Published in final edited form as:

J Am Soc Mass Spectrom. 2020 November 04; 31(11): 2338–2347. doi:10.1021/jasms.0c00276.

Accelerating Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry Imaging Using a Subspace Approach

Yuxuan Richard Xie^{1,4}, Daniel C. Castro^{2,4}, Fan Lam^{1,4,*}, Jonathan V. Sweedler^{1,2,3,4,*}

¹Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

²Department of Molecular and Integrative Physiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

³Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801

⁴Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Abstract

We present a subspace methodology that accelerates data acquisition using Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometry imaging (MSI). For MSI of biological tissue samples, there is a finite number of heterogeneous tissue types with distinct chemical profiles that introduces redundancy in the high-dimensional measurements. Our subspace model exploits the redundancy in data measured from whole-slice tissue samples by decomposing the transient signals into linear combinations to form a set of basis transients with the desired spectral resolution. Decomposing transient signals allowed us to design a strategy that acquires a subset of long transients for basis determination and short transients for the remaining pixels, drastically reducing acquisition time. The computational reconstruction strategy can maintain high mass resolution and spatial resolution MSI while providing a ten-fold improvement in throughput. We validated the capability of the subspace model using a rat sagittal brain slice imaging dataset. Comprehensive evaluation of the quality of the mass spectral and ion images demonstrated that the reconstructed data produced by the reported method required only 15% of the typical acquisition time and exhibited both qualitative and quantitative consistency when compared to the original data. Our method enables either higher sample throughput or higher resolution images at similar acquisition lengths, providing greater flexibility in obtaining FT-ICR MSI measurements.

Graphical Abstract

*Corresponding Authors: Fan Lam, Postal: 405 N. Mathews Avenue, Urbana, IL 61801, fanlam1@illinois.edu, Jonathan V. Sweedler, Postal: 600 S. Mathews Avenue, Urbana, IL 61801, jsweedle@illinois.edu.

Supporting Information:

Supporting methods: Experimental details for MS imaging preparation; computational details for the subspace model. Supporting figures: Figure S1. The evaluation pipeline; Figure S2. The basis transients from 4000 sampled long transients; Figure S3. Peak features recovered in the truncated transients; Figure S4. Spectral reconstruction evaluation; Figure S5. Spectra reconstructed using different numbers of basis transients; Figure S6. Distribution of the Pearson correlation between ion images; Figure S7. Visualizations of ion images with low correlations; Figure S8. Scatter plot of the color-coded PCA scores; Figure S9. K-means segmentation masks (PDF).

INTRODUCTION

Mass spectrometry imaging (MSI) is widely used for visualizing and mapping multiple chemical species by determining their ion intensity distribution across a sample surface.^{1–4} As an untargeted approach, MSI does not require prior knowledge of the chemical species to be analyzed, a capability that has been exploited for many applications, including clinical diagnostics,^{5,6} drug metabolism studies,^{7,8} and localization/characterization of biomolecules within tissue samples.⁹ Among the available ionization methods for MSI, matrix-assisted laser desorption/ionization (MALDI) is frequently used because its soft ionization allows analysis of intact biomolecules over a large range of molecular weights,¹⁰ including metabolites and lipids.^{11,12} These measurements are aided by the advancement of high-resolution mass spectrometers. Fourier transform (FT)-based mass analyzers offer better performance in terms of mass resolution and mass accuracy compared to other routinely used mass analyzers, such as time-of-flight (TOF). Currently, FT-ion cyclotron resonance (ICR) mass spectrometry (MS) provides the highest mass resolution and mass accuracy available.^{13,14}

Mass resolution, mass accuracy, and acquisition time are all interrelated when using FT instruments because the longer the acquisition time, the more precisely one can determine the frequency of the ion cyclotron motion.¹⁵ Not surprisingly, higher mass resolution and mass accuracy aid in confident analyte identification.^{14,16,17} Recent developments in absorption mode processing using a 9.4 Tesla FT-ICR system achieved >250 000 mass resolution (full width half maximum) for multiple lipid species,¹⁸ and a 21 Tesla FT-ICR instrument exceeded 500 000 mass resolution while retaining sub-ppm mass accuracy.¹⁹ Mass resolution is improved with increasing magnetic field strength when the transient length of the acquired time domain signal is constant. Increasing transient length also improves mass resolution and mass accuracy without requiring a larger magnet. However, there are practical limits to how long a transient can be acquired because ions will collide with residual molecules inside the ICR cell, restricting the duration for which a transient can be recorded before the signal dies out.^{15,16} Furthermore, for an MSI experiment where oftentimes tens of thousands of spectra are obtained, acquiring the data with high mass resolution, mass accuracy, spatial resolution, and large tissue coverage leads to exceedingly long acquisition times. Although MSI using FT-ICR instruments can achieve significantly better mass resolution and mass accuracy than other approaches, it is limited in throughput if the goal is to obtain the previously mentioned properties simultaneously. Here we present a subspace approach to accelerate data acquisition for FT-ICR-based MSI, increasing its throughput. The method maintains high mass resolution and mass accuracy through a subspace-based scanning strategy and a computational reconstruction method with minimal loss of fidelity. In addition to improving throughput, shortening the acquisition time is

desirable when working with volatile MALDI matrices because the matrix-coated samples may not remain stable under vacuum throughout an extended imaging experiment.²⁰

Multiple approaches have been developed to improve MSI acquisition. One technique is to acquire higher-resolution mass spectra at a lower spatial resolution during the initial acquisition, and then use a complementary high spatial-resolution imaging modality to enhance the final spatial resolution of the ion images; these complementary modalities include optical microscopy,^{21,22} electron microscopy,²³ and infrared spectroscopy.²⁴ Another approach is to acquire high spatial-resolution data with lower resolution mass spectra at each pixel (i.e., shorter transients for FT-ICR) and use signal processing techniques to enhance the mass spectral resolution. Representative examples include methods that impose parametric models on the temporal signals or on the spectra to achieve spectral “super-resolution” by estimating the model parameters from short transient data and synthesizing higher-resolution spectra using the parameter estimates. In particular, the filter diagonalization method (FDM) can provide higher mass resolution for shorter transients, ultimately speeding up high-resolution FT-ICR MS experiments.²⁵ However, applying FDM on the temporal signal requires selection of specific frequency windows and time-consuming, nonlinear parameter fitting of each pixel. Therefore, for imaging chemically complex tissues with thousands of pixels, each with a large set of frequency peaks, it may become computationally impractical to apply FDM.

Recently, several techniques from other imaging modalities have been used for MSI, such as compressed sensing (CS)^{26,27} and super-resolution reconstruction.²⁸ The CS-based method randomly samples a partial number of pixels and reconstructs hyperspectral images by solving nonlinear optimization problems, thus effectively reducing the data acquisition time. These methods have yet to take full advantage of the inherent redundancy in the high-resolution, high-dimensional MSI data. Some implementations focus on the instrumentation and data processing to improve the data acquisition efficiency for FT-ICR MS,^{29,30} but require highly modified systems that may hinder their wide-spread application.

As illustrated in Figure 1, we present an approach to accelerate FT-ICR MSI by exploiting the low-rank property of the high-dimensional hyperspectral function, i.e., the ensemble of spatially resolved temporal transients from a tissue imaging dataset to form an approximately low-rank matrix. This low-rank property indicates that the desired long temporal transients, although high dimensional, reside in a lower-dimensional subspace, meaning they can be accurately approximated using a small number of basis elements, each of which captures a unique set of frequency components. As a result, if the basis can be predefined, determination of the transient at each pixel can be formulated as the estimation of a small set of weighting coefficients that linearly combine with the basis to synthesize the desired high-resolution transient. This enables the use of special acquisition and reconstruction strategies to achieve fast FT-ICR MSI with simultaneous high mass and high spatial resolution. Specifically, we sample long transients with the desired mass resolution from a randomly selected subset of pixels on the tissue sample (e.g., 10% of the total number of pixels). The remaining pixels are sampled with much shorter transients (e.g., 5% of the original long transient length). During reconstruction, the measured long transients are used to estimate the basis transients of the subspace model, which will then be used to

reconstruct long transients from the acquired short transients through a least-squares fitting for the entire tissue sample.

The capability of the subspace-based method to accelerate high-resolution MSI data acquisition on whole-slide tissue samples is demonstrated on a rat sagittal brain slice with 39 775 pixels at a spatial resolution of 50 μm acquired using a 7T FT-ICR mass spectrometer. The MSI data from the conventional long transient acquisition, which serves as the gold standard, versus the reconstructed data using accelerated acquisition, exhibited both qualitative and quantitative consistency. Most importantly, the required time for the acquisition is eight-fold shorter than the original standard acquisition approach. Our subspace method achieved considerable improvement over a standard acceleration strategy that acquires uniformly short transients for all pixels with an equivalent acquisition time. Furthermore, the mass spectra and ion images produced by our method, at times, exhibited improved signal-to-noise ratios (SNRs) and higher contrast, which may be due to the denoising effects introduced by the dimensionality reduction of the subspace model. Our subspace approach can also be combined with CS-based methods or other spatial resolution-enhancing techniques (such as using complementary imaging modalities) to achieve further improvements by introducing a spatial regularization term into the reconstruction formulation. We expect the acceleration capability provided by our subspace method to be useful for increasing the throughput of high-resolution MSI without compromising mass and spatial resolution, or for obtaining higher-resolution images without substantially increasing imaging time.

SUBSPACE THEORY

Subspace Modeling of FT-ICR MSI Data.

For FT-ICR MS, the individual pixels are acquired as temporal signals called transients. The transients are combinations of many sinusoidal components corresponding to individual m/z values that can be resolved after Fourier transformation. Let N_r denote the number of pixels to scan over the tissue sample of interest, and T the duration of each transient, which depends on the sampling time interval determined by the mass range of interest and the number of sampled points per transient N_T , to achieve a specific spectral resolution, making the total acquisition time for a standard uniform scanning paradigm $N_r \times T$. Thus, the acquisition time can be prohibitively long if very high-resolution imaging over a large tissue section is desired. However, the unique physical property of temporal transient-based sampling enables the use of a special subspace modeling strategy for faster scanning and reconstruction. Specifically, we adapted a subspace approach previously used for magnetic resonance spectroscopic imaging.^{31–33} For MSI, this model assumes each transient, though containing many frequency components, can be accurately approximated by a linear combination of a substantially smaller number of basis functions, denoted by $\{\phi_l(t)\}_{l=1}^L$, which are also referred to as the temporal basis transients. Thus, if we denote the high-dimensional MSI function of interest as with $s(r, t)$, with $s(r_n, t)$, $n = 1, 2, \dots, N_r$, the transient of interest with the desired mass spectral resolution at a specific pixel r_n , we can mathematically express our subspace model of $s(r_n, t)$ as:

$$s(r_n, t) = \sum_{l=1}^L c_l(r_n) \phi_l(t) \quad (\text{Eq.1})$$

where $\{c_l(r_n)\}_{l=1}^L$ are a set of coefficients that specify the pixel-dependent contributions of the L basis functions to construct the transients of interest, and L is the model order, which can be defined by users and expected to be much smaller than N_T . This model implies that if $\phi_l(t)$ can be predetermined, we can transform the imaging problem into the estimation of $c_l(r)$ instead of sampling and reconstructing the long transients for individual pixels independently. Since the degrees of freedom of $c_l(r)$ are dramatically smaller than the original high-dimensional $s(r, t)$ (from $N_r \times N_T$ to $N_r \times L$), our subspace model enables a potentially significant accelerated acquisition by reducing the number of temporal samples needed for each pixel from the tissue section because the spectral resolution here is specified by the predefined $\phi_l(t)$. For a discretized representation of the MSI data, the Casorati matrix $\mathbf{S}$ composed of temporal samples of all the pixels will be a low-rank matrix with rank = L (see the Supporting Information for detailed formulations). Therefore, it allows us to design new approaches for spatial scanning and reconstruction for MSI with simultaneous high-spatial and -spectral resolution, as described in the following sections.

Transient Subspace Estimation and Whole Tissue MSI Reconstruction.

Based on the model in Eq. (1), the imaging problem becomes two decoupled steps. The first estimates the transient basis $\{\phi_l(t)\}$ and the second estimates the spatial coefficients $\{c_l(r)\}$. Moreover, only a small subset of transients is needed for determining $\{\phi_l(t)\}_{l=1}^L$ because the number of basis vectors L required to capture the majority of the energy of the data matrix is small (e.g., in a test data set, a rank of 100 can already capture more than 95% of the signal variations for a large set of long transients, as illustrated in Figure 2). To this end, we propose the following sampling strategy that acquires two complementary datasets,

$S_1 = \{s_1(r_d, t_i)\}_{d,i=1}^{D, N_T}$ with long transients of duration T for a random subset of D pixels to

determine the basis transients $\{\phi_l(t_i)\}_{i=1}^{N_T}$ with the desired spectral resolution, and

$S_2 = \{s_2(r_k, t_j)\}_{k,j=1}^{K, N_{T'}}$ with K short transients of duration T' covering the rest of the tissue slice with high spatial resolution (note that $K \gg D$ and $T' \ll T$). The reconstruction is then performed as follows. The data in S_1 are used to form a $D \times N_T$ Casorati matrix, and singular value decomposition (SVD) is applied to this matrix to extract the basis from the right singular vectors (see the Supporting Information). Based on the SVD, L was chosen such that the truncation error was small ($< 3\%$). With $\phi_l(t)$ estimated, the coefficients can be determined from the short transients S_2 by a regularized least-squares fitting:

$$\hat{\mathbf{C}} = \arg \min_{\mathbf{C}} \|\mathbf{S}_2 - \mathbf{C}\Phi\|_{\text{F}}^2 + \alpha \|\mathbf{C}\|_{\text{F}}^2 \quad (\text{Eq.2})$$

where the first term is a data consistency measures, Φ is a matrix with each row being a basis transient evaluated up to T' to match the transient length of the data in S_2 , and $\mathbf{C}$ is a matrix with each row containing the spatial coefficients for a pixel. The second term in the

equation is a regularization term to stabilize the fitting with α being a small regularization parameter. This problem can be solved with a closed-form solution:

$$\hat{\mathbf{C}} = (\mathbf{\Phi}^H \mathbf{\Phi} + \alpha \mathbf{I})^{-1} \mathbf{\Phi}^H \mathbf{S}_2. \quad (\text{Eq.3})$$

Finally, the long transients at pixels r_k , $k = 1, 2, \dots, K$ can be reconstructed as a linear combination of $\{\phi_l(t_i)\}$ with a desired transient duration T , weighted by the estimated coefficients $\hat{\mathbf{C}}$ as follows:

$$\hat{s}_2(r_k, t_i) = \sum_{l=1}^L \hat{c}_l(r_k) \phi_l(t_i). \quad (\text{Eq.4})$$

METHODS

Sample Preparation, Matrix Application, and MSI.

Experimental details about animals, tissue sectioning, and matrix application can be found in the Supporting Information. The rat sagittal brain slice was prepared and imaged to generate 39 775 pixels. MSI was performed on a solarix 7T MALDI FT-ICR mass spectrometer (Bruker Corp., Billerica, MA) equipped with an atmospheric pressure ionization dual ESI/MALDI source operating at ~ 2.65 mbar. A mass window of m/z 150–1200 was selected, yielding transients with 0.734 s duration, which resulted in a resolving power of 93 000 at m/z 700. MALDI mass spectra were acquired in positive-mode using a smartbeam-II UV laser (Bruker Corp.) in ‘minimum’ mode with a 50- μm raster width. Each MALDI acquisition consisted of 400 laser shots at a frequency of 1000 Hz. External calibration was performed using PepMix II (Bruker Corp.). The total time for data acquisition was 8 h and the entire imaging experiment took 15 h, with an ~ 0.85 s measurement overhead per pixel (e.g., time spent on multiple laser shots, ion accumulation, stage movement, and some online processing). The acquired dataset was used as the gold standard to evaluate the subspace approach. We would like to note that these measurement overheads are instrument- and experiment-dependent, and can be further reduced by optimizing various experimental parameters, but are not the focus of this paper.

Evaluation of Reconstruction.

We evaluated the subspace model by randomly sampling 10% of the transients with the long transient duration (0.734 s) to construct the $\mathbf{S}_1$ matrix from all pixel locations in order to cover most of the heterogeneous tissue types. Fast decay of the singular values for $\mathbf{S}_1$ supports the use of our low-dimensional subspace model (Figure 2A). The selection of L is data adaptive by observing the singular value decays of $\mathbf{S}_1$ and the total signal energy retained (usually $>97\%$ as the threshold) by keeping the singular vectors associated with the dominant singular values. For the results presented below, we chose the number of basis transients as $L = 250$ based on the procedure described above, which was specific to the dataset to minimize truncation or modeling errors. The remaining 90% of the transients were truncated to only 5% of the original duration (0.037 s) to produce $\mathbf{S}_2$. We then applied our subspace model to estimate the coefficients $\{c_l(r_k)\}$ using the predetermined basis from $\mathbf{S}_1$ and the substantially truncated transients in $\mathbf{S}_2$. The estimated coefficients and the basis were

used to reconstruct transients to the original duration (0.734 s) to achieve the desired spectral resolution.

We evaluated the quality of the reconstruction by quantitatively comparing the original spectra and the reconstructed spectra using the described method. First, fast Fourier transform (FFT) and mass calibration were performed to convert both transients from the time domain to the m/z domain. Next, a peak detection algorithm was applied to generate peak lists for both sets of spectra. Pairwise comparison of the generated peak lists was carried out using three metrics: the percentage of m/z features that were correctly recovered, the peak intensity Pearson correlation, and the average relative peak error. Given a set of m/z features $P_i = \{p_1, p_2 \cdots p_a\}$ for an original spectrum i , and another set of m/z features $Q_i = \{q_1, q_2 \cdots q_b\}$ for the corresponding reconstructed spectrum, the percentage of peaks recovered is calculated as:

$$PQ_{i, \text{shared}} = \frac{|P_i \cap Q_i|}{|P_i|} \quad (\text{Eq.5})$$

which is between 0 and 1. If all m/z features in the original peak list P_i are contained in Q_i , $PQ_{i, \text{shared}}$ equals to 1. Otherwise, if none of the features in P_i are contained in Q_i , it equals to 0. Peak intensity Pearson correlation measures how well the reconstructed peak intensities correlate linearly with the original peak intensities for the shared peaks. The average relative peak error was defined as the mean of the relative peak intensity errors for the shared peaks within a spectrum.

We also examined the consistency between ion image pairs. Peak lists were first aligned nonuniformly to combine them into a multivariate matrix, with each row as a spectrum and each column as an m/z feature. Pearson correlation was used to measure the fidelity of the reconstructed ion images to the original ion images. The ion image sets were also subjected to principal component analysis (PCA) and k-means clustering to further evaluate the model's capability of accurately recovering chemical information through the reconstruction. The scores and coefficients for the first three principal components (PCs) that retained most of the data variance were compared, and hyperspectral visualizations were made through translating the scores of the low-dimensional projections to RGB color coding.^{34,35} For k-means, $k = 6$ was selected to group pixels with similar spectral profiles together in an unsupervised manner. The segmentation masks were visually inspected to determine if groupings based on tissue chemical profiles were similar between the original and the reconstructed ion image data. The evaluation pipeline workflow is provided in Figure S1.

RESULTS AND DISCUSSION

Low-Rank Properties of Experimental Transient Signals.

For data that are low rank, a lower-dimensional subspace exists compared to the original data. Various applications, such as data compression,³⁶ signal reconstruction,³⁷ and signal denoising,³⁸ exploit the low-rank properties of the data, providing more efficient ways for data manipulation and data acquisition. We assumed the FT-ICR temporal signals that were

measured from the tissue slice to be low rank. In other words, their transient data can be projected onto a lower-dimensional subspace with a minimal loss of fidelity. As shown in Figure 2A, the Casorati matrix formed by the S_1 data with 4000 randomly sampled long transients measured from the sagittal brain slice (see Methods section) has rapidly decaying singular values, indicating that these transients reside in a low-dimensional subspace. These rapidly decaying singular values support our assumption that the transients can be accurately approximated by the obtained subspace, with a model order (number of basis vectors) determined by thresholding the normalized singular values. By visualizing the first few basis transients, we found varying signal signatures with distinct frequency components (Figure 2B, C). The basis transients were then converted to the m/z domain through FFT and mass conversion/calibration (Figure 2D) to visualize the basis spectra. We noticed the presence of different spectral features in those basis spectra due to the finite heterogeneity of the measured data (Figure 2E). Since many tissue subregions with distinct chemical profiles were covered in a sagittal brain slice, SVD essentially preserved the biological variations and transformed them into an orthonormal basis. Interestingly, by projecting transient data onto a lower-dimensional space through rank truncation, background noise can be efficiently removed, leading to the denoising effect once back-projected onto the original data space. This denoising effect can be observed in the top-ranked basis transients corresponding to larger singular values (high-energy components) versus the ones corresponding to smaller singular values (low-energy components), as the top basis transients give rise to nearly noiseless signals and the later basis transients are more contaminated by noise (Figure S2). Based on the degree of homogeneity of the measured data, such that pixels from the same tissue type likely have similar chemical contents, a subset of transient data can be used to approximate the basis without using all of the data points. A similar concept was applied to compressing large MSI datasets³⁶ over the spectrum in the m/z domain rather than in the temporal domain. The choice of the number of long transients to be sampled affects the representativeness of the extracted basis transients. Long transients need to span the entire subspace for accurate subspace estimation and the more samples that are available to SVD results in a better subspace approximation. Theoretically, the lower bound should be equal to the rank, i.e., the subspace model order L . For our test data, the 4000 long transients were sufficient to cover the tissue heterogeneity present through the sagittal slice; beyond 4000 long transients, the subspace estimation improvement is minimal. The number of long transients can be further optimized for different applications, which results in a tradeoff between long transient acquisition time and subspace estimation accuracy.

Typically, computing the transformation needed to represent the transient data in a subspace requires storing them onto the computer's memory. As the number of the sampled transients increases, using SVD to determine the basis transients becomes computationally expensive. Since a small subset of transients was used to obtain the basis, we were able to fit this group onto 64 Gb RAM. For experiments with more data points per transient, one may need to consider decreasing the number of sampled transients, depending on the amount of memory available. Workarounds are available to compute the basis more efficiently.³⁹

Accurate Estimation of Experimental Transients Using the Predefined Subspace.

Our results from performing SVD on sampled transients demonstrate that we were able to find a lower-dimensional subspace for the data. As stated previously, the actual measurements are somewhat homogeneous with the addition of random noise due to the similar chemical profiles of pixels from the similar tissue types. With such redundancy, it may not be necessary to acquire all transients at the same transient length. To validate the subspace approach, we randomly sampled 4000 transients to obtain the basis for the brain imaging dataset and truncated the remaining 35 775 transients to keep only the first 5% of the signals. We applied least squares to find the best fit between each truncated transient and the top basis transients at $L = 250$ using the same truncated length (see Methods section). The fitted coefficients were subsequently used as weights to linearly combine the full-length basis transients and generate the reconstructed temporal signals with the desired spectral resolution for all pixels. As the total data acquisition time for an FT-ICR MSI experiment is limited by the transient length and the total number of pixels scanned, our new strategy was to reduce the total data acquisition time by approximately 85% (~8 h down to ~1 h). To better demonstrate the unique capability of our method, we also simulated a situation in which all transients were uniformly truncated to achieve the same reduced acquisition time (0.11 s per transient) without using our method for comparison.

The reconstructed and truncated transients were converted to the m/z domain through FFT and mass calibration, allowing spectrum-wise evaluations against the original data. The reconstructed mass spectra showed high fidelity compared to the original spectra (from the long acquisition, Figure 3A), with detailed peak features well recovered and an improved SNR, whereas the spectra from purely truncated transients displayed a much degraded mass resolution caused by shorter transients. We quantified the reconstruction quality by comparing peak features. Specifically, we performed peak detection on the reconstructed and original spectrum pairs to test how effective the reconstruction was at recovering peak features, which was evaluated by three metrics: the percentage of peaks that were correctly recovered, the Pearson correlation coefficients, and the average relative peak error (defined in the Methods section). The mutual peaks between the original spectrum and the reconstructed spectrum are linearly correlated (Figure 3B), showing accurate estimations of peak intensity across each whole spectrum. Performing the above evaluations on the entire sagittal brain dataset demonstrated that >90% of peak features are recovered correctly (Figure 3C, left) while achieving high correlation values for intensities of the shared peaks (Figure 3C, middle). The average relative peak error is <25% for most spectra (Figure 3C, right), with the main errors likely due to the least-squares fitting favoring high-intensity peaks. This error should be tolerable as the relative intensities within a spectrum are consistent, shown by high Pearson correlation values. Significant improvement was achieved, which can be visualized by contrasting the distribution of the percent peak features recovered in uniformly truncated transients. As shown in Figure S3, a limited number of features were correctly matched, even though the m/z values were rounded to two decimal places. In addition, fewer peak features were detected in the truncated transients due to loss of spectral resolution, whereas more peak features were detected in the reconstructed transients using our subspace model (Figure S4). As described in the Methods section, we empirically chose the model order L to minimize the truncation errors by inspecting the

singular value decay of the Casorati matrix formed by the long transients. To further investigate the effects of L , we have performed reconstructions of the same data with different L s, with the results shown in Figure S5. As can be observed, fitting with a very low order will likely produce accurate reconstruction of the dominant spectral features (signals with higher energy) but also can introduce structured errors in the results. As finer features and instrumental noise are predominantly captured by the higher order basis, reconstruction using too high an order can result in an increase in spectral noise. It is also worth emphasizing that the resolution of the reconstructed spectra is entirely determined by the basis transients. The basis transients are equivalent to that of the sampled long transients (a unique feature of the subspace-based fitting), regardless of the persistency of the transient signals as long as the subspace captures the majority of the spectral features to be recovered.

Evaluation of Ion Images.

An ion image is a visual representation of the localization and distribution of specific molecular species. The peak lists, or m/z and peak intensity pairs, obtained from peak detection are aligned with a predefined mass tolerance to generate a multivariate intensity matrix, with columns and rows representing m/z features and pixel locations, respectively. Each column vector of an intensity matrix can then be mapped to a two-dimensional ion image, forming hyperspectral image data. For this study, the peak lists from the original spectra and reconstructed spectra were aligned to generate intensity matrices, and 480 ion images were formed for qualitative and quantitative comparison. Pearson correlation coefficients between the original and reconstructed image pairs of the same m/z feature were calculated. Higher correlation values indicate more faithful spatial reconstruction results when compared to the ion images generated by the original long acquisition.

The distribution of the correlation values of the ion images for the brain imaging dataset (Figure 4B, Figure S6) shows that the reconstructed ion images from m/z 150–1000 have a median 0.64 Pearson correlation. Almost identical ion images were reconstructed from peaks with strong signal intensities using the described method (m/z 753.589, m/z 820.523, Figure 4A), which correspond to lipids that are enriched in specific brain regions. Not surprisingly, ion image pairs with lower average intensities had lower correlation values compared to the ones with higher average intensities (Figure 4C). We found that the information from small peaks was less dominant in the predefined subspace to estimate the transients. For example, the ion image at m/z 688.394 shows poor reconstruction (Pearson = 0.0551); in the original image, this appears to be chemical noise without a spatial colocalization (Figure S7). Interestingly, in some cases our method yielded improved ion image visualization with higher image contrast and more defined spatial features compared to the original images (m/z 784.562, m/z 785.567, m/z 811.602 in Figure 4A; and m/z 772.559, m/z 825.557, m/z 828.548 in Figure S7). Enhanced ion image visualization can be attributed to the denoising ability of the subspace model due to the dimensionality reduction. Similar fittings (coefficients) that were obtained from nearby tissue locations with closely related chemical profiles led to a smoothing effect.

To further evaluate the capability of our method to preserve important chemical information for subsequent tissue sample analysis and data mining, we applied PCA on the original data

and compared it against the reconstructed data using a total of 480 images. Unsupervised, multivariate analysis, such as PCA and k-means clustering, play irreplaceable roles in exploring and interpreting high-dimensional hyperspectral MSI data.^{40,41} Evaluation of the first three PCs (Figure 5A) resulted in similar scores and coefficients, indicating that the majority of the variance (90% variance explained by first three PCs) was accurately captured by our method while still reducing the data acquisition time by 85%. Similar hyperspectral visualizations were achieved by color coding the scores of the first three PCs (Figure 5B, Figure S8). K-means clustering was used to further demonstrate that our method can adequately recover chemical information to perform unbiased spatial segmentation (Figure S9). By grouping similar spectra together, physical interpretations about tissue organization based on the spectral features can be obtained in a much shorter experimental time frame.

CONCLUSIONS

We have developed a subspace approach for the acquisition and computational reconstruction of FT-ICR-based MSI data. Our method allows for significantly faster imaging experiments without compromising mass or spatial resolution while obtaining exceptionally similar chemical information, validated using a high-resolution rat sagittal brain imaging dataset. In addition, the approach does not require customized systems and can be implemented with commonly available computational resources. Future directions include extending the method to three-dimensional MSI, which is time consuming due to the imaging of multiple tissue slices at high spatial resolutions. We also expect the approach to be readily combined with alternative computational imaging modalities (e.g., CS), and generalized to other FT-based instruments such as Orbitrap MS platforms that also acquire data as time-domain transients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This project was supported by the National Institute on Drug Abuse under Award No. P30DA018310 and the National Human Genome Research Institute under award No. R01HG010023. The content is solely the responsibility of the authors and does not necessarily represent the official views of the awarding agencies.

REFERENCES

1. Rae Buchberger A; DeLaney K; Johnson J; Li L Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal. Chem* 2018, 90, 240–265. [PubMed: 29155564]
2. Gessel MM; Norris JL; Caprioli RM MALDI Imaging Mass Spectrometry: Spatial Molecular Analysis to Enable a New Age of Discovery. *J. Proteomics* 2014, 107, 71–82. [PubMed: 24686089]
3. Š upáková K; Balluff B; Tressler C; Adelaja T; Heeren RMA; Glunde K; Ertaylan G Cellular Resolution in Clinical MALDI Mass Spectrometry Imaging: The Latest Advancements and Current Challenges. *Clin. Chem. Lab. Med* 2020, 58, 914–929.
4. Gemperline E; Chen B; Li L Challenges and Recent Advances in Mass Spectrometric Imaging of Neurotransmitters. *Bioanalysis* 2014, 6, 525–540. [PubMed: 24568355]

5. Norris JL; Caprioli RM Imaging Mass Spectrometry: A New Tool for Pathology in a Molecular Age. *Proteomics: Clin. Appl* 2013, 7, 733–738. [PubMed: 24178781]
6. Vaysse P-M; Heeren A, A. RM; Porta T; Balluff B Mass Spectrometry Imaging for Clinical Research – Latest Developments, Applications, and Current Limitations. *Analyst* 2017, 142, 2690–2712. [PubMed: 28642940]
7. Cornett DS; Frappier SL; Caprioli RM MALDI-FTICR Imaging Mass Spectrometry of Drugs and Metabolites in Tissue. *Anal. Chem* 2008, 80, 5648–5653. [PubMed: 18564854]
8. Rubakhin SS; Jurchen JC; Monroe EB; Sweedler JV Imaging Mass Spectrometry: Fundamentals and Applications to Drug Discovery. *Drug Discov. Today* 2005, 10, 823–837. [PubMed: 15970265]
9. Chughtai K; Heeren RMA Mass Spectrometric Imaging for Biomedical Tissue Analysis. *Chem. Rev* 2010, 110, 3237–3277. [PubMed: 20423155]
10. Aichler M; Walch A MALDI Imaging Mass Spectrometry: Current Frontiers and Perspectives in Pathology Research and Practice. *Lab. Investig* 2015, 95, 422–431. [PubMed: 25621874]
11. Shariatgorji M; Nilsson A; Fridjonsdottir E; Vallianatou T; Källback P; Katan L; Sävmarker J; Mantas I; Zhang X; Bezard E; Svenningsson P; Odell LR; Andrén PE Comprehensive Mapping of Neurotransmitter Networks by MALDI–MS Imaging. *Nat. Methods* 2019, 16, 1021–1028. [PubMed: 31548706]
12. Soltwisch J; Ketting H; Vens-Cappell S; Wiegmann M; Müthing J; Dreisewerd K Mass Spectrometry Imaging with Laser-Induced Postionization. *Science* 2015, 348, 211–215. [PubMed: 25745064]
13. Hendrickson CL; Quinn JP; Kaiser NK; Smith DF; Blakney GT; Chen T; Marshall AG; Weisbrod CR; Beu SC 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom* 2015, 26, 1626–1632. [PubMed: 26091892]
14. Shaw JB; Lin T-Y; Leach FE; Tolmachev AV; Toli N; Robinson EW; Koppelaar DW; Paša-Toli L 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox. *J. Am. Soc. Mass Spectrom* 2016, 27, 1929–1936. [PubMed: 27734325]
15. Scigelova M; Hornshaw M; Giannakopoulos A; Makarov A Fourier Transform Mass Spectrometry. *Mol. Cell. Proteomics* 2011, 10, M111.009431.
16. Xian F; Hendrickson CL; Marshall AG High Resolution Mass Spectrometry. *Anal. Chem* 2012, 84, 708–719. [PubMed: 22263633]
17. Nagornov KO; Gorshkov MV; Kozhinov AN; Tsybin YO High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Increased Throughput for Biomolecular Analysis. *Anal. Chem* 2014, 86, 9020–9028. [PubMed: 25140615]
18. Smith DF; Kilgour DPA; Konijnenburg M; O'Connor PB; Heeren RMA Absorption Mode FTICR Mass Spectrometry Imaging. *Anal. Chem* 2013, 85, 11180–11184. [PubMed: 24175640]
19. Bowman AP; Blakney GT; Hendrickson CL; Ellis SR; Heeren RMA; Smith DF Ultra-High Mass Resolving Power, Mass Accuracy, and Dynamic Range MALDI Mass Spectrometry Imaging by 21-T FT-ICR MS. *Anal. Chem* 2020, 92, 3133–3142. [PubMed: 31955581]
20. Astigarraga E; Barreda-Gómez G; Lombardero L; Fresnedo O; Castaño F; Giralt MT; Ochoa B; Rodríguez-Puertas R; Fernández JA Profiling and Imaging of Lipids on Brain and Liver Tissue by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using 2-Mercaptobenzothiazole as a Matrix. *Anal. Chem* 2008, 80, 9105–9114. [PubMed: 18959430]
21. Van de Plas R; Yang J; Spraggins J; Caprioli RM Image Fusion of Mass Spectrometry and Microscopy: A Multimodality Paradigm for Molecular Tissue Mapping. *Nat. Methods* 2015, 12, 366–372. [PubMed: 25707028]
22. Patterson NH; Tuck M; Van de Plas R; Caprioli RM Advanced Registration and Analysis of MALDI Imaging Mass Spectrometry Measurements through Autofluorescence Microscopy. *Anal. Chem* 2018, 90, 12395–12403. [PubMed: 30272960]
23. Vollnhals F; Audinot J-N; Wirtz T; Mercier-Bonin M; Fourquaux I; Schroepel B; Kraushaar U; Lev-Ram V; Ellisman MH; Eswara S Correlative Microscopy Combining Secondary Ion Mass Spectrometry and Electron Microscopy: Comparison of Intensity–Hue–Saturation and Laplacian Pyramid Methods for Image Fusion. *Anal. Chem* 2017, 89, 10702–10710. [PubMed: 28901122]

24. Neumann EK; Comi TJ; Spegazzini N; Mitchell JW; Rubakhin SS; Gillette MU; Bhargava R; Sweedler JV Multimodal Chemical Analysis of the Brain by High Mass Resolution Mass Spectrometry and Infrared Spectroscopic Imaging. *Anal. Chem* 2018, 90, 11572–11580. [PubMed: 30188687]
25. Kozhinov AN; Tsybin YO Filter Diagonalization Method-Based Mass Spectrometry for Molecular and Macromolecular Structure Analysis. *Anal. Chem* 2012, 84, 2850–2856. [PubMed: 22376180]
26. Bartels A; Dülk P; Trede D; Alexandrov T; Maaß P Compressed Sensing in Imaging Mass Spectrometry. *Inverse Probl.* 2013, 29, 125015.
27. Gao Y; Zhu L; Norton I; Agar NYR; Tannenbaum A Reconstruction and Feature Selection for Desorption Electrospray Ionization Mass Spectroscopy Imagery In Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling; International Society for Optics and Photonics, 2014; Vol. 9036, p 90360D.
28. Tang F; Bi Y; He J; Li T; Abliz Z; Wang X Application of Super-Resolution Reconstruction of Sparse Representation in Mass Spectrometry Imaging. *Rapid Commun. Mass Spectrom* 2015, 29, 1178–1184. [PubMed: 25981548]
29. Kooijman PC; Nagornov KO; Kozhinov AN; Kilgour DPA; Tsybin YO; Heeren RMA; Ellis SR Increased Throughput and Ultra-High Mass Resolution in DESI FT-ICR MS Imaging through New-Generation External Data Acquisition System and Advanced Data Processing Approaches. *Sci. Rep* 2019, 9, 1–11. [PubMed: 30626917]
30. Park S-G; Anderson GA; Bruce JE Parallel Detection of Fundamental and Sixth Harmonic Signals Using an ICR Cell with Dipole and Sixth Harmonic Detectors. *J. Am. Soc. Mass Spectrom* 2020, 31, 719–726.
31. Liang Z Spatiotemporal Imaging with Partially Separable Functions. In 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2007; pp 988–991.
32. Haldar JP; Liang Z-P Spatiotemporal Imaging with Partially Separable Functions: A Matrix Recovery Approach. In 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2010; pp 716–719.
33. Lam F; Liang Z-P A Subspace Approach to High-Resolution Spectroscopic Imaging. *Magn. Reson. Med* 2014, 71, 1349–1357. [PubMed: 24496655]
34. Fonville JM; Carter CL; Pizarro L; Steven RT; Palmer AD; Griffiths RL; Lalor PF; Lindon JC; Nicholson JK; Holmes E; Bunch J Hyperspectral Visualization of Mass Spectrometry Imaging Data. *Anal. Chem* 2013, 85, 1415–1423. [PubMed: 23249247]
35. Smets T; Verbeeck N; Claesen M; Asperger A; Griffioen G; Tousseyn T; Waelput W; Waelkens E; De Moor B Evaluation of Distance Metrics and Spatial Autocorrelation in Uniform Manifold Approximation and Projection Applied to Mass Spectrometry Imaging Data. *Anal. Chem* 2019, 91, 5706–5714. [PubMed: 30986042]
36. Palmer AD; Bunch J; Styles IB Randomized Approximation Methods for the Efficient Compression and Analysis of Hyperspectral Data. *Anal. Chem* 2013, 85, 5078–5086. [PubMed: 23534867]
37. Qu X; Mayzel M; Cai J-F; Chen Z; Orekhov V Accelerated NMR Spectroscopy with Low-Rank Reconstruction. *Angew. Chem., Int. Ed* 2015, 54, 852–854.
38. Chiron L; van Agthoven MA; Kieffer B; Rolando C; Delsuc M-A Efficient Denoising Algorithms for Large Experimental Datasets and Their Applications in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A* 2014, 111, 1385–1390. [PubMed: 24390542]
39. Halko N; Martinsson PG; Tropp JA Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Rev.* 2011, 53, 217–288.
40. Verbeeck N; Caprioli RM; Van de Plas R Unsupervised Machine Learning for Exploratory Data Analysis in Imaging Mass Spectrometry. *Mass Spectrom. Rev* 2020, 39, 245–291.
41. Alexandrov T MALDI Imaging Mass Spectrometry: Statistical Data Analysis and Current Computational Challenges. *BMC Bioinformatics* 2012, 13, S11.

Figure 1.

Subspace-based imaging for FT-ICR MSI. (Top) Data acquisition that consists of the measurement of two complementary datasets: the yellow triangles indicate the acquisition of transient set S_1 with the desired spectral resolution (long transient duration T) at random tissue locations that should cover most heterogeneous tissue types. The blue circles indicate the larger transient set S_2 acquired with a shorter transient duration T' ($T' \ll T$) acquired at the rest of tissue locations with the desired spatial resolution. (Bottom) The reconstruction strategy that consists of an estimation of basis transients (the subspace) with the desired spectral resolution from S_1 through singular value decomposition, and fitting of spatial coefficients for individual transients in S_2 through least-squares estimation using the predefined subspace. The fitted coefficients are then used as weights for the basis transients to reconstruct long transients for all pixels with the desired spectral resolution. This acquisition and reconstruction strategy significantly improve the data acquisition efficiency without compromising the spectral resolution and mass accuracy, while providing high reconstruction fidelity to recover accurate chemical information from the whole tissue section.

Figure 2.

SVD was performed on 4000 transients randomly sampled from the rat brain imaging dataset. (A) The rapid decay of the sorted singular values was shown approaching a <3% truncation error for models with an order >80, capturing most of the energy within the sampled signals. (B) Only the first three basis transients were visualized, and (C) distinct frequency compositions were found within a short time window, indicated by the black arrow. (D) The basis transients were converted to the m/z domain, where (E) distinct spectral features were present from m/z 768 to m/z 770 in the first three basis spectra.

Figure 3.

(A) Spectra for the original (black), reconstructed (red), and truncated (blue) transients. The inset shows the spectra from m/z 844 to m/z 848. The reconstructed spectrum displays high fidelity to the original spectrum with a high SNR and peak features effectively recovered, whereas the spectrum from the truncated, short transient has a much lower spectral resolution that obscures spectral details and lowers mass accuracy. Peak detection was performed for the original and the reconstructed spectrum, and (B) reconstructed peak intensities correlate with the original intensities. (C) The distributions of percent peak features recovered, peak intensity correlation, and the average relative peak error for the reconstructed spectra in the brain imaging dataset.

Figure 4.

Pairwise comparison between the original and reconstructed ion images for the brain imaging dataset. (A) Ion image pairs, the original image (top) and the reconstructed image (bottom), for selected *m/z* values. (B) The distribution of Pearson correlation coefficients for 480 ion images. (C) Plot of the average intensity of the ion image versus the Pearson correlation demonstrates that ion images with lower average signal intensities tend to have lower coefficients for reconstruction. Color bar represents the normalized mean square error.

Figure 5.

PCA results from the original data and the reconstructed data using our method. (A) Nearly identical scores and coefficients were obtained for the first three PCs (~90% explained variance). (B) Hyperspectral visualizations by translating PC scores to RGB channels also produced similar results. The reconstructed data retained the variation for most of the chemical information, and the total acquisition time needed was only 15% of the original acquisition time.