DEFECT DETECTION FROM COMPRESSED 3-D ULTRASONIC FREQUENCY MEASUREMENTS

Sebastian Semper[*], *Jan Kirchhof*[†], *Christoph Wagner*[*], *Fabian Krieg*[†], *Florian Roemer*[†], *Giovanni Del Galdo*[◇,*]

[*]Technische Universität Ilmenau, Institute for Information Technology, Germany
[†]Fraunhofer Institute for Nondestructive Testing IZFP, Saarbrücken, Germany
[◇]Fraunhofer Institute for Integrated Circuits IIS, Ilmenau, Germany

ABSTRACT

In this paper, we propose a compressed sensing scheme for volumetric synthetic aperture measurements in ultrasonic nondestructive testing. The compression is achieved by limiting the measurement to a subset of the Fourier coefficients of the full measurement data, where we also address the issue of a suitable hardware architecture for the task. We present a theoretic analysis for one of the proposed schemes in terms of the Restricted Isometry Property and derive a scaling law for the lower bound of the number of necessary measurements for stable and efficient recovery. We verify our approach with reconstructions from measurement data of a steel specimen that was compressed synthetically in software. As a side result, our approach yields a variant of the 3-D Synthetic Aperture Focusing Technique which can deal with compressed data.

Index Terms— 3D ultrasonic imaging, Sparse Signal Recovery, Compressive Sensing, SAFT

I. INTRODUCTION

It is now well established that based on Compressed Sensing (CS) theory [1] the number of measurement samples to reconstruct the relevant signal information can be reduced significantly compared to the Nyquist rate, provided that the signal is sparse in some known domain.

In ultrasonic Nondestructive Testing (NDT), a widespread approach is to acquire several pulse-echo measurements using a single transducer moved along an equidistant grid on the surface of the specimen. Typically, these measurements, called A-scans, are sampled at rates much higher than the Nyquist rate, since imaging is performed using the Synthetic Aperture Focusing Technique (SAFT) [2]. SAFT is based on a Delay-and-Sum (DAS) scheme in time-domain for which a dense temporal sampling is beneficial [3, Sec. 3.4.3]. To give an example: For an ultrasonic pulse with center frequency $4\,\mathrm{MHz}$ and a bandwidth of $3\,\mathrm{MHz}$ a sampling rate of $80\,\mathrm{MHz}$ to $120\,\mathrm{MHz}$ is common.

In our previous work, we showed that using the model underlying the SAFT as a forward model for Sparse Signal Recovery (SSR) can improve the image quality compared to SAFT images [4], [5]. Equivalent results have also been reported for the Total Focusing Method using a phased array instead of a single transducer [6]. Further, we showed that this can be implemented matrix-free enabling 3-D reconstruction of ultrasonic NDT data [5] using the Python package described in [7]. The code is freely available at [8].

Still, these results are based on measurement data sampled at a sampling rate that is typical for SAFT although the occupied bandwidth is much lower. In medical ultrasound applications it was already shown by [9] that beamforming in the frequency domain allows sampling at much lower rate. As oversampling is then no

longer beneficial to beamformer performance it is substituted by employing I/Q-sampling which is efficient for bandpass signals such as the ultrasonic pulse. In [9] the amount of data is further reduced after sampling by the application of CS theory [1], exploiting sparsity in the frequency domain.

In this paper, we extend the SSR scheme from [5] to a CS scheme based on subsampling in the frequency domain in Sec. II. A theoretic analysis of the recovery performance in terms of the Restricted Isometry Property (RIP) of this subsampling scheme is presented in Sec. III. Based on this, three different strategies for the frequency subsampling are derived and compared: A randomized approach following standard CS theory and two schemes that take into account *a priori* knowledge about the spectrum of the measurements. Using measurement data of a steel specimen containing several flat bottom holes, it is shown in Sec. IV that the randomized scheme performs best in terms of focusing along the depth axis, which is due to the fact that it results in the largest sampled bandwidth. However, it leads to artifacts and false negatives in the x- and y-direction that do not appear when using the two schemes based on *a priori* knowledge. As a side result, we show that frequency-domain subsampling allows us to carry out a reconstruction similar to 3-D SAFT based on the compressed data. Finally, Sec. V concludes the findings.

II. DATA MODEL

Consider the pulse-echo model defined in [4], [5] for a specimen that is homogenous and isotropic with constant speed of sound c_0 and has a flat surface. The specimen contains an unknown number of I defects located at the unknown arbitrary positions (x_i, y_i, z_i). The discrete measurement samples, sampled at a rate $\frac{1}{t_s}$, can be described in the noise-free case as

$$b_{x,y}(t) = \sum_{i=1}^{I} a_i \cdot g_{x,y}(x_i, y_i, z_i) \cdot h(t - \tau_{x,y}(x_i, y_i, z_i)) \tag{1}$$

where the i-th reflector leads to a trace $\tau_{x,y}(x_i, y_i, z_i) = \frac{1}{c_0}\sqrt{(x - x_i)^2 + (y - y_i)^2 + z_i^2}$ in the measurements from positions $(x, y, 0)$,

$$g_{x,y}(x_i, y_i, z_i) = e^{-\alpha_{z_i}((x-x_i)^2 + (y-y_i)^2)}, \quad \alpha_{z_i} = \frac{1}{(\tan(\alpha) \cdot z_i)^2} \tag{2}$$

accounts for the characteristic of a transducer with beam opening angle α [5] and $h(\cdot)$ describes the transmit pulse emitted into the medium by the transducer in time domain. So, in order to make these measurements $h(\cdot)$ has to be chosen appropriately, which also means that it is a *known* property of the measurement setup.

Introducing a suitable grid [4], a linear model can be deduced from (1) as

$$\mathrm{vec}\,\boldsymbol{b} = \boldsymbol{H}\boldsymbol{a}, \tag{3}$$

where the vector $\boldsymbol{a} \in \mathbb{R}^{N_x N_y N_z}$ contains the defect amplitudes a_i at the indices i corresponding to their defect positions (x_i, y_i, t_i). Accordingly, each column of $\boldsymbol{H}$ contains the vectorized modeled

volumetric data set of a single reflector located at that position. In this work, we aim at determining the $i \in \mathbb{N}$ where $\boldsymbol{a}_i \neq 0$, i.e. the positions where a defect is located in the specimen. To do so, we assume that all defects lie on a grid, which introduces a model error that is negligible as long as we chose the grid fine enough.

The size of the dictionary matrix $\boldsymbol{H}$ depends on the number of scan positions N_x, N_y and the number of measurement samples N_t per A-scan, corresponding to N_z depth positions (z_i), as $\boldsymbol{H} \in \mathbb{R}^{N \times N}$ with $N = N_x N_y N_t$. For realistic scenarios, its size becomes infeasibly large. However, the underlying shift-invariance of the model, when sampled uniformly in x and y directions [4], allows for an efficient transform of $\boldsymbol{H}$, which means there exist low complexity algorithms to calculate $\boldsymbol{H} \cdot \boldsymbol{x}$ and $\boldsymbol{H}^{\mathrm{H}} \cdot \boldsymbol{x}$ without building $\boldsymbol{H}$ explicitly. Most SSR algorithms admit a formulation, where $\boldsymbol{H}$ only appears via these products, resulting in so-called matrix-free and thus practically suitable algorithms [5].

By observing linear combinations of the signal the standard CS [1] model can be applied to extend (3) under the assumption that the measured signal is sparse in some known basis. Then, the signal of interest $\boldsymbol{a}$ can be reconstructed directly from the observations taken, which resemble a collection of linear functionals, according to the model

$$\boldsymbol{y} = \boldsymbol{\Psi} \cdot \operatorname{vec} \boldsymbol{b} + \boldsymbol{n} = \boldsymbol{\Psi H a} + \boldsymbol{n}, \qquad (4)$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{M \times N}$ with $M \ll N$ is the so-called measurement or compression matrix representing the aforementioned linear functionals and the vector $\boldsymbol{n} \in \mathbb{R}^M$ accounts for additive measurement noise.

One of the cornerstones of the CS theory is that if $\boldsymbol{\Psi}$ is drawn randomly from a suitable ensemble then there exist efficient (in terms of runtime and memory) means of recovering $\boldsymbol{a}$ from $\boldsymbol{y}$, if the true $\boldsymbol{a}$ has only S non-zero entries. For instance, it is sufficient to draw the entries of $\boldsymbol{\Psi}$ i.i.d. from a sub-Gaussian distribution with suitably chosen variance factor and $M \sim S \log(N/S)$ for efficient recovery to be possible with overwhelming probability. In this case the term efficient recovery means that the generally NP-hard problem

$$\min_{\boldsymbol{a}} \|\boldsymbol{a}\|_0 \text{ s. t. } \|\boldsymbol{y} - \boldsymbol{\Psi H a}\|_2^2 \leq \epsilon$$

can be solved with the following proxy

$$\min_{\boldsymbol{a}} \|\boldsymbol{y} - \boldsymbol{\Psi H a}\|_2^2 + \lambda \|\boldsymbol{a}\|_1$$

for suitably chosen λ, which can be solved in polynomial runtime. However, this random sampling has two drawbacks: Due to the size of N, a random and thus unstructured $\boldsymbol{\Psi}$ would be prohibitively large in terms of memory consumption and computing time and hardware feasibility. Further, an unstructured $\boldsymbol{\Psi}$ would require the complete measurement $\boldsymbol{b}$ for the compression stage which is impractical for synthetic aperture pulse-echo measurements.

However, since the measurements are performed A-scan-wise, an intuitive choice is to implement the same compression for each A-scan, i.e. $\boldsymbol{\Psi} = \boldsymbol{A} \otimes \boldsymbol{I}_{N_x N_y}$, where $\boldsymbol{I}_{N_x N_y}$ is the identity matrix and $\otimes$ denotes the Kronecker or outer product. Ultimately, this results in $M = m N_x N_y \ll N$ if and only if $m \ll N_t$ with m being the number of observations taken for each A-scan. Note that matrix-vector products of the kind $(\boldsymbol{A} \otimes \boldsymbol{B}) \boldsymbol{z}$ can be implemented efficiently, since

$$\boldsymbol{A} \otimes \boldsymbol{B} = (\boldsymbol{A} \otimes \boldsymbol{I}) \cdot (\boldsymbol{I} \otimes \boldsymbol{B}).$$

If now $\boldsymbol{A}$ and $\boldsymbol{B}$ share the same property as $\boldsymbol{H}$ in (4), i.e. they allow for an efficient matrix-vector multiplication, the complexity of calculating $(\boldsymbol{A} \otimes \boldsymbol{B}) \boldsymbol{z}$ decreases even further. For the given $\boldsymbol{\Psi} = \boldsymbol{A} \otimes \boldsymbol{I}_{N_x N_y}$ the second factor $\boldsymbol{I}_{N_x N_y}$ trivially has an efficient matrix-vector product.

It has been shown in [10] that Fourier matrices $\boldsymbol{\mathcal{F}}_J \in \mathbb{C}^{m \times N_t}$ premultiplied with a random sign matrix $\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\xi})$ for a Rademacher vector $\boldsymbol{\xi} \in \{-1, +1\}^{N_t}$, where one selects the rows of $\boldsymbol{\mathcal{F}}$ indexed by the set J, are suitable compression matrices in terms of recovery performance. To this end, one selects J uniformly at random from all possible subsets of magnitude m

for some appropriately chosen m. When assuming N_t to be drawn from a uniform sampling grid the Fast Fourier Transform may be applied to offer an efficient matrix-vector product when we choose $\boldsymbol{A} = \boldsymbol{\mathcal{F}}_J$. Combining these thoughts results in an efficient matrix-vector product for the whole matrix $\boldsymbol{\Psi H}$, which is represented by the following two linear functions

$$f_{\boldsymbol{\Psi H}} : \mathbb{R}^N \to \mathbb{C}^N, \ \boldsymbol{x} \mapsto f_{\boldsymbol{\Psi H}}(\boldsymbol{x}) = \boldsymbol{\Psi H x}$$

and

$$b_{\boldsymbol{\Psi H}} : \mathbb{C}^N \to \mathbb{R}^N, \ \boldsymbol{y} \mapsto b_{\boldsymbol{\Psi H}}(\boldsymbol{y}) = (\boldsymbol{\Psi H})^{\mathrm{H}} \boldsymbol{y}.$$

So, by choosing $\boldsymbol{A}_J = \boldsymbol{\mathcal{F}}_J$, in each measurement, we measure $|J|$ Fourier coefficients of $b_{x,y}$ in (1) in each measurement instead of N_t time samples. Rewriting (1) in the frequency domain, we get

$$\hat{b}_{x,y}(m\omega_0) = g_{x,y}(x_i, y_i, z_i)\hat{h}(m\omega_0) \sum_{i=1}^{I} \mathrm{e}^{jm\omega_0 \tau_{x,y}(x_i, y_i, z_i)}, \quad (5)$$

with $\omega_0 = \frac{2\pi}{N_t \cdot t_s}$. From this, we define $\hat{\boldsymbol{h}} \in \mathbb{C}^{N_t}$ as the vector containing the N_t discrete Fourier coefficients of the inserted pulse.

II-A. Implementation considerations

To implement this observation strategy in hardware, a natural idea is to perform detection of the Fourier coefficients selected in $\boldsymbol{\mathcal{F}}_J$ directly in the frequency domain. A bank of narrow-band filters, consisting of either N_t fixed-frequency or m tunable filters, performs selection of m narrow-band bandpass signals resembling the desired coefficients. The output of those filters will then be processed by m detectors determining the amplitude and phase of each filter's output signal. Further, a $2m$-channel Analogue-to-Digital Converter (ADC) is required to sample the data at a very low rate, corresponding to the measurement rate of A-scans. To save some hardware effort this ADC may also be implemented at a rate $> 2m$ higher if $2m$ sample-and-hold amplifiers and a $2m$-to-1 multiplexer is incorporated. It is clear that such an implementation is restricted to observing m coefficients at most and is highly inflexible. If implemented using discrete components, a ludicrous amount of design effort, tuning and calibration is necessary to achieve sufficient performance. However, this can be mitigated to some degree with an integrated solution due to better scaling and matching of repetitive subcircuit elements.

Therefore, the more implementation-efficient approach, by uniformly obtaining N_t time-domain samples and detecting the Fourier coefficients by computing the Fast Fourier Transform (FFT) of this ensemble, is advised instead. Depending on the application different implementation strategies based on modular Commercial-Off-The-Shelf (COTS) components can be followed. An ADC converts the transducer signal to a digital representation with Nyquist rate. A microprocessor or a Field Programmable Gate Array (FPGA) may then be used to compute the FFT and to perform the compression step by selecting the relevant coefficients. However, if many channels are to be considered an FPGA offers advantages in hauling the data at bulk. As the A-scan shot rate is small compared to common processing clock rates pipelining can provide benefits. An extensive overview on different implementation approaches is given in [11]. As the task is fully determined at design-time and unlikely to change in the field, implementation in form of an Application Specific Integrated Circuit (ASIC) based on Complementary Metal Oxide Semiconductor (CMOS) technology is promising and offers advantages in cost, power dissipation and performance [12].

Another very interesting approach is introduced in [13], where an analogue/mixed signal implementation of an FFT processor, realised in integrated silicon technology, is shown. Using sample-and-hold amplifiers as means of storing analogue signal values, the structure of digital FFT processors can be applied to analogue signal processing. The adders and multipliers required can be implemented with relatively few transistors compared to their digital counterparts, saving power at the cost of computation accuracy.

I/Q-sampling is utilized in [9] to reduce the sampling rate. Although the mathematics of homodyne sampling promises ideal

performance, equivalent to Nyquist-rate baseband sampling, real-world implementations suffer from hardware imperfections such as Local Oscillator (LO)-leakage and I/Q path imbalance, as commonly known from their use in communication systems. Also, in practice a measurement system has to support a wide range of transducer specifications with a multitude of different pulse bandwidth and center-frequency values, thus requiring a high degree of flexibility when applying I/Q-sampling. As neither the ADC and the FFT processing system is close to technological limits we propose to instead implement a simple baseband-sampler operating at rate sufficiently high to support different transducers. As the compression method is able to compensate oversampling, this approach allows to achieve high linearity at a reasonably low hardware effort, thus achieving good implement efficiency while maintaining full flexibility.

The next section will outline how to choose J in theory and what recovery guarantees one can expect from this choice.

III. MEASUREMENT DESIGN

III-A. Random Fourier Sampling

The recovery performance of sparsity exploiting algorithms and the used compression strategy can be measured in terms of the RIP, if a signal y is sparse in an orthonormal basis H, which means that in $b = Hx$ the vector x is sparse. If on the other hand x is sparse, but H is not a basis anymore, but only an overcomplete dictionary, then the so-called H-RIP [14] yields the natural framework for reconstruction guarantees.

If we consider a single A-scan vector $b_{x,y}$ recorded at an arbitrary but fixed measurement position (x, y) the pulse-echo model in (1) results in $b_{x,y}$ being a linear superposition of shifted versions of the inserted pulse. It follows that $b_{x,y} = G\alpha_{x,y}$ and according to our compression scheme $y_{x,y} = \mathcal{F}_J \Sigma G\alpha_{x,y}$, where the columns of G are the discretized and shifted versions of the inserted pulse.

Now, since the total number of defects in the specimen is assumed to be small, i.e. a in (1) being sparse, we have that each $\alpha_{x,y}$ is sparse as well. This implies that each $b_{x,y}$ is sparse in the dictionary G. Let now $\hat{s} = \max_{x,y} \|\alpha_{x,y}\|_0$. However, the sparsity is not prevalent with respect to a basis but generally to an overcomplete dictionary G. This sparks the need for a modified reconstruction guarantee presented in the next definition.

Definition III.1 (G-RIP, [14]). Let U_k be the union of all subspaces spanned by all subsets of k columns of G. The measurement matrix Φ obeys the restricted isometry property adapted to G with constant δ, if
$$(1 - \delta)\|y\|_2^2 \leqslant \|\Phi y\|_2^2 \leqslant (1 + \delta)\|y\|_2^2 \quad \text{for all} \quad y \in U_k.$$
We call the smallest δ for which above chain of inequalities holds the restricted G-isometry constant (G-RIC) δ_k^G. $\square$

The following result calculates the H-RIC for the matrix $A \otimes I_{N_x N_y}$ and as such it delivers necessary conditions for efficient, stable and robust recovery to happen.

Lemma III.1. *For each $k \in \mathbb{N}$ it holds that that the G-RIC δ_k^G of the matrix A and the H-RIC of $\Psi = A \otimes I_{N_x N_y}$ are equal.*

Proof. From the definition of δ_k^G and the vectors $y_{x,y}$ we have for every (x, y) that
$$(1 - \delta_{\hat{s}}^G)\|y_{x,y}\|_2^2 \leqslant \|Ay_{x,y}\|_2^2 \leqslant (1 + \delta_{\hat{s}}^G)\|y_{x,y}\|_2^2.$$
Together with the definition of Ψ and $y_{x,y}$ and the properties of the Kronecker product the statement directly follows by applying
$$\|a\|_2^2 = \sum_{i=1}^N \|a_i\|_2^2 \quad \text{for} \quad a = [a_1^T, \ldots, a_N^T]^T$$
twice. $\square$

The lemma above illuminates how, given the proposed sensing scenario, only the properties of the A-scans and their sparse representation influence the recovery performance. Additionally,

only the worst A-scan in the sense that it is the least sparse one determines the worst case performance. Now, since randomly sub-selected Fourier matrices are known to have a low G-RIC (as outlined below), Lemma III.1 shows that the recovery performance and also the number of necessary measurements follows. Moreover, one is able to determine the magnitude $|J|$ such that the G-RIP of the appropriate order holds with a high probability, which delivers stable, robust and efficient recovery. This is also formalized in [14], where it is shown that $\delta_{\hat{s}}^G < 0.08$ is a sufficient condition for stable recovery to happen in every A-scan. As outlined in [10], for a uniformly randomly generated set J with
$$|J| \geqslant C\hat{s}\delta^{-2} \log^2(1/\delta) \log N_z \log(\hat{s}) \log^2(\hat{s}/\delta)$$
the matrix $\mathcal{F}_J$ satisfies the standard (not G) RIP with constant δ with high probability for sparsity order $\hat{s}$. Now, [15] shows that if the matrix $\mathcal{F}_J$ satisfies the RIP of order k with constant δ, the matrix $\mathcal{F}_J\Sigma$ is a 4δ-almost isometry on any set $\mathcal{S}$ of fixed points with $|\mathcal{S}| = C \exp(k)$, meaning
$$(1 - 4\delta)(\|x\|_2^2) \leqslant \|\mathcal{F}_J\Sigma x\| \leqslant (1 + 4\delta)(\|x\|_2^2)$$
for all $x \in \mathcal{S}$. This directly implies together with $\delta_{\hat{s}}^H < 0.08$ and Lemma III.1 applied for $A = \mathcal{F}_J\Sigma$ the minimal number of measurements to satisfy
$$|J| \geqslant \hat{C} \log N_z \hat{s} \log(\hat{s}) \log^2(50\hat{s}).$$
We would like to stress the fact that in the above analysis we derived a performance bound for the 3D reconstruction process, while only dealing with the restricted isometry constants associated to the dictionary of the single dimensional A-scans. In other words, since we are only compressing / subsampling in one dimension, this dimension alone determines the reconstruction performance. Further, the choice of $|J|$ in a practical setup can be based on a worst-case number of defect echoes that are expected to appear in a single A-scan. Since usually the number of defects is small, even some safety margin will still result in large compression.

III-B. Knowledge Aware Fourier Sensing

So far, we only considered the question of how many Fourier coefficients need to be sampled such that recovery is guaranteed when pursuing a random sampling scheme. However, this does not take any prior knowledge about the signals we compress into account. Since we know what waveform $h(\cdot)$ we insert into the specimen, which we have to supply to the transducer in order to excite the medium, we have a rough estimate of the spectrum of the received data after traveling through the medium and being reflected. So, the above presented sensing mechanism, which we call *random uniform* from now on, does not exploit all prior knowledge about the system at hand.

Since we approximately know the energy distribution of the measurements in the spectral domain from the inserted pulse $h(\cdot)$, we can use this approximate spectral density to maximize the quantity
$$\frac{\|y\|_2}{\|\operatorname{vec} b\|_2}$$
which measures how much energy is conserved / lost during the compression step. This is an approximation because attenuation and interference can lead to some parts being weaker than others compared to the way we excited. To this end, we propose two new sensing schemes, which are able to account for the knowledge of $\hat{h}$ in (5). Since both exploit knowledge about the spectrum of the $b_{x,y}$ we set $\Sigma = I_{N_z}$ to avoid any change in the spectrum. The first one makes the sensing design a deterministic process by selecting the Fourier coefficients of each $b_{x,y}$ indexing the k largest (in absolute value) elements in $\hat{h}$. We call this procedure the *max* approach. The second sensing scheme in some sense mediates between *random uniform* and *max* by still randomly selecting the set J, but now according to the spectral density $|\hat{h}|/\|\hat{h}\|_1$ of the inserted pulse $h(\cdot)$, which we readily interpret as the density for the sampling of the Fourier coefficients. Since in our model (1) the A-scans contain shifted versions of the pulse $h(\cdot)$, the spectrum of the observed $b_{x,y}$
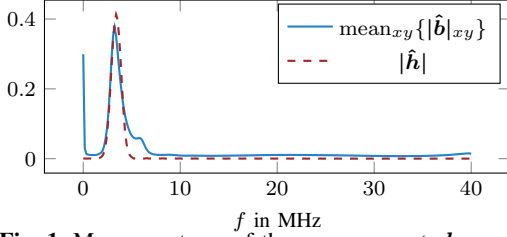
Fig. 1. Mean spectrum of the measurements $\boldsymbol{b}_{xy}$ and spectrum of the modeled pulse $\hat{\boldsymbol{h}}$. The latter is used as pdf for the *energy based* strategy.
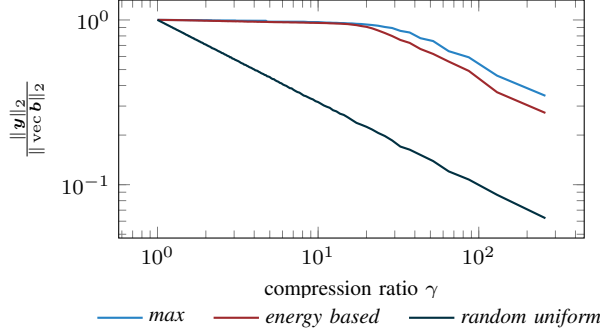
Fig. 2. Normalized energy of the measured frequency samples depending of the chosen strategy for different compression ratios γ.

can be approximated by $|\hat{\boldsymbol{h}}|$. With this *energy based* approach the probability for a Fourier coefficient to be in the set J is determined by the amplitude we expect it to have.

The following section deals with the comparison of these three methods in terms of energy conservation and reconstruction accuracy.

IV. RECONSTRUCTION RESULTS

We now carry out numerical simulations based on prerecorded fully sampled measurement data of a steel specimen. The material parameters are the same as in [5][Table II]. The compression is simulated in software [16], [17] by applying $\boldsymbol{\Psi}$ to $\mathrm{vec}(\boldsymbol{b})$, where $\boldsymbol{b}$ is the 3D measurement data of the above specimen. We define $\gamma = N_t/(2|J|)$ as the compression ratio, where the factor $1/2$ is due to the fact that we choose J only from the non redundant part of the spectrum of the real valued $\boldsymbol{b}$.

To illustrate the sampling strategies further, Fig. 1 depicts the average over (x, y) of the absolute value of the spectrum of the measurement data $\boldsymbol{b}_{x,y}$. As a comparison, we plot the Fourier coefficients $\hat{\boldsymbol{h}}$, which we know from the inserted pulse. Their amplitudes are used for both our proposed strategies. As such, $\hat{\boldsymbol{h}}$ constitutes the prior knowledge we assume to have during the measurement design process. Qualitatively, our estimated $\hat{\boldsymbol{h}}$ succeeds in approximating the spectrum of the real data, which indicates that basing our measurement design on $\hat{\boldsymbol{h}}$ is a valid approach. Fig. 2 shows the normalized remaining signal energy after compression for the three strategies. For the two random strategies we averaged over 100 realizations. We see that we can increase compression up to $\gamma = 10$ until the two proposed approaches start to reduce the energy in the measurement, while due to the nature of uniform sampling, the *random uniform* approach immediately decreases the energy with increasing compression.

The goal is now to reconstruct the sparse vector $\boldsymbol{a}$ given $\boldsymbol{y}$ using one of the above strategies. We compare two different approaches:

$$\boldsymbol{a}_{\mathrm{SAFT}} = b_{\boldsymbol{\Psi H}}(\boldsymbol{y}) \tag{6}$$

$$\min_{\boldsymbol{a}} \|\boldsymbol{y} - f_{\boldsymbol{\Psi H}}(\boldsymbol{a})\|_2^2 + \lambda \|\boldsymbol{a}\|_1, \tag{7}$$

In the uncompressed case SAFT constitutes to applying the backward $\boldsymbol{H}^{\mathrm{T}}$ to the measurement data $\mathrm{vec}\,\boldsymbol{b}$. Interestingly, this still holds true if we apply the linear compression $\boldsymbol{\Psi}$. Accordingly, (6) results in a SAFT reconstruction of the compressed measurements.

To approximate a solution to (7) we use the STELA algorithm [18]. STELA requires the Euclidean norm of the columns of $\boldsymbol{\Psi H}$. Since we do not compute the elements of $\boldsymbol{\Psi H}$ explicitly, we need to approximate the norm of its columns. From (2), (5) and the definition of $\boldsymbol{H}$ it can be seen that

$$\|[\boldsymbol{\Psi H}]_k\|_2^2 = \| \left[[\mathcal{F}_J \otimes \boldsymbol{I}_{N_x N_y}] \boldsymbol{H} \right]_k \|_2^2$$

$$\leqslant \|\hat{\boldsymbol{h}}_J\|_2^2 \sum_{x,y} g_{x,y}(x_k, y_k, z_k)$$

$$\approx \|\hat{\boldsymbol{h}}_J\|_2^2 \frac{2\pi}{(\frac{1}{2}\tan(\alpha)z_k)^2 \Delta x \Delta y} \tag{8}$$

where $[\boldsymbol{\Psi H}]_k$ is the k-th column of the matrix $\boldsymbol{\Psi H}$ and (x_k, y_k, z_k) is the defect position corresponding to the k-th grid point. To compute the third line, we use that $\sum_{x,y} f(x, y) = \Delta x \Delta y \int f(x, y) \mathrm{d}x \mathrm{d}y$ if for $x^2 + y^2 \to \infty$ it holds that $f(x, y) \to 0$ fast enough. Further, we chose λ in (7) as $\lambda = \max |b_{\boldsymbol{\Psi H}}(\boldsymbol{y})| \cdot \beta$ with $0 < \beta < 1$.

Figure 3 and 4 show the volumetric reconstruction results in two different projections for all mentioned reconstruction schemes. The code to reproduce these figures is freely available at [8]. In all cases we only use 2% of the data. Fig. 3 shows a C-scan view (projection on the x-y plane) of the complete specimen. A sketch of the ground truth is in the top figure. Fig. 4 shows a B-scan view (projection on the x-z plane) zoomed to the oblong hole on the very left of the specimen. As a first observation from Fig. 3 it can be seen that a SAFT reconstruction from the compressed data is possible. All defects are clearly visible in the SAFT image. However, compared to the sparsity-based reconstruction, SAFT produces images that are much less focused. The edges of the defect are much better preserved in the sparsity-based reconstruction. Comparing the different sampling strategies, it can be seen that the *random uniform* approach produces artifacts and false negatives. Two of the small flat bottom holes on the right are missing. The C-scans of the sparsity-based reconstruction using the *max* and the *energy-based* strategies show almost no differences.

However, moving to the B-scan views in Figure 4, it can be seen that the focusing along the z-axis varies more depending on the sampling strategy. Further, sparsity-based reconstruction results in a better focusing compared to SAFT. The *random uniform* approach leads to the best focusing in depth, since due to the whitening introduced through $\boldsymbol{\Sigma}$ the bandwidth is larger compared to the other two strategies. For the *max* approach the bell-shaped density of the pulse model in frequency domain for higher compression leads to a smaller sampled bandwidth. This reduces the resolution in z-direction in the reconstruction. However, choosing the $|J|$ largest coefficients is superior to the *random uniform* approach regarding the performance along the x-y-plane. So, our simulations support the claim that the *energy based* strategy mediates between those two strategies leading to the best overall results.

V. CONCLUSION

In this paper, we showed that the conventional Fourier based approach can be improved by including prior knowledge during the sampling process. To this end, we incorporated the prior knowledge of the spectrum by proposing two novel strategies, i.e. a deterministic and a probability based subsampling strategy. As we showed numerically, for ultrasonic NDT measurements fairly high compression ratios still lead to good reconstruction results. Moreover, the focusing along the z-axis highly depends on
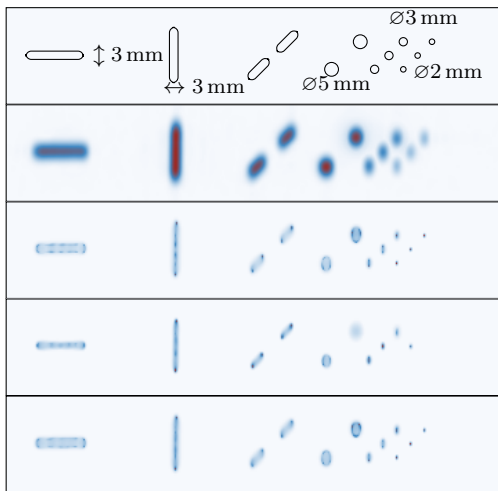
Fig. 3. Reconstruction results in x-y plane (C-scan view). The top figure shows a sketch of the specimen with dimensions $180.5\,\text{mm} \times 35\,\text{mm}$ and the depicted flat bottom holes. Second from top: SAFT with $\gamma = 50$ and strategy *max*. The remaining three show STELA reconstructions with parameters: $\beta = 0.2$, $\gamma = 50$ after 10 steps using the compression strategies *max*, *random uniform* and *energy based*, respectively.
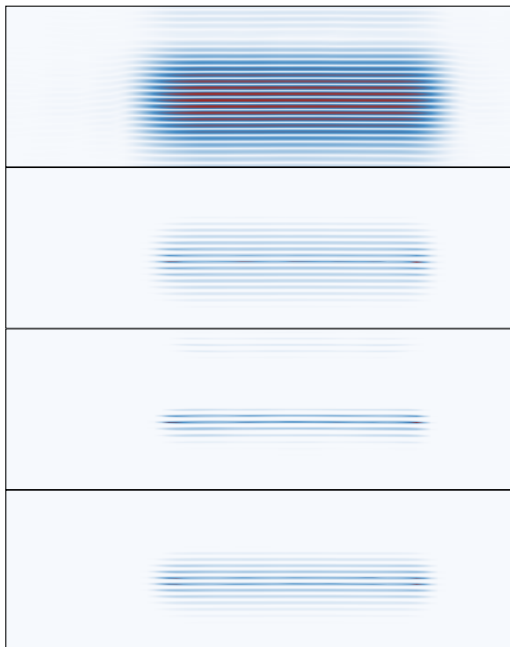
Fig. 4. Reconstruction results in x-z plane (B-scan view) zoomed to the oblong hole on the very left. Top: SAFT with $\gamma = 50$ and strategy *max*. The remaining three show STELA reconstructions with parameters: $\beta = 0.2$, $\gamma = 50$ after 10 steps using the compression strategies *max*, *random uniform* and *energy based*, respectively.

the sampling strategy chosen. Especially the hybrid *energy-based* approach unifies the advantages of the two other strategies, i.e. good resolution across all spatial dimensions. Additionally, due to this probabilistic approach it can also deal with irregularities in the medium that are not captured by the model, thus making it more robust to unexpected effects during wave propagation.

Further, we showed how formulating SAFT as a matrix-vector product makes it applicable to compressed Fourier measurements, which extends this well studied method to more sophisticated measurement approaches. We discussed our proposed architecture with regards to multiple hardware implementation strategies, outlining a path to a cost- and energy-efficient implementation in integrated mixed-signal CMOS technology.

VI. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[2] M. Spies, H. Rieder, A. Dillhöfer, V. Schmitz, and W. Müller, "Synthetic aperture focusing and time-of-flight diffraction ultrasonic imaging—past and present," *Journ. of NDE*, vol. 31, pp. 310–323, 2012.

[3] M. H. Skjelvareid, "Synthetic aperture ultrasound imaging with application to interior pipe inspection," Ph.D. dissertation, University of Tromsø, 2012.

[4] S. Semper, J. Kirchhof, C. Wagner, F. Krieg, F. Römer, A. Osman, and G. Del Galdo, "Defect detection from 3D ultrasonic measurements using matrix-free sparse recovery algorithms," in *EUSIPCO*, Sep. 2018.

[5] J. Kirchhof, S. Semper, and F. Römer, "GPU-accelerated matrix-free 3D ultrasound reconstruction for nondestructive testing," in *IEEE IUS*, Oct. 2018.

[6] N. Laroche, E. Carcreff, S. Bourguignon, J. Idier, and A. Duclos, "An inverse approach for ultrasonic imaging by total focusing point for close reflectors separation," in *IEEE International Ultrasonics Symposium (IUS)*, Oct 2018, pp. 1–4.

[7] C. Wagner and S. Semper, "Fast linear transformations in python," *arXiv preprint arXiv:1710.09578*, 2017.

[8] S. Semper, "Ultrasonic Multilevel Transforms in Cuda," 2019. [Online]. Available: https://doi.org/10.5281/zenodo.2562741

[9] T. Chernyakova and Y. Eldar, "Fourier-domain beamforming: the path to compressed ultrasound imaging," *IEEE Trans. on UFFC*, vol. 61, no. 8, pp. 1252–1267, Aug. 2014.

[10] I. Haviv and O. Regev, "The Restricted Isometry Property of Subsampled Fourier Matrices," in *Proc. of the 27th Annual ACM-SIAM Symp. on Discrete Algorithms*, 2015, pp. 288–297.

[11] M. A. Kumar and D. A. Chakrapani, "An survey of low power FFT processor for signal processing applications," *J. of Adv. Res. in Dynamical and Control Systems*, pp. 633–641, 11 2017.

[12] C. Yang, C. Wei, Y. Xie, H. Chen, and C. Ma, "Area-efficient mixed-radix variable-length fft processor," *IEICE Electronics Express*, vol. 14, 04 2017.

[13] M. Lehne, "An analog/mixed signal FFT processor for ultra-wideband OFDM wireless transceivers," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2008.

[14] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *App. and Comp. Harm. Anal.*, vol. 31, no. 1, pp. 59–73, 2011.

[15] F. Krahmer and R. Ward, "New and Improved Johnson–Lindenstrauss Embeddings via the Restricted Isometry Property," *SIAM J. on Math. Anal.*, vol. 43, no. 3, 2011.

[16] P. S. Foundation. Python. [Online]. Available: http://www.python.org

[17] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with CUDA," *Queue*, vol. 6, no. 2, pp. 40–53, 2008.

[18] Y. Yang and M. Pesavento, "A unified successive pseudoconvex approximation framework," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3313–3328, July 2017.