Coordinate-based Speed of Sound Recovery for Aberration-Corrected Photoacoustic Computed Tomography

Supplementary Material

S1. Experiment Details

S1.1. Numerical Phantoms

In this section, we describe the simulation details of our numerical phantoms.

We use the same IP image shown in Fig. 4a for all numerical phantoms but use five SOS with different variations (see Fig. S5a). SOS values range from 1490.0 to 1650.0 m/s in our numerical dataset. The structures in the IP and SOS images are inspired by the mouse body[1] while the SOS values are based on realistic measurements[2]. For example, the small circles with high SOS in samples 1 and 3 correspond to bones, the orange circles in samples 1 and 2 correspond to the liver, and the circle with small SOS in samples 1, 4, and 5 correspond to the stomach lumen.

The background water SOS is set to 1499.4 m/s (assuming a 26°C water temperature based on [21]) outside the sample. The PA signals $p(n,t)$ are obtained by 2D numerical simulation using the k-Wave toolbox [32] assuming an evenly distributed 512-transducer ring array with a diameter of 10 cm (same setting for the leaf phantom and *in vivo* mouse liver experiments). Similar to [5], we take the derivative of the PA signals to simulate transducer effects: $S(n,t) = -2\frac{\partial p(n,t)}{\partial t}$. The images are cropped into 75% overlapping 3.2mm × 3.2mm patches, as for all reported results.

We show the IP and SOS reconstructions on all five numerical phantoms using our methods and APACT in Fig. S5 b,c,d. Results are qualitatively similar to the example discussed in the text, and numbers reported in tables are averaged over these five phantoms.

S1.2. Leaf Phantom

We use the experimentally collected phantom data from [5], which features a leaf in a 2.5 cm diameter agarose cylinder with a water-filled hole in the middle for acoustic heterogeneity (see photograph in Fig. 5a). The leaf phantom is surrounded by water with a temperature of 26°C, and the wavelength of laser illumination is 700 nm in the experiment. The constant background SOS is set to 1499.4 m/s according to [21].

Due to nonideal transducer alignment and transducer impulse responses, the collected PA signals are calibrated using measured ring error and transducer electrical impulse response (EIR) before feeding into the algorithms. The same calibration is applied to the *in vivo* experiment.

S1.3. In vivo Mouse Liver

The PA signals in our *in vivo* experiments were collected from a nude mouse liver at 1064 nm [3]. The water temperature was 31°C during the experiment, thus we used a background SOS of 1511.4 m/s. See Fig. S2 to see the alignment between recovered SOS and IP, which matches better in our method than others.

S1.4. Transducer EIR and System MTF

We measured the transducer electrical impulse response (EIR) in the setup we used to collect the data of the leaf phantom and the *in vivo* mouse liver. We show that our experimentally collected transducer EIR and its Fourier transform is a band-pass filter, shown in Fig. S1a . We calculate the system PSF by projecting this EIR from all 512 transducers into the image space using DAS and calculate the system modulation transfer function (MTF) by taking the 2D Fourier transform of this PSF, shown in Fig. S1b. The ring structure in the MTF is caused by the projection arcs of DAS. We apply a phase only EIR deconvolution on the collected PA signals before feeding them into our algorithm as a calibration in the phantom and the *in vivo* experiment.

S1.5. Image Stitching

We adopted the same strategy used for APACT [5] in IP reconstruction. This process differs from most existing im-

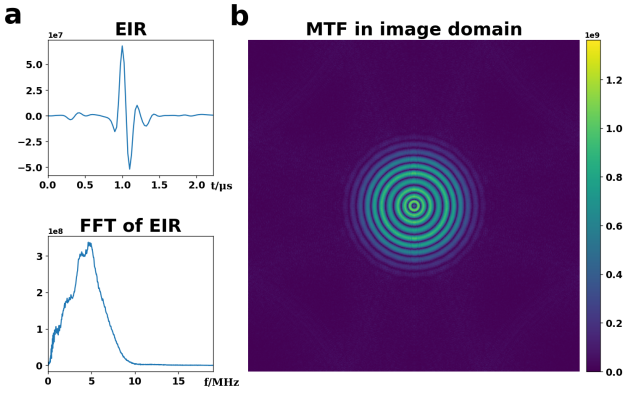

Figure S1. (a) Experimentally measured transducer EIR and its Fourier transform. (b) The system MTF is the Fourier transform of the PSF, which is the projection of EIR into image space.

[1] https://www.imaios.com/en/vet-anatomy/mouse/mouse-whole-body

[2] https://itis.swiss/virtual-population/tissue-properties/database/acoustic-properties/

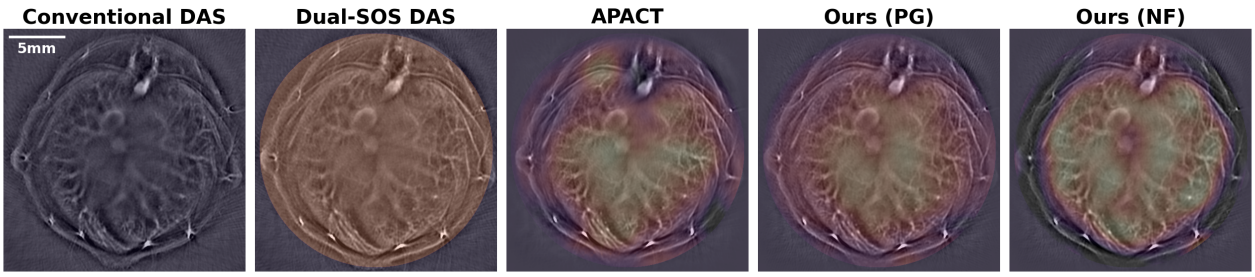

Figure S2. **Overlaid initial pressure image and SOS of *in vivo* mouse liver.** Our reconstructed SOS with NF (right) shows a superior match to the liver's anatomical shape in the initial pressure image.

age reconstruction frameworks because the image is reconstructed patch-by-patch and stitched to form the final result. Images are cropped using 3.2 mm × 3.2 mm Gaussian windows with an FWHM of 1.5 mm before being fed to our algorithm. The patch centers are arranged on a Cartesian grid, and adjacent patches overlap by 75%. With our proposed method for solving the speed of sound distribution by utilizing information from all patches, we obtain the wavefront correction model for each isoplanatic patch. This simultaneously yields the deconvolved, aberration-corrected image patches. These patches are then translated to their corresponding center positions and summed. The 75% overlap ensures a smooth transition during image stitching. Finally, the result is adjusted to remove the modulation pattern due to the Gaussian windows by dividing by the sum of the Gaussian window weights.

S2. Difference Between Our Wavefront Model and APACT's Wavefront Model

One critical contribution of our work is the improved wavefront model compared to the previous work APACT [5]. As explained in the main text, the wavefront model represents the wavefront advancement or retardation induced by a nonuniform speed of sound in tissue relative to an ideal spherical wave in a homogeneous medium. Mathematically, the wavefront function $w(\theta)$, which depicts the propagation length difference relative to the spherical wave when received by the transducers, is a function of the propagation direction θ. This wavefront function can be expanded using a Fourier series:

$$w(\theta) = C + \sum_{n=1}^{\infty} A_n \cos(n\theta) + B_n \sin(n\theta).$$

In APACT, only three low order harmonics were used, namely DC (constant C), $\cos(2\theta)$, and $\sin(2\theta)$. This is because, firstly, the first order terms $\cos(\theta)$ and $\sin(\theta)$ result in a shifted PSF, making them non-identifiable from the reconstructed images when the true IP is unknown. Secondly, the authors omitted the higher order terms, i.e. $\cos(n\theta)$ and $\sin(n\theta)$ for $n \geq 3$, to avoid the exponential increase in computation time during the exhaustive search for the optimal wavefront. However, as shown in Fig. 2, the actual wavefront function can be complex, which necessitate those high order terms. In our work, the wavefront functions corresponding to different isoplanatic patches are obtained from the learned SOS map. There is no explicit restriction on their profiles such as the assumption in APACT that they are slowly varying.

S3. Ablation Studies

We conducted ablation studies using our synthetic numerical phantoms on an NVIDIA RTX A6000 GPU. All results are averaged over 5 numerical phantoms, which share the same IP but have different SOS.

S3.1. Number of Delays

We show the performance of our method (with and without SOS oracle), pixel grid (with and without TV regularization), and APACT with respect to numbers of delays used

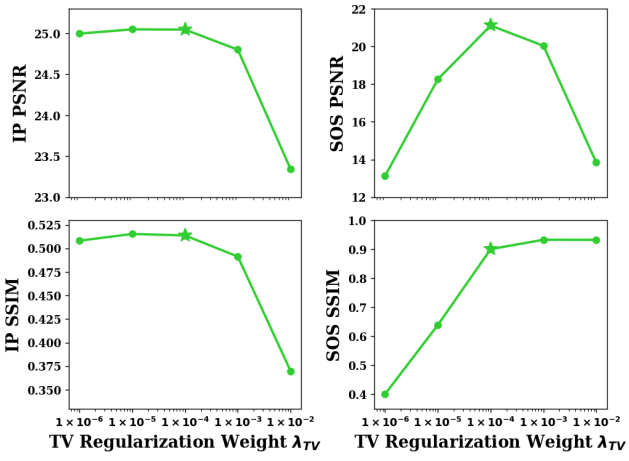

Figure S3. **Ablation results on TV regularization weight λ in Eq. (8).** We select $\lambda = 1 \times 10^{-4}$ for the best IP and SOS reconstruction (marked with a green star).

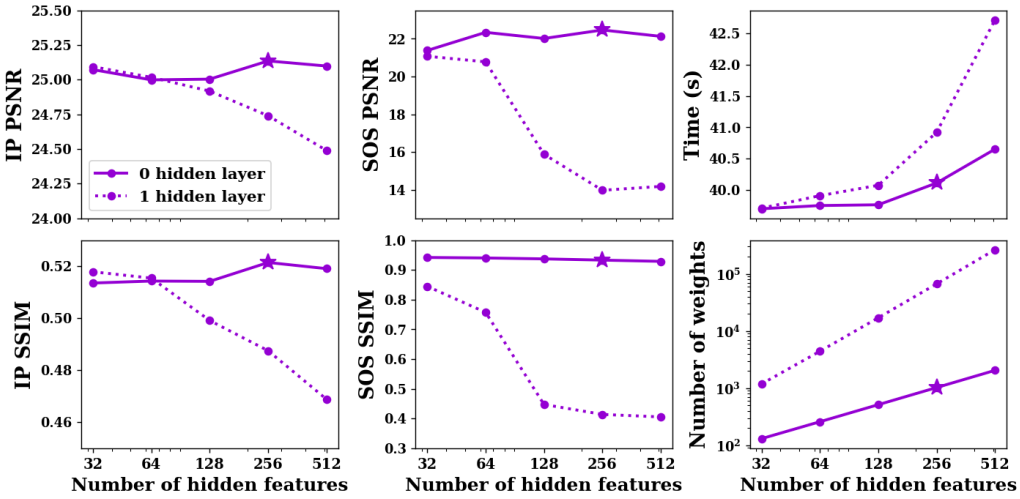

Figure S4. **Ablation study on the network size of NF (averaged over 5 numerical phantoms).** Larger network size decreases the implicit smoothness regularization on the SOS, resulting in noisier SOS and IP reconstructions. We select 0 hidden layer with 256 hidden features (marked with a purple star).

in Fig. 6. We keep the range of the delay distances fixed at [-0.8 mm, 0.8 mm] and uniformly sample delay distances within this range based on the number of delays used. We show that more delay channels leads to increased performance and computation time for all methods. This originates from Eq. (7), where more delay channels promote the performance of pseudo-inverse deconvolution, provide more redundant information for SOS estimation, and require more computation time. Removing TV regularization in pixel grid (dashed green line) has little impact on the quality of the IP image but significantly decreases the quality of the SOS. Considering the performance and computation time, we choose 16 delays (marked with stars). We also show that the efficient parameterization of SOS gives NF a slight time advantage compared to PG (see lower right subplot), which takes more iterations to reach convergence.

S3.2. TV regularization for PG

In Fig. S3, we show the effects of TV regularization weight λ with our pixel grid SOS representation. We observe that the IP image quality drops when $\lambda > 1\times10^{-4}$, while the SOS accuracy reaches optimum around $\lambda = 1 \times 10^{-4}$. Therefore, we select $\lambda = 1 \times 10^{-4}$ (marked with a green star in Fig. S3) for all reconstructions in our implementation of PG.

S3.3. Network Size of NF

We compared the performance of our method with NF with different network sizes ranging from 131 to 264,707 parameters, shown in Fig. S4. We show that the implicit smoothness regularization of the MLP decreases as the network becomes larger, resulting in noisier SOS reconstructions and hence worse IP images. This can be observed from the comparison between no hidden layer (solid line) and one hidden layer (dashed line) and the overall decreasing trend of the two curves. An increase in network size also leads to a small rise in computation time. We select 0 hidden layer and 256 hidden features after an overall consideration of the performance and computation time (marked with a purple star).

S4. Supplementary Videos

We provide six videos showing the convergence process of the IP image and SOS for our methods with our numerical phantom (`nf_numerical_phantom.mp4` and `pg_numerical_phantom.mp4`), leaf phantom (`nf_leaf_phantom.mp4` and `pg_leaf_phantom.mp4`), and *in vivo* mouse liver (`nf_in_vivo_mouse_liver.mp4` and `pg_in_vivo_mouse_liver.mp4`).

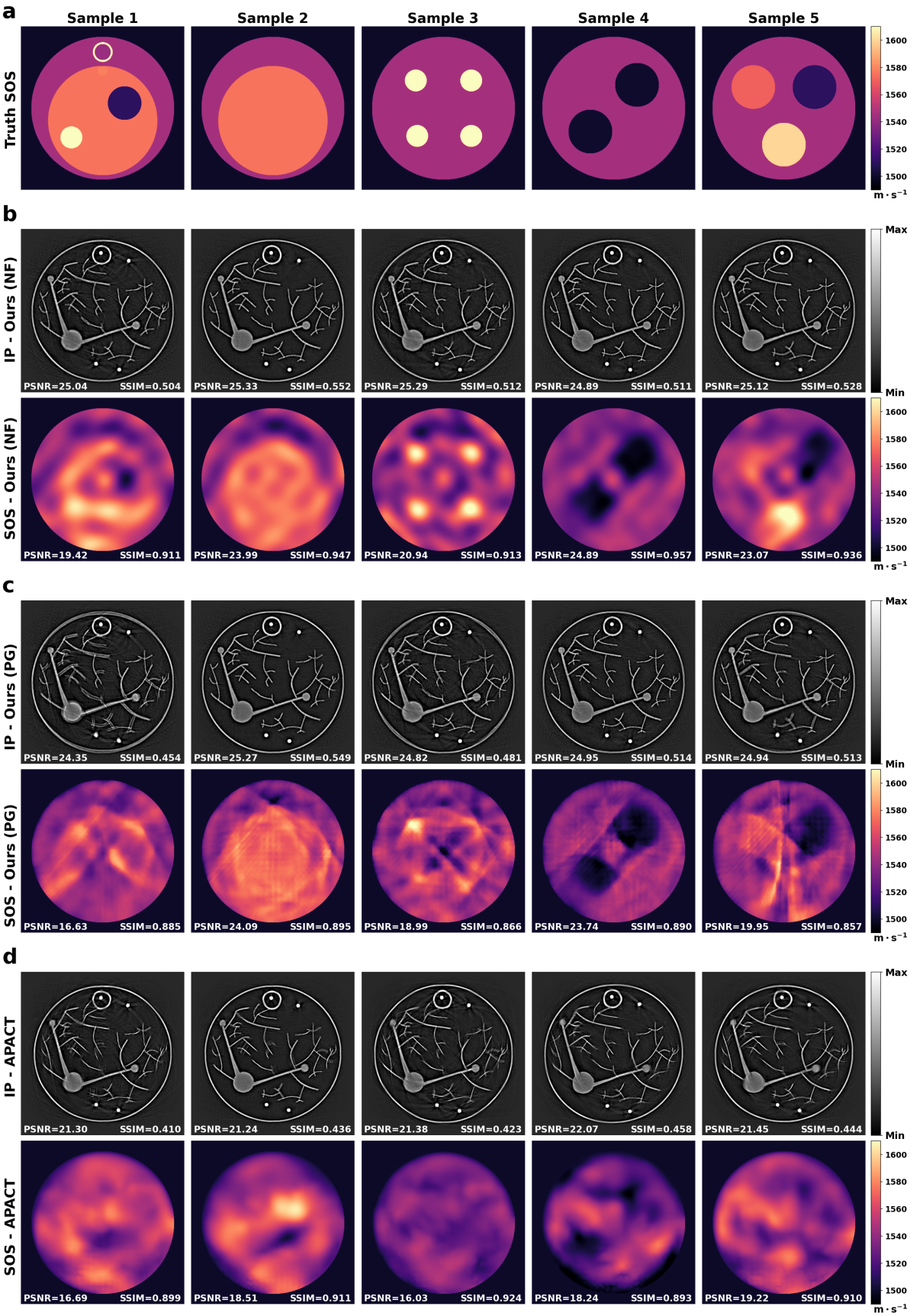

Figure S5. **All five numerical phantoms and results.** (a) The SOS maps used in our numerical phantoms. (b) Reconstructions of our method with NF. (c) Reconstructions of our method with PG. (d) Reconstructions of APACT [5].