Compressive ultrafast pulse measurement via time-domain single-pixel imaging: supplement

Jiapeng Zhao,1,* Jianming Dai,2 Boris Braverman,3 Xi-Cheng Zhang,1 and Robert W. Boyd1,3

1The Institute of Optics, University of Rochester, Rochester, New York 14627, USA
2Center for Terahertz Waves and School of Precision Instruments & Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China
3Department of Physics, University of Ottawa, Ottawa K1N 6N5, Canada

*Corresponding author: jzhao@rochester.edu

This supplement published with The Optical Society on 30 August 2021 by The Authors under the terms of the Creative Commons Attribution 4.0 License in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

Supplement DOI: https://doi.org/10.6084/m9.figshare.15108678

Parent Article DOI: https://doi.org/10.1364/OPTICA.431455
Ultrafast pulse measurement via
time-domain single-pixel imaging:
supplemental document

In this supplemental document, we provide the experimental details, theoretical model, additional experimental results, numerical simulation data and more discussions.

1. EXPERIMENTAL DETAILS

Here we include additional details of our experimental setup. As shown in Fig. 2(a), a 800 nm NIR laser pulse (Maitai oscillator laser from Spectra-Physics) with 80 fs pulse duration and 80 MHz repetition rate is split by a 45:55 non-polarizing beam splitter (BS1). The average power is 1.02 W which corresponds to 12.75 nJ pulse energy. The diameter of the probe pulse is expanded to 30 mm by two lenses with focal lengths equal to 75 mm and 500 mm, respectively (not shown in figure). One rectangular aperture stop (3.7 mm by 6.6 mm, not shown in figure) is used to select the central part of the expanded probe pulse, providing a relatively uniform intensity distribution. The selected probe pulse is then incident on the DMD (Texas Instruments DLP3000). Note that, our DMD chip is the most basic one which has a medium micromirror size, a small amount of micromirrors and a small total size of the mirror array. Therefore, by using a more advanced commercially available DMD with a smaller mirror size (a shorter $\Delta \tau$, e.g. $\sim 11$ fs), a larger $N$ (e.g. 2560) and a larger temporal field-of-view (a larger $T$, e.g. $\sim 28$ ps), one can easily extend the potential applications of our scheme. A much large temporal field-of-view can be achieved by combining DMD and a mechanical stage, and synchronizing the temporal sampling between them. After DMD, a 250-mm-focal-length lens (L1) is used to perform the Fourier transform, and the detection ZnTe crystal ($10 \times 10 \times 3$ mm) is placed at the Fourier plane. Although the preparation of pulse trains comes from the focusing in the horizontal direction, the quadratic phase from the diffraction can only be eliminated if the Fourier transform is performed.

In the signal arm, a delay stage (not shown in the figure) with two mirrors is used to measure THz waveforms (red curves in Fig. 4 and 5) for comparison. The NIR pulse is then focused onto an 1-mm-thick ZnTe crystal using a 75-mm focal length lens (L2) to generate THz pulse. Two 4-inch-focal-length parabolic mirrors (PM1 and PM2) are used to collimate and refocus THz pulses onto the detection ZnTe. An ITO-coated glass plate is used as a dichroic mirror to combine THz and probe beams. After the detection crystal, a conventional electro-optic sampling detection unit is used, which consists of a quarter-wave plate, a Wollaston prism and a balanced photodiode detector (Hamamatsu Photonics S2386-8K). The signal from the balanced photodiode detector is finally measured by a lock-in amplifier (Stanford Research Systems SR830). The acquisition time of each measurement (one pattern on DMD or one temporal position on delay stage) in either TSPI or delay stage raster scanning is the same (100 ms). However, the integration time (time constant) on lock-in amplifier is different. The delay stage raster scanning has a integration time of 100 ms while the TSPI data is 30 ms to avoid the rolling average on lock-in amplifier. Therefore, the SNR of each measurement in delay stage raster scanning is 1.83 times higher than the TSPI data. All experimental raw data and numerical codes are available upon reasonable request.

Compared to the minimal $\Delta \tau$ of the pulse train, our laser pulse has a relatively long pulse duration of 80 fs. Therefore, adjacent replicas will be highly overlapped in time. Due to the fact that TFO replicas are mutually coherent, this overlap will lead to multi-pulse interference and the adjacent copies will coherently add up to form a new pulse. Thus, we can combine every 4 DMD columns together as effective columns to prepare pulse trains with a $\Delta \tau = 64$ fs. For the same reason, to avoid the multi-pulse interference between adjacent TFO replicas, we represent each row of Hadamard matrix $H$ as the sum of four sub-rows. Letting the Nth row be defined as $\{a_i\}$, we set the sub-rows as $\{b_i = a_i, \text{ for } i=0,1,2,3 \mod 4 \}$. For example, the 2nd row of $H_8$ is:

$$H_{8,2} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$  (S1)
We break it into a 4 by 8 matrix consisting of 4 sub-rows:

\[
H_{8,2} = \begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}.
\]

This procedure will quadruple the acquisition time but provide a precise pulse train with less interference between adjacent TFO replicas, and hence a more accurate measurement. One can skip this procedure when a short pulse duration laser is used, and hence a shorter data acquisition time. The corresponding simulation results can be found in Supplementary Section 5. As a comparison, if a pulse with a longer pulse duration is used, one has to break each row of Hadamard matrix into more sub-rows, which will lead to an increased data acquisition time. It is noteworthy that the original Hadamard matrix consists of both 1 and -1 elements. However, since our DMD can only encode non-negative values, we use shifted Hadamard matrix in our temporal encoding: \( H_{\text{shift}} = (H + 1)/2 \).

The sampling rate of our system in is 10 Hz (for data in Fig. 4 and 5). Even though the sampling speed is faster than our delay stage (~2 Hz), we do not attain the potential speed that could be attained with TFO gate. To achieve kHz-level speed, the hardware of our system, including a faster data acquisition system and a more sensitive THz detector, has to be upgraded with details given in Supplementary Section 7.

2. THEORETICAL MODEL AND COMPRESSIVE SENSING

Based on Eqn. (1) to (3) in the main text, we successfully build a TFO gate using a DMD and a lens L1 within the Rayleigh range of the focus. Although at the cost of a lower pulse power, one can further remove the spatial dependence by coupling these TFO replicas into a single mode fiber. Hence then, at the focus of L1, we get programmable pulse trains containing temporal replicas of the original input ultrafast pulse, and each replica has its individual temporal position and amplitude under deterministic control. Meanwhile, a much larger \( T \) is available if one uses the output of a frequency comb as the input source of our TFO gate.

Our THz measurement is based on the free-space electro-optic (EO) effect [1–3]. THz pulses introduce birefringence in EO crystals which change the polarization of the probe beam. By using the quarter wave plate (QWP) and Wollaston prism, the probe beam will be separated into two orthogonal polarizations with equal intensity. If THz pulses do not interact with probe pulses, balanced detectors will yield no response. Otherwise, polarization state of probe pulses will be modified leading to the unequal intensities of two polarizations, which further results in a signal \( P_s \) on balanced detectors:

\[
P_s = \int \int \frac{1}{c} |E_{\text{TFO}}(x, y, \omega_0, t)|^2 \omega_0 L n_0(\omega_0) r_{41} E_{\text{THz}}(x, y, \omega_{\text{THz}}, t) dx dy dt,
\]

where \( c \) is the speed of light in vacuum. \( L \) is the thickness of the detection crystal, and \( \omega_0 \) is the frequency of probe. \( n_0(\omega_0) \) is the refractive index of the detection crystal at frequency \( \omega_0 \), and \( r_{41} \) is the EO efficient of the detection crystal. From Eqn. (S3), we can see that the THz pulse can only be detected when that part of the THz pulse is temporally coinciding with a TFO replica. Therefore, temporally modulating the structure of TFO output \( E_{\text{TFO}}(x, y, z, \omega_0, t) \) is equivalent to temporally modulating the THz field in the same manner but leaving the probe field unchanged. However, temporally encoding the probe beam using our ultrafast TFO gate can be straightforwardly implemented using commercially available NIR DMDs, and thus remove the requirement of ultrafast THz temporal modulator, which, to the best of our knowledge, has not been reported yet.

As what we discussed in the main text, we assume that our THz pulse is temporally sampled \( N \) times, and the measured pulse is represented by a \( N \) dimensional vector \( E_d(t) \). The measurement basis is \( S \) and the corresponding coefficient vector is \( \phi \). The basic idea of compressive sensing comes from the fact that, for a \( N \) dimensional vector \( E_d(t) \), it is possible to find a basis \( R \) that most coefficients in the coefficient vector \( \phi \) are zero or very small. What the CS does is to exploit the sparsity of the signal vector, through optimization, to recover most information by only
measuring those nonzero coefficients. Therefore, one can beat the Nyquist sampling limit by sub-sampling the signal without losing too much information. To successfully implement CS, two conditions have to be fulfilled: the sparsity of the signals of interest, and the incoherence of sensing basis $S$ and recovery basis $R$ [4].

Sparsity pertains to the signals of interest that is independent of the choice of sensing basis. Luckily, most signals in nature are sparse if they are represented in appropriate bases [4]. Therefore, in most cases, we don’t have to worry about the first condition. The second condition requires that the sampling and recovery bases are mutually incoherent. The coherence between two bases is defined as the largest correlation between any two elements in $S$ and $R$:

$$\mu(S, R) = \sqrt{N} \times \max\{|\langle S_k, R_j \rangle|\}$$  \hspace{1cm} (S4)

where $N$ is the dimension of the basis and $1 \leq k, j \leq N$. Therefore, two bases are maximally incoherent when $\mu(S, R) = 1$. One example of maximally incoherent bases is the Fourier conjugate bases (in our case the conjugate bases are time and frequency). This explains why random basis, no matter in spatial single-pixel imaging or temporal ghost imaging, can always work as the sampling basis in CS. However, we will show that random basis is usually not the ideal choice of the sampling basis.

We first assume that the signal $E_d(t)$ can be measured using basis $S$ with the measurement result coefficient vector $\phi$: $\phi = SE_d(t)$. Now we assume that there is a basis $R$ which is incoherent with $S$. If we assume that most coefficients under this representation $R$ are close or equal to zero, we can recover the signal $E_d(t)$ by $\ell_1$-norm minimization. If we assume the system is in the ideal case without noise, the solution coefficient vector $\phi'$ will be found by solving the convex optimization problem [4, 5]:

$$\min \|RE_d(t)\|_{\ell_1}, \text{subject to } SE_d(t)' = \phi'$$  \hspace{1cm} (S5)

where $\|\cdot\|_{\ell_1}$ is the $\ell_1$-norm. If the coherent coefficient $\mu$ between two bases $Q$ and $T$ is very small, i.e. two bases are incoherent, the THz field vector $E_d(t)$ can be reconstructed with $M \geq O[K\log(N)]$ measurements, where $K$ is the number of nonzero components of vector $RE_d(t)$ [6]. Therefore, we can surpass the Nyquist rate but still recover the field with a high fidelity.

![Fig. S1. Fidelity and RMSE of recovered THz pulses as functions of CR under different input pulse durations.](image)

3
From what we describe above, one can see that the goal of using CS is to achieve high fidelity signal recovery via sub-sampling the signal. Therefore, we need to carefully reconsider the use of random basis. Even though random basis is always a choice of sampling basis, no matter in spatial or temporal measurement, it is not ideal in most cases for the following reason. Due to the randomness in the sampling process, the recovered signal is usually noisy if one strictly sub-samples the signal. Therefore, when the random basis is used, one usually has to oversample the signal to average out noise, which loses the point of implementing CS. This has been demonstrated in spatial ghost imaging and one can also find the existence of over-sampling in previous temporal ghost imaging systems [7, 8]. Thus, a fixed basis with deterministic sampling vectors is a better choice in real applications.

From a practical point of view, we use the Total Variation Minimization by Augmented Lagrangian and Alternating Direction Algorithms (TVAL3) package, which provides 4 different total variation based minimization models to solve the minimization problem. Since the noise always exists in practice, we are using TV/L2 model for denoising. The package was provided by Chengbo Li, Wotao Yin and Yin Zhang at Rice University [9]. Fig. S1 shows the simulated fidelity and RMSE as functions of compression ratio (CR) under different pulse durations. We show that the performance of our TSPI only slightly depends on the pulse duration of the input pulse of TFO gate, which demonstrates that our approach is practical to different pulse durations. Similar to the experimental data shown in Fig. 4(h), turning points in fidelity and RMSE are found for all pulse durations. We speculate that this phenomenon comes from the joint effect of two facts. Firstly speaking, even though the row vectors of the Walsh-ordered Hadamard matrix represent different frequencies in an increasing order, the information measured in each row vector is different because each row vector can contain multiple frequency terms. Secondly, the spectra of the object is not uniform. Therefore, when we use some row vectors to sample a specific object, we may measure little information when the object carries little information at those frequency terms which can be intensively sampled by these row vectors, or the object carries rich information at those frequency terms which can be barely sampled by these row vectors. Since we are sampling the object using each row vector sequentially, those row vectors, as what we describe above, happen to have the CR at each turning point. It is worthy noting that we do not have substantial data to support our speculation of turning points. Therefore, further experiments are required to verify our hypothesis.

3. ARBITRARILY STRUCTURED TFO OUTPUTS

Here we would like to show some examples of TFO outputs under different modulations. Similar as the spatial fan-out, TFO can produce multiple coherent copies of the input field which are separated in the time domain, and the intensity of each replica can be tuned by turning on different amounts of mirrors. In Fig. S2(a), since all the TFO replicas are coherent, they interfere to each other and form a pulse with a much longer pulse duration (9.73 ps). The top curve in Fig. S2(a) has the same envelope as what we use in the experiment by turning on all DMD mirrors (Fig. 2(c)). After turning different amounts of mirrors in each column of DMD, we can prepare a nearly-uniform top-hat envelope so that the ratio between the maximum and minimum is only 1.28, which is shown as the bottom curve in Fig. S2(a). To clearly show other modulations, we use this top-hat curve as the reference, which are the top curves in Fig. S2(b) and (c). Based on this nearly-uniform intensity envelope, we can further add a temporal structure to it, the same structure shown in the bottom of Fig. 2(c), so that all TFO replicas have similar intensity (Fig. S2(b)). This actually mimics the generation of frequency comb in the temporal domain, and our device is more flexible and can produce more complicated temporal structures. Moreover, as what we show in Fig. S2(c), we can also generate a top-hat pulse with a different pulse duration (3 ps) by turning off all mirrors outside the central part. Last but not least, we can add a complex intensity envelope as well. For example, in Fig. S2(d), the envelope consists of a V shape envelope with its valley at 6.7 ps and 7 Gaussian distributions centered at 2 ps, 3.7 ps, 5.4 ps, 6.7 ps, 8.4 ps, 9.8 ps and 11.2 ps respectively. Even though the complex structure is not as good as the theoretical design due to the nonuniform illumination of the input pulse, it shows the possibility of arbitrarily engineering the temporal structure of TFO replicas. For all the structures shown above, one can further spend more time to optimize them by tuning the amount of mirrors in each column to compensate the defects from nonuniform illumination. However, this would beyond the scope of our work.
Fig. S2. Examples of TFO outputs. (a): The top curve is the measured envelope with all DMD mirrors on, while the bottom envelope is the modified envelope to have a nearly-uniform intensity by turning off different amounts of mirror in each DMD column. (b): The top envelope is the bottom curve in (a). The bottom curve is the modulated temporal structure based on the top curve by turning off some DMD columns. (c): The top envelope is the bottom curve in (a), and the bottom curve is a top-hat pulse with 3 ps pulse duration by turning off all outside DMD columns. (d): The top curve is a complex intensity envelope by turning on different amounts of micromirrors in each DMD column. Bottom curve is one example of the temporal structure based on the top curve. All the top curves are normalized and shifted by 1 unit to provide a better comparison to the bottom curves.

4. MEASUREMENT OF THE TEMPORAL INTERVAL OF PULSE TRAIN

As what we show in the main paper, the theoretical temporal interval, which is the temporal resolution in our TSPI system, can be calculated as: \( \Delta \tau = \sin(2\alpha) d/c \). Therefore, we can find a theoretical value of 14.66 fs. To experimentally measure the temporal interval \( \Delta \tau \), we first prepare a pulse train which is the same as what show in the middle figure of Fig. 2(c). Each effective TFO replica consists of 5 DMD columns, which means 5 columns are set to ‘on’ state. There are 15 DMD columns are set to ‘off’ state between two effective replicas, and each DMD column corresponds to one temporal interval \( \Delta \tau \). Therefore, the separation between the peaks of two adjacent effective TFO replicas is 20 DMD columns, which is 20\( \Delta \tau \) in the time domain. Since there are 30 effective replicas at the ‘on’ state, the separation between the first effective replica to the last effective replica is 580 DMD columns, which is 580\( \Delta \tau \). We then calculate the temporal separation \( \Delta T \) = 9.28 ps by measuring the position of the first effective replica and the last effective replica in 9 measurements. The temporal interval can be found as: \( \Delta \tau = \Delta T / 580 = 16.00 \) fs with a standard deviation of 0.01 fs. Note that 0.01 fs is the average uncertainty of each micromirror, and the uncertainty of the entire pulse train is 6.08 fs, which corresponds to an uncertainty of 1.82 \( \mu \)m in spatial separation between the first and last effective replica. The 1.82 \( \mu \)m spatial uncertainty is a reasonable number according to the performance of our translation stage. Therefore, we believe that the temporal resolution induced by each micromirror is 16.00\( \pm \)0.01 fs.

5. SIMULATION

A Matlab code is developed to simulate our TFO gate and TSPI based on the theoretical model in Supplementary Information Section 2. We first simulate the results shown in Fig. 4 in the main manuscript. As shown in Fig. S3, our experimental data precisely matches our simulation results. When the CR is 20\%, both experimental and simulation data show the lack of high frequency components above 1.5 THz. Meanwhile, as what we discuss in the main manuscript, the weak side lobe at 2.4 THz, which corresponds to the sampling frequency of the 25th row of the Walsh-ordered Hadamard Matrix (WOHM), is also verified by simulation. When the CR goes to 30\%, the main part of the spectrum has been accurately recovered except for the weak side
lobe, which is also observed in our experimental data. When the CR goes to 40%, the side lobe disappears and the spectrum of the CS recovered pulse is almost identical to original spectrum. Therefore, these simulation results indicate that our experimental data has a high agreement with the theoretical model.

We then focus on the role of pulse duration of the input pulse in TFO, which has a critical impact to TFO outputs. Basically, all TFO replicas are coherent unless the input pulse is spatially incoherent. Therefore, due to the long pulse duration, the interference between adjacent two or more sub-pulses will result in temporal structures under control. For example, in Fig. S2(a), when the input pulse has a duration of 80 fs and all columns are on, we can get a 9.73 ps long pulse in the time domain. As shown in Fig. S4(a)-(c), even when we separate two effective TFO copies, each one is the coherence interference of 4 TFO copies, with 192 fs, one still cannot get a unit contrast of pulse trains when the pulse duration is longer than 60 fs. If we further reduce the separation to 64 fs, we can only get pulse trains with unit contrast when the input pulse duration is around 16 fs. When no separation is added between two effective TFO copies, a 8.19 ps long top-hat pulse is generated in the time domain even the input pulse duration is 40 fs. Pulse trains with a unit contrast are only available with input pulse duration is equal to 10 fs if there is no additional separation between adjacent replicas.

Then we turn to figure out the impact of pulse duration on recovered THz signals. In fact, we find that pulse duration, which leads to imperfections in pulse trains, only significantly affect the quality of pulses recovered by WOHM directly. Pulses recovered by CS, even both methods use the same encoding method, have similar fidelity and RMSE (difference less than 0.5%) under different pulse durations. As shown in Fig. S5(a)-(c), when input pulse duration is long and pulse trains have poor contrasts, THz signals recovered by WOHM have both over-estimations and under-estimations compared to original signals. This effect will get weaker when the pulse duration decreases.
duration becomes short and the contrast of pulse trains gets close to unit. However, when pulse duration becomes ultra-short, for example 16 fs, one would expect the pixelization effect in the recovered THz signal as shown in Fig. S5(g) and (h). As a comparison, THz signals recovered by CS, shown in Fig. S5(d)-(f) and (i), are much less affected by pulse duration and the quality of pulse trains, which demonstrates the robustness of our TSPI. Therefore, our TSPI is more robust against temporal fluctuations of the input pulse of TFO gate.

Next, we will demonstrate that our TSPI is also robust against temporal distortions which comes from the imperfect TFO gate. We first prepare a simple distortion by making the envelope of pulse trains asymmetric. The ratio between the maximum and minimum in Fig. S6(a) is about 10, which is the same as the pulse train shown in Fig. 5(b). As one can see in Fig. S6(b), asymmetric pulse trains lead to the over-estimation at the tail of THz pulse and the under-estimation at the beginning of THz pulse when we use WOHM recovery directly. However, the defects from the distorted TFO gate do not affect the performance of TSPI (Fig. S6(c)). Then we make the distortion more complicated by moving the Fourier plane of DMD 25 mm away from the detection crystal. Due to the spread of k-vectors of different TFO replicas, each copy will intersect with different transverse position of the THz pulse. Considering the Gaussian distribution of the THz pulse and the asymmetric structure of pulse trains, the measured envelope of pulse trains at the detection crystal plane have an irregular envelope as shown in Fig. S6(d). The abnormal envelope of pulse trains leads to the over- and under-estimation of THz pulses. Nevertheless, our THz pulse recovered by CS is still robust to this more complicated distortion. The experimental verification and explanation of these two distortions have been shown in Fig. 5 of the main manuscript. In fact, the envelope of pulse trains plays a much less important role in CS recovery than WOHM or random recovery. One can even use a highly distorted TFO gate generating random intensity envelope of pulse trains to recover THz signals using our TSPI approach. The results are shown in Fig. S6(g)-(i). One can figure out that the random intensity envelope of pulse trains leads to a THz pulse with randomly distributed under- and over-estimations when WOHM recovery is used. However, in Fig. S6(i), the THz signal recovered by CS only gets slightly affected. Therefore, based on both experimental verification and theoretical analysis, we can claim that, as long as the correlation between temporal position of each TFO replica and intensity envelope of the entire pulse train is not destroyed, the recovered THz pulse will be robust against distortions on TFO. 

Fig. S4. (a)-(c): Pulse trains with different input pulse duration: 100 fs, 80 fs and 60 fs respectively. The temporal separation between two effective columns is equal to 192 fs. The contrast in (a) is 53.59% while it becomes 91.97% in (b), which matches our experimental observation. (d)-(f): Pulse trains with different input pulse duration: 60 fs, 40 fs and 16 fs respectively. The temporal separation between two effective columns is equal to 64 fs. The contrast in (d) is 19.43% which is too poor to use. The contrast becomes 53.52% in (e) which is good enough for both TSPI and WOHM recovery. Pulse trains with different input pulse duration: 40 fs, 16 fs and 10 fs respectively. The temporal separation between two effective columns is equal to 16 fs. The contrast in (h) is 86.06%.
Fig. S5. (a)-(c): Recovered THz pulses using WOHM directly. The encoding process is the same as what we do in the experiment, i.e. each row of WOHM is represented by the sum of 4 sub-rows, which results in the corresponding pulse trains shown in Fig. S4(a)-(c) respectively. (d)-(f): Recovered THz pulses using CS and pulse trains in Fig. S4(a)-(c) respectively. CR is 40% for all three recoveries. (g): Recovery using WOHM directly. The pulse duration is 16 fs. (h): The zoom-in figure of 4-6 ps in (g). (i): CS recovery with a pulse duration of 16 fs.

gate.

6. MACHINE LEARNING

A simple 12-layer CNN is developed to classify the compressive signals of different samples under weak SNR, and the architecture is shown in Fig. S7. The input signal layer (1×128) is connected to the first convolution layer with 16 1×4 filters. Then a batch normalization layer and a ReLU layer are implemented sequentially, which normalizes the activations and is the nonlinear activation function respectively. After the first stage, a max pooling layer with a 1×2 pooling size is used to localize the features. Then the second stage starts with another convolution layer consisting of 8 1×4 filters. Similar to the first stage, a batch normalization and a ReLU layer are used. After this, we enter the classification stage with two fully connected layers (8 neurons and 3 neurons respectively). At last, a softmax layer, which normalizes the output from the fully connected layer, and a classification layer are adopted.

Five samples are used to train this CNN. Each sample has 1000 sets of data from our ultrafast THz TDS system so that in total 5000 sets of data are used for training. Each set of data is measured under a sampling rate of 20 Hz or 50 Hz with an integration time (time constant) of 3 ms, which corresponds to acquisition times for each DMD pattern of 50 ms and 20 ms respectively. However, due to the limited data readout time between lock-in amplifier and computer (~15 ms), there are only 3 data points inside each acquisition time of 50 ms. For the same reason, only 1 data point for the acquisition time equal to 20 ms. Therefore, the effective measurement time of each DMD pattern in ML is 9 ms (for 50 ms acquisition time) and 3 ms (for 20 ms acquisition time). Since our THz pulse is attenuated to 3 times weaker, the final SNR of these data sets is about 10 (for 20 Hz sampling rate) or 17 (for 50 Hz sampling rate) times weaker than the data shown in Fig. 4 and 5. Due to the poor SNR, the accuracy of the successful classification without training is less than 35% at 20 Hz sampling rate. For example, 34.20% in the confusion matrix shown in Fig. S7(b).

The training process shuffles these sets of data at the beginning of every epoch and 6 epochs are used. In order to provide a more stable network, a batch size of 150 sets of data is used, which further leads to the maximal number of iterations at 60. The entire training process takes about 400 s with a constant training rate of 0.01. We then use 2000 sets of data (400 sets of each sample) to validate the performance of our CNN, and the successful classification probability is about 97.5% at 40% CR as shown in Fig. S7(c).
7. PERFORMANCE LIMITATIONS AND CORRESPONDING SOLUTIONS

As what we mentioned in the main text, one advantage of our technique over raster scanning is the use of CS algorithm and DMD, which provide a better SNR for weak signals, a faster sampling speed, a better sampling accuracy and a higher flexibility. In theory, the measurement time of raster scanning using delay stage scales up linearly with the measurement points $N$ in the 1-D temporal measurement, and scales up with $N^2$ in spatial characterization. Therefore, if one wants to fully characterize the THz pulse in both space and time domain, the measurement time will scale up with $N^3$, which leads to a huge time consumption. However, CS can fundamentally enhance the measurement efficiency by a factor more than 3 as what we shown in the main paper. That is to say, for a 3-D characterization problem, the time consumption will be only 1/27 of the raster scanning method using delay stage. Therefore, if we combine our previous CS method on THz spatial sampling with the current temporal CS approach, the measurement efficiency will be significantly improved [8]. From a practical point of view, as what we mention in the main paper, our system is faster than a delay stage (4 Hz) [10], which is working at 10 Hz. However, we does not fully utilize the advantage of DMD in sampling speed, and there are two limiting factors stop us from fast measurements at kHz level, which is the potential speed limit of the system. The first limiting factor is the poor SNR of our THz pulse. The signal becomes noisy when the integration time (time constant) on lock-in amplifier is below 10 ms and hence the recovered pulse has a poor quality (as shown in Fig. 5 in the paper). The second limiting factor is the communication speed between lock-in amplifier and our data acquisition program. The data read-out time from lock-in amplifier to our computer program is about 15 ms. Therefore, the maximal sampling rate with this lock-in amplifier is about 66.67 Hz, which is far below the speed limit of DMD. Hence, to improve the measurement speed to kHz level, not only a stronger THz pulse is required, a faster digital signal processing card with synchronization is also necessary. Luckily, both requirements are not hard to accomplish if the hardware are upgraded. For example, researchers have demonstrated the possibility of using a commercially available fiber-based THz system to generate high SNR THz signals, and perform fast measurement over 2 kHz with a fast data acquisition card [11]. Even though due to the long response time of silicon, the speed of their scheme cannot go beyond 4 kHz, our methods, including our previous spatial CS approach, have no such limitation [8]. Therefore, we believe that, after upgrading the hardware, kHz-level sampling rate is applicable to our approach, and it is even possible to reach the speed limit of DMD at 20 kHz.

Another limitation is the temporal field of view. As what we discuss in the main text, the temporal field of view is limited to 28 ps. However, this limitation comes from the commercially
Fig. S7. (a): Architecture of our CNN. Conv stands for convolution layer while BN and ReLU stand for batch normalization layer and rectified linear unity layer respectively. MP is the max-pooling layer and FC is the fully connected layer. SM is the softmax layer. (b): Confusion matrix before training. Due to the poor SNR, the classification accuracy is only 34.2%. The green numbers are accuracy of each sample while the red numbers are failure probabilities. The CR is 40%. (c): Confusion matrix after training. The CR is 40%.

available DMDs. One can further fabricate a customized DMD, or similar devices as long as it can introduce a spatial-temporal chirp, with a larger field of view, or cascade multiple DMDs to introduce more chirps to achieve a better field of view. One could also combine DMD and a translation stage with a long scanning range to extend the temporal field of view. This will not cost too much time since the stage only need to scan a few times. Therefore, this limitation should not be an obstacle.

REFERENCES

1. Q. Wu and X.-C. Zhang, “Free-space electro-optic sampling of terahertz beams,” Appl. Phys. Lett. 67, 3523–3525 (1995).
2. Y.-S. Lee, Principles of terahertz science and technology, vol. 170 (Springer Science & Business Media, Berlin, 2009).
3. X.-C. Zhang and J. Xu, Introduction to THz wave photonics, vol. 29 (Springer, New York City, 2010).
4. E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” IEEE Signal Process. Mag. 25, 21–30 (2008).
5. J. Romberg, “Imaging via compressive sampling,” IEEE Signal Process. Mag. 25, 14–20 (2008).
6. E. Candes and J. Romberg, “Sparsity and incoherence in compressive sampling,” Inverse Probl. 23, 969 (2007).
7. P. Ryczkowski, M. Barbier, A. T. Friberg, J. M. Dudley, and G. Genty, “Ghost imaging in the time domain,” Nat. Photonics 10, 167–170 (2016).
8. J. Zhao, E. Yiwen, K. Williams, X.-C. Zhang, and R. W. Boyd, “Spatial sampling of terahertz
fields with sub-wavelength accuracy via probe-beam encoding,” Light. Sci. & Appl. 8, 1–8 (2019).
9. C. Li, “Compressive sensing for 3d data processing tasks: applications, models and algorithms,” Tech. rep. (2011).
10. “Newport Precision Long-Travel & High-Speed Motorized Actuators,” https://www.newport.com/mam/celum/celum_assets/resources/LTA - _User_s_Manual.pdf?1 (2018).
11. R. I. Stanchev, X. Yu, T. Blu, and E. Pickwell-MacPherson, “Real-time terahertz imaging with a single-pixel detector,” Nat. Commun. 11, 1–8 (2020).