

JULY 23 2025

## Digital twins enable full-reference quality assessment of photoacoustic image reconstructions

Janek Gröhl ; Leonid Kunyansky ; Jenni Poimala ; Thomas R. Else ; Francesca Di Cecio; Sarah E. Bohndiek ; Ben T. Cox ; Andreas Hauptmann 



*J. Acoust. Soc. Am.* 158, 590–601 (2025)

<https://doi.org/10.1121/10.0037188>



View  
Online



Export  
Citation

### Articles You May Be Interested In

Measurement and simulation of steered acoustic fields generated by a multielement array for therapeutic ultrasound

*JASA Express Lett.* (January 2021)

Modelling power-law ultrasound absorption using a time-fractional, static memory, Fourier pseudo-spectral method

*J. Acoust. Soc. Am.* (March 2025)

# Digital twins enable full-reference quality assessment of photoacoustic image reconstructions

Janek Gröhl,<sup>1,2,3,a)</sup>  Leonid Kunyansky,<sup>4</sup>  Jenni Poimala,<sup>5</sup>  Thomas R. Else,<sup>6</sup>  Francesca Di Cecio,<sup>1,2</sup>  
Sarah E. Bohndiek,<sup>1,2</sup>  Ben T. Cox,<sup>7</sup>  and Andreas Hauptmann<sup>8,9,b)</sup> 

<sup>1</sup>Department of Physics, University of Cambridge, Cambridge, United Kingdom

<sup>2</sup>Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom

<sup>3</sup>ENI-G, a Joint Initiative of the University Medical Center Göttingen and the Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany

<sup>4</sup>Department of Mathematics, University of Arizona, Tucson, Arizona 85721, USA

<sup>5</sup>Department of Technical Physics, University of Eastern Finland, Kuopio, Finland

<sup>6</sup>Department of Bioengineering, Imperial College London, London, United Kingdom

<sup>7</sup>Department of Medical Physics and Biomedical Engineering, University College London, London, United Kingdom

<sup>8</sup>Research Unit of Mathematical Sciences, University of Oulu, Oulu, Finland

<sup>9</sup>Department of Computer Science, University College London, London, United Kingdom

## ABSTRACT:

Quantitative comparison of the quality of photoacoustic image reconstruction algorithms remains a major challenge. No-reference image quality measures are often inadequate, but full-reference measures require access to an ideal reference image. While the ground truth is known in simulations, it is unknown *in vivo* or in phantom studies, as the reference depends on both the phantom properties and the imaging system. This paper tackles this problem by using numerical digital twins of tissue-mimicking phantoms and the imaging system to perform a quantitative calibration to reduce the *simulation gap*. The contributions of this paper are twofold: First, this digital-twin framework is used to compare multiple state-of-the-art reconstruction algorithms. Second, among these is a Fourier transform-based reconstruction algorithm for circular detection geometries, which is tested on experimental data for the first time. The results demonstrate the usefulness of digital phantom twins by enabling assessment of the accuracy of the numerical forward model and enabling comparison of image reconstruction schemes with full-reference image quality assessment. This paper shows that the Fourier transform-based algorithm yields results comparable to those of iterative time reversal, but at a lower computational cost.

© 2025 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0037188>

(Received 12 March 2025; revised 30 May 2025; accepted 24 June 2025; published online 23 July 2025)

[Editor: Juan Tu]

Pages: 590–601

## I. INTRODUCTION

Photoacoustic (PA) imaging (PAI) is a medical imaging modality that promises advances in multiple clinical applications, such as diagnosis and staging of Crohn's disease (Knieling *et al.*, 2017), neuromuscular degenerative disease (Regensburger *et al.*, 2019), breast cancer (Manohar and Dantuma, 2019), and multiple others (Assi *et al.*, 2023; Lin and Wang, 2022). The potential of PAI in this context comes from its ability to resolve molecular imaging contrast based on the optical absorption coefficient of the tissue, while simultaneously enabling multi-scale imaging at depths of up to several centimetres in tissue (Beard, 2011).

From an engineering standpoint, PAI is a coupled-physics modality, combining the high resolution of ultrasound

techniques with the high sensitivity of electromagnetic waves to the optical properties of biological tissues. The region of interest (e.g., a woman's breast in mammography) is irradiated with a short laser pulse. Light is partially absorbed by tissues, which raises the temperature of the medium. By a process referred to as thermoelastic expansion, an acoustic wave emerges that is measured by transducers on the object boundary. The propagation of the pressure wave  $p(t, x)$  can be modelled by the wave equation

$$\begin{cases} p_{tt} = c^2 \Delta p, & t \geq 0, \quad x \in \mathbb{R}^D, \\ p(0, x) = p_0(x), \quad p_t(0, x) = 0, \\ g(t, y_j) = p(t, y_j), \quad j = 1, \dots, N_d, \end{cases} \quad (1)$$

where  $p_0(x)$  is the initial pressure in the tissues,  $c$  is the speed of sound,  $y_j$  are the locations of  $N_d$  point-like transducers,  $g(t, y_j)$  are the measurements made by these transducers, and

<sup>a)</sup>Email: j.groehl@eni-g.de

<sup>b)</sup>Email: Andreas.Hauptmann@oulu.fi, a.hauptmann@ucl.ac.uk

$D$  is the dimension of the space (in this case,  $D = 3$ ). This simplified model assumes that light absorption happens instantaneously and that the acoustic wave propagates in the open space, i.e., without reflecting from transducers or other parts of the acquisition scheme. It also neglects the absorption and dispersion of acoustic waves in tissues and the frequency and directional responses of real transducers. The *acoustic inverse problem* of PAI consists of reconstructing the initial pressure  $p_0$  from the measurements  $g(t)$  at the tissue boundary. The *optical inverse problem* then describes reconstructing the optical absorption coefficient  $\mu_a$  from  $p_0$ . Accurate acoustic inversion is thus a fundamental step towards reproducible and quantitative PA data analysis. It is one of the most researched topics in the field of PAI (Poudel *et al.*, 2019), and many methods to solve the acoustic inverse problem have been proposed: See, for example, Xu and Wang (2005), Perrot *et al.* (2021), Xu and Wang (2002), Tarvainen and Cox (2023), Treeby *et al.* (2010), and Arridge *et al.* (2016).

From a mathematical standpoint, in order to reconstruct  $p_0(x)$ , which is a function of a three-dimensional (3D) variable, measurements of pressure  $p(t, x)$  should be performed on a two-dimensional (2D) surface, at least partially surrounding the source of the acoustic wave (Kuchment and Kunyansky, 2015). Most PAI image reconstruction algorithms presented in the literature are based on this assumption. However, in modern biomedical practice most often one has to deal with the situation where the data are collected by a sparse arrangement of detectors and the illuminated region is only a part of the whole object. Such is the case with our experimental setup, which uses the MSOT inVision-256TF pre-clinical PAI system (iThera Medical GmbH, Munich, Germany), which has an angular coverage of only  $270^\circ$  (see the detailed description in Sec. II A 2). Accurate image reconstruction of a 3D image from 2D data is not possible: Thus, all existing algorithms can produce only approximations of  $p_0(x)$ . While all exact reconstructions are alike, each approximate reconstruction is approximate in its own way, thus introducing unique reconstruction artefacts (Rietberg *et al.*, 2025). One of the goals of this paper is to compare such approximations using both simulated and real data.

One particularly promising acoustic inversion algorithm is a fast Fourier transform (FFT)-based reconstruction algorithm that is optimised specifically for circular detection geometries and was first presented by Kunyansky in 2012 (Kunyansky, 2012). [Note this is different from the well-established FFT reconstruction scheme for planar geometries (Jaeger *et al.*, 2007).] All computationally non-trivial steps of this algorithm are done using one-dimensional (1D) or 2D FFTs, which makes this algorithm very fast. Furthermore, it is straightforward to implement within frameworks like PyTorch and so is available to use in learned reconstruction frameworks, such as, e.g., Hauptmann *et al.* (2018), Boink *et al.* (2020), and Hsu *et al.* (2023). The objective of the work presented in this paper is to investigate how the algorithm compares to other state-of-the-art approaches, particularly time-reversal (TR)

(Treeby *et al.*, 2010) or model-based (MB) image reconstruction (Rosenthal *et al.*, 2010). We present a concise summary of the algorithm as part of this study and make it available open source. Additionally, for the first time the FFT-based algorithm for circular geometries is tested on experimental data. This is particularly relevant, since the model assumptions of ideal 2D wave propagation are violated in the experimental system.

The quantitative evaluation and comparison with other reconstruction algorithms are challenging on experimentally acquired data, however, as the initial pressure distribution that represents the ground truth is usually unknown (Gröhl *et al.*, 2021). This restricts researchers to simulation studies or no-reference image quality assessment (IQA) measures. While common no-reference IQA measures, such as the full-width at half maximum (FWHM) or the signal-to-noise-ratio (SNR), are well-established to estimate general-purpose image quality indicators in natural images, they only have limited applicability to medical images (Breger *et al.*, 2024) and are thus not necessarily indicative of the accuracy of the acoustic inversion.

To use the more informative full-reference IQA measures, a reference image of  $p_0$  is required to provide a reliable performance measure for the reconstruction quality. In pure simulation studies,  $p_0$  is known and can be used as a ground truth, but in experimental settings, this is typically infeasible. State-of-the-art approaches normally use acoustic simulations of binary acoustic pressure maps for the comparison of reconstruction algorithms (Warbal and Saha, 2022). For experimental settings, point sources are often considered (Paltauf *et al.*, 2007), which can be evaluated without knowledge of  $p_0$  by characterising the resolution (Van de Sompel *et al.*, 2016). In almost all cases, however, the underlying optical properties are unknown and the light fluence is only approximated.

This is even true in phantom studies—however carefully the phantom is fabricated and characterised— $p_0$  cannot be known directly as it depends not just on the phantom itself but on the illumination pattern used by the imaging system. If we had a perfect set of measurement data (i.e., complete data) of the phantom, then an exact algorithm could be used to recover the ground truth  $p_0$  as a reference to compare against. However, we do not have that data in this case and in most other measurement scenarios. Therefore, to know what the ground truth  $p_0$  is, it is necessary to model the phantom and the imaging system (cf. Fig. 1).

As such, there is an unmet need to test both the applicability of full-reference IQA algorithms for PAI, as well as the comparison of image reconstruction algorithm performance with incomplete data in experimental settings. In this work, we propose a framework based on digital twins of test objects to tackle these problems. We simulate numerical representations of well-characterised imaging phantoms. This allows the comparison of reconstruction results and numerical simulations of  $p_0$  as a means to quantify the remaining *simulation gap* after calibration. Because of the

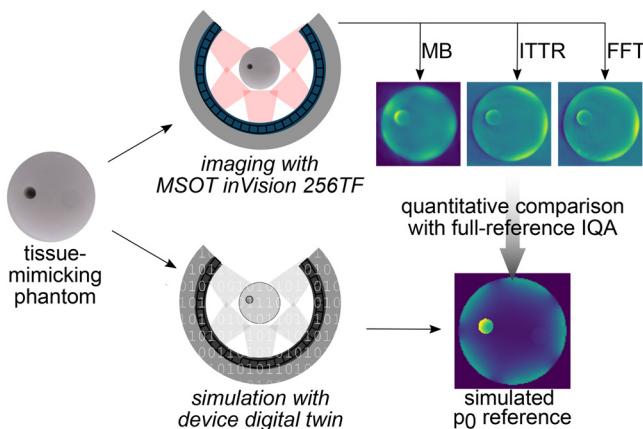


FIG. 1. Overview of the proposed evaluation strategy for full-reference IQA of reconstruction algorithms for photoacoustic images. A tissue-mimicking phantom is measured by the imaging device, and a reference  $p_0(x)$  is simulated using digital twins of the phantom and the device. Reconstructions of different algorithms can then quantitatively be compared against the simulated reference.

availability of paired data in this digital twin framework, we can directly compare simulated acoustic pressure measurement data with actual acoustic pressure measurements and perform a quantitative calibration of the simulation pipeline. The quantitative measures can then give an essential indication of how accurate the simulated  $p_0$  is as the reference—a step that is missing in the experimental comparison of reconstruction algorithms across the state of the art (Choi *et al.*, 2020). This allows for the derivation of a quantifiable measure to judge the reliability of full-reference IQA measures to determine the reconstruction quality of the algorithms.

We use these digital twins as an evaluation framework to compare the performance of the FFT-based reconstruction algorithm to other state-of-the-art image reconstruction schemes (Fig. 1) and show that its performance is on par while offering computational advantages.

## II. METHODS

### A. Data

#### 1. Phantom data

We used  $N=30$  cylindrical phantoms with a diameter of 27.5 mm, a height of 65–80 mm, and an approximate volume of 40–50 ml. They were fabricated based on a previously published protocol (Hacker *et al.*, 2021) and are a subset of the phantoms used in a prior publication (Gröhl *et al.*, 2024). Each phantom is piecewise constant and consists of a background cylinder into which imaging targets are added. Rectangular samples with a length of 5.9 cm, a width of 1.8 cm, and a thickness ranging between 2 mm and 4 mm were prepared from each material for optical characterisation. Sample thicknesses were determined at five distinct locations using digital vernier callipers.

### 2. PAI

All data were acquired using the MSOT inVision-256TF pre-clinical PAI system (iThera Medical GmbH, Munich, Germany). It has 256 transducer elements with a 5-MHz centre frequency and 60% bandwidth, arranged in a circular array of radius 40.5 mm and angular coverage of 270°. More details on the measurement setup can be found in a prior publication (Joseph *et al.*, 2017). We reconstructed the images using various reconstruction schemes, with their details outlined later. The images were reconstructed into a 300 × 300 pixel grid with a field of view of 32 × 32 mm.

### 3. Digital twin simulations

Our phantoms were created in such a way that they had a piecewise-constant material distribution. This allowed us to take samples of each material and characterise them with an in-house double integrating sphere (DIS) system (Hacker *et al.*, 2021) based on the system developed by Pickering *et al.* (1993) and determine the optical parameters, absorption  $\mu_a$  and reduced scattering  $\mu'_s$ , in a wavelength range of 600–950 nm.

We implemented numerical phantoms by assigning the optical measurements and acoustic reference values (Hacker *et al.*, 2021; Gröhl *et al.*, 2024) to a manually created segmentation mask that delineates the different material regions of the real phantoms. Specifically, we simulate homogeneous phantoms with a density of 1000 g/cm<sup>3</sup> and a sound speed of 1468 m/s. In the coupling medium, we set a sound speed of 1489 m/s and a density of 1000 g/cm<sup>3</sup>. We assume negligible acoustic attenuation. We created multi-label segmentation masks using the MEDICAL IMAGING INTERACTION TOOLKIT (MITK) (Nolden *et al.*, 2013), based on delay-and-sum (DAS) reconstructions of the measured time series data.

We simulated device-specific initial pressure distributions  $p_0(x)$  and measurement data  $g(t,y)$ , based on the numerical phantoms. We first used a Monte Carlo model of light transport [MCX (Fang and Boas, 2009)] to simulate the light fluence  $\phi$  and calculated the expected initial pressure distribution using  $p_0(x) = \Gamma \cdot \mu_a(x) \cdot \phi(x)$ , assuming a constant Grüneisen parameter  $\Gamma$ . We then use the  $k$ -space pseudospectral method implemented in k-WAVE (Treeby and Cox, 2010) as the 3D acoustic forward model to generate measurement data  $g(t,y)$  (see Fig. 2).

We used a digital twin of the MSOT InVision-256TF that we implemented in the SIMPA toolkit (Gröhl *et al.*, 2022). The computational model of the device was built using hardware geometry details provided by the vendor. We implemented a custom device class in MCX and added a custom kWaveArray definition of the detection elements of the MSOT InVision, which interpolates the toroidal surface of each transducer element with 707 points. We optimised the illumination and detector design parameters to best match our experiments based on  $N=15$  calibration phantoms. We found good agreement between vendor specifications and the manually identified parameters and further found that a Gaussian illumination profile led to a good

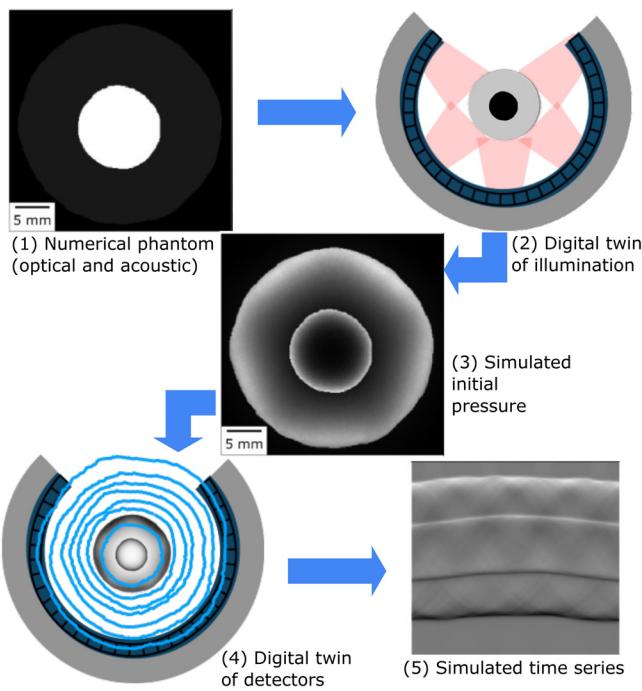


FIG. 2. Overview of the digital twin simulation pipeline. (1) A numerical phantom with optical and acoustic properties matching a corresponding real phantom is created. (2) With a digital twin of the illumination geometry, the light fluence is computed using the Monte Carlo method. (3) The computed fluence is multiplied with the absorption coefficients to obtain the initial pressure distribution. (4) With a digital twin of the detection geometry, sound propagation is simulated using the  $k$ -space pseudospectral method, which leads to simulated measurements that correspond to experimental measurements.

match between simulations and the experimental radiant exposure. The simulations were run using the open-source SIMPA toolkit (Gröhl *et al.*, 2022), which essentially acts as an orchestration layer and provides adapters to both MCX and K-WAVE.

We calibrated the proposed digital twin pipeline against experimental data using three components: a linear scaling of the simulated signal amplitudes, the introduction of a noise model based on an experimental measurement of the device noise, and the consideration of the impulse response of the imaging system. We optimised the following linear parameters,  $a$ ,  $b$ , and  $c$ , based on  $N=15$  calibration phantoms using a least-squares optimisation scheme,

$$g(t, y)_{\text{exp}} = a + (b \cdot g(t, y)_{\text{sim}}) * \text{IRF}(t, y) \\ + c \cdot \text{noise}(t, y),$$

where  $g(t, y)_{\text{exp}}$  are the experimental measurements,  $g(t, y)_{\text{sim}}$  are the outputs of the SIMPA simulation pipeline, a convolution with  $\text{IRF}(t, y)$ , the impulse response function of the imaging system that the manufacturer provided, and  $\text{noise}(t, y)$  is a noise measurement taken in an empty water bath. We used the digital twin framework to compare the accuracy of this calibrated forward model to a naive scaling of the SIMPA simulation, where the scaling factor was empirically determined to be 10.

## B. Acoustic inversion

In the MSOT inVision system, the illuminated region is concentrated around the plane  $\Pi$  containing the detectors. The detectors cover three fourths of a circle,  $S$ , of radius  $R = 40.5$  mm lying in the plane. Our goal is to reconstruct the values of  $p_0(x)$  for  $x \in \Pi$  and restricted to the interior of circle  $S$ .

Some of the known algorithms model this situation by assuming that the region is infinitely thin and use the 3D wave equation to describe the propagating wave. Others use the 2D model of wave propagation; this corresponds to the assumption that the region is extended in the direction orthogonal to  $\Pi$  and that  $p_0(x)$  is invariant in that direction. In the following, we describe the reconstruction methods we use in this paper.

### 1. Circular FFT-based reconstruction

The Fourier series algorithm we investigate here experimentally is based on the method first proposed in Kunyansky (2011), with additional recent correction based on the technique of Eller *et al.* (2020). The method generates a theoretically exact inversion under the assumption that the direct problem is accurately modelled by a 2D wave equation, and the data are acquired by a full circle of detectors. In the situation when a certain subset of detectors is missing (as is the case with the MSOT scanner we use), the technique of Eller *et al.* (2020) allows us to significantly reduce the arising artefacts. Additionally, the experimental system violates the ideal 2D assumption, and hence, an evaluation on experimental data is crucial to establish the algorithm's usefulness for application.

For simplicity of presentation, we scale the variables as  $\hat{t} = ct/R$  and  $\hat{x} = x/R$  and introduce a function,  $u(\hat{t}, \hat{x}) = p(t, x)$ . The function  $u$  is a solution of the 2D wave equation in  $\hat{t}$  and  $\hat{x}$ , with a unit speed of sound; also,  $u(0, \hat{x})$  is supported in the unit circle, so that  $p_0(x) = u(0, R\hat{x})$ . The algorithm consists of the following steps:

1. Zero-pad data  $g(t, y)$  in  $t$  by a factor of 2 or more, by adding zeros after the actually measured values.
2. Using the FFT, compute the Fourier transform of the zero-padded data  $g(ct/R, y(\theta))$  in  $t$  and the Fourier series in  $\theta$ , where  $y(\theta) = R(\cos \theta, \sin \theta)$ :

$$\hat{g}_k(\rho) = \frac{c}{2\pi R} \int_{\mathbb{R}} \left[ \int_0^{2\pi} g(ct/R, y(\theta)) e^{-ik\theta} d\theta \right] e^{it\rho} dt.$$

The computed values are defined on a computational grid in  $\rho \in [-\rho_{\text{Nyq}}, \rho_{\text{Nyq}}]$ ,  $k \in [-N_d/2, N_d/2]$ , where  $\rho_{\text{Nyq}}$  is the Nyquist frequency of discretization in  $\hat{t}$  and  $N_d$  is the number of the detectors, assuming that they fill the whole circle. (We are only using positive values of  $\rho$ .)

3. For each  $k$  and each positive value of  $\rho$  in the grid, compute coefficients  $b_k(\rho)$  by the formula

$$b_k(\rho) = \frac{4i^{|k|}}{\rho H_{|k|}^{(1)}(\rho)} \hat{g}_k(\rho),$$

- where  $H_m^{(0)}$  is the Hankel's function of order  $m$ .  
 4. For each value of  $\rho$  in the grid, using the FFT, sum the Fourier series, thus obtaining function  $\hat{B}(\rho, \varphi)$ :

$$\hat{B}(\rho, \varphi) = \sum_{k=-N_d/2}^{N_d/2} b_k(\rho) e^{ik\varphi}.$$

Function  $\hat{B}(\rho, \varphi)$  represents a theoretically exact polar grid representation to the Fourier transform  $\hat{u}(0, \xi)$  of the function  $u(0, \hat{x})$  we seek, assuming that  $\xi = |\xi|(\cos \varphi, \sin \varphi)$ ,  $\xi \in (0, \rho_{\max})$ ,  $\varphi \in [0, 2\pi]$ .

5. Interpolate  $\hat{B}(\rho, \varphi)$  from the polar grid to a Cartesian grid in  $\xi$ , producing an approximation to  $\hat{u}(0, \xi)$ . Here, we utilize bilinear interpolation.  
 6. Additional correction: The key observation of [Eller et al. \(2020\)](#) is that, if a segment of detectors is absent, roughly half of the approximation to  $\hat{u}(0, \xi)$  will be severely distorted, while the other half will be reconstructed quite accurately. Since  $u(0, \hat{x})$  is a real function, its Fourier transform has the property  $\hat{u}(0, \xi) = \overline{\hat{u}(0, -\xi)}$ . The additional correction we deploy is to replace the “bad” half of values  $\hat{u}(0, \xi)$  by values of  $\overline{\hat{u}(0, -\xi)}$ . Speaking crudely, this means that every point in the reconstruction image will receive its high spatial frequency information from the detectors it “sees” in the  $180^\circ$  angular range. For points lying further away from the absent transducers, this will mean fewer transducers; for points lying closer this will mean almost all of them. A more clear and rigorous explanation of this correction technique can be found in [Eller et al. \(2020\)](#).  
 7. Using a 2D FFT, we compute  $u(0, \hat{x})$  from  $\hat{u}(0, \xi)$  and reconstruct  $p_0(x) = u(0, R\hat{x})$ .

This FFT-based method, in its current implementation, is asymptotically fast, meaning that it requires  $\mathcal{O}(n^2 \log n)$  floating point operations (flops) for an  $(n \times n)$  image, assuming that the data contains  $\mathcal{O}(n^2)$  values. Other fast algorithms are only known for the cases of linear or planar arrangement of transducers: See, for example, [Jaeger et al. \(2007\)](#). FFT reconstruction algorithms utilizing K-WAVE ([Treeby and Cox, 2010](#)) are not asymptotically fast in the sense that just one time step of such an algorithm in 2D already requires  $\mathcal{O}(n^2 \log n)$  flops. The present FFT-based method is available within the PATATO toolbox ([Else et al., 2024](#)) and also on GitHub ([Kunyansky, 2024](#)).

## 2. Delay-and-sum

In DAS reconstruction, the approximation  $\tilde{p}_0(x)$  to  $p_0(x)$  is calculated by summing all signals based on the time,  $\tau(x, y_j)$ , it takes for the sound to travel from the spatial origin  $x$  to the respective detector element  $y_j$  on  $S$  ([Kirchner et al., 2018](#)),

$$\tilde{p}_0(x) = \sum_{j=1}^{N_d} g(\tau(x, y_j), y_j),$$

where  $y_j$  is the location of the  $j$ -th detector element on  $S$  and  $N_d$  is the number of these elements. We use the DAS implementation of the PATATO toolbox ([Else et al., 2024](#)).

## 3. Filtered backprojection (FBP)

The FBP implementation in PATATO applies the DAS reconstruction scheme to the time-derivative of the detected acoustic signals  $g(t, y)$  after applying a low-pass filter,  $h$ , that partially compensates for the frequency response of the imaging system:

$$p_0(x) = \sum_{j=1}^{N_d} \frac{\partial}{\partial t} g(\tau(x, y_j), y_j) * h(t).$$

One should note that the universal backprojection as defined by [Xu and Wang \(2005\)](#) contains a detector-dependent image weighting factor that was not considered in this implementation.

## 4. Model-based reconstruction

Instead of using a simplified delay operator mapping  $g(t, y_j)$  into an approximation to  $p_0(x)$ , MB reconstruction techniques explicitly incorporate a solution to the wave equation. We use an MB algorithm provided in the PATATO toolbox ([Else et al., 2024](#)), which implements the approach by [Rosenthal et al. \(2010\)](#).

[Rosenthal et al. \(2010\)](#) approximate the solution to the 3D wave Eq. (1) by implicitly assuming that the initial condition  $p_0(x)$  is concentrated in an infinitely thin region lying in the plane  $\Pi$ . This leads to the following expression for the measurements  $g(t, y_j), j = 1, \dots, N_d$ ,

$$g(t, y_j) \approx \frac{1}{c} \frac{\partial}{\partial t} \int_{\mathbb{R}^2} \frac{p_0^{\text{interp}}(x) \delta(|y_j - x| - t)}{4\pi |y_j - x|} dx,$$

where  $p_0^{\text{interp}}(x)$  is a function of a 2D variable representing an approximation to  $p_0(x)$  (with  $x \in \mathbb{R}^2$ ). Function  $p_0^{\text{interp}}(x)$  is defined by interpolating the values of  $p_0(x_k)$  given at the nodes  $x_k$  of the discretization grid. This leads to a system of linear equations relating measurements  $g(t, y_j)$  to the values of  $p_0(x_k)$  at the nodes, which is solved with any standard approach, such as, for example, least squares.

## 5. Time reversal

Our TR reconstruction is based on a 2D approximation to the forward problem. It is represented by Eq. (1), with  $D = 2$ . Such a model implicitly assumes that the initial pressure  $p_0(x)$  is invariant in the direction normal to plane  $\Pi$ , and the detectors measure values of the pressure at the points sampling the whole circle  $S$ . By restricting the solution  $p(t, x)$  of the forward problem to the interior  $B$  of circle  $S$ , one can see that  $p$  is a solution of the initial/boundary

value problem in the time-space cylinder  $Z = (0, T) \times B$ , with initial conditions  $p(0, x) = p_0, p_t(0, x) = 0$  in  $B$ , and boundary values  $p(t, y) = g(t, y)$  at all  $(t, y) \in (0, T) \times S$ . One also observes that, in practice, for a sufficiently large  $T$ , values of  $p(T, x)$  and  $p_t(T, x)$  are quite small and can be approximated by 0. The idea of the TR (Agranovsky and Kuchment, 2007; Kunyansky, 2007; Hristova *et al.*, 2008) is to solve the wave equation backwards in time, in the time-space cylinder  $(t, x) \in (0, T) \times B$ , with the boundary conditions  $g$  and zero initial conditions at  $t = T$ . In other words, one would like to find the solution  $u(t, x)$  to the following problem,

$$\begin{cases} u_{tt} = c^2 \Delta u, & (t, x) \in (0, T) \times B, \\ u(T, x) = 0, & u_t(T, x) = 0, \\ u(t, y) = g(t, y), & (t, y) \in (0, T), \end{cases} \quad (2)$$

with the approximation  $\tilde{p}_0$  to  $p_0$  computed as

$$\tilde{p}_0(x) = u(0, x), \quad x \in B.$$

Ideally, if  $g(t, y)$  is known exactly on  $(0, T) \times S$ , in the limit  $T \rightarrow \infty$ , the approximation  $\tilde{p}_0$  becomes exact, i.e.,  $p_0(x) = u(0, x)$ . When a part of the circle  $S$  does not contain the detectors, the corresponding values are replaced by 0 and  $\tilde{p}_0$  represents a crude approximation to  $\tilde{p}_0$ . Using the solution of problem (2), one defines an operator,  $A^\triangleleft$ , mapping boundary values  $g$  into  $\tilde{p}_0$ :

$$A^\triangleleft : g \rightarrow \tilde{p}_0.$$

So, our TR technique consists of computing  $\tilde{p}_0 = A^\triangleleft g$ , by solving numerically problem (2). For the present study, we utilize the implementation of TR described in the Appendix of Arridge *et al.* (2016), which, in turn, is based on the use of the 2D version of the k-WAVE toolbox (Treeby and Cox, 2010).

## 6. Iterative time reversal (ITTR)

The ITTR method (Stefanov and Uhlmann, 2009) uses TR on each iteration step, to refine the current approximation to  $p_0$ . The next approximation,  $\tilde{p}_0^{(k+1)}(x)$ , is defined through the previous approximation  $\tilde{p}_0^{(k)}(x)$  by the formula

$$\tilde{p}_0^{(k+1)} = \tilde{p}_0^{(k)} + A^\triangleleft(g - A\tilde{p}_0^{(k)}), \quad k = 0, 1, 2, 3, \dots, \quad (3)$$

with  $\tilde{p}_0^{(0)} = 0$  and where  $A$  is the operator that solves problem (1).

In this case,  $\tilde{p}_0^{(1)} = A^\triangleleft g$  is just the approximation obtained by the TR algorithm. As in TR, the application of the operator  $A^\triangleleft$  is equivalent to solving problem (2) with the boundary condition  $g - A\tilde{p}_0^{(k)}$ . We utilize the k-WAVE toolbox to implement each step of this algorithm.

## C. Image quality assessment measures

We use five full-reference IQA measures and one no-reference measure to compare the reconstruction performance

of the different reconstruction methods. We chose the IQA measures to capture different properties and relationships between the images. Each is described in the following, where  $N$  is the number of pixels in the image:

### 1. The Pearson correlation coefficient ( $R$ )

$R$  quantifies the strength and direction of the linear relationship between two variables. It ranges from  $-1$  to  $1$ , where  $1$  indicates a perfect positive linear relationship,  $-1$  indicates a perfect negative linear relationship, and  $0$  indicates no linear relationship. We use the coefficient to quantify the strength of the linear correlation between the reconstructed signal  $\tilde{p}_0$  and the simulated initial pressure distribution  $p_0$  (Benesty *et al.*, 2009):

$$R = \frac{\sum_{i=1}^N (\tilde{p}_{0,i} - \bar{\tilde{p}}_0)(p_{0,i} - \bar{p}_0)}{\sqrt{\sum_{i=1}^N (\tilde{p}_{0,i} - \bar{\tilde{p}}_0)^2} \sqrt{\sum_{i=1}^N (p_{0,i} - \bar{p}_0)^2}}.$$

### 2. Mean absolute error (MAE)

The MAE quantifies the average magnitude of errors between predicted and actual values, without considering their direction. It is calculated by taking the average of the absolute differences between  $\tilde{p}_0(x)$  and  $p_0(x)$ , where  $x$  represents the pixel position:

$$\text{MAE} = \frac{1}{N} \sum_{x=1}^N |\tilde{p}_0(x) - p_0(x)|.$$

### 3. Structural similarity index measure (SSIM)

The SSIM evaluates the similarity between two images by jointly comparing their structural information, luminance, and contrast. The SSIM between  $\tilde{p}_0$  and  $p_0$  can be calculated using the mean, variance, and covariance of the images (Wang *et al.*, 2004),

$$\text{SSIM} = \frac{(2 \cdot \bar{\tilde{p}}_0 \cdot \bar{p}_0 + C_1)(2\sigma_{\tilde{p}_0 p_0} + C_2)}{(\bar{\tilde{p}}_0^2 + \bar{p}_0^2 + C_1)(\sigma_{\tilde{p}_0}^2 + \sigma_{p_0}^2 + C_2)},$$

where  $\bar{x}$  denotes the mean,  $\sigma^2$  denotes the variance, and  $\sigma_{\tilde{p}_0 p_0}$  the covariance of  $\tilde{p}_0$  and  $p_0$ .  $C_1$  and  $C_2$  are stabilising constants. We use the TorchMetrics default parameters of  $C1 = 0.01$  and  $C2 = 0.03$  (Detlefsen *et al.*, 2022).

### 4. Jensen–Shannon divergence (JSD)

The JSD quantifies the similarity between two probability distributions. It is a symmetric and smoothed version of the Kullback–Leibler divergence and provides a bounded value between 0 (when the distributions are identical) and 1 (when the distributions are maximally different) (Menéndez *et al.*, 1997). We perform  $z$ -score normalization for all pixel

values in  $\tilde{p}_0$  and  $p_0$  and create histograms P and Q, with  $N=100$  bins ranging from  $-3\sigma$  to  $3\sigma$  to extract the probability distributions. We use a PYTHON implementation of the JSD, available in the SCIPY (v1.10.1) package (Virtanen et al., 2020),

$$\text{JSD}(P \parallel Q) = \frac{1}{2} \text{KL}(P \parallel M) + \frac{1}{2} \text{KL}(Q \parallel M),$$

where

$$M = \frac{1}{2}(P + Q)$$

and

$$\text{KL}(P \parallel M) = \sum_{i=1}^N P(i) \log \frac{P(i)}{M(i)},$$

$$\text{KL}(Q \parallel M) = \sum_{i=1}^N Q(i) \log \frac{Q(i)}{M(i)}.$$

### 5. Haar wavelet-based perceptual similarity index (HaarPSI)

HaarPSI assesses the visual similarity between two images in the interval [0, 1]. It leverages the Haar wavelet transform to capture the local phase coherence and contrast of the images. HaarPSI between the reconstructed signal  $\tilde{p}_0$  and the simulated initial pressure distribution  $p_0$  is calculated based on the responses of the Haar wavelet transform applied to both images (Reisenhofer et al., 2018). We use the PYTHON implementation of the method available on GitHub (Research Group Computational Data Analysis, University Bremen, 2018),

$$\text{HaarPSI} = l_\alpha^{-1} \left( \frac{\sum_x \sum_{k=1}^2 \text{HS}_{p_0, \tilde{p}_0}^{(k)}[x] \cdot W_{p_0, \tilde{p}_0}^{(k)}[x]}{\sum_x \sum_{k=1}^2 W_{p_0, \tilde{p}_0}^{(k)}[x]} \right)^2,$$

where  $\alpha > 0$  is a free parameter (here  $\alpha = 4.2$ ),  $l_\alpha(x) = 1/(1 + e^{-\alpha x})$ ,  $W$  is a weight map derived from the response of a single low-frequency Haar wavelet filter  $k$ , and HS is the local similarity measure, which is based on the first two stages of a 2D discrete Haar wavelet transform.

### 6. Full width at half maximum

The FWHM is a no-reference IQA measure and quantifies the width of signal peaks. It is defined as the distance between the points on a signal curve at which the signal amplitude falls to half of its maximum value (Castleman, 1996). We use it as a measure to quantify the sharpness of the reconstruction by computing the FWHM of the absolute gradient along a manually defined line profile through the object boundaries in  $\tilde{p}_0$ . We compute the FWHM using the following steps:

1. Extract a line profile through the image that includes both background and target boundaries.
2. Compute the absolute value of the gradient of the line profile.
3. Identify the peaks in the absolute gradient.
4. For each peak, locate the first positions to the left and right where the absolute gradient drops to half of the peak value.
5. Compute the width by subtracting the left position from the right position.

### D. Computational footprint estimation

We evaluated the computational footprint, including both time and memory usage, for the specific algorithm implementations. To measure execution time, we used timing functions in MATLAB and PYTHON. For the MATLAB script, we recorded the peak memory consumption of the graphics processing unit (GPU), as reported by k-WAVE. In PYTHON, we used the *tracemalloc* module around the execution block to track the peak memory consumption of the central processing unit (CPU) and the *nvidia-smi* tool to monitor peak GPU memory consumption. Each algorithm was tested by reconstructing a single data point three times. We computed the total script execution time of each run, which includes any data loading operations as well as the initialisation of any PYTHON or MATLAB modules. We run these experiments on an Intel® Core™ i9-10900KF CPU (Intel, Santa Clara, CA) @ 3.70 GHz with 64 GB random access memory (RAM) and an Nvidia RTX3090 graphics card (Nvidia, Santa Clara, CA).

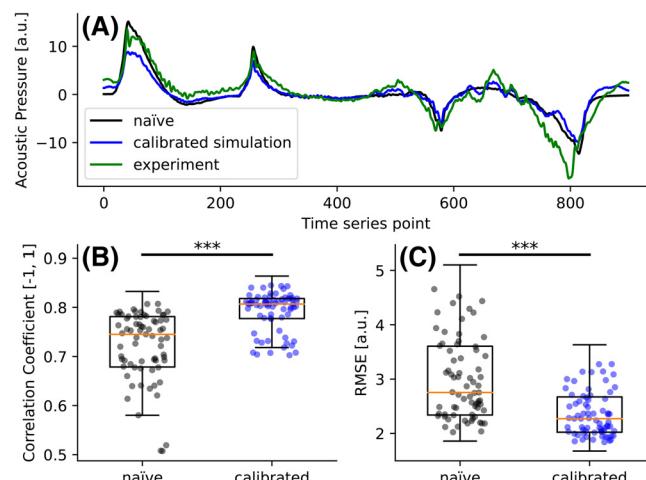


FIG. 3. Our calibration scheme achieves higher agreement with the experiment compared to naive scaling of the simulation. We first show a qualitative comparison of the measurement data compared to naive and calibrated simulations of a representative measurement of a single detector of a randomly chosen test phantom (A). We then use the corresponding measurement pairs between simulation and experiment to show that we can evaluate the accuracy of the forward model, for example, in terms of the Pearson correlation coefficient (B) or the root mean squared error (C). Both of these evaluations show that a careful calibration scheme that includes noise modelling and modelling of the impulse response significantly outperforms a naive scaling of the k-WAVE simulations. \*\*\*,  $p < 0.0001$  using a Wilcoxon signed-rank test.

### III. RESULTS

#### A. With the digital twin pipeline, the accuracy of the forward model can be evaluated and compared

To investigate the accuracy of our forward model, we compared the simulated measurement data  $g(t, y)_{\text{sim}}$  with the experimental measurements  $g(t, y)_{\text{exp}}$  in the time domain. Visual inspection of the data shows generally good agreement between the two [Fig. 3(A)], where one can observe peaks and valleys of the pressure waves in the same locations and with similar amplitudes. Also, the noise patterns have a high degree of similarity.

In addition to qualitative evaluation, we can use the  $g(t, y)_{\text{sim}}$  and  $g(t, y)_{\text{exp}}$  data pairs to quantify the difference and to compare the goodness of fit of different forward models. To this end, we compared two calibration approaches: (1) a naive scaling of the SIMPA simulated  $g(t, y)_{\text{sim}}$ , empirically determined as  $g(t, y)_{\text{naive}} = g(t, y)_{\text{sim}}/10$ , and (2) an optimised calibration:

$$g(t, y)_{\text{cal}} = 4.5 + (0.068 \cdot g(t, y)_{\text{sim}}) * \text{IRF}(t, y) + 0.89 \cdot \text{noise}(t, y).$$

With this calibration, we could improve the Pearson correlation coefficient from 0.75 to 0.81 [Fig. 3(B)] and reduce the root mean squared error from 2.75 to 2.27 a.u. [Fig. 3(C)].

#### B. Qualitative and quantitative performance comparison reveals similar reconstruction quality of the FFT-based method

We first qualitatively show the reconstruction results of a representative test phantom (cf. Fig. 4). We show the results of the idealised simulations [panels (A), (C), and (E)] side-by-side with the experiments [panels (B), (D), and (F)]. The results show that reconstruction artefacts are more pronounced in the simulations across all methods [see especially Figs. 4(A), 4(C), and 4(E)] and that there is a significant level of noise in the experimental images. Furthermore, the qualitative comparison suggests that the circular FFT-based method produces results that are similar to the computationally expensive ITTR method.

For quantitative comparison, we compute full-reference IQA measures that relate the reconstructed image to the *in silico* initial pressure distribution, which was obtained from numerical simulation of the digital twins. In addition, we compute the FWHM as a no-reference IQA measure that is indicative of the sharpness of the reconstructions. Our results show that TR and ITTR perform nearly equivalently and best across nearly all measures (cf. Table I). The only notable exceptions are HaarPSI, which favours MB reconstruction and FBP, and the FWHM, which indicates that the FFT-based method can reconstruct the sharpest object boundaries—an observation that is corroborated by the qualitative results.

Notably, the FFT-based method only performs slightly worse compared to the TR and ITTR methods and significantly outperforms MB reconstruction and FBP. DAS

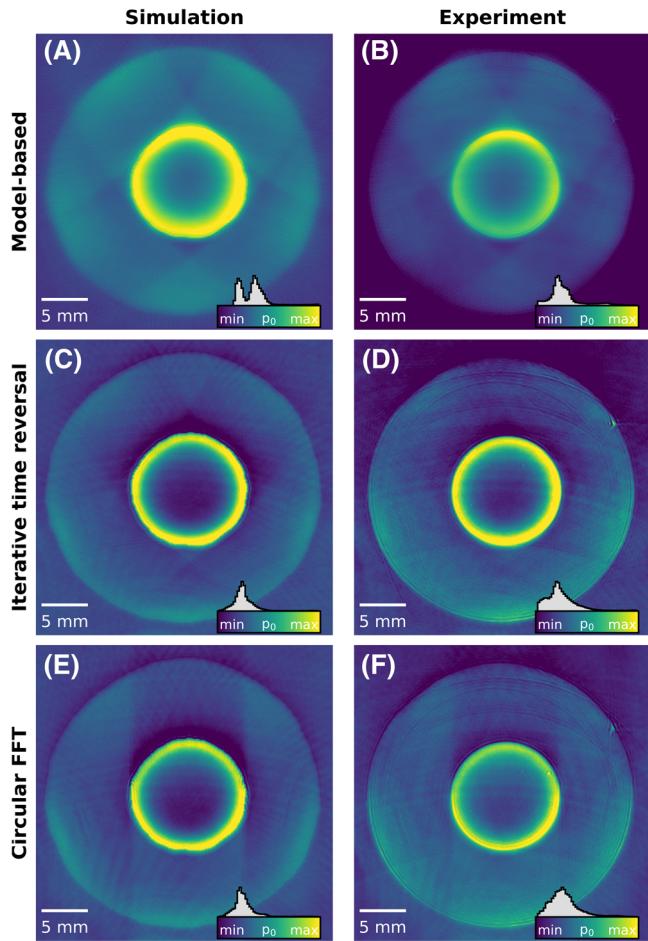


FIG. 4. Qualitative visualisation of an example phantom with a circular imaging target for simulated and experimentally acquired data. The figure shows reconstruction results for three algorithms: model-based reconstruction [(A) and (B)], iterative time reversal [(C) and (D)], and the FFT-based method [(E) and (F)]. The reconstructions are shown for simulated data [(A), (C), and (E)] and experimentally acquired data [(B), (D), and (F)]. Each panel shows a value histogram with the reconstructed values relative to the simulated initial pressure distribution.

reconstruction without signal processing performs significantly worse than all other methods. Our results show that on experimental data, the FFT-based algorithm achieves performance measures that are worse compared to the TR algorithms, but better than backprojection and FBP algorithms.

TABLE I. Quantitative IQA reveals algorithm performance differences on experimental data. The table shows the quantitative results for all measure/algorithm combinations. ↑ denotes that higher values are better, and ↓ denotes that lower values are better. Bold values denote the best-performing algorithm. The relative performance of the algorithms is also denoted by the cell colour, where the top three algorithms are labelled in shades of green and the bottom three are labelled in shades of orange.

Experiment	DAS	FBP	MB	TR	ITTR	FFT
R ↑	0.24	0.57	0.57	0.76	0.77	0.61
MAE ↓	199.84	163.18	156.39	109.43	110.91	140.17
SSIM ↑	0.72	0.72	0.69	0.77	0.78	0.75
JSD ↓	0.48	0.43	0.44	0.36	0.37	0.41
HaarPSI ↑	0.33	0.50	0.46	0.37	0.31	0.32
FWHM ↓	8.20	7.35	6.42	4.79	4.92	4.55

### C. Reconstruction algorithms show differences when moving from a full-view to a limited-view setting

We evaluate the change in performance of the MB, ITTR, and FFT methods when moving from a full-view scenario ( $360^\circ$  angular coverage) to a limited-view setting ( $270^\circ$  angular coverage, as encountered in the MSOT InVision scanner). We report the average results on 15 simulated phantoms at 800 nm randomly chosen from the calibration and testing data.

We can make two immediate observations when visually inspecting the reconstruction results: (1) The MB method does not seem to notably change visually [cf. Figs. 5(A) and 5(B)], whereas (2) the ITTR and FFT methods seem to deteriorate in a similar fashion (cf. Figs. 5(C)–5(F)]. These findings are corroborated by the quantitative results (Table II). For the MB method,  $R$ , MAE, SSIM, and JSD only slightly change and HaarPSI even shows improvements. For ITTR and FFT, on the other hand, all measures, except for SSIM, show a deterioration of the reconstruction performance on the order of 10% (varying from 7% to

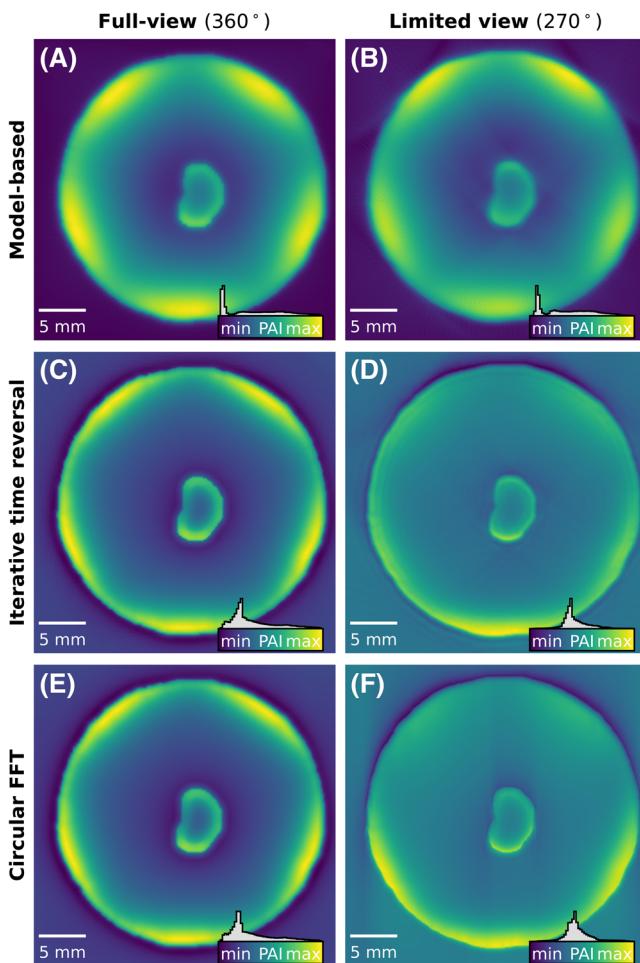


FIG. 5. Comparisons of full-view versus limited-view reconstructions reveal the emergence of reconstruction artefacts in the limited-view case. The figures show a representative example phantom simulation in a full-view setting [(A), (C), and (E)] and a limited-view setting [(B), (D), and (F)]. We show the results of the model-based reconstruction [(A) and (B)], iterative time reversal [(C) and (D)], and the FFT-based method [(E) and (F)].

15%). One should note that the PATATO implementation of the MB method was exclusively tested against limited-view data from the MSOT InVision 256-TF, which might have introduced a bias that could constitute a reason why it might not deteriorate similarly compared to the others.

A second important observation is that the SSIM measure does not seem to reflect the changes to the same extent as the other measures. Some reconstruction artefacts around the central absorber in the image are visible for the MB method, but only the JSD measure predicts a deterioration of the image quality in all three cases. This indicates that there are stark differences in the type of data and changes that the different measures are sensitive to.

### D. Computation analysis confirms high computational efficiency of the FFT-based method

The full-pipeline run times for DAS, FBP, and TR were on the order of 2–3 s, whereas the run times of MB and ITTR were on the order of 1 min per image. The FFT-based algorithm took an average of 1.4 s per image. A noteworthy observation is that the FFT-based algorithm only required a peak memory allocation of 200 MB, whereas all other methods required 1.5 GB and more. That being said, it must be noted that the algorithms' run times and memory requirements are heavily implementation-dependent. Our versions of DAS, FBP, and MB are implemented in PYTHON in the research toolkit PATATO (Else *et al.*, 2024) and use JAX to run on the GPU. TR and ITTR are implemented in MATLAB and use the K-WAVE toolbox, which uses CUDA (Nvidia, Santa Clara, CA)-compiled binaries. The FFT-based algorithm is implemented in PYTHON and runs on the CPU. It has been demonstrated that through meticulous optimisation and, e.g., the precomputation of delays or model matrices, the computational performance can be heavily reduced and single-stage reconstruction schemes can typically be applied in real time (Yuan *et al.*, 2013). The values of the Hankel functions  $H_{|k|}(\rho)$  at the nodes of the equispaced Cartesian grid in

$$\rho \in [-\rho_{Nyq}, \rho_{Nyq}], \quad k \in [-N_d/2, N_d/2]$$

TABLE II. Reconstruction algorithms show performance differences in limited view settings. We compare the performance measures of the model-based reconstruction (MB) with iterative time reversal (ITTR) and the Fourier transform-based method (FFT). For each, we show the measure scores for the full-view case ( $360^\circ$ ) and the relative performance change in percent ( $\Delta\%$ ) when switching to a limited view setting with  $270^\circ$  coverage.  $\uparrow$  denotes that higher values and positive changes are better, and  $\downarrow$  denotes that lower values and negative changes are better.

Experiment	MB		ITTR		FFT	
	360°	$\Delta\%$	360°	$\Delta\%$	360°	$\Delta\%$
R $\uparrow$	0.89	+1.6%	0.94	-6.6%	0.93	-10.6%
MAE $\downarrow$	43.0	-3.4%	88.38	+8.2%	91.88	+10.7%
SSIM $\uparrow$	0.87	+1.3%	0.86	-1.4%	0.85	-2.4%
JSD $\downarrow$	0.24	-2.3%	0.29	+14.6%	0.28	+11.0%
HaarPSI $\uparrow$	0.57	+10.4%	0.69	-12.3%	0.65	-13.6%

within the FFT-based method can be precomputed, which could significantly speed up the processing time.

#### IV. DISCUSSION

Based on our results, we believe that evaluating image reconstruction algorithms with full-reference IQA measures based on digital twins of the test objects and imaging system is an objective way to ensure comparability of reconstruction algorithms that tackle the acoustic inverse problem in PAI. No-reference IQA measures do not directly indicate the accuracy of the initial pressure reconstruction, though they can be a valuable complement to full-reference IQA measures. Since the ground truth  $p_0$  is not known *in vivo*, a high-confidence reference  $p_0$  simulation for a real-world test object is therefore an ideal middle ground that enables objective comparison of acoustic inversion schemes using full-reference IQA measures and experimentally acquired data.

The proposed phantom materials by [Hacker \*et al.\* \(2021\)](#) are accurately characterisable with a DIS measurement system, and it is straightforward to manufacture complex piecewise-constant test objects. We can build digital twins of the phantoms and run simulations using the proposed digital twin simulation pipeline within SIMPA. By comparing the simulated versus experimentally determined measurement data, we can quantitatively evaluate the fidelity of the implemented forward model and then use intermediate simulation outputs, such as the initial pressure distribution, as high-fidelity data for full-reference IQA measures. Other non-linear calibration functions are also valid choices, but in our experiments, none of the tested functions yielded improvements significant enough to favour over a simple linear model. We believe this is because the impulse response and a measured noise profile of the imaging system have the biggest influence on the simulation fidelity—in conjunction with scaling the simulation to match the experimental amplitudes. Our approach is generalizable to any photoacoustic system, by manufacturing and measuring custom phantoms and modelling the PAI device in question in, e.g., MCX and k-WAVE. We further believe that our proposed approach can generalise to any type of complexity in the phantoms. While more complex phantoms could lead to more representative performance analyses, the added complexity in manufacturing and segmentation might undermine the general feasibility of the approach.

We demonstrate the usefulness of the ability to apply full-reference IQA measures by comparing an FFT-based image reconstruction algorithm that was conceived to work with circular detection geometries to several state-of-the-art image reconstruction techniques. With the proposed framework, we can show that the method exhibits similar performance to the ITTR algorithm while being considerably faster. We further show that we can use the digital twin framework to evaluate the performance changes of the reconstruction schemes when moving from a full-view to a

limited-view setting. We find that the MB reconstruction algorithm behaves drastically differently compared to ITTR and the FFT-based method. At the same time, we also find that SSIM does not manage to pick up on the performance changes to the same extents that the other IQA measures under investigation did. This might indicate that SSIM is not a good fit to evaluate PA image reconstruction algorithms.

One has to keep in mind that the simulated  $p_0$  is only a reference and not an exact ground truth. The exact quantitative results of the IQA measures depend on the correctness of the forward model, which hinges mainly on the approximating assumptions done in the mathematical models, but also on the representation of hardware constraints, such as noise and the IRF. One can see this *simulation gap* quantitatively when comparing the simulated and experimental measurement data. We show that careful calibration of the simulations with the reconstructions can correct some modelling errors, but systematic non-linear changes could not be captured. Furthermore, the piecewise-constant nature of the proposed test objects is not sufficient to capture the complexity of *in vivo* tissue and sophisticated fabrication strategies, such as 3D printing approaches, would be required to achieve the spatial heterogeneity needed. The complementary information of the FWHM measure and the lack of descriptive power of the SSIM measure in this context also show that the suite of IQA measures should probably be tailored to the target application. Ideally, the PAI community should identify common use cases for PAI and define a ranked list of the most useful IQA measures that capture all relevant aspects.

#### V. CONCLUSION

We demonstrate that digital twins can enable full-reference quality assessment of reconstructed PA images by simulating numerical equivalents of tissue-mimicking phantoms and the imaging system. With these, we propose a unique evaluation framework that can compare algorithm performance both within simulated and experimental settings and we quantitatively evaluate different image reconstruction algorithms. We believe the approach can provide a robust and objective method for evaluating acoustic inversion schemes, enhancing reproducibility and accuracy in PAI. Furthermore, our study highlights successful application of an FFT-based reconstruction method that performs comparably to state-of-the-art approaches with high computational efficiency (up to 100 times faster) on the experimental data. This is especially promising for developing MB learned reconstruction methods, which necessitate using highly efficient models.

#### ACKNOWLEDGMENTS

J.G. is funded by the Deutsche Forschungsgemeinschaft [DFG (German Research Foundation)] under Project Nos. GR 5824/1 and GR 5824/2. L.K. is partially supported by the NSF, through the Award No. NSF/DMS-2405348. T.R. E. is funded by Cancer Research UK (A29580). A.H.

acknowledges support from the Research Council of Finland with the Flagship of Advanced Mathematics for Sensing Imaging and Modelling Project No. 359186, Centre of Excellence of Inverse Modelling and Imaging Project No. 353093, and the Academy Research Fellow Project No. 338408. J.P. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 101001417—QUANTOM), the Research Council of Finland (Centre of Excellence in Inverse Modelling and Imaging Grant No. 353086 and Flagship of Advanced Mathematics for Sensing Imaging and Modelling Grant No. 358944), and the Finnish Ministry of Education and Culture's Pilot for Doctoral Programmes (Pilot project Mathematics of Sensing, Imaging and Modelling). B.T.C. acknowledges support from the Engineering and Physical Sciences Research Council, UK (EPSRC), Grant Nos. EP/W029324/1 and EP/T014369/1. F.D.C. acknowledges support from the Physics of Life Network (PoLNET3) summer 2023 student bursary. The authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support during the "Rich and Nonlinear Tomography" Programme, supported by EPSRC Grant No. EP/R014604/1.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

## DATA AVAILABILITY

All data and code to reproduce this work are publicly available on Zenodo at <https://doi.org/10.5281/zenodo.15388429>. The code is available on GitHub at <https://github.com/BohndiekLab/CircularFFTReconPA>.

Agranovsky, M., and Kuchment, P. (2007). "Uniqueness of reconstruction and an inversion procedure for thermoacoustic and photoacoustic tomography with variable sound speed," *Inverse Prob.* **23**(5), 2089–2102.  
 Arridge, S. R., Betcke, M. M., Cox, B. T., Lucka, F., and Treeby, B. E. (2016). "On the adjoint operator in photoacoustic tomography," *Inverse Prob.* **32**(11), 115012.  
 Assi, H., Cao, R., Castelino, M., Cox, B., Gilbert, F. J., Gröhl, J., Gurusamy, K., Hacker, L., Ivory, A. M., Joseph, J., Knieling, F., Leahy, M. J., Lilaj, L., Manohar, S., Meglinski, I., Moran, C., Murray, A., Oraevsky, A. A., Pagel, M. D., Pramanik, M., Raymond, J., Singh, M. K. A., Vogt, W. C., Wang, L., Yang, S., Members of IPASC, and Bohndiek, S. E. (2023). "A review of a strategic roadmapping exercise to advance clinical translation of photoacoustic imaging: From current barriers to future adoption," *Photoacoustics* **32**, 100539.

Beard, P. (2011). "Biomedical photoacoustic imaging," *Interface Focus* **1**(4), 602–631.  
 Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*, Springer Topics in Signal Processing, Vol. 2 (Springer, Berlin), pp. 1–4.  
 Boink, Y. E., Manohar, S., and Brune, C. (2020). "A partially-learned algorithm for joint photo-acoustic reconstruction and segmentation," *IEEE Trans. Med. Imaging* **39**(1), 129–139.  
 Breger, A., Karner, C., Selby, I., Gröhl, J., Dittmer, S., Lilley, E., Babar, J., Beckford, J., Else, T. R., Sadler, T. J., Shahipasand, S., Thavakumar, A., Roberts, M., and Schönlieb, C.-B. (2024). "A study on the adequacy of common IQA measures for medical images," in *Proceedings of 2024 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2024)*, edited by R. Su, A. F. Frangi, and Y. Zhang, Lecture Notes in Electrical Engineering, Vol. 1372 (Springer, Singapore), pp. 451–462.

- Castleman, K. R. (1996). *Digital Image Processing* (Prentice Hall Press, Englewood Cliffs, NJ).  
 Choi, W., Oh, D., and Kim, C. (2020). "Practical photoacoustic tomography: Realistic limitations and technical solutions," *J. Appl. Phys.* **127**(23), 230903.  
 Detlefsen, N. S., Borovec, J., Schock, J., Harsh, A., Koker, T., Di Liello, L., Stanci, D., Quan, C., Grechkin, M., and Falcon, W. (2022). "TorchMetrics—Measuring reproducibility in PyTorch," *J. Open Source Software* **7**(70), 4101.  
 Eller, M., Hoskins, P., and Kunyansky, L. (2020). "Microlocally accurate solution of the inverse source problem of thermoacoustic tomography," *Inverse Prob.* **36**(8), 085012.  
 Else, T. R., Gröhl, J., Hacker, L., and Bohndiek, S. E. (2024). "PATATO: A Python photoacoustic tomography analysis toolkit," *J. Open Source Software* **9**(93), 5686.  
 Fang, Q., and Boas, D. A. (2009). "Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units," *Opt. Express* **17**(22), 20178–20190.  
 Gröhl, J., Dreher, K. K., Schellenberg, M., Rix, T., Holzwarth, N., Vieten, P., Ayala, L., Bohndiek, S. E., Seitel, A., and Maier-Hein, L. (2022). "SimpA: An open-source toolkit for simulation and image processing for photonics and acoustics," *J. Biomed. Opt.* **27**(8), 083010.  
 Gröhl, J., Else, T. R., Hacker, L., Bunce, E. V., Sweeney, P. W., and Bohndiek, S. E. (2024). "Moving beyond simulation: Data-driven quantitative photoacoustic imaging using tissue-mimicking phantoms," *IEEE Trans. Med. Imaging* **43**, 1214–1224.  
 Gröhl, J., Schellenberg, M., Dreher, K., and Maier-Hein, L. (2021). "Deep learning for biomedical photoacoustic imaging: A review," *Photoacoustics* **22**, 100241.  
 Hacker, L., Joseph, J., Ivory, A. M., Saed, M. O., Zeqiri, B., Rajagopal, S., and Bohndiek, S. E. (2021). "A copolymer-in-oil tissue-mimicking material with tuneable acoustic and optical characteristics for photoacoustic imaging phantoms," *IEEE Trans. Med. Imaging* **40**(12), 3593–3603.  
 Hauptmann, A., Lucka, F., Betcke, M., Huynh, N., Adler, J., Cox, B., Beard, P., Ourselin, S., and Arridge, S. (2018). "Model-based learning for accelerated, limited-view 3-D photoacoustic tomography," *IEEE Trans. Med. Imaging* **37**(6), 1382–1393.  
 Hristova, Y., Kuchment, P., and Nguyen, L. (2008). "Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media," *Inverse Prob.* **24**(5), 055006.  
 Hsu, K.-T., Guan, S., and Chitnis, P. V. (2023). "Fast iterative reconstruction for photoacoustic tomography using learned physical model: Theoretical validation," *Photoacoustics* **29**, 100452.  
 Jaeger, M., Schüpbach, S., Gertsch, A., Kitz, M., and Frenz, M. (2007). "Fourier reconstruction in optoacoustic imaging using truncated regularized inverse k-space interpolation," *Inverse Prob.* **23**(6), S51.  
 Joseph, J., Tomaszewski, M. R., Quiros-Gonzalez, I., Weber, J., Brunker, J., and Bohndiek, S. E. (2017). "Evaluation of precision in optoacoustic tomography for preclinical imaging in living subjects," *J. Nucl. Med.* **58**(5), 807–814.  
 Kirchner, T., Sattler, F., Gröhl, J., and Maier-Hein, L. (2018). "Signed real-time delay multiply and sum beamforming for multispectral photoacoustic imaging," *J. Imaging* **4**(10), 121.  
 Knieling, F., Neufert, C., Hartmann, A., Claussen, J., Uriach, A., Egger, C., Vetter, M., Fischer, S., Pfeifer, L., Hagel, A., Kielisch, C., Görtz, R. S., Wildner, D., Engel, M., Röther, J., Uter, W., Siebler, J., Atreya, R., Rascher, W., Strobel, D., Neurath, M. F., and Waldner, M. J. (2017). "Multispectral optoacoustic tomography for assessment of Crohn's disease activity," *N. Engl. J. Med.* **376**(13), 1292–1294.  
 Kuchment, P., and Kunyansky, L. (2015). "Mathematics of photoacoustic and thermoacoustic tomography," in *Handbook of Mathematical Methods in Imaging*, 2nd ed., Vol. 1 (Springer, New York), pp. 1117–1167.  
 Kunyansky, L. (2011). "Reconstruction of a function from its spherical (circular) means with the centers lying on the surface of certain polygons and polyhedra," *Inverse Prob.* **27**(2), 025012.  
 Kunyansky, L. (2012). "Fast reconstruction algorithms for the thermo-acoustic tomography in certain domains with cylindrical or spherical symmetries," *Inverse Probl. Imaging* **6**(1), 111–131.

- Kunyansky, L. (2024). "Fast\_Hankel-based\_solver," GitHub, [https://github.com/leondak/Fast\\_Hankel-based\\_solver](https://github.com/leondak/Fast_Hankel-based_solver).
- Kunyansky, L. A. (2007). "A series solution and a fast algorithm for the inversion of the spherical mean Radon transform," *Inverse Prob.* **23**(6), S11–S20.
- Lin, L., and Wang, L. V. (2022). "The emerging role of photoacoustic imaging in clinical oncology," *Nat. Rev. Clin. Oncol.* **19**(6), 365–384.
- Manohar, S., and Dantuma, M. (2019). "Current and future trends in photoacoustic breast imaging," *Photoacoustics* **16**, 100134.
- Menéndez, M. L., Pardo, J. A., Pardo, L., and Pardo, M. C. (1997). "The Jensen-Shannon divergence," *J. Franklin Inst.* **334**(2), 307–318.
- Nolden, M., Zelzer, S., Seitel, A., Wald, D., Müller, M., Franz, A. M., Maleike, D., Fangerau, M., Baumhauer, M., Maier-Hein, L., Maier-Hein, K. H., Meinzer, H.-P., and Wolf, I. (2013). "The Medical Imaging Interaction Toolkit: Challenges and advances: 10 years of open-source development," *Int. J. CARS* **8**, 607–620.
- Paltauf, G., Nuster, R., Haltmeier, M., and Burgholzer, P. (2007). "Experimental evaluation of reconstruction algorithms for limited view photoacoustic tomography with line detectors," *Inverse Prob.* **23**(6), S81.
- Perrot, V., Polichetti, M., Varray, F., and Garcia, D. (2021). "So you think you can DAS? A viewpoint on delay-and-sum beamforming," *Ultrasonics* **111**, 106309.
- Pickering, J. W., Prahl, S. A., Van Wieringen, N., Beek, J. F., Sterenborg, H. J., and Van Gemert, M. J. (1993). "Double-integrating-sphere system for measuring the optical properties of tissue," *Appl. Opt.* **32**(4), 399–410.
- Poudel, J., Lou, Y., and Anastasio, M. A. (2019). "A survey of computational frameworks for solving the acoustic inverse problem in three-dimensional photoacoustic computed tomography," *Phys. Med. Biol.* **64** (14), 14TR01.
- Regensburger, A. P., Fonteyne, L. M., Jüngert, J., Wagner, A. L., Gerhalter, T., Nagel, A. M., Heiss, R., Flenkenthaler, F., Qurashi, M., Neurath, M. F., Klymiuk, N., Kemter, E., Fröhlich, T., Uder, M., Woelfle, J., Rascher, W., Trollmann, R., Wolf, E., Waldner, M. J., and Knieling, R. (2019). "Detection of collagens by multispectral optoacoustic tomography as an imaging biomarker for Duchenne muscular dystrophy," *Nat. Med.* **25**(12), 1905–1915.
- Reisenhofer, R., Bosse, S., Kutyniok, G., and Wiegand, T. (2018). "A Haar wavelet-based perceptual similarity index for image quality assessment," *Signal Process. Image Commun.* **61**, 33–43.
- Research Group Computational Data Analysis, University Bremen (2018). "haarpsi," GitHub, <https://github.com/rgcda/haarpsi>.
- Rietberg, M. T., Gröhl, J., Else, T. R., Bohndiek, S. E., Manohar, S., and Cox, B. T. (2025). "Artifacts in photoacoustic imaging: Origins and mitigations," *arXiv:2504.12772*.
- Rosenthal, A., Razansky, D., and Ntziachristos, V. (2010). "Fast semi-analytical model-based acoustic inversion for quantitative optoacoustic tomography," *IEEE Trans. Med. Imaging* **29**(6), 1275–1285.
- Stefanov, P., and Uhlmann, G. (2009). "Thermoacoustic tomography with variable sound speed," *Inverse Prob.* **25**(7), 075011.
- Tarvainen, T., and Cox, B. (2023). "Quantitative photoacoustic tomography: Modeling and inverse problems," *J. Biomed. Opt.* **29**(S1), S11509.
- Treeby, B. E., and Cox, B. T. (2010). "k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields," *J. Biomed. Opt.* **15**(2), 021314.
- Treeby, B. E., Zhang, E. Z., and Cox, B. T. (2010). "Photoacoustic tomography in absorbing acoustic media using time reversal," *Inverse Prob.* **26** (11), 115003.
- Van de Sompel, D., Sasportas, L. S., Jokerst, J. V., and Gambhir, S. S. (2016). "Comparison of deconvolution filters for photoacoustic tomography," *PLoS One* **11**(3), e0152597.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., and SciPy 1.0 Contributors (2020). "Scipy 1.0: Fundamental algorithms for scientific computing in Python," *Nat. Methods* **17**(3), 261–272.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612.
- Warbal, P., and Saha, R. K. (2022). "Performance comparison of commonly used photoacoustic tomography reconstruction algorithms under various blurring conditions," *J. Mod. Opt.* **69**(9), 487–501.
- Xu, M., and Wang, L. V. (2002). "Pulsed-microwave-induced thermo-acoustic tomography: Filtered backprojection in a circular measurement configuration," *Med. Phys.* **29**(8), 1661–1669.
- Xu, M., and Wang, L. V. (2005). "Universal back-projection algorithm for photoacoustic computed tomography," *Phys. Rev. E* **71**(1), 016706.
- Yuan, J., Xu, G., Yu, Y., Zhou, Y., Carson, P. L., Wang, X., and Liu, X. (2013). "Real-time photoacoustic and ultrasound dual-modality imaging system facilitated with graphics processing unit and code parallel optimization," *J. Biomed. Opt.* **18**(8), 086001.