Detecting and Removing Data Artifacts in Hadamard Transform Ion Mobility-Mass Spectrometry Measurements

Spencer A. Prost[1], Kevin L. Crowell[2], Erin S. Baker[2], Yehia M. Ibrahim[2], Brian H. Clowers[3], Matthew E. Monroe[2], Gordon A. Anderson[2], Richard D. Smith[2], Samuel H. Payne[2]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

[2]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA

[3]Department of Chemistry, Washington State University, Pullman, WA

Pacific Northwest NATIONAL LABORATORY

Overview

- The Hadamard transform can be utilized in ion-mobility spectrometry-MS to improve the signal-to-noise (SNR) ratio[1].
- The demultiplexing of Hadamard transform encoded data can result in artifacts that appear as pervasive noise
- Artifacts such as spurious peaks have been observed in other types of Hadamard transform spectrometry[2,3].
- A novel algorithm has been developed that preserves true signal while eliminating artifacts

Introduction

- The Hadamard transform has been applied to a variety of instruments, including ion-mobility mass spectrometers, which increases the SNR.
- Data artifacts resulting from the application of the Hadamard transform can increase the likelihood of false discovery of peptides in downstream analysis.

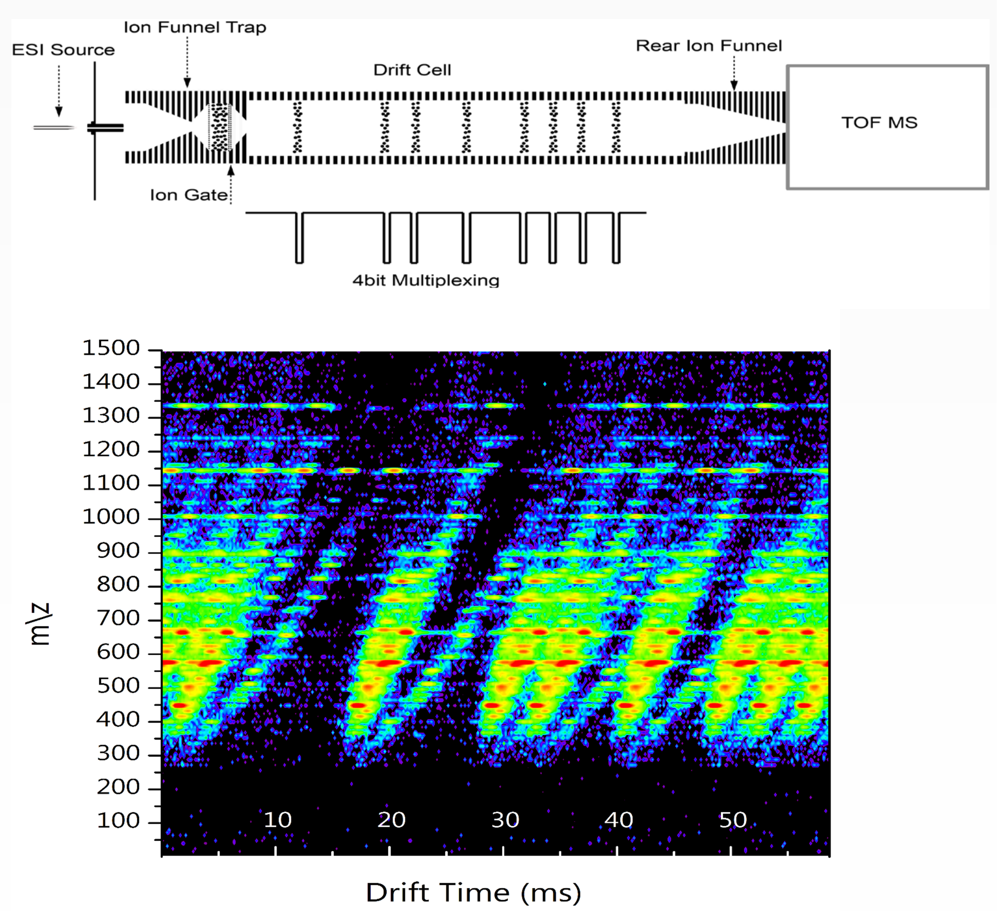

Figure 1. Schematic diagram of LC-IMS-TOF MS platform (top) and an example of a single LC frame of multiplexed data (bottom).

Algorithmic Methods

Input: Multiplexed segment and demultiplexed segment
Output: Demultiplexed segment containing only validated intensity values

1. Take the sum of the demultiplexed segment. If a unique value exists that equals the sum, this is the location of the true signal.
   a. This only applies to segments with a single true signal. If there are multiple values that equal the sum (and the negation of the sum) there does not exist any true signal in the segment.
   b. If no value equals the sum (or its negation) then the algorithm proceeds as there is likely more than one signal in the segment.
2. For each positive value in the demultiplexed segment, compute the sum of the values in the multiplexed segment starting with the assumption that the selected index is the start of the pseudo-random sequence (PRS). This is the scoring portion of the algorithm.
3. For all non-negative summed values (all values will be positive on the first pass) pick the one with the greatest sum for the candidate index of the true signal.
4. For all values that align with a "1" in the PRS with the selected index of the location of the true signal, subtract from the multiplexed segment the multiplexed intensity at the index of the true signal.
5. Using this updated multiplexed segment, repeat from step 2 in order to eliminate the possibility of multiple true signals in the multiplexed segment.

Step 1

Multiplexed	Demultiplexed
9163	2288
2619	13272
867	2976
244	2604
0	33048
0	1064
14	-2356
12306	51448
47	-32676
2	-13560
11944	-4300
5672	4400
35	-3028
8851	-2472
120	-824
Sum	51884

Steps 2, 3

PRS → 12306	Multiplexed Data
1	9163
1	2619
1	867
1	244
0	0
0	0
0	0
1	12306
0	0
0	0
1	11944
1	5672
0	0
1	8851
0	0
Sum	51666

Steps 4,5

Original Demultiplexed	Multiplexed Segment After Subtraction
2288	-3143
13272	-9687
2976	-11439
2604	-12062
33048	0
1064	0
-2356	14
51448	0
-32676	47
-13560	2
-4300	-362
4400	-6634
-3028	35
-2472	-3455
-824	120

Results

Investigating Cause of Artifacts

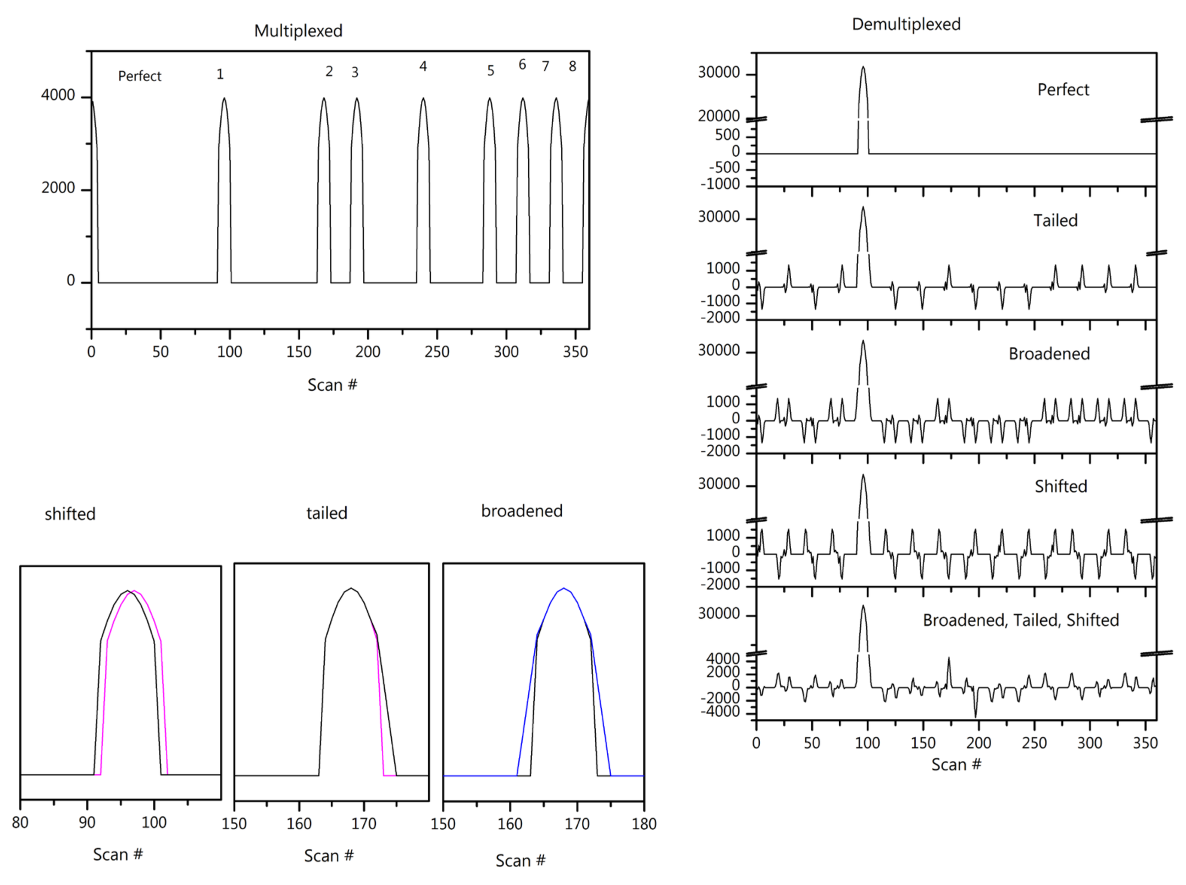

Figure 2. Exemplification of the different causes behind the formation of the artifacts. Typically shifted, tailed, or broadened peaks will result in multiplexed segments that do not all contain the same value, which leads to the generation of artifacts. The greater the shift from the uniform value, the greater the size of the resulting artifact.

Recovering Multiple Signals

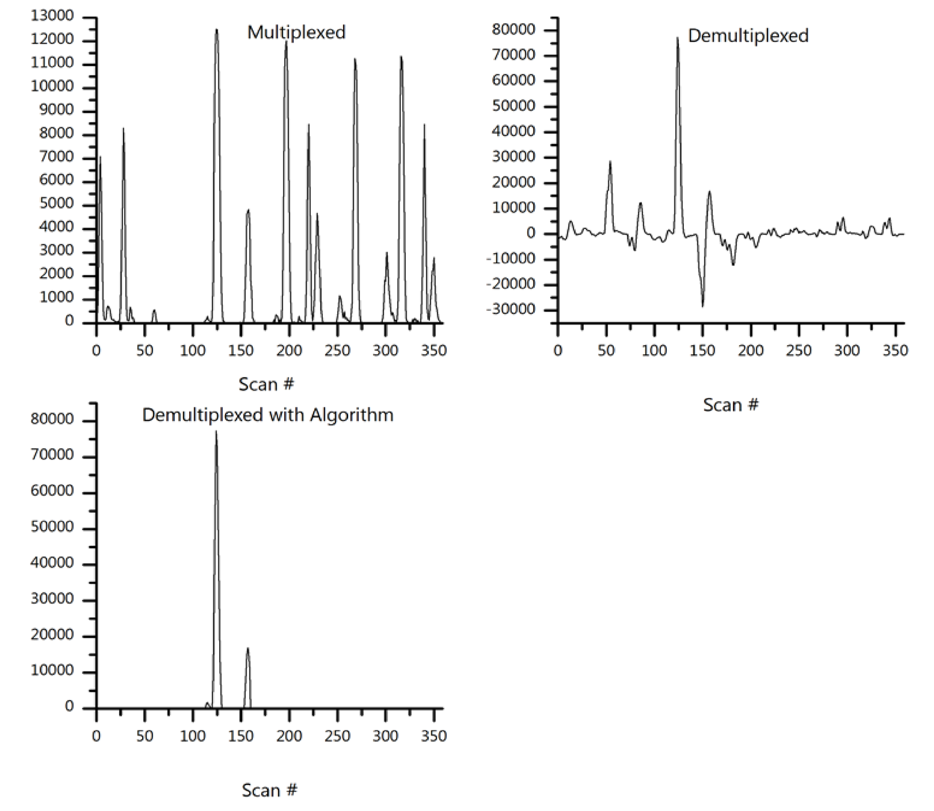

Figure 3. A single m/z slice with three signals. Note that the demultiplexed values that are greater than the true signal, and therefore might obscure the location of the other true signals in this slice.

Resolving Artifacts

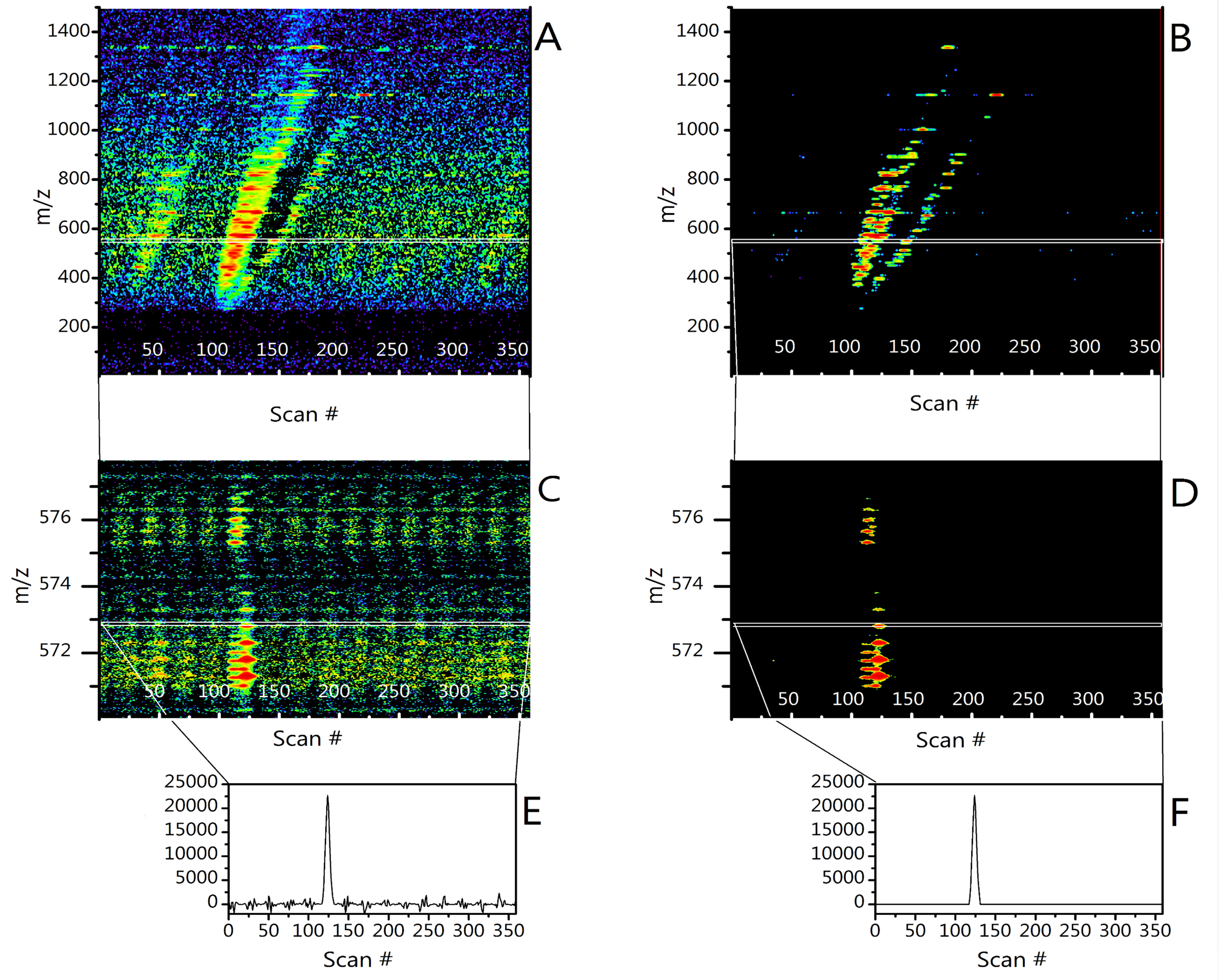

Figure 5. 3A is the demultiplexed LC frame of Figure 1. 3B is after the application of the algorithm on the entire frame. 3C is a zoomed in portion of 3A between 570 and 578 m/z. Figure 3D is the same area as 3A after the algorithm. 3E is a single slice of m/z space at 573. 3F is the same slice of m/z space after the algorithm has been applied.

- The algorithm eliminated artifacts present in the LC-Frame in Figure 3A.
- Low abundance ions and low signal is preserved, as can be observed in Figure 3.
- Three signals are present in Figure 3.
  - The artifacts present in Figure 4 obscure the true signal located around scan number 155.
  - There is a very low intensity signal that precedes the primary peak which is also obscured by the artifacts.
  - The algorithm is capable of preserving multiple true signals due to the scoring described in the algorithmic methods.
- The cause of the artifacts as exemplified in Figure 2 originate from two primary contributing factors:
  - Temporal differences in arrival times of the multiplexed ion packets by 1 or 2 ion mobility scans
  - Variation in signal intensity across the multiplexed ion packets

Conclusions

- A novel algorithm has been developed that has greatly improved the IMS-MS data SNR obtained from the application of the Hadamard transform.
- The nature of the types of artifacts observed in IMS data have been characterized, which will allow for further investigation.

Acknowledgements

The authors thank Richard Zare, Craig Aspinwall, Facundo Fernandez and Ignacio Zuleta for valuable correspondence and Hadamard transformed data utilizing a large number of bits. SAP was sponsored by the U.S. Department of Energy Science Undergraduate Laboratory Internships (SULI) program. SHP acknowledges funding from a U.S. Department of Energy Early Career award. The development of the IMS-MS platform was provided through the National Institute of General Medical Sciences Proteomic Center at PNNL (P41 GM103493 and R21 GM103497), and other portions of this research were supported by grants from the National Institute of General Medical Sciences (P41 GM103493 and R21 GM103497), National Cancer Institute (R21-CA12619-01, U24-CA-160019-01, and Interagency Agreement Y01-CN-05013-29), National Institute of Environmental Health Sciences of the National Institutes of Health (R01ES022190), the Laboratory Directed Research and Development Program at PNNL and by the U.S. Department of Energy Office of Biological and Environmental Research Genome Sciences Program under the Pan-omics project. Work was performed in the Environmental Molecular Science Laboratory, a U.S. Department of Energy (DOE) national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, WA. Battelle operates PNNL for the DOE under contract DE-AC05-76RLO01830.

References

1. Clowers, B.H., Belov, M.E., Prior, D.C., Danielson, W.F., Ibrahim, Y., and Smith, R.D.: Pseudorandom sequence modifications for ion mobility orthogonal time-of-flight mass spectrometry. *Analytical chemistry*. **80**, 2464-2473 (2008)
2. Zeppenfeld, P., Krzyzowski, M., Romainczyk, C., and David, R.: On the origin of spurious peaks in pseudorandom time-of-flight analysis. Rev. Sci. Instrum. 64, 1520-1523 (1993)
3. Kimmel, J.R., Fernández, F.M., and Zare, R.N.: Effects of modulation defects on hadamard transform time-of-flight mass spectrometry (ht-tofms). *J. Am. Soc. Mass. Spectrom.* **14**, 278-286 (2003)

CONTACT: Spencer Prost
Instrument Development Laboratory
Pacific Northwest National Laboratory
E-mail: spencer.prost@pnnl.gov

Career Opportunities: For potential openings in the Omics Separations and Mass Spectrometry Department at PNNL please visit http://omics.pnl.gov/careers.

www.omics.pnl.gov