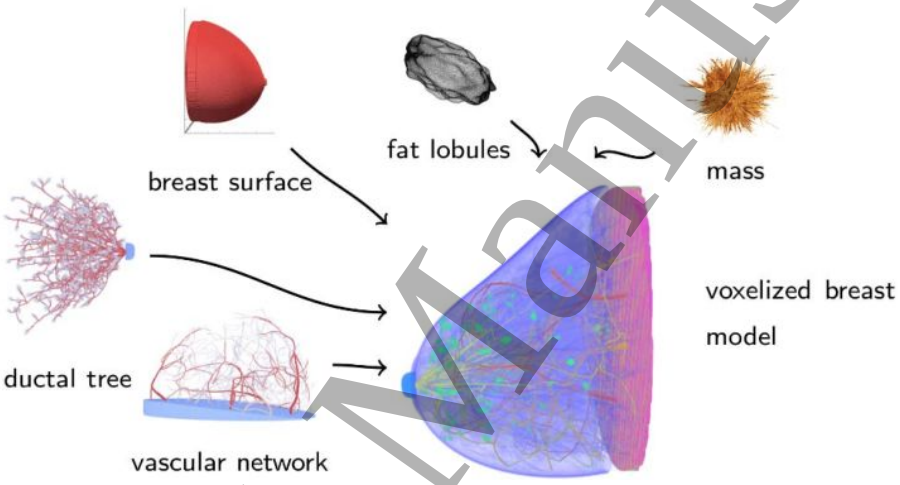


Figure 3. Anatomical components of virtual breast model (from Badano *et al* (Badano *et al.*, 2018))

2.4.3 Alternative phantom methods

The OPTIMAM phantom (Elangovan *et al.*, 2017) is a full breast generated at 100 μm³ resolution and then cropped into 30×30×30 mm³ volumes of interest (VOI) for use in reader studies. Volumetric density is set to values between 17% and 19%. For the CatSim platform, Li *et al* (Li *et al.*, 2016) described a solid 3D breast texture model, based on segmented UC Davis BCT datasets (Lindfors *et al.*, 2008) in which a stochastic geometry was used to mathematically model small and medium scale fibro-glandular and adipose tissue shapes. This analytical breast model was voxelized at 100 μm³ resolution and imaged in a study using the CatSim framework (Li *et al.*, 2018). Similar to the OPTIMAM studies, 50×50×50 mm³ VOIs were extracted from the generated volumes for the reader study. The model has variable density; Li *et al* (Li *et al.*, 2018) used average glandular densities of 27% and 55% to simulate Breast Imaging Reporting and Data System (BI-RADS) categories 2 and 3, respectively.

The UC Davis BCT dataset (Lindfors *et al.*, 2008) has been the starting point for a number of breast voxel models. Erickson *et al* (Erickson *et al.*, 2016) produced 224 virtual phantoms from this dataset. Mean volumetric breast density was 25.3% ± 13.2% and mean breast volume was 716.3 ml ± 386.5 ml. These data were used as input for the development of physical phantoms (Rossman *et al.*, 2019) and were used by (Kiarashi *et al.*, 2016) in a VCT to study the influence of breast tissue density and heterogeneity on the detection of irregular masses in DBT.

A further set of 150 3D breast phantoms have been derived from the UC Davis BCT data within the INFN AGATA project and are available in a public repository for research purposes (Sarno *et al.*, 2021). In-plane pixel sizes in the original reconstructed coronal slices range between 0.1938 mm and 0.4274 mm, and between 0.1907 mm and 0.2375 mm in the axial direction. A subset of 60 phantoms were compressed to simulate DM and DBT geometry. The resulting average glandular fraction for this subset was 12.1% (minimum of 1.2%, maximum of 28.7%). Compressed breast thickness varied between 35 mm and 89 mm, with a mean value of 61 mm.

2.5 Breast lesions

One of the advantages of a VCT methodology is that lesion location is known exactly in the signal present images yielding a ground truth, which allows a paired comparison of modalities using identical targets (Abadi *et al.*, 2020). Current VCTs focus on detection studies using microcalcifications and mass-lesions, yet there are considerable differences in the lesion models used across the frameworks.

2.5.1 Microcalcifications

OpenVCT currently uses two methods for calcification cluster modelling. A geometric method, in which between one and four polycubes are used, each with a side of 100 μm (i.e. the resolution of the voxel phantom) (Bakic, Barufaldi, Pokrajac, Lago, *et al.*, 2018). Alternatively in OpenVCT, clusters can be simulated as a collection of single microcalcifications with selected compositions, locations and sizes varying between 300 μm and 600 μm (Bakic, Barufaldi, Pokrajac, Weinstein, *et al.*, 2018). These individual microcalcifications are approximately spherical in shape. The microcalcifications in the studies using the VICTRE pipeline were simulated as a cluster of 5 lesions of diameter 171, 179 or 195 μm randomly located within a 5 mm^3 volume (Badano *et al.*, 2018). The precise shape of an individual calcification is not specified. The Leuven platform utilizes the microcalcification clusters segmented from breast biopsies containing malignant and benign lesions, acquired on a micro-CT system (Shaheen *et al.*, 2011). Figure 4 shows a photograph of the biopsy specimen after extraction from the patient, together with an example of a full 3D model of microcalcification cluster. A total of 23 clusters are available in the dataset, and while calcification diameter differs within a cluster, detailed information on individual lesion diameters is not given. The study into calcification detection using the OPTIMAM platform utilized a single calcification extracted from the database of Shaheen *et al* (Shaheen *et al.*, 2011) that was replicated and rotated 5 times to generate a cluster. These clusters were situated in a 2.5 mm^3 volume, and resampled to produce clusters with individual calcifications ranging in diameter from 110 μm to 275 μm (Hadjipanteli *et al.*, 2017). The CatSim based

study of (Li *et al.*, 2018) simulated calcifications as ellipsoids with a maximum deviation of 5% of lesion diameter for each axis; average diameters of the simulated lesions was 100μm, 200μm, 400μm and 600μm. Perlin noise was used to create 20 μm irregularities on the surface of the ellipsoid. Figure 5 shows examples of microcalcification lesions for the different platforms.

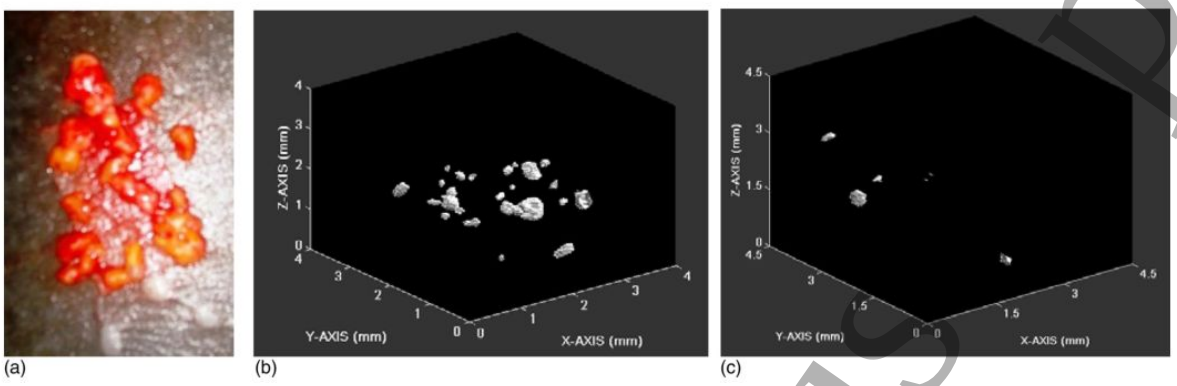


Figure 4. (a) A photograph of a biopsy specimen containing microcalcification clusters after extraction from a patient. (b) An example of a 3D model of microcalcification cluster from the database built during this study and considered as full cluster (voxel size was 0.02 mm). (c) An example of a 3D model of microcalcification cluster from the database and considered as subcluster (voxel size was 0.03 mm). (from Shaheen *et al* (Shaheen *et al.*, 2011)).

The chemical composition assumed for the calcification plays an important role in determining contrast and ultimately the visibility of the calculation in the simulated images. Calcification composition varies across the platforms, often with a weighting factor to control the visibility. OpenVCT has used calcium hydroxyapatite with a weighting factor (0.2 to 1.0) to control the contrast (Bakic, Barufaldi, Higginbotham, *et al.*, 2018). Current VICTRE simulations (Badano *et al.*, 2018) have adopted the method used in the OPTIMAM platform (Warren *et al.*, 2013; Hadjipanteli *et al.*, 2017), using calcium oxalate with a mass density weighted by a factor 0.84 (i.e. 1.78 g/cm^2). The Leuven group currently uses calcium oxalate (unweighted) while Li *et al* (Li *et al.*, 2018) used Al to simulate calcification, with mass density scaled by 20%, 40% and 60%.

2.5.2 Mass lesions

Diverse methods have been used to simulate the mass lesions used in these platforms, as illustrated in Figure 6. Geometric shapes such as oblate spheroids with diameter 7 mm and thickness ranging from 0.5 mm to 2 mm to change the contrast have been used in OpenVCT (Bakic, Barufaldi, Higginbotham, *et al.*, 2018). Concentric shells composed of different simulated materials are used to form the lesion, with different attenuation properties for each shell, enabling lesion blending with the local background (Bakic, Barufaldi, Pokrajac, Weinstein, *et al.*, 2018). Non-spheroidal lesions can be formed from a union of four non- concentric ellipsoids (Bakic, Barufaldi, Pokrajac, Weinstein, *et al.*, 2018). The large study performed using the VICTRE platform used a 5 mm spiculated mass lesion generated using the method of de Sisternes *et al* (de Sisternes *et al.*, 2015) in which an iterative fractal branching algorithm is used to add spicule structures to a central lesion produced with a modified Gaussian random sphere model. Masses with ill-

defined borders are currently used in the OPTIMAM platform, generated stochastically using a method known as diffusion limited aggregation (A Rashidnasab *et al.*, 2013). The study into threshold diameter for masses varied mass diameter from 4.7 mm to 10.3 mm (Hadjipanteli *et al.*, 2019). The irregular (non-spiculated) masses used in the Leuven platform originate from segmentations of magnetic resonance imaging (MRI) masses; the method of de Sisternes *et al* (de Sisternes *et al.*, 2015) is then applied to these irregular lesions to form spiculated masses (Shaheen *et al.*, 2014). Although CatSim has not been explicitly used to simulate mass lesions in DBT, Sánchez de la Rosa (Sánchez de la Rosa, 2019) describes a method to generate 3D lesions for contrast enhanced digital mammography (CEDM), using the aggregation of several analytical surfaces, each associated with specific linear attenuation coefficient, with methods that allow control over the shape, margin and contrast uptake distribution. Attenuation coefficients for mass-lesions are assumed to be identical to that of glandular tissue for the OPTIMAM and Leuven platforms, while factors of 1.04 and 1.02 are applied to glandular attenuation to simulate lesion attenuation for the OpenVCT and VICTRE platforms, respectively.

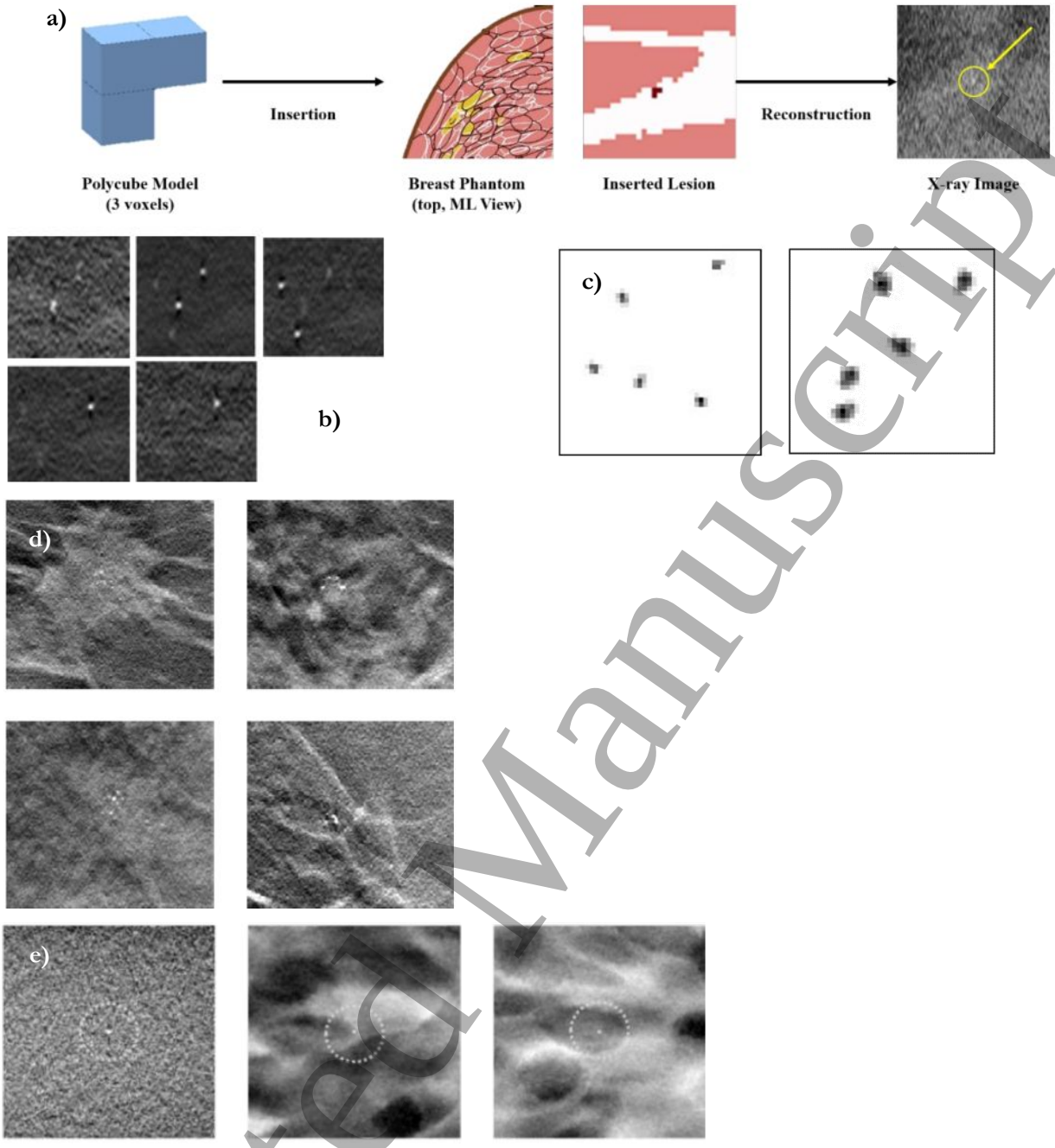


Figure 5. Examples of microcalcifications used in the simulation platforms.

a) Polycubes used to simulate microcalcifications in OpenVCT, also showing the insertion and reconstruction steps (Bakic, Barufaldi, Higginbotham, *et al.*, 2018) b) 5 DBT planes showing a reconstructed microcalcification cluster in VICTRE platform (Badano *et al.*, 2018) c) 2D projection images of (2.5 × 2.5 × 2.5 mm^3 cubic volume) of clusters in OPTIMAM platform two different microcalcification diameters before insertion: 125 μm and 250 μm. (Hadjipanteli *et al.*, 2017) d) False positive cases (i.e. simulated microcalcification clusters thought to be real cases by at least three radiologists) in tomosynthesis study by the Leuven group with the in-focus plane presented (Shaheen *et al.*, 2011) e) Examples of 2.5cm × 2.5cm regions of interest (ROIs) of simulated DBT central reconstructed slices from the uniform, BI-RADS 2 and BI-RADS 3 background test objects. A single microcalcification of diameter 400 μm and attenuation coefficient μAl60 simulated using the CatSim platform (Li *et al.*, 2018).

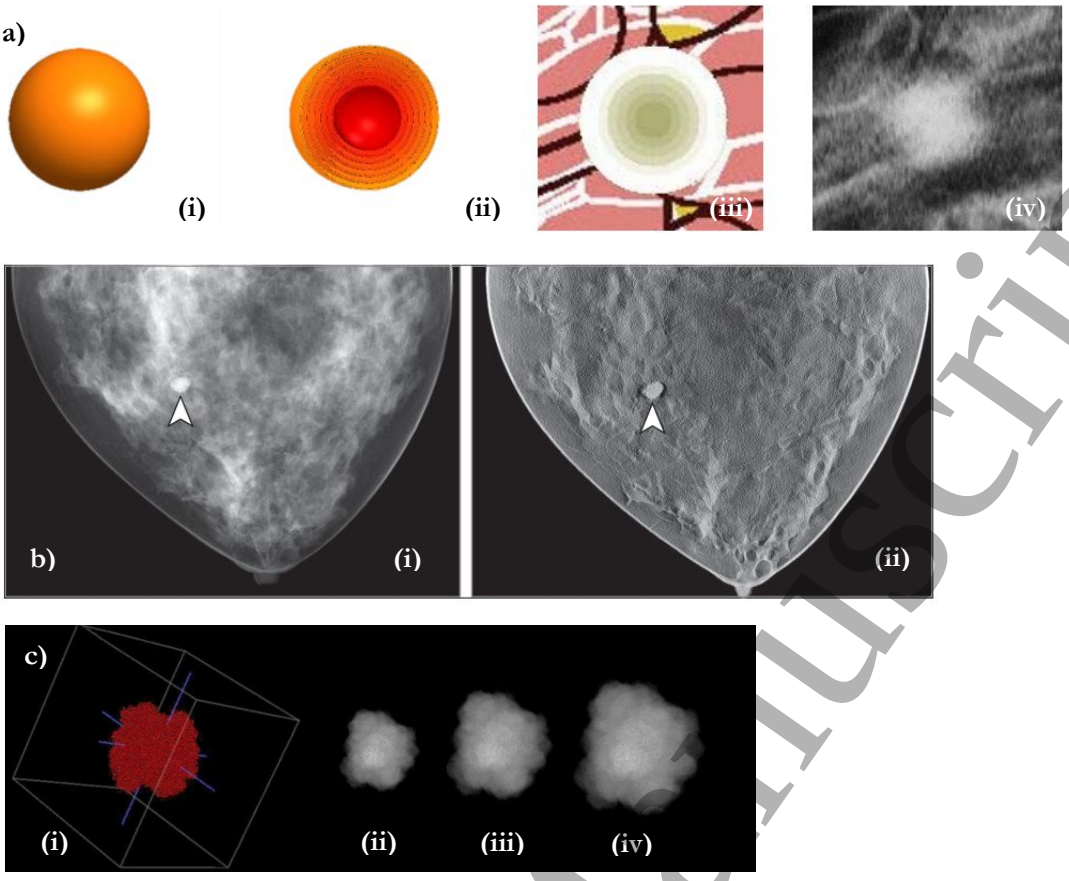
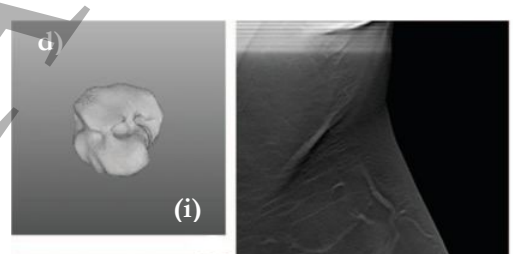


Figure 6. Examples of mass lesions used in the simulation platforms. a) OpenVCT : (i), (ii) Models of spherical lesion (concentric shells). (iii) Example of lesion insertion in a VOI in the breast phantom (longitudinal section) and (iv) cropping of the ROI in the reconstructed image (central slice) (Bakic, Barufaldi, Pokrajac, Weinstein, *et al.*, 2018).

b) VICTRE: Digital mammography (i) and selected DBT slice (ii) of a case corresponding to a breast with scattered areas of fibroglandular density containing a spiculated mass lesion (arrowheads). Lesions have been made more conspicuous for display purposes by artificially increasing their radiography attenuation during image acquisition (Badano *et al.*, 2018).

c) OPTIMAM: Simulated mass: a) 3D rendering of mass; 2D projection images of isolated masses of average diameters (a) 6.6 mm, (b) 8.4 mm and (c) 10.3 mm before insertion into the mathematical breast phantom. (Hadjipanteli *et al.*, 2019)

d) Leuven: Example of well simulated nonspiculated and spiculated mass models in the tomosynthesis study. The nonspiculated had a mean realism score of 4.8 among all observers. (i) The 3D nonspiculated mass model. (ii) The template [central projection of (i) before insertion]. (iii) The mass in (ii) after insertion in magnification view. (iv) The nonspiculated mass in the in-focus plane in DBT with an arrow pointing to the location of insertion. (Shaheen *et al.*, 2014)



2.6 Lesion insertion

Lesion insertion relates to the selection of a location and the actual means of including the lesion in the phantom model or projection image. In detection studies, lesion location will clearly influence observer detection rates, depending on the magnitude of the local breast structures. The strategy adopted by Mackenzie *et al* (Mackenzie, Thomson, *et al.*, 2021) for a human reader VCT study was to use randomized insertion sites that covered the complete breast volume. Furthermore, the site had to be more than 5 mm from the skin line, constrained to a height within the central 60% of the compressed breast thickness. For the insertion calculation, average breast density at the insertion location was estimated using the Volpara software (Youk *et al.*, 2016).

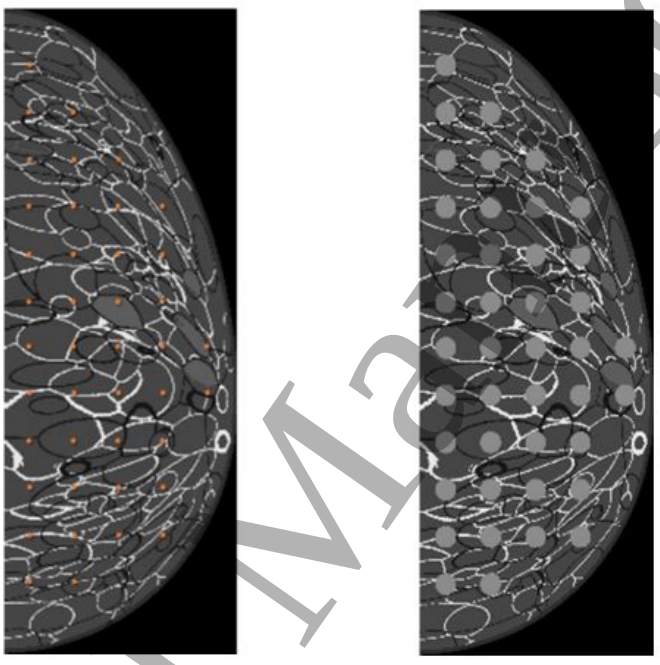


Figure 7. Locations of inserted lesions in virtual breast phantoms. Shown are breast phantom sections through the nipple, with inserted microcalcifications (left; insertion locations magnified and indicated in red for visibility) and inserted masses (right). (from Bakic *et al* (Bakic, Barufaldi, Higginbotham, *et al.*, 2018)).

In the VICTRE platform, locations are selected randomly from regions containing terminal duct lobular units, which is a common site for carcinogenesis. The location was constrained to be within phantom boundaries and could not overlap with air, muscle, nipple, and skin, and lesions already inserted (Badano *et al.*, 2018). It should be noted that lesion characteristics were adjusted during a number of pre-pilot stages so DM performance reached values reported in the literature, however how this was done is not specified. Breast phantom voxels at the lesion site are replaced with voxels labelled as glandular tissue (mass lesion) or weighted calcium oxalate (microcalcification). In OpenVCT, insertion is performed by a dedicated module that contains information on lesion type and dimensions, centre position, composition, and the boundaries. In the published work so far, voxel values at the insertion location have been replaced with values for the lesion material; in the study by Bakic *et al* (Bakic, Barufaldi, Higginbotham, *et al.*, 2018) a total of 42 lesions were inserted into each phantom; each lesion was extracted using 150×150 pixel regions of

interest (ROIs) for image interpretation (i.e. evaluation by model observer). This ensures that a range of local background breast structures are sampled in a given clinical study (Figure 7

The Leuven platform uses attenuation templates whose value ranges from 1.0 (background) to 0.0 (total attenuation within the lesion); these are multiplied with the primary component of the projection image, after accounting for the signal due to scatter (Section 2.3). The template represents the primary x-ray contrast and therefore must be applied to the primary signal content at the insertion site. Prior to insertion, the scatter fraction is estimated and subtracted from the pixel values, and the template is multiplied with the primary signal. The scatter signal is then added back to the primary signal with lesion. Insertion locations were manually chosen based on clinical knowledge of the distribution of breast carcinoma (Shaheen *et al.*, 2011). A similar insertion method is used for the templates used in the partial simulation method described for the OPTIMAM platform (Elangovan *et al.*, 2014). A correction is also applied to the noise level using a noise model (Mackenzie *et al.*, 2012), as the template changes signal level can affect the local noise level. Insertion in to voxel breast phantoms is done using voxel replacement. When inserting microcalcifications, super sampling was applied to a cubic region containing background tissue voxels at the insertion position to hold the microcalcification clusters. This was to account for the difference in voxel size of the phantom (0.1 mm) and the voxel size of the microcalcifications (13.8 μm – 34.4 μm).

A number of insertion steps are applied in the CatSim framework in order to model the local replacement of the phantom background during microcalcification insertion (Li *et al.*, 2018). First, given that the phantom is voxelized while the calcifications are held in mesh file format, the phantom and calcification are projected separately. To combine these images, the average attenuation coefficient of the background phantom voxels at the microcalcification position was subtracted from the microcalcification attenuation coefficient. The x-ray detector properties were then applied.

2.7 X-ray detector/physics

The current implementation of OpenVCT assumes an ideal detector model (Barufaldi *et al.*, 2018) and therefore the VCT generated images do not simulate the sharpness and noise levels that would be present in the projection images if a real x-ray detector were to be used. Work is ongoing to include realistic noise with the correct magnitude and texture for a given x-ray detector type (Borges *et al.*, 2017, 2019). These methods consider signal-independent electronic noise and quantum noise that accounts for correlation within the noise arising from detector crosstalk. This is characterized using the power spectral density (PSD) or noise power spectrum (NPS) of uniformly exposed images in which quantum noise contains the highest fraction (Borges *et al.*, 2017, 2019).

The VICTRE platform uses MC-GPU to implement a very complete model of the a-Se-based x-ray detector; greater detail can be found in (Badal *et al.*, 2021). X-rays are tracked until the first photoelectric interaction so that detector detection efficiency is modelled. The simulation also includes effects such as re-absorption of K x-rays and geometric distortion of the PSF due to oblique entry (see Figure 8). Electronic noise is included by adding a randomly sampled signal to each pixel from a Poisson distribution with a

mean value of 5200 electrons. A number of physical processes are not included (Badal *et al.*, 2021), including Compton interactions inside the detector, signal spread due to charge sharing across pixels and temporal effects such as lag. A limitation of this method is that CsI scintillator-based detectors such as that used in the GEHC Pristina are not currently simulated, although MC methods have been used previously to model these detectors (Badano and Sempau, 2006).

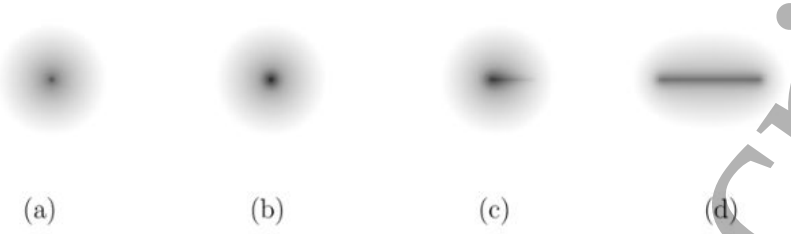


Figure 8. Visualization (in logarithmic scale) of the simulated focal spot using an ideal pinhole camera and a 200-μm-thick Se detector with 1 μm pixels: (a) ideal point focal spot, showing the effect of fluorescence spread in the detector layer; (b) 300 μm Gaussian focal spot, normal incidence; (c) 300 μm Gaussian focal spot, 21 degree incidence (depth-of-interaction effects on the point spread function visible); (d) 300 μm Gaussian focal spot, 0.24° source motion in horizontal direction (taken from Badal *et al* (Badal *et al.*, 2021))

An alternative method is used in the OPTIMAM platform, whereby the detector response function, MTF and NPS are used to quantify the sharpness and noise of a real imaging system and then applied to the simulated images. The output of the ray-tracing stage is a noise-free projection image containing only geometric blurring (no detector blurring). In the first step, this image is blurred with the detector presampling MTF of the target system. The total NPS is assumed to be separable and formed from three noise types (electronic, quantum and fixed pattern/structured NPS), each with different dose dependences and textures, in a model initially applied to diagnostic computed radiography (CR) detectors (Mackenzie and Honey, 2007). NPS coefficients are used generate three real noise images corresponding to the three noise sources, using a method proposed for quantum noise by Bochud *et al* (Bochud *et al.*, 1995) and Båth *et al* (Båth *et al.*, 2005). Each noise image is scaled corresponding to the exposure in the simulated image, the three noise sources are summed and then added pixel-wise to the simulated DBT projection image (Mackenzie *et al.*, 2012; Elangovan *et al.*, 2014). Oblique entry blurring and temporal effects are not modelled.

In the Leuven partial simulation platform, only the inserted template must be adapted to the imaging system/detector properties. The signal due to the inserted lesion is quantified by calculating the energy absorbed in the detector material (currently a-Se), from which a lesion template is formed. This template is then Fourier transformed and multiplied by an MTF that includes the focus size and focus motion blurring relevant to the insertion height of the lesion and detector presampling MTF (Shaheen *et al.*, 2011).

Recent changes to the modelling of the CsI-based mammography x-ray detectors in CatSim are described in Sánchez de la Rosa (Sánchez de la Rosa, 2019). The sharp, noiseless primary image that is the output of the ray tracing step is blurred using an MTF that quantifies sharpness due to optical scattering processes in the scintillator, the square pixel aperture and the x-ray scatter field. Three stochastic processes are included

to model the detector: the effect of the quantization ramp, additive electronic noise and quantum noise, which is described as a parametrized Poisson process. Although the magnitude of the noise is correctly modelled as a function of signal intensity, the NPS of the simulated images differs from that of the real system. In contrast to the other platforms, Sánchez de la Rosa (Sánchez de la Rosa, 2019) incorporated temporal (memory) effects in the simulation of contrast enhanced DBT. An impulse response was used to represent the memory term for each subsequent projection, with parameters estimated empirically from measurements of the evolution of signal intensity during a DBT acquisition sequence.

2.8 Dosimetry

Breast doses reported for patients undergoing a clinical study are calculated using a breast model (Svahn *et al.*, 2015), either that of Dance *et al* (D. R. Dance *et al.*, 2000; Dance, 2011) or the American College of Radiology (ACR) method (Sechopoulos *et al.*, 2014) which utilizes the model of Wu *et al* (Wu, Barnes and Tucker, 1991). Direct estimates of the dose to glandular tissue are available from the MC transport in the VICTRE study and could be calculated for the different breast types (extremely dense, heterogeneously dense, scattered fibroglandular densities and fatty) (Badano *et al.*, 2018) – see Figure 9. The average of the entire VICTRE population was then compared to the average glandular dose (AGD) for the comparative trial population, which was calculated using the method of Dance *et al* (D R Dance, Thilander, *et al.*, 2000; Dance, 2011). The OPTIMAM group also use the Dance model in combination with the incident air kerma to estimate breast dose in the virtual studies performed with the platform (Hadjipanteli *et al.*, 2017, 2019). Image generation in OpenVCT requires an "exposure" level to be set, along with the tube voltage and A/F setting, allowing the generation of low and high exposure images (Sahu *et al.*, 2019) however an explicit mean glandular dose is not currently given, although this is clearly possible by assuming a dosimetry model. Dosimetry within CatSim is performed using the MC engine to calculate dose to fibroglandular tissues, from which normalized glandular dose coefficients (DgN) are calculated. A full description is available in Carvalho (Carvalho, 2014), including a validation against the DgN coefficients of Wu *et al* (Wu, Barnes and Tucker, 1991) and Boone (Boone, 2002).

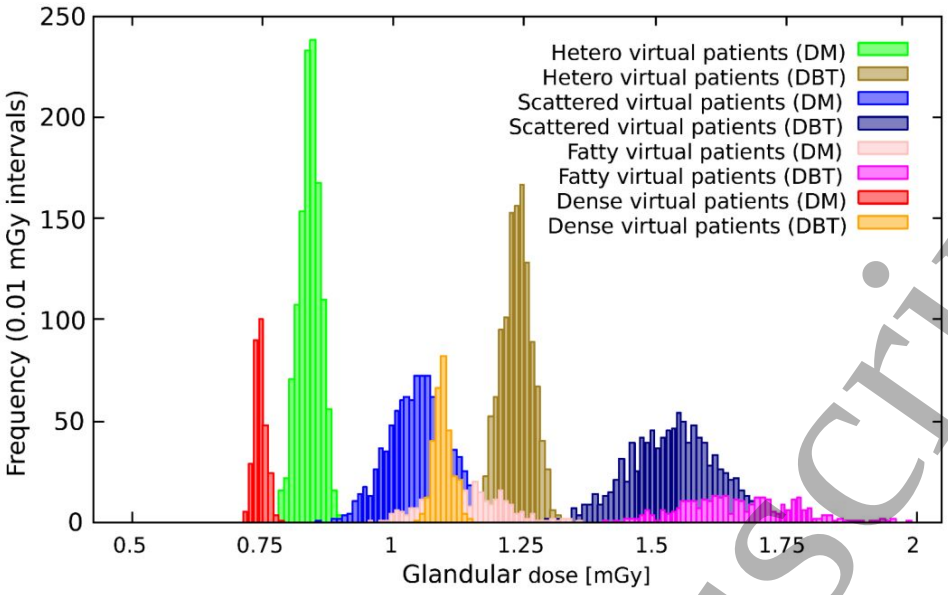


Figure 9. Radiation dose distributions in the VICTRE trial population. Glandular dose for all virtual patients was calculated and included in this histogram for digital mammography (DM) and digital breast tomosynthesis (DBT) and for each of the four breast sizes and radiographic densities considered. (taken from Badano *et al* (Badano *et al.*, 2018))

2.9 Reconstruction and image processing

Once the projections have been generated and adapted to the imaging system, these images have to be reconstructed and this step in the VCT can present some difficulties. Vendors design reconstruction algorithms for the images produced by their DBT systems, i.e. with a characteristic angular range, dose/projection, x-ray source blurring, and with images reflecting the blurring and noise characteristics of the x-ray detector used. These algorithms require detailed knowledge of the imaging system, considerable time and resources to develop and are often proprietary. The VCT described by Badano *et al* (Badano *et al.*, 2018) simulated a Siemens Inspiration DBT device which uses a proprietary Siemens reconstruction. As an alternative, a filtered backprojection (FBP) algorithm was used, with a smoothing filter applied to give a reasonable balance between sharpness and noise in the final images (Fessler, 2018; Sengupta *et al.*, 2018). Both the OpenVCT and OPTIMAM platforms use the commercially available Briona library (Briona Standard, Real Time Tomography, LLC, Villanova, PA, USA) (Kuo *et al.*, 2011), which can be configured for different geometries and scan angles. These platforms have been used to study the impact of angular range on mass lesion and microcalcification detection, with images scored either using computational observers or human observers (Hadjipanteli *et al.*, 2017, 2019; Barufaldi, Bakic and Maidment, 2019; Barufaldi *et al.*, 2020). The Leuven platform currently simulates Siemens devices (Inspiration and Revelation), and reconstruction is performed offline using software provided by Siemens which implements the earlier FBP based reconstruction (Mertelmeier *et al.*, 2006; Shaheen *et al.*, 2014) and 'EMPIRE', the latest clinical algorithm. The study by Li *et al* (Li *et al.*, 2018) using CatSim, implemented the algorithm from the GEHC Pristina ("ASIR^{DBT}"), which can generate planes at 0.5 mm and 1.0 mm spacing, and also 10 mm slabs.

An important point to consider when comparing DBT against DM imaging is the availability of clinical image processing for DM images. The VICTRE platform produces 'For Processing' images (Badal *et al.*, 2021), which are then evaluated by a computer algorithm (Badano *et al.*, 2018; Zeng *et al.*, 2020). If the images are to be used in a human observer study, then 'For Processing' images would likely be sub-optimal (especially in terms of window settings) and some form of image processing should be applied to help the human reader extract the relevant features quickly. Image processing has been shown to influence human observer lesion detection performance in DM images (Zanca *et al.*, 2009; Warren *et al.*, 2014) and therefore must be considered when designing human reader studies with simulation frameworks. This is achieved using the "Adara" library (Real Time Tomography, LLC, Villanova, PA, USA) for the OPTIMAM studies and Siemens proprietary software e.g. Opview2 (Shaheen *et al.*, 2014). Whether the lack of image processing applied to images used in computer readout stage of VCTs is influencing study performance has not been studied. Clearly, VCTs that compare, for example DBT+DM with DBT+SM require access to the manufacturer's algorithm in order to generate relevant SM images (Mackenzie, Thomson, *et al.*, 2021).

One could question whether the use of a generic reconstruction or a reconstruction that is not optimized for a given geometry will bias these types of study, for example comparing angular range. This topic has been studied in detail in two papers by Zeng *et al* (Zeng *et al.*, 2015; Zeng, Badano and Myers, 2017). A simulation framework was used to project the Bakic phantom (Bakic, Zhang and Maidment, 2011) containing lesions simulated by 4 mm diameter spheres, positioned at a height of 31.5 mm above the detector. Projection data were generated for angular spans covering 10° to 70° with 5 or 9 projections and then the number of projections were varied from 3 to 15 for angular spans of 20° and 50°. The projections were then reconstructed using the analytical FBP method and three iterative methods: simultaneous algebraic reconstruction technique (SART), the maximum-likelihood method (ML) and the total-variation regularized least-square reconstruction method (TVLS) (Zeng *et al.*, 2015). Lesion detectability was quantified using 2D and 3D channelized Hotelling observers (CHO) that implemented Laguerre-Gauss (LG) channels (Zeng *et al.*, 2015). The study found that the optimal ranges for angular span and number of views were the same for the reconstruction algorithms, suggesting that the choice of reconstruction algorithm may not be critical for optimizing the DBT acquisition parameters, at least for the DBT/lesion/phantom simulated studied. The work also emphasized that for a given geometry (angular range/number of projections), there were differences in performance between the different algorithms, and that algorithms implemented for a chosen geometry should be optimized. The later study by Zeng *et al* (Zeng, Badano and Myers, 2017) extended the work to include human readers, with CHOs designed to replicate human performance for the 4 mm spherical lesion detection task. Results from this study confirmed that optimization of system geometry can be considered to be independent of reconstruction algorithm used.

2.10 Image interpretation

Image interpretation in current platforms is performed either by computational model observer (MO) or human readers. Literature on MO methods is extensive – several reference works are available covering theory, application and use in VCTs (Van Metter, Beutel and Kundel, 2000; Barrett and Myers, 2003; He and Park, 2013; Samei and Krupinski, 2018; Abadi *et al.*, 2020). The tasks currently simulated are of the type "signal known exactly/background known exactly", where the readers are forced to make a binary decision (present/absent) or select from e.g. 4 alternative locations in an alternative forced choice (AFC) task. Spatial domain rather than Fourier domain MOs are used, as a large number of images can be simulated with and without the tasks to generate the decision variables. Furthermore, the use of statistical, spatial domain methods does not make strong assumptions about system linearity or the stationarity of the image statistics. Fourier-based MOs, on the other hand, use the NPS to characterize correlations in the anatomical noise for the detectability calculation, and the accuracy of this is limited (Barrett and Myers, 2003).

Some aspects of CHO implementation in the platforms are now described, illustrated with examples taken from Zeng *et al* (Zeng *et al.*, 2020), who developed the CHO currently used in the VICTRE platform. The main steps are:

- Selection of channel type – efficient or anthropomorphic
    - Selection of channel functions
    - Application of the channel functions to the image data (multi-slice (n2D) or 3D)
    - Channel tuning
- Template generation / "training"
- Application of the template to an independent (fresh) set of images / "testing"

Selection of channel type and application to images

All CHOs require sets of signal present and signal absent images; these are generated by the simulation platform. The physical characteristics of the target influence channel selection. A number of different channel functions have been described in the literature, including Difference of Gaussians (DOGs) (Abbey and Barrett, 2001), Gabor functions (Eckstein *et al.*, 2003) and Laguerre Gauss (LG) functions (Gallas and Barrett, 2003). Channels that extract as much information as possible about the object are known as 'efficient' channels and CHOs using these channels tend to the performance of an optimal linear observer. Alternatively, 'anthropomorphic' channels can be selected that try to mimic human observer performance (Barrett and Myers, 2003). Given the approximately round spiculated masses simulated in the VICTRE study (Badano *et al.*, 2018), Zeng *et al* (Zeng *et al.*, 2020) selected LG channels, which are circularly symmetric and therefore expected to be efficient for this task. The set of 5 calcification specks used to simulate microcalcification clusters were however spread within a 5 mm^3 volume and the group itself had no circular symmetry and thus the LG channels were therefore not suitable. Instead, a convolutional LG method was applied, first described by Diaz *et al* (Diaz *et al.*, 2015), whereby LG channel functions are convolved with the signal to produce a multi-focal set of channels that can be used to quantify performance (Zeng *et al.*, 2020).

For DBT volumes, a choice has to be made between a 2D or 3D implementation of the CHO (Platisa *et al.*, 2011). Zeng *et al* (Zeng *et al.*, 2020) examined both approaches, forming the 3D CHO by concatenating the 2D-channels for each slice. Given the slow change in the cross-section of the mass in the z-direction, the same LG channel was used for each slice, while the 3D convolutional channel set varied according to the mean signal through the calcification cluster. Zeng *et al* (Zeng *et al.*, 2020) also compared 3D and 2D versions of the CHO by converting the VOI to a 2D image using the central slice, the mean of the slab and the maximum intensity of the VOI. Overall, the 3D CHOs were found to give higher area under curve (AUC) values for DBT images and these are currently implemented in VICTRE.

Tuning of channel parameters

The LG functions have a number of parameters which, for an efficient set of channels, are selected to maximize lesion detectability – this is the tuning step. To do this, the number and width of the Gaussian function in the channels is varied to maximise AUC; Zeng *et al* (Zeng *et al.*, 2020) used 5×LG channels and found a Gaussian width of 1.5 pixels for the calcifications and widths of 30 and 25 pixels for DM and DBT images, respectively. For an anthropomorphic set of channels, the channel parameters are tuned so that the CHO performance comes close to human reader performance for the tasks (Zeng, Badano and Myers, 2017; Petrov *et al.*, 2019). Note that the tuning and template generation steps (see below) are not separate; tuning requires a template with a covariance matrix.

Template generation / "training"

The covariance matrix needed for the CHO template has to be estimated and this is done in the training stage. The covariance matrix is calculated by applying the selected channel functions to a set of signal present and signal absent images and combined with the mean signal to form an estimate of the CHO template. Applying the template to an ROI extracted from the image being evaluated produces a scalar value called a decision variable. Applying the template to many signal present and signal absent ROIs produces two distributions of decision variables, from which an ROC with an associated AUC can be established (Barrett and Myers, 2003). The AUC is the metric used to quantify task performance; the aim of the training is estimate the number of images required for a stable AUC. To do this, the number of images in the training sample is increased and AUC calculated. The number of images required will depend on the number of channels (in turn, linked to the targets) and the complexity of the background. Zeng *et al* (Zeng *et al.*, 2020) found that just ~100 image pairs (signal present/absent) were sufficient to give a stable AUC, a consequence of the small number of channels used.

Application of the template / "testing"

Once the template has been built, the final stage is to apply the template to an independent (fresh) set of signal present and signal absent images, generate distributions of decision variables and again calculate the AUC. This is the "testing" stage, where the computational observer reads or interprets the VCT images. Figure 10 illustrates the application of the trained model observer template to the dataset containing the

images to be read. A dot product of the template with an image in the reading dataset produces a scalar value (the decision variable). This is done for the signal present and signal absent reading images, resulting in distributions of decision variables for signal present and signal absent. The AUC is then calculated from these distributions.

OpenVCT also implements a CHO as the virtual reader, but fewer published details are available. The CHO uses 15 LG channels, with spreads of 15, 20, 22, 25, 26, 31 pixels depending on the application (Bakic, Barufaldi, Higginbotham, *et al.*, 2018; Barufaldi, Bakic and Maidment, 2019; Barufaldi *et al.*, 2022). These spreads are used for both calcification and mass-lesion object. Between 252 and 500 image pairs were used for the training stage. The calcification detection study by Li *et al* (Li *et al.*, 2018) used a CHO with LG function channels; between 1 and 29 channels were studied while Gaussian spread ranged from 100μm to 800μm. For training, 200 pairs of images (signal present/absent) were used. The CHO was applied to the in-focus DBT slice.

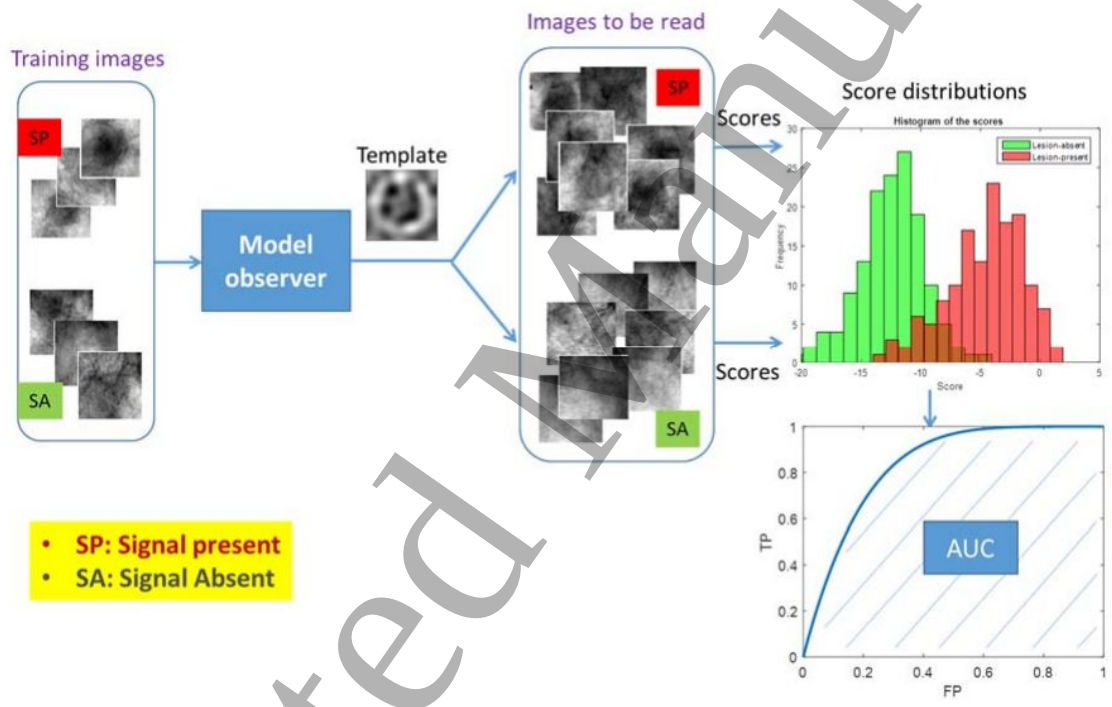


Figure 10. Flowchart of the image interpretation process by the computer readers in VICTRE. (from Badano *et al* (Badano *et al.*, 2018).

Both the OPTIMAM and Leuven platforms currently use human readers for image interpretation. Methodologies used include ROC and free-response ROC (FROC) studies for DBT and DM imaging (Shaheen *et al.*, 2014; Salvagnini *et al.*, 2016) while an AFC method has been used to establish a minimum diameter for detection of microcalcifications and masses (Hadjipanteli *et al.*, 2017, 2019).

3. VCT platform validation

Verification and validation of VCTs is discussed in (Abadi *et al.*, 2020). Two levels of validation can be considered when trying to establish the extent to which a platform is accurately modelling the performance of a real imaging system. The first level of validation assesses whether the important physical imaging

properties of the imaging system have been included in the platform and are accurately modelled. This can be done by acquiring images of well-defined test objects, which are also simulated by the platform. A range of parameters are then computed from these images and compared to establish accuracy of implementation.

Detailed first level validation of the VICTRE platform is described in Badal *et al* (Badal *et al.*, 2021), and covers the PENELOPE MC code, along with MTF, NPS and detective quantum efficiency (DQE) data which are compared to measurements, including the influence of focus size on MTF. Small, self-contained VCTs were performed examining how AUC changed as radiation dose, scattered radiation handling (grid in/out, grid properties) and focus size were varied. Elangovan *et al* (Elangovan *et al.*, 2014) describe physical validation of the OPTIMAM where two main methods were used: PMMA blocks (10 mm to 70 mm thickness) containing a 0.2 mm thick Al square for signal difference to noise (SDNR) measurements and a phantom containing gold discs of diameter 0.13 mm to 0.8 mm from which the contrast degradation factor (CDF) was calculated. Acquisition factors for the simulation were taken from the real acquisitions set by the imaging system, such that tube voltage and mAs were varied. The accuracy of the Leuven partial simulation method was assessed in a study by Vancoillie *et al* (Vancoillie *et al.*, 2020), also using PMMA blocks (20 mm to 70 mm) with objects to assess SDNR and contrast. In addition to varying PMMA thickness, tube voltage and tube current-time product (mAs) were also varied for 40 mm PMMA and both MTF and ASF were assessed. Differences between image parameters measured in simulated compared to real images were typically within 10% (Elangovan *et al.*, 2014; Vancoillie *et al.*, 2019). Explicit experimental validation for OpenVCT has not currently been published. While initial validation results for CatSim were for CT systems (De Man *et al.*, 2007) detailed validation for mammography applications is presented by Carvalho (Carvalho, 2014), including accuracy of the source modelling, scatter, breast dose and x-ray detector (evaluated via MTF, NNPS).

While these validations confirm that the physical processes occurring in the imaging chain components are correctly modelled, a second, essential level of VCT validation requires accurate and relevant implementation of the patient phantom, clinical tasks and the image interpretation stages. This should result in similar AUC values as found in the real clinical task/patient sample being modelled (Barufaldi, Maidment, *et al.*, 2021). However, if the case mix within the simulated patient group does not reflect that seen for the group of patients imaged on real imaging system, then the virtual population and the lesions can be changed so that the shape of the ROC curves more closely match (Barufaldi, Maidment, *et al.*, 2021). This was also the case for the VICTRE study, where lesion characteristics were adjusted in order to better match lesion detection results from the patient study after running a virtual pilot study (Badano *et al.*, 2018). The calibration of soft tissue lesions which produce imaging cases that match the case difficulty found in human reader studies is discussed in depth by Barufaldi *et al* (Barufaldi *et al.*, 2022).

4. VCT applications in breast imaging

4.1 Studies performed with OpenVCT

In the studies described by the UPENN group (Bakic, Barufaldi, Higginbotham, *et al.*, 2018; Barufaldi, Maidment, *et al.*, 2021), lesion detectability modelled using OpenVCT was compared to the clinical results of (Rafferty *et al.*, 2013), a study that compared the use of DM only to DM combined with DBT. As part of the work, polycubes containing one to three voxels (0.1 mm^3) of hydroxyapatite, with attenuation coefficient weighted by a factor of 0.8 were found to give a similar detection range as the real microcalcification data in (Rafferty *et al.*, 2013). A VCT using an admixture of these polycubes resulted in a good agreement between the in-silico and real data in terms of AUC and ROC curve shape (Bakic, Barufaldi, Higginbotham, *et al.*, 2018). A similar approach was applied to the modelled masses and found that thicknesses of 1.1 mm and 1.25 mm were in the detection range for the real data; the resulting VCT performed with these mass thicknesses closely matched the AUC of the clinical study. Projection and application of reconstruction/image processing of the mass lesions shown in Figure 7 results in the virtual DM and DBT images shown in Figure 11. Overall, when comparing the use of DM+DBT to just DM, Barufaldi *et al* (Barufaldi, Maidment, *et al.*, 2021) found a change in AUC of -0.003 (+0.025 for the clinical data) for calcifications, and a change of +0.106 for non-calcification lesions (+0.096 for the clinical data). While there are some differences in the shape of the ROC curves for the VCT and real study, AUC values for microcalcifications and masses were within 4%, a result that took just 4 days of GPU computation time to realize.

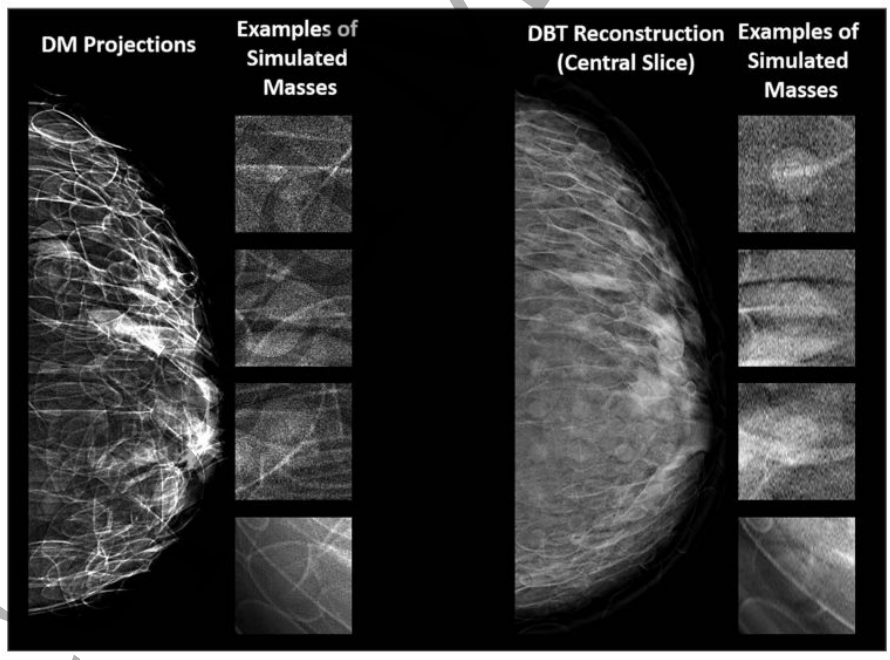


Figure 11. Examples of synthetic breast images with simulated masses, generated using OpenVCT software. (From (Barufaldi, Maidment, *et al.*, 2021))

OpenVCT has been used to study factors affecting calcification detection in DBT using the polycube lesions in a number of preliminary studies (Barufaldi, Bakic and Maidment, 2019). System parameters for

the Hologic Selenia Dimensions (15° angular range) were assumed and the VCT compared 70 μm and 140 μm detector element size, x-y voxel size of the reconstructed planes (70 μm vs 100 μm) and step and shoot versus continuous motion. The factor that most impacted calcification detection was voxel size, followed by source motion and then detector element size. A reduction in AUC of ~6% was seen for the smallest polycube. The better performance of smaller reconstructed voxels is consistent with the earlier work describing super-resolution for reconstructed DBT planes (Acciavatti and Maidment, 2012). The influence of angular range on calcification detection was also examined by (Barufaldi *et al.*, 2020) for continuous tube motion and step and shoot regimes. For all the angular ranges considered (15°, 30° and 50°), switching from step and shoot to continuous tube motion led to the largest reduction in AUC, at ~10% for the smallest polycube. For a step and shoot method, angular range did not have a significant influence on AUC, while for continuous tube motion, there was a reduction in AUC of ~3% for the 50° angular range compared to 15°. No information is given on the exposure time per projection used or how this was implemented in the ray tracing

4.2 Studies performed with the VICTRE platform

The VICTRE platform was used to evaluate the potential for DBT to replace DM imaging by re-running an in-silico version (Badano *et al.*, 2018) of the clinical study reported by (Georgian-Smith *et al.*, 2019) using the Siemens Inspiration DBT system. A total of 2986 patients were simulated of which 1944 contained lesions. The distribution of densities across the patients was 9.6% extremely dense, 40.2% for both the heterogeneously and scattered fibroglandular dense classes and 10% fatty breasts. Approximately 30,000 DM and 30,000 DBT cases were extracted from the simulated patient images and evaluated in the computational reader detection study. The change in AUC was +0.0587 (0.0062 standard error (SE)) in favour of DBT compared to DM, averaged for calcifications and masses, which was consistent with the change of +0.043 (0.017 SE) for the real study for masses and calcifications combined (Georgian-Smith *et al.*, 2019). The real study found a +0.065 (0.017 SE) change in AUC for DBT for masses, which was echoed in the in-silico trial (+0.0903 (0.008 SE)). Surprisingly, VICTRE found superior detection for microcalcifications for DBT versus DM (+0.0268 (0.004 SE)), where there was a small but non-significant reduction in AUC for microcalcifications in the real study (Georgian-Smith *et al.*, 2019) (-0.047 (0.032 SE)). Figure 12 shows the results of the VICTRE VCT, illustrating the change in AUC compared to the study of (Georgian-Smith *et al.*, 2019).

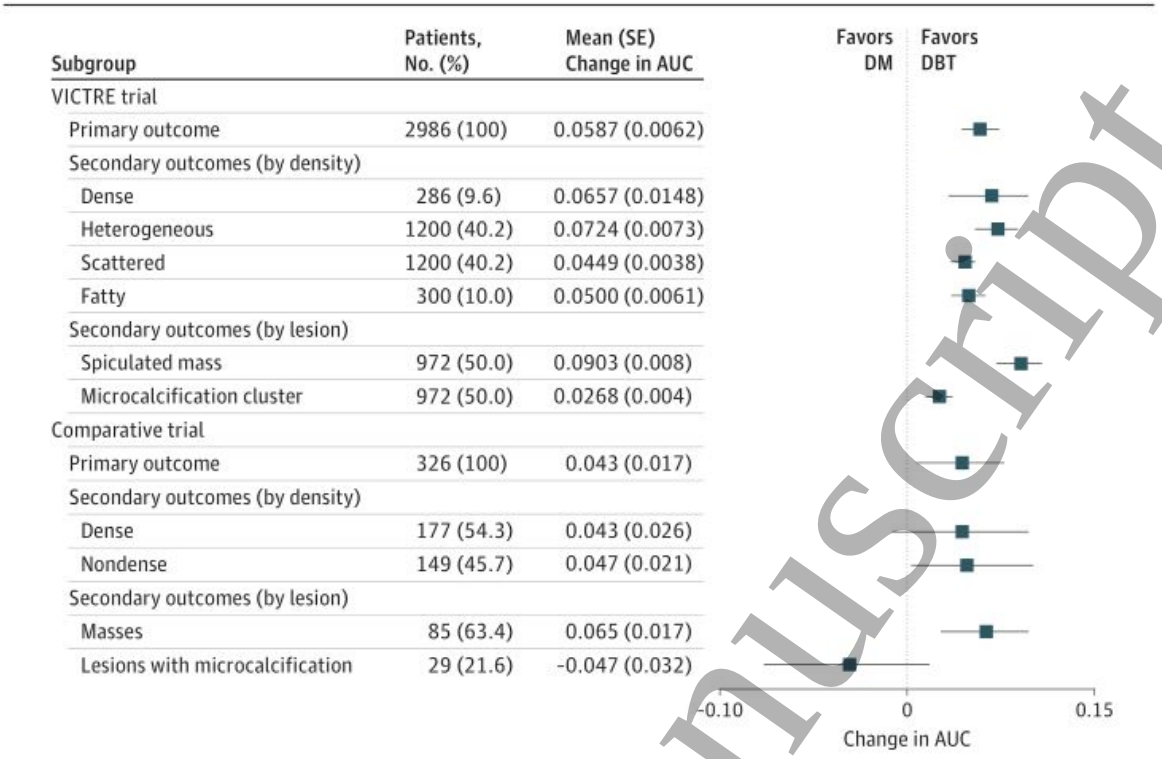


Figure 12. Trial primary and secondary outcomes. (AUC indicates area under the receiver operating characteristic curve; DBT, digital breast tomosynthesis; DM, digital mammography; VICTRE, Virtual Imaging Clinical Trial for Regulatory Evaluation; and error bars, standard errors. (taken from Badano *et al* (Badano *et al.*, 2018))

4.3 Studies performed with the OPTIMAM platform

The OPTIMAM group has performed a number of studies using the framework outlined in Table 1. In contrast to OpenVCT and VICTRE, these studies do not re-run specific clinical studies but instead try to establish the threshold diameter for detection of a given lesion type (Hadjipanteli *et al.*, 2017, 2019). A further difference is the use of human readers, generally physicists, in a 4-AFC reader study, rather than a computational reader, and then generates an AUC. Hadjipanteli *et al* (Hadjipanteli *et al.*, 2017) examined threshold diameter for microcalcification detection in simulated DM and narrow angle (15°/15 projections) and wide angle (50°/25 projections) DBT systems. Microcalcification clusters containing 5 specks of the same size (but rotated) were scaled to give clusters ranging between 110 μm to 275 μm. For a 6 cm thick breast, at a fixed breast dose of 2.5 mGy, the threshold calcification diameters were 165 ± 9 μm for DM, 211 ± 11 μm for narrow angle DBT and 257 ± 14 μm for wide angle DBT. In addition, the height of the calcification cluster above the table did not influence threshold diameter. The effect of dose was studied for the DM and narrow angle DBT systems and found to have an influence, however the effect was smaller than the influence of geometry/modality. This work was repeated for masses, again for a 6 cm breast, at 2.5 mGy. Threshold diameter for masses was 10.2 ± 1.4 mm for DM, which was significantly larger than the results for the narrow and wide angle DBT systems. No significant difference in threshold diameter

was seen between the narrow angle (6.0 ± 1.1 mm) and wide angle (5.6 ± 1.2 mm) DBT systems. Using a similar lesion set and methodology to Hadjipanteli *et al* (Hadjipanteli *et al.*, 2017, 2019), Mackenzie *et al* (Mackenzie, Kaur, *et al.*, 2021) studied the influence of breast glandularity on lesion detection for DM, DBT and synthetic mammography (SM) images. Breast voxel phantoms of thickness 5.3 cm were simulated with volumetric glandularities of 9%, 18% and 30%. Simulated images were generated for the Siemens Inspiration DBT system, using a prototype reconstruction algorithm to generate the DBT and SM images. For all three modalities, glandularity had only a small effect on calcification detection while threshold diameter for messes was significantly larger for the higher glandularity images for all three image types.

Using a different method, (Mackenzie, Thomson, *et al.*, 2021) compared lesion detection in DM, DBT and SM images. In the study, calcification clusters acquired at Leuven (Shaheen *et al.*, 2011) and masses with round, ill-defined borders (A Rashidnasab *et al.*, 2013) were generated and simulated into a dataset of 300 real breast images of all BI-RADS classes, acquired on a Siemens Inspiration DBT system. A partial simulation framework (Elangovan *et al.*, 2014) was used to generate signal present images containing subtle calcification clusters and masses for use in a reader study with three arms: DM alone, DBT+DM and DBT+SM, read by five experienced radiologists. Lesion detection using DBT was significantly better than in DM alone, while detection of subtle calcification clusters was slightly reduced but not significantly different between the DM and DBT+SM. Figure 13 illustrates these results via the lesion detection fraction. These results support the conclusions of a number of studies in which SM combined with DBT does not significantly change recall rates and cancer detection rates in comparison to DBT+DM (Houssami, 2017; Zuckerman *et al.*, 2020; Abdullah *et al.*, 2021).

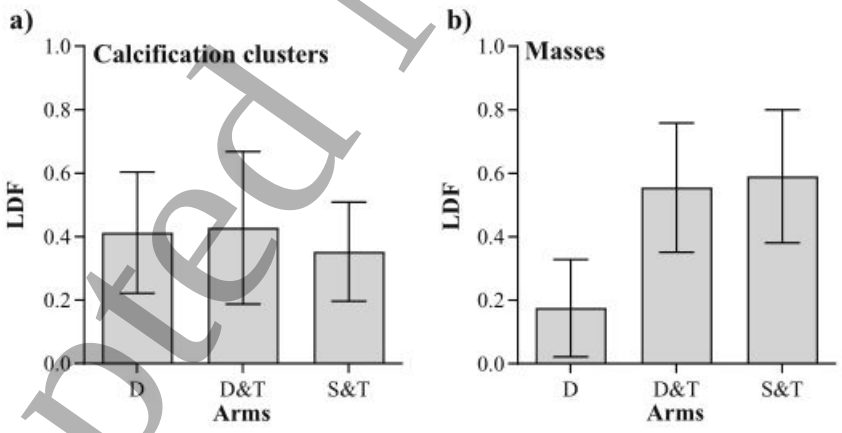


Figure 13. Lesion detection fraction (LDF) for lesion localisation marks for (a) calcification clusters and (b) masses. Error bars indicate 95% confidence interval. (taken from Mackenzie *et al* (Mackenzie, Thomson, *et al.*, 2021))

4.4 Studies performed with the CatSim and Leuven platforms

After the initial development and validation of the Leuven platform by Shaheen *et al* (Shaheen *et al.*, 2011, 2014) studies focused on DM applications. The study by Salvagnini *et al* (Salvagnini *et al.*, 2016) used the platform to examine a regime for a DM device in which automatic exposure control (AEC) programming

was changed. The dose to the detector was varied in response to increasing breast thickness in order to maintain a technical measure of lesion detectability constant. Both microcalcification and mass lesions were simulated in sets of breast images acquired before and after the dose augmentation. The work found that an increase in dose for breast thicknesses ≥50 mm brought lesion detectability to a constant level. An early DBT study by Shaheen *et al* (Shaheen, Marshall and Bosmans, 2011) found an increase in peak contrast for simulated microcalcifications for step and shoot motion compared to continuous motion the x-ray focus for DBT imaging, however this was a technical study using simple spherical objects to simulate calcifications. As part of a PhD thesis on methods to evaluate SM images, Vancoillie (Vancoillie, 2022) has performed a VCT in which microcalcifications (Shaheen *et al.*, 2011) were simulated in DBT and SM images and different dose acquisition strategies were examined (this work currently under submission). The study found reduced detectability, in terms of human reader AUC, for SM compared to DBT images, but a relative increase in the dose to the central projections increased AUC for the SM images.

Li *et al* (Li *et al.*, 2018) used CatSim to compare detectability in DBT and DM images of single microcalcifications simulated in a 5 cm SSBT (Li *et al.*, 2016) background using a simulation of the GEHC Pristina system. Calcification diameters ranged between 100 μm and 600 μm, simulated with an Al attenuation coefficient weighted by 20%, 40% and 60%. Evaluated using a CHO computational reader to generate an AUC value, DBT gave higher AUC values for the 200 μm diameter calcifications than DM. For the 100 μm calcifications, there was no difference in AUC performance between DM and DBT.

5. Discussion

This section discusses a number of aspects relating to the limitations the platforms and looks at some potential future directions for VCTs in breast imaging.

5.1 Detector modelling

While the structure of each platform is similar, dictated to a large extent by current DBT system architecture, there are notable differences in the depth or degree to which physics processes are modelled and methods used. This is in part also due to the differences between the different DBT systems, as evidenced by the range of results seen in the technical evaluation data in the companion paper and also in detailed characterization studies such as those described by Mackenzie *et al* (Mackenzie *et al.*, 2017). The method and extent to which these are incorporated in the simulations varies, an example being the x-ray detector modelling where OpenVCT uses an ideal detector model while VICTRE implements a detailed MC model. The level of physical realism required is not yet known and will to some extent depend on which factors in the imaging chain are limiting reader performance or potentially biasing the results of VCTs. Badano (Badano, 2021) refers to this as "overmodelling" in which complex system properties are incorporated at the ground level, rather than being added in a later step (for example a filtering stage) that requires additional validation. Where two imaging modalities such as DM and DBT are compared on the same device with the same x-ray detector then an ideal detector model will suffice, provided differences in pixel spacing between imaging modes are modelled (Bakic, Barufaldi, Higginbotham, *et al.*, 2018). When the influence of detector

characteristics on system performance is the object of the study, then detailed modelling of the different detectors will be needed. The OPTIMAM group have shown that x-ray detector technology influences both human reader calcification detection and technical image quality for DM imaging (Mackenzie *et al.*, 2012; Warren *et al.*, 2012), using the CDMAM contrast-detail test object (Karssemeijer and Thiijssen, 1996) to quantify technical image quality. The use of technical tests may be one method of detecting any shortcomings in the system modelling. Insufficient accuracy in modelling factors that affect the level of quantum noise and system sharpness are likely to bias microcalcification results, while errors in modelling breast structure and anatomical noise may affect the simulation of mass detection.

5.2 Dosimetry

A recent paper by Mettivier *et al* (Mettivier *et al.*, 2022) has examined the potential for speed up in breast dosimetry when switching to GPU-based simulation. Three VCT platforms were compared: Agata from the University & INFN Napoli, which is a central processing unit (CPU) implementation of Geant4, XRMC MC code developed by the University of Cagliari (U Cagliari), which also runs on CPU hardware but uses variance reduction techniques and finally the gCTD MC code from the University of Texas Southwestern (Jia *et al.*, 2012), which is written in CUDA and runs on a GPU. Computation time was reduced by a factor of up to 10^4, which may ultimately enable real-time patient dosimetry for volumetric breast datasets (Mettivier *et al.*, 2022). Future large scale VCTs may benefit from such methods.

The breast dosimetry model is currently being revised by a joint European Federation of Organisations for Medical Physics (EFOMP) and American Association of Physicists in Medicine (AAPM) Task Group ("Development of a new universal breast dosimetry method (TG282)"). The new model, which covers DM, DBT, contrast enhanced mammography and BCT modalities, reflects improved knowledge of the spatial position of the glandular parenchyma within the breast (Fedon *et al.*, 2021). In developing the model, a total of 88 BCT datasets were acquired at a site in the Netherlands. A mechanical compression algorithm was applied and distributions of fibroglandular density measured in the axial, coronal and sagittal directions of the compressed breasts (Fedon *et al.*, 2021). The fibroglandular tissue was found to be concentrated in the anterior and caudal areas, distributed symmetrically in the medio-lateral direction for CC views. For MLO views, this distribution was shifted by ~10% towards the lateral direction. The shape of the distributions was approximately independent of breast size, thickness and overall glandular fraction (Fedon *et al.*, 2021). 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). This information could be used to increase the realism of the physical distribution of fibroglandular tissue in the breast models currently used in these platforms (Bakic *et al.*, 2002; Pokrajac, Maidment and Bakic, 2012; Graff, 2016; Li *et al.*, 2016; Elangovan *et al.*, 2017), and hence the accuracy of the dosimetry in VCTs.

5.3 Breast phantoms and realism

Anthropomorphic breast phantoms act as realistic substitutes for breasts in these studies, yet the question remains, what degree of realism is required. One means of quantifying realism is to compare parameters

thought to be indicative of realism calculated for a patient image dataset and for the virtual images. Parameters used include the power spectrum (Burgess, Jacobson and Judy, 2001; Cockmartin, Bosmans and Marshall, 2013; Elangovan *et al.*, 2017) and Laplacian fractional entropy (Abbey *et al.*, 2014; Barufaldi, Abbey, *et al.*, 2021). For example, Badano *et al* (Badano *et al.*, 2018) calculated power spectra for VICTRE images of the Graff phantom from which the term (β) was determined by fitting the power law equation over the low spatial frequency range (i.e. ~0.1 to 0.7 mm^{-1}):

$$PS(f) = \frac{\mathrm{K}}{f^{\beta}} \quad (1)$$

The term β is used to characterize the 'texture' of image structures, while κ is used as measure of magnitude of the image power (Burgess, Jacobson and Judy, 2001). Results were compared to the β values calculated for DM and DBT patient datasets by Cockmartin *et al* (Cockmartin, Bosmans and Marshall, 2013) and close agreement was generally found.

Another option involves the use of human observers to evaluate image or lesion realism by showing simulated and real cases and asking radiologists to rate realism on a scale. Receiver Operating Characteristic (ROC) analysis is applied to establish whether reader response for the real and simulated datasets is significantly different (Shaheen *et al.*, 2011; Elangovan *et al.*, 2017). This type of experiment can be referred to as a "fool the radiologist" study (Badano, 2017). An alternative is to assess realism in terms of the purpose for which the image has been generated rather than simply whether the image simply looks realistic, even to an expert observer (Badano, 2017). Badano suggests that "functional realism" is required (Ferwerda, 2003; Badano, 2017), where an image is defined as realistic if it provides the same visual information as the original scene. For x-ray imaging, we could interpret this as follows: a simulated image is considered realistic if an observer has the same task performance using these images as when using real breast images i.e. is receiving and interpreting the same visual information. Following this, it is likely to be easier to generate realistic images for simple tasks (e.g. detection task) compared to higher order tasks (estimation, characterization). It is a working hypothesis that the phantoms above generate sufficiently realistic structure such that detection performance data measured in the study for DM and DBT modalities predict system performance within a (limited) class of real patients.

5.4 Lesion modelling

All platforms show significantly improved mass lesion detection for DBT versus DM, and most find similar or slightly inferior microcalcification detection performance for DBT imaging. The exception to this are the OPTIMAM results for threshold diameter of microcalcifications which are significantly worse for DBT compared to DM. While it is clear from clinical studies that the addition of DBT to DM results in improved detection of masses and distortions (Ciatto *et al.*, 2013; Skaane *et al.*, 2019), calcification detection results are more mixed. A number of clinical studies describe the potential underestimation of calcification clusters (Spangler *et al.*, 2011; Tagliafico *et al.*, 2014; Gilbert *et al.*, 2015), although generally not significant.

Regarding simulation of lesions, some studies have selected calcification clusters such that the clusters cover only a small diameter range, yielding subtle clusters that can be used to find the absolute diameter for threshold detection (Hadjipanteli *et al.*, 2017; Mackenzie, Kaur, *et al.*, 2021). This is similar to the use of five calcification specks of a certain diameter to form clusters in a phantom (Cockmartin *et al.*, 2017; Ikejimba *et al.*, 2021). While useful for establishing absolute performance differences between modalities, clinical calcification clusters are generally composed of a range of sizes, with some finer and some coarser (brighter) by which the cluster will often be detected. One of the challenges for VCTs investigating screening performance lies in accurately simulating the range of calcification clusters actually found in the screening population. This involves knowledge of the extent and shape of the overall cluster and of the size distribution and morphology of the calcifications within the cluster (Demetri-Lewis, Slanetz and Eisenberg, 2012; Horvat *et al.*, 2019); image databases with associated clinical data prove to be a valuable resource in this respect (Halling-Brown *et al.*, 2020). Improved lesion insertion methods will also help. A problem associated voxel replacement is the potential for negative contrast lesions when inserted on certain backgrounds leading to an unrealistic appearance. To overcome this, Barufaldi *et al* (Barufaldi *et al.*, 2022) describe the use of partial volumes to blend phantom and lesion materials via a weighted addition of mass attenuation coefficients. Accurate lesion modelling and insertion should largely obviate the need for pre-pilot studies in which lesion characteristics are adjusted or the use of microcalcification weighting to match the detection performance for the benchmark modality. Similar observations will apply to mass-lesions. This knowledge can also help to improve the relevance of physical test objects used for QC performance testing, provided tissue substitutes that accurately mimic these lesions can be found.

An aspect that can increase the realism of VCTs investigating breast screening is the incorporation of tumour growth models, potentially with the aim of evaluating screening programme efficiency and the ability of a modality to detect interval cancers. While this is a very broad field (Edelman, Eddy and Price, 2010; Jeanquartier *et al.*, 2016), specific applications relevant to VCTs have begun (Sengupta, Sharma and Badano, 2021; Tomic *et al.*, 2021). In the model developed by Sengupta *et al* (Sengupta, Sharma and Badano, 2021), pressure fields determined from adjacent anatomical structures govern tumour growth by allowing the lesion to develop in a given directions. Changing pressure maps generated anisotropic lesions that were seen in clinical cases. The aim of the study by Tomic *et al* (Tomic *et al.*, 2021) was to develop a set of growing tumours to evaluate multiple screening rounds with growing tumours. A fit to the probability distribution for tumour volume doubling times (TVDT) was applied to clinical data and used to generate growing tumours in 30 virtual breasts. Two successive screening rounds for each virtual breast were simulated. Measured TVDT from simulated mammograms was not significantly different from the real clinical TVDT.

The platforms discussed here currently utilise simplified geometric or procedural methods (A Rashidnasab *et al.*, 2013; de Sisternes *et al.*, 2015; Bakic, Barufaldi, Higginbotham, *et al.*, 2018) or lesions segmented from datasets acquired on a high resolution modality, for example micro-CT (Shaheen *et al.*, 2011). Future methods are likely to involve the use of artificial intelligence (AI) as a means of generating and inserting

realistic breast lesions. Ayalfi *et al* (Alyafi, Diaz and Marti, 2020) used a Deep Convolutional Generative Adversarial Network (DCGAN) to generate mass lesions using training data taken from the OPTIMAM database (Halling-Brown *et al.*, 2020). The paper by Shen *et al* (Shen *et al.*, 2021) also used DCGANs to generate mass lesions, trained with data from the Digital Database for Screening Mammography (DDSM) (Heath *et al.*, 2001), with the aim of producing lesions that include information of the shape, margin and (local) context. While the focus of this work is currently to provide data to augment training data and improve computer aided detection (CAD) algorithms, these methods have potential application in VCT lesion generation.

5.5 Image interpretation

Both OpenVCT and VICTRE use multiple insertion of lesions in a single breast to improve simulation efficiency when generating signal present images. This results in lesions evenly distributed across the breast which are then extracted in ROIs for a signal known exactly task. Increasing realism in the task modelling may involve simulating lesions at locations where certain lesion types are more likely to be found. An extension to VCTs may include the use of MOs that include search (Lau, Das and Gifford, 2013; Gifford, Liang and Das, 2016; Lago, Abbey and Eckstein, 2021b, 2021a). This may be necessary as Lago *et al* (Lago, Abbey and Eckstein, 2021a) have shown that human observer performance in a 3D location known exactly detection task does not always reflect performance in more clinically realistic tasks that involve 3D search tasks. The search task also induces some inefficiency that is not evaluated in a location known exactly method; failure to include these sources may not lead to an accurate evaluation of 3D imaging performance (Lago, Abbey and Eckstein, 2021a). Whether this outweighs the advantage of performing a search/detection task in 3D (DBT), with reduced levels of anatomical noise, compared to 2D (DM) has not been studied.

A further question related to the use of both human observers and channelized MOs for the image interpretation stage, relates to MO generalizability. If, for example, a manufacturer generates an image dataset using the VICTRE platform for a VCT to support the introduction of a new reconstruction (B) in place of some current reconstruction algorithm (A), can an efficient MO tuned to maximise AUC for reconstruction A be applied to dataset B without re-tuning, or should the AUC be maximized for each reconstruction. If this has to be supported by human reader results then much of the benefit in terms of the resources saved is lost. An anthropomorphic MO could be developed but tuning separately to human reader data acquired for both reconstructions is necessary in this case.

5.6 Validation datasets

One means of encouraging cross-platform validation or comparison could be the publication of test datasets, akin to the datasets published MC simulations (Sechopoulos *et al.*, 2015). In order to facilitate validation of the correct physical modelling of important aspects of the VCT chain, measured projection image data of well-defined, relevant test objects acquired with specific x-ray spectra could be listed along the technical data for the imaging system and x-ray detector required to perform a simulation. This might

include the MTF, NPS, focus size and shape, antiscatter grid parameters, angular velocity of the x-ray source etc. These would be input to the simulation platform and the results compared against the measured data, thus allowing various groups to compare simulation platforms and methods. Extending this to a validation of the full VCT chain would be more involved, requiring sets of lesions, for example voxel templates of microcalcifications or mass lesions, a breast model and a reconstruction algorithm. Comparison or validation could be made in terms of AUC derived for a standard MO.

6. Conclusion

As can be seen from the studies performed so far, VCTs can greatly help in bridging the gap between technical measurements and the results of clinical studies. *In-silico* trials have the potential to replace expensive and time consuming physical trials that recruit human subjects, and their use as part of regulatory evaluations is almost certain to increase (Badano, 2021). An advantage of these methods is that studies can be performed for theoretical devices or used to explore system configurations at a design stage (Sánchez de la Rosa, 2019). While improved modelling and computational power will lead to increased realism, VCTs have already produced useful results, despite limitations on the modelling of anatomy, task and imaging chain. Rigorous validation and transparent description of the frameworks should ensure progress in this area, something that could be helped with the availability of projection image datasets of test objects along with measured system data. As has already been done for the VICTRE and OpenVCT platforms, dissemination of platforms via internet hosting services should improve collaboration and also help in standardization. Physical characterization and test object methods will remain indispensable for technical assessment and QC of DBT and DM systems in the field. The rapid development of virtual methods applied to x-ray breast imaging and DBT in particular suggests that medical physicists, researchers and manufacturers will increasingly rely on simulation and virtual methods for system design and performance evaluation.

References

A Rashidnasab, A. *et al.* (2013) 'Simulation and assessment of realistic breast lesions using fractal growth models', *Physics in Medicine and Biology*, 58, pp. 5613–5627. doi: 10.1088/0031-9155/58/16/5613.

Abadi, E. *et al.* (2019) 'DukeSim: A realistic, rapid, and scanner-specific simulation framework in computed tomography', *IEEE Transactions on Medical Imaging*. IEEE, 38(6), pp. 1457–1465. doi: 10.1109/TMI.2018.2886530.

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.* (2014) 'Non-Gaussian statistical properties of virtual breast phantoms', 9037, pp. 90370G-90370G–8. doi: 10.1117/12.2044446.

Abbey, C. K. and Barrett, H. H. (2001) 'Human- and model-observer performance in ramp-spectrum noise :

effects of regularization and object variability', *Journal of the Optical Society of America A*, 18(3), pp. 473–488.

Abdullah, P. *et al.* (2021) 'Synthetic 2D mammography versus standard 2D digital mammography: A diagnostic test accuracy systematic review and meta-analysis', *American Journal of Roentgenology*, 217(2), pp. 314–324. doi: 10.2214/AJR.20.24204.

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

Agostinelli, S. *et al.* (2003) 'GEANT4 - A simulation toolkit', *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3), pp. 250–303. doi: 10.1016/S0168-9002(03)01368-8.

Alyafi, B., Diaz, O. and Marti, R. (2020) 'DCGANs for realistic breast mass augmentation in x-ray mammography', in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE 11314*, p. 1131420. doi: https://doi.org/10.1117/12.2543506.

Badal, A. *et al.* (2021) 'Mammography and breast tomosynthesis simulator for virtual clinical trials', *Computer Physics Communications*. Elsevier B.V., 261, p. 107779. doi: 10.1016/j.cpc.2020.107779.

Badano, A. (2011) 'In silico imaging: Definition, possibilities and challenges', *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Elsevier, 648(SUPPL. 1), pp. S276–S280. doi: 10.1016/j.nima.2010.11.054.

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. *et al.* (2018) 'Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial', *JAMA network open*, 1(7), p. e185474. doi: 10.1001/jamanetworkopen.2018.5474.

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 Sempau, J. (2006) 'MANTIS: combined x-ray, electron and optical Monte Carlo simulations of indirect radiation imaging systems.', *Physics in medicine and biology*, 51(6), pp. 1545–61. doi: 10.1088/0031-9155/51/6/013.

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.

Bakic, P. R., Barufaldi, B., Pokrajac, D., Lago, M. A., *et al.* (2018) 'Developing populations of software breast phantoms for virtual clinical trials', (July), p. 73. doi: 10.1117/12.2318473.

Bakic, P. R., Barufaldi, B., Pokrajac, D., Weinstein, S., *et al.* (2018) 'Optimized simulation of breast anatomy

for virtual clinical trials', in *Proc. SPIE 10718 14th International Workshop on Breast Imaging (IWBI 2018)*. SPIE, Bellingham, WA, pp. 107181X-1–10. doi: 10.1117/12.2318525.

Bakic, P. R., Barufaldi, B., Higginbotham, D., *et al.* (2018) 'Virtual clinical trial of lesion detection in digital mammography and digital breast tomosynthesis', in *Proc. SPIE 10573, Medical Imaging 2018: Physics of Medical Imaging*. SPIE, Bellingham, WA, p. 1057306 (9 March 2018). doi: 10.1117/12.2294934.

Bakic, P. R., Zhang, C. and Maidment, A. D. A. (2011) 'Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm', 19104(March), pp. 3165–3176. doi: 10.1118/1.3590357.

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.* (2018) 'OpenVCT: a GPU-accelerated virtual clinical trial pipeline for mammography and digital breast tomosynthesis', (July), p. 194. doi: 10.1117/12.2294935.

Barufaldi, B. *et al.* (2020) 'Determining the optimal angular range of the X-ray source motion in tomosynthesis using virtual clinical trials', (March 2020), p. 18. doi: 10.1117/12.2549600.

Barufaldi, B., Abbey, C. K., *et al.* (2021) 'Computational Breast Anatomy Simulation Using Multi-scale Perlin Noise', *IEEE Transactions on Medical Imaging*, xx(X), pp. 1–11. 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.

Barufaldi, B. *et al.* (2022) 'Computer simulations of case difficulty in digital breast tomosynthesis using virtual clinical trials', *Medical Physics*, 49(4), pp. 2220–2232. doi: 10.1002/mp.15553.

Barufaldi, B., Bakic, P. R. and Maidment, A. D. A. (2019) 'Multiple-reader, multiple-case ROC analysis for determining the limit of calcification detection in tomosynthesis', (March 2019), p. 22. doi: 10.1117/12.2512884.

Båth, M. *et al.* (2005) 'Method of simulating dose reduction for digital radiographic systems', *Radiation Protection Dosimetry*, 114(1–3), pp. 253–259. doi: 10.1093/rpd/nch540.

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

Berger, M. J. *et al.* (2005) *XCOM: Photon Cross Sections Database*. doi: https://dx.doi.org/10.18434/T48G6X.

Birch, R. and Marshall, M. (1979) 'Computation of bremsstrahlung X-ray spectra and comparison with

spectra measured with a Ge(Li) detector', *Physics in medicine and biology*, 24(3), pp. 505–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/461510.

Bliznakova, K. *et al.* (2010) 'Experimental validation of a radiographic simulation code using breast phantom for X-ray imaging.', *Computers in biology and medicine*. Elsevier, 40(2), pp. 208–14. doi: 10.1016/j.compbiomed.2009.11.017.

Bliznakova, K. *et al.* (2012) 'BreastSimulator: A software platform for breast x-ray imaging research', *Journal of Biomedical Graphics and Computing*, 2(1), pp. 1–14. doi: 10.5430/jbgc.v2n1p1.

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.

Bochud, F. O. *et al.* (1995) 'Detectability of radiological images: the influence of anatomical noise', in *SPIE Medical Imaging 1995: Image Perception*. SPIE, Bellingham, WA, pp. 156–164. doi: 10.1117/12.206845.

Boone, J. M. *et al.* (2000) 'Scatter/primary in mammography: comprehensive results.', *Medical physics*, 27(10), pp. 2408–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11099211.

Boone, J. M. (2002) 'Normalized glandular dose (DgN) coefficients for arbitrary x-ray spectra in mammography: Computer-fit values of Monte Carlo derived data', *Medical Physics*, 29(5), p. 869. doi: 10.1118/1.1472499.

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.

Borges, L. R. *et al.* (2017) 'Method for Simulating Dose Reduction in Digital Breast Tomosynthesis', *IEEE Transactions on Medical Imaging*, 36(11), pp. 2331–2342. doi: 10.1109/TMI.2017.2715826.

Borges, L. R. *et al.* (2019) 'Technical Note: Noise models for virtual clinical trials of digital breast tomosynthesis', *Medical Physics*, 46(6), pp. 2683–2689. doi: 10.1002/mp.13534.

Bujila, R., Omar, A. and Poludniowski, G. (2020) 'A validation of SpekPy: A software toolkit for modelling X-ray tube spectra', *Physica Medica*. Elsevier, 75(June), pp. 44–54. doi: 10.1016/j.ejmp.2020.04.026.

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.

Caballo, M. *et al.* (2022) 'Patient-derived heterogeneous breast phantoms for advanced dosimetry in mammography and digital breast tomosynthesis', *Medical Physics*, Online ahe. doi: 10.1002/mp.15785.

Carton, A. *et al.* (2003) 'Development and validation of a simulation procedure to study the visibility of micro calcifications in digital mammograms', *Medical Physics*, 30(8), pp. 2234–2240. doi: 10.1118/1.1591193.

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

Chan, H. P. and Doi, K. (1983) 'The validity of Monte Carlo simulation in studies of scattered radiation in diagnostic radiology.', *Physics in Medicine and Biology*, 28(2), pp. 109–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6867102.

Chen, H. *et al.* (2015) 'On image quality metrics and the usefulness of grids in digital mammography', *Journal of Medical Imaging*, 2(1), pp. 013501-1–7. doi: 10.1117/1.jmi.2.1.013501.

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.

Cockmartin, L. *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., 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.

Cooper, V. N. *et al.* (2000) 'An edge spread technique for measurement of the scatter-to-primary ratio in mammography.', *Medical physics*, 27(5), pp. 845–853. doi: 10.1118/1.598950.

Cranley, K. *et al.* (1997) *IPEM Report 78 Catalogue of diagnostic X-ray spectra & other data*. York.

Cullen, D. E., Hubbell, J. H. and Kissel, L. (1997) *EPDL97 The Evaluated Photon Data Library, 097 Version, Technical Report UCRL-50400*. Livermore, CA.

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. 3225–40.

Dance, D. R. *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, a K., *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.', *The British journal of radiology*, 73(874), pp. 1056–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11271898.

Dance, D. R. (2011) 'Estimation of mean glandular dose for breast tomosynthesis : factors for use with the UK , European and IAEA breast dosimetry protocols', 453. doi: 10.1088/0031-9155/56/2/011.

Day, G. J. and Dance, D. R. (1983) 'X-ray transmission formula for antiscatter grids', *Phys. Med. Biol.*, 28(12),

pp. 1429–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6665036.

Demetri-Lewis, A., Slanetz, P. J. and Eisenberg, R. L. (2012) 'Breast calcifications: The focal group', *American Journal of Roentgenology*, 198(4), pp. 325–343. doi: 10.2214/AJR.10.5732.

Diaz, I. *et al.* (2015) 'Derivation of an Observer Model Adapted to Irregular Signals Based on Convolution Channels', *IEEE Transactions on Medical Imaging*, 34(7), pp. 1428–1435. doi: 10.1109/TMI.2015.2395433.

Diaz, O. *et al.* (2014) 'Estimation of scattered radiation in digital breast tomosynthesis.', *Physics in medicine and biology*, 59(15), pp. 4375–4390. doi: 10.1088/0031-9155/59/15/4375.

Diaz, O. *et al.* (2019) 'Simple method for computing scattered radiation in breast tomosynthesis', *Medical Physics*, 46(11), pp. 4826–4836. doi: 10.1002/mp.13760.

Ducote, J. L. and Molloi, S. (2010) 'Scatter correction in digital mammography based on image deconvolution', 1295. doi: 10.1088/0031-9155/55/5/003.

Dustler, M. *et al.* (2015) 'Application of the fractal Perlin noise algorithm for the generation of simulated breast tissue', in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, p. 94123E. doi: 10.1117/12.2081856.

Eckstein, M. *et al.* (2003) 'Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks.', *Optics express*, 11(5), pp. 460–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19461753.

Edelman, L. B., Eddy, J. A. and Price, N. D. (2010) 'In silico models of cancer', *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(4), pp. 438–459. doi: 10.1002/wsbm.75.

Elangovan, P. *et al.* (2014) 'Development and validation of a modelling framework for simulating 2D-mammography and breast tomosynthesis images.', *Physics in medicine and biology*, 59(15), pp. 4275–4293. doi: 10.1088/0031-9155/59/15/4275.

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.* (2016) *Protocol for the Quality Control of the Physical and Technical Aspects of Digital Breast Tomosynthesis Systems v1.01*.

Erickson, D. W. *et al.* (2016) 'Population of 224 realistic human subject-based computational breast phantoms', *Med. Phys.*, 43(1), pp. 23–32. doi: 10.1118/1.4937597.

Fahrig, R., Rowlands, J. A. and Yaffe, M. J. (1996) 'X-ray imaging with amorphous selenium: Optimal spectra for digital mammography', *Medical Physics*, 23(4), pp. 557–567.

Fedon, C. *et al.* (2021) 'Fibroglandular tissue distribution in the breast during mammography and tomosynthesis based on breast CT data: A patient-based characterization of the breast parenchyma', *Medical Physics*, 48(3), pp. 1436–1447. doi: 10.1002/mp.14716.

Feng, S. S. J. and Sechopoulos, I. (2012) 'Clinical digital breast tomosynthesis system: Dosimetric characterization', *Radiology*, 263(1), pp. 35–42. doi: 10.1148/radiol.11111789.

Ferwerda, J. A. (2003) 'Three varieties of realism in computer graphics', in *SPIE Human Vision and Electronic Imaging VIII*. SPIE, Bellingham, WA, pp. 290–297. doi: 10.1117/12.473899.

Fessler, J. A. (2018) 'Michigan Image Reconstruction Toolbox'.

di Franco, F. *et al.* (2020) 'GEANT4 Monte Carlo simulations for virtual clinical trials in breast X-ray imaging: Proof of concept', *Physica Medica*. Elsevier, 74(May), pp. 133–142. doi: 10.1016/j.ejmp.2020.05.007.

Frangi, A. F., Tsaftaris, S. A. and Prince, J. L. (2018) 'Simulation and Synthesis in Medical Imaging', *IEEE Transactions on Medical Imaging*. IEEE, 37(3), pp. 673–679. doi: 10.1109/TMI.2018.2800298.

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

Gallas, B. D. and Barrett, H. H. (2003) 'Validating the use of channels to estimate the ideal linear observer.', *Journal of the Optical Society of America. A, Optics, image science, and vision*, 20(9), pp. 1725–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12968645.

Georgian-Smith, D. *et al.* (2019) 'Can Digital Breast Tomosynthesis Replace Full-Field Digital Mammography? A Multireader, Multicase Study of Wide-Angle Tomosynthesis', *American Journal of Roentgenology*, 212, pp. 1393–1399.

Gifford, H. C., Liang, Z. and Das, M. (2016) 'Visual-search observers for assessing tomographic x-ray image quality', *Medical Physics*, 43(3), pp. 1563–1575. doi: 10.1118/1.4942485.

Gilbert, F. J. *et al.* (2015) 'Accuracy of Digital Breast Tomosynthesis for Depicting Breast Cancer Subgroups in a UK Retrospective Reading Study', *Radiology*, 277(3), pp. 697–706.

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.

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.

De Greef, M. *et al.* (2009) 'Accelerated ray tracing for radiotherapy dose calculations on a GPU', *Medical Physics*, 36(9), pp. 4095–4102. doi: 10.1118/1.3190156.

Hadjipanteli, A. *et al.* (2017) 'The effect of system geometry and dose on the threshold detectable

calcification diameter in 2D-mammography and digital breast tomosynthesis', *Physics in Medicine and Biology*. IOP Publishing, 62(3), pp. 858–877. doi: 10.1088/1361-6560/aa4f6e.

Hadjipanteli, A. *et al.* (2019) 'The threshold detectable mass diameter for 2D-mammography and digital breast tomosynthesis', *Physica Medica*. Elsevier, 57(November 2018), pp. 25–32. doi: 10.1016/j.ejmp.2018.11.014.

Halling-Brown, M. D. *et al.* (2020) 'OPTIMAM Mammography Image Database: A Large-Scale Resource of Mammography Images and Clinical Data', *Radiology: Artifical Intelligence*, 3(1). doi: https://doi.org/10.1148/ryai.2020200103.

He, X. and Park, S. (2013) 'Model Observers in Medical Imaging Research', *Theranostics*, 3(10), pp. 774–786. doi: 10.7150/thno.5138.

Heath, M. *et al.* (2001) 'The Digital Database for Screening Mammography', in Yaffe, M. J. (ed.) *in Proceedings of the Fifth International Workshop on Digital Mammography*. Medical Physics Publishing, pp. 212–218.

Hernandez, A. M. *et al.* (2017) 'Generation and analysis of clinically relevant breast imaging x-ray spectra', *Medical Physics*, 44(6), pp. 2148–2160.

Horvat, J. V. *et al.* (2019) 'Calcifications at digital breast tomosynthesis: Imaging features and biopsy techniques', *Radiographics*, 39(2), pp. 307–318. doi: 10.1148/rg.2019180124.

Houssami, N. (2017) 'Evidence on Synthesized Two-dimensional Mammography Versus Digital Mammography When Using Tomosynthesis (Three-dimensional Mammography) for Population Breast Cancer Screening', *Clinical Breast Cancer*. Elsevier Inc., 18(4), pp. 255-260.e1. doi: 10.1016/j.clbc.2017.09.012.

ICRU (1989) *ICRU Report 44: Tissue substitutes in radiation dosimetry and measurement*.

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.

Jeanquartier, F. *et al.* (2016) 'In silico modeling for tumor growth visualization', *BMC Systems Biology*. BMC Systems Biology, 10(1), pp. 1–15. doi: 10.1186/s12918-016-0318-8.

Jia, X. *et al.* (2012) 'Fast Monte Carlo simulation for patient-specific CT/CBCT imaging dose calculation', *Physics in Medicine and Biology*, 57(3), pp. 577–590. doi: 10.1088/0031-9155/57/3/577.

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

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.

Kopans, D. B. (2021) 'Design, implementation, and pitfalls of TMIST', *Clinical Imaging*. Elsevier Inc., 78(May), pp. 304–307. doi: 10.1016/j.clinimag.2021.06.011.

Kuo, J. *et al.* (2011) 'Dynamic reconstruction and rendering of 3D tomosynthesis images', in *Medical Imaging 2011: Physics of Medical Imaging*. SPIE, Bellingham, WA, p. 796116. doi: 10.1117/12.878910.

Lago, M. A., Abbey, C. K. and Eckstein, M. P. (2021a) 'Foveated Model Observers for Visual Search in 3D Medical Images', *IEEE Transactions on Medical Imaging*, 40(3), pp. 1021–1031. doi: 10.1109/TMI.2020.3044530.

Lago, M. A., Abbey, C. K. and Eckstein, M. P. (2021b) 'Medical image quality metrics for foveated model observers', *Journal of Medical Imaging*, 8(4), p. 041209. doi: https://doi.org/10.1117/1.JMI.8.4.041209.

Lau, B. A., Das, M. and Gifford, H. C. (2013) 'Towards visual-search model observers for mass detection in breast tomosynthesis', in *Proc. SPIE 8668, Medical Imaging 2013: Physics of Medical Imaging*. SPIE, Bellingham, WA, p. 86680X.

Lazos, D. *et al.* (2003) 'An integrated research tool for X-ray imaging simulation', 70, pp. 241–251.

Lazos, D., Kolitsi, Z. and Pallikarakis, N. (2000) 'A software data generator for radiographic imaging investigations.', *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 4(1), pp. 76–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10761778.

Leon, S. M., Brateman, L. F. and Wagner, L. K. (2014) 'Characterization of scatter in digital mammography from physical measurements', *Medical Physics*, 41(6), pp. 061901-1–8. doi: 10.1118/1.4873321.

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.

Li, Z. *et al.* (2018) 'Comparison of microcalcification detectability in FFDM and DBT using a virtual clinical trial', in *Proc. SPIE 10577, Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment*. SPIE, Bellingham, WA, p. 105770D (7 March 2018). doi: 10.1117/12.2293619.

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

Mackenzie, A. *et al.* (2012) 'Conversion of mammographic images to appear with the noise and sharpness characteristics of a different detector and x-ray system', 39(May), pp. 2721–2734.

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.

Mackenzie, A., Kaur, S., *et al.* (2021) 'Effect of glandularity on the detection of simulated cancers in planar, tomosynthesis and synthetic 2D imaging of the breast using a hybrid virtual clinical trial', *Medical Physics*.

doi: 10.1002/mp.15216.

Mackenzie, A., Thomson, E. L., *et al.* (2021) 'Virtual clinical trial to compare cancer detection using combinations of 2D mammography, digital breast tomosynthesis and synthetic 2D imaging', *European Radiology*. European Radiology. doi: 10.1007/s00330-021-08197-x.

Mackenzie, A. and Honey, I. D. (2007) 'Characterization of noise sources for two generations of computed radiography systems using powder and crystalline photostimulable phosphors', *Medical Physics*, 34(8), pp. 3345–3357. doi: 10.1118/1.2750973.

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.

Marchessoux, C., Kimpe, T. and Bert, T. (2008) 'A virtual image chain for perceived and clinical image quality of medical display', *IEEE/OSA Journal of Display Technology*, 4(4), pp. 356–368. doi: 10.1109/JDT.2008.2001164.

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.

Mertelmeier, T. *et al.* (2006) 'Optimizing filtered backprojection reconstruction for a breast tomosynthesis prototype device', in *Medical Imaging 2006: Physics of Medical Imaging*. SPIE, Bellingham, WA, pp. 61420F-61420F–12. doi: 10.1117/12.651380.

Van Metter, R., Beutel, J. and Kundel, H. L. (eds) (2000) *Handbook of Medical Imaging, Volume 1. Physics and Psychophysics*. SPIE, Bellingham, WA.

Mettivier, G. *et al.* (2017) 'Evaluation of the BreastSimulator software platform for breast tomography', *Phys. Med. Biol.* IOP Publishing, 62, pp. 6446–6466.

Mettivier, G. *et al.* (2019) 'The Napoli-Varna-Davis project for virtual clinical trials in X-ray breast imaging', *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2019*. IEEE, pp. 0–4. doi: 10.1109/NSS/MIC42101.2019.9059828.

Mettivier, G. *et al.* (2022) 'Virtual Clinical Trials in 2D and 3D X-ray Breast Imaging and Dosimetry: Comparison of CPU-Based and GPU-Based Monte Carlo Codes', *Cancers*, 14(1027), pp. 1–14. doi: https://doi.org/10.3390/cancers14041027.

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.

Omar, A., Andreo, P. and Poludniowski, G. (2020a) 'A model for the energy and angular distribution of x rays emitted from an x-ray tube. Part I. Bremsstrahlung production', *Medical Physics*, 47(10), pp. 4763–4774.

doi: 10.1002/mp.14359.

Omar, A., Andreo, P. and Poludniowski, G. (2020b) 'A model for the energy and angular distribution of x rays emitted from an x-ray tube. Part II. Validation of x-ray spectra from 20 to 300 kV', *Medical Physics*, 47(9), pp. 4005–4019. doi: 10.1002/mp.14360.

Perlin, K. (1985) 'An image synthesizer', *Computer Graphics (ACM)*, 19(3), pp. 287–296. doi: 10.1145/325165.325247.

Petersson, H. *et al.* (2016) 'Validation of a simulation procedure for generating breast tomosynthesis projection images', *Radiation Protection Dosimetry*, 169(1), pp. 386–391. doi: 10.1093/rpd/ncv555.

Petrov, D. *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.

Platisa, L. *et al.* (2011) 'Channelized Hotelling observers for the assessment of volumetric imaging data sets', *J. Opt. Soc. Am. A*, 28(6), pp. 18–20.

Pokrajac, D. D., Maidment, A. D. A. and Bakic, P. R. (2012) 'Optimized generation of high resolution breast anthropomorphic software phantoms', *Medical Physics*, 39(April), pp. 2290–2302.

Poludniowski, G. *et al.* (2009) 'SpekCalc : a program to calculate photon spectra from tungsten anode x-ray tubes', *Phys. Med. Biol.*, 54, pp. N433–N438. doi: 10.1088/0031-9155/54/19/N01.

Poludniowski, G. *et al.* (2021) 'Technical Note: SpekPy v2.0—a software toolkit for modeling x-ray tube spectra', *Medical Physics*, 48(7), pp. 3630–3637. doi: 10.1002/mp.14945.

Poludniowski, G. G. (2007) 'Calculation of x-ray spectra emerging from an x-ray tube. Part II. X-ray production and filtration in x-ray targets', *Medical Physics*, 34(6), p. 2175. doi: 10.1118/1.2734726.

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

Rafferty, E. A. *et al.* (2013) 'Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: Results of a multicenter, multireader trial', *Radiology*, 266(1), pp. 104–113. doi: 10.1148/radiol.12120674.

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.

Sahu, P. *et al.* (2019) 'Using virtual digital breast tomosynthesis for de-noising of low-dose projection images', *Proceedings - International Symposium on Biomedical Imaging*. IEEE, 2019-April(Isbi), pp. 1647–1651. doi: 10.1109/ISBI.2019.8759408.

Salvagnini, E. *et al.* (2012) 'Quantification of scattered radiation in projection mammography : Four practical methods compared', *Medical Physics*, 39(6), pp. 3167–3180.

Salvagnini, E. *et al.* (2016) 'Impact of compressed breast thickness and dose on lesion detectability in digital mammography: FROC study with simulated lesions in real mammograms', *Medical Physics*, 43(9), pp. 5104–5116. doi: 10.1118/1.4960630.

Salvat, F., Fernandez-Varea, J. M. and Sempau, J. (2006) *PENELOPE - 2006 : A Code System for Monte Carlo Simulation of Electron and Photon Transport*.

Samei, E. and Krupinski, E. A. (eds) (2018) *The Handbook of Medical Image Perception and Techniques*. 2nd Editio. Cambridge University Press.

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.

Sarno, A. *et al.* (2021) 'Dataset of patient-derived digital breast phantoms for in silico studies in breast computed tomography, digital breast tomosynthesis, and digital mammography', *Medical Physics*, 48(5), pp. 2682–2693. doi: 10.1002/mp.14826.

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

Sechopoulos, I. *et al.* (2015) 'Monte Carlo reference data sets for imaging research: Executive summary of the report of AAPM Research Committee Task Group 195', *Medical Physics*, 42(10), pp. 5679–5691. doi: 10.1118/1.4928676.

Sempau, J., Badal, A. and Brualla, L. (2011) 'A PENELOPE-based system for the automated Monte Carlo simulation of clinacs and voxelized geometries—application to far-from-axis fields', *Medical Physics*, 38(11), p. 5887. doi: 10.1118/1.3643029.

Sengupta, A. *et al.* (2018) 'The first freely available, open source software package for performing 3D image reconstruction for digital breast tomosynthesis', in *Medical Imaging 2018: Physics of Medical Imaging*. SPIE, Bellingham, WA. doi: https://doi.org/10.1117/12.2293146.

Sengupta, A., Sharma, D. and Badano, A. (2021) 'Computational model of tumor growth for in silico trials', in *Volume 11595, Medical Imaging 2021: Physics of Medical Imaging*. SPIE, Bellingham, WA, p. 115954S.

Shaheen, E. *et al.* (2010) *Realistic simulation of microcalcifications in breast tomosynthesis*, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/978-3-642-13666-5_32.

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.

Shaheen, E., Marshall, N. and Bosmans, H. (2011) 'Investigation of the effect of tube motion in breast tomosynthesis: continuous or step and shoot?', *Proceedings of SPIE 7961; Medical Imaging 2011: Physics of Medical Imaging*, 7961(November 2015), p. 79611E. doi: 10.1117/12.877348.

Shen, T. *et al.* (2021) 'Mass Image Synthesis in Mammogram with Contextual Information Based on GANs', *Computer Methods and Programs in Biomedicine*. Elsevier B.V., 202, p. 106019. doi: 10.1016/j.cmpb.2021.106019.

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.

Siddon, R. L. (1985) 'Fast calculation of the exact radiological path for a three-dimensional CT array', *Med Phys*, 12(2), pp. 252–255.

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.* (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.

Spangler, M. L. *et al.* (2011) 'Detection and Classification of Calcifications on Digital Breast Tomosynthesis and 2D Digital Mammography: A Comparison', (February), pp. 320–324. doi: 10.2214/AJR.10.4656.

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.

Tagliafico, A. *et al.* (2014) 'Characterisation of microcalcification clusters on 2D digital mammography (FFDM) and digital breast tomosynthesis (DBT): does DBT underestimate microcalcification clusters? Results of a multicentre study', *European Radiology*, 25(1), pp. 9–14. doi: 10.1007/s00330-014-3402-8.

Tomic, H. *et al.* (2021) 'Assessment of a tumour growth model for virtual clinical trials of breast cancer screening', in *Proceedings of SPIE 11595, Medical Imaging: Physics of Medical Imaging*. SPIE, Bellingham, WA, p. 115954Q.

Vancoillie, L. *et al.* (2019) 'Verification of the accuracy of a partial breast imaging simulation framework for virtual clinical trial applications', in *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. doi:

10.1117/12.2513406.

Vancoillie, L. *et al.* (2020) 'Verification of the accuracy of a hybrid breast imaging simulation framework for virtual clinical trial applications', *Journal of Medical Imaging*, 7(4), pp. 042804-1–17. doi: 10.1117/1.jmi.7.4.042804.

Vancoillie, L. (2022) *Technical and Task-Based Performance Evaluation of Synthetic Mammograms*. KU Leuven.

Warren, L. M. *et al.* (2012) 'Effect of image quality on calcification detection in digital mammography', 39(June), 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. *et al.* (2014) 'The effect of image processing on the detection of cancers in digital mammography', *American Journal of Roentgenology*, 203(2), pp. 387–393. doi: 10.2214/AJR.13.11812.

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.

Youk, J. H. *et al.* (2016) 'Automated volumetric breast density measurements in the era of the BI-RADS fifth edition: A comparison with visual assessment', *American Journal of Roentgenology*, 206(5), pp. 1056–1062. doi: 10.2214/AJR.15.15472.

Zanca, F. *et al.* (2009) 'Evaluation of clinical image processing algorithms used in digital mammography', *Medical Physics*, 36(3), p. 765. doi: 10.1118/1.3077121.

Zeng, R. *et al.* (2015) 'Evaluating the sensitivity of the optimization of acquisition geometry to the choice of reconstruction algorithm in digital breast tomosynthesis through a simulation study', *Physics in Medicine and Biology*. IOP Publishing, 60(3), pp. 1259–1288. doi: 10.1088/0031-9155/60/3/1259.

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.

Zeng, R., Badano, A. and Myers, K. J. (2017) 'Optimization of digital breast tomosynthesis (DBT) acquisition parameters for human observers: Effect of reconstruction algorithms', *Physics in Medicine and Biology*. IOP Publishing, 62(7), pp. 2598–2611. doi: 10.1088/1361-6560/aa5ddc.

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.

Zuckerman, S. P. *et al.* (2020) 'Multicenter evaluation of breast cancer screening with digital breast tomosynthesis in combination with synthetic versus digital mammography', *Radiology*, 297(3), pp. 545–553.

doi: 10.1148/radiol.2020200240.