Computer Networks 254 (2024) 110800

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Data Article

Data signals for deep learning applications in Terahertz communications

Duschia Bodet [a,*], Jacob Hall [b], Ahmad Masihi [a], Ngwe Thawdar [b], Tommaso Melodia [a], Francesco Restuccia [a], Josep M. Jornet [a]

[a] Northeastern University, 360 Huntington Ave, Boston, MA 02115, USA
[b] Air Force Research Lab, 26 Electronic Parkway, Rome, NY 13441, USA

ARTICLE INFO

Keywords:
Broadband communications
Machine learning
sub-THz
Experimental dataset

ABSTRACT

The Terahertz (THz) band (0.1–10 THz) is projected to enable broadband wireless communications of the future, and many envision deep learning as a solution to improve the performance of THz communication systems and networks. However, there are few available datasets of true THz signals that could enable testing and training of deep learning algorithms for the research community. In this paper, we provide an extensive dataset of 120,000 data frames for the research community. All signals were transmitted at 165 GHz but with varying bandwidths (5 GHz, 10 GHz, and 20 GHz), modulations (4PSK, 8PSK, 16QAM, and 64QAM), and transmit amplitudes (75 mV and 600 mV), resulting in twenty-four distinct bandwidth-modulation-power combinations each with 5,000 unique captures. The signals were captured after down conversion at an intermediate frequency of 10 GHz. This dataset enables the research community to experimentally explore solutions relating to ultrabroadband deep and machine learning applications.

Specifications Table

Subject	Computer Networks and Communications
Specific subject area	Terahertz Communications for 6 G Networks and Beyond
Type of data	Modulated time-domain signals
Data collection	All waveforms were acquired using the TeraNova Testbed described in [1,2,3], and [4]
Data source location	Institution: Northeastern University City: Boston, MA Country: United States
Data accessibility	Repository name: Northeastern University Digital Repository Service (DRS) Follow the *Orange (2024)* link provided from the data set's web page to access the data files.
Related research article	J. Hall, J. M. Jornet, N. Thawdar, T. Melodia and F. Restuccia, "Deep Learning at the Physical Layer for Adaptive Terahertz Communications," in IEEE Transactions on Terahertz Science and Technology, vol. 13, no. 2, pp. 102–112, March 2023, doi: 10.1109/TTHZ.2023.3237697

1. Value of the Data

- Many works are exploring the possibility of using deep learning and neural networks for applications in Terahertz Communications. However, several researchers have identified the need for an extensive experimental dataset of Terahertz data signals to train and test algorithms [5,6]. With 60,000 data signals of varying modulation orders and bandwidths, this is the first dataset of its kind to be available to the research community.
- THz communication researchers who want to test Deep Learning techniques on a real sub-THz data will benefit from this data set, especially given the volume of unique data captures that are available as well as the ultrabroadband nature of the signals (up to 20 GHz of bandwidth). Especially researchers who do not have access to hardware required to generate THz signals or who do not have the resources to perform a data collection campaign of this scope will benefit from this data.
- The diversity of signals in this dataset can enable various applications to further experiments. Processing these results could enable experimental testing of deep learning solutions for channel estimation, equalization, demodulation techniques, and much more.

2. Background

Initially, this data was collected to explore the ability to identify the

* Corresponding author.
E-mail address: bodet.d@northeastern.edu (D. Bodet).

https://doi.org/10.1016/j.comnet.2024.110800
Received 30 July 2024; Accepted 9 September 2024
Available online 12 September 2024

bandwidth and modulation of a sub-THz link. Thus, the data was collected for a static scenario varying the modulation, bandwidth, and signal-to-noise ratio (SNR) observed by the receiver. This dataset, however, can be used to explore other deep learning applications for sub-THz communications. By making it available, we allow other THz communications researchers to perform their own studies and analysis with the dataset.

3. Data Description

The signals are saved as SigMF data files using the file format detailed by the National Telecommunications and Information Administration (NTIA). Each signal has a meta data file accompanying it. The dataset is organized by the bandwidth of the transmitted signal as shown in Fig. 1. Within the subgroup of each bandwidth, there are folders for given modulation type and transmit voltage used. Each of these folders has 5000 captured data signals and one noise capture for reference. Specific instructions on how to read and decode the data can be found in a ReadMe file uploaded with the data files.

4. Experimental Design, Materials and Methods

This data was collected using the TeraNova Testbed described thoroughly in [4]. Raw data bits were modulated to an intermediate frequency (IF) of 20 GHz in MATLAB with a sampling frequency of 90 GHz. The IF signals were then converted to analog signals by Keysight's Arbitrary Waveform Generator (AWG) M8196A before being fed into an up-converter chain to bring it to a radio frequency (RF) of 165 GHz. 40-dBi lens antennas designed by Anteral were used at both the transmitter and receiver to improve the signal strength. Both the up and down-converters were custom-designed by Virginia Diodes (VDI) for D-Band experiments. Each multiplies its Local Oscillator (LO) signal by four before mixing the signal with the IF/RF signal. For these experiments, the LO at the transmitter was set to 33.75 GHz and at the receiver it was 31.25 GHz for an IF of 10 GHz observed by Keysights Digital Storage Oscilloscope (DSO) DSOZ632A. The DSO sampled and saved each capture with a sampling frequency of 160 GHz.

These sampled IF signals originally contained multiple copies of each received header, pilot, and data portion of the frame stored in a .mat files. For ease of use, we have parsed the captured signals to only contain the pilot and data portion of a frame before converting them to SigMF files following the NTIA format requirements.

Limitations

The largest limitation of the data set is that all the data is taken with the transmitter and receiver in the same relative positions. In other words, given the static nature of the experimental set-up, this entire dataset essentially experienced the same channel conditions.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief. We confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT authorship contribution statement

Duschia Bodet: Data curation, Validation, Writing – original draft, Writing – review & editing. **Jacob Hall:** Investigation, Methodology, Software, Validation. **Ahmad Masihi:** Validation. **Ngwe Thawdar:** Conceptualization, Funding acquisition, Supervision. **Tommaso Melodia:** Methodology, Supervision. **Francesco Restuccia:** Conceptualization, Methodology, Supervision. **Josep M. Jornet:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision.

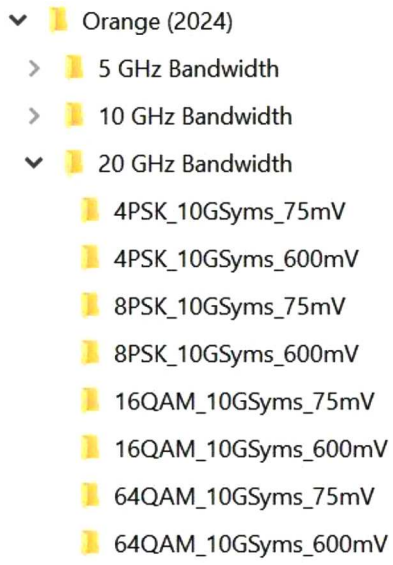

Fig. 1. File Organization for Orange (2024).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Dataset is linked in the submitted manuscript

Acknowledgements

This work was supported in part by the US Air Force Research Laboratory [Grant FA8750-20-1-0200] and in part by the US National Science Foundation [Grant CNS-2011411]. The views expressed are those of the authors and do not reflect the official guidance or position of the United States Government, the Department of Defense or of the United States Air Force. The experimental dataset was collected by Northeastern University team and will be disseminated per DoD memorandum on Fundamental Research dated 24 May 2010. Approved for public release. AFRL-2024-3608.

References

[1] P. Sen, V. Ariyarathna, A. Madanayake, J.M. Jornet, A versatile experimental testbed for ultrabroadband communication networks above 100 GHz, Comput. Netw. 193 (2021) 108092.
[2] P. Sen, D. Pados, S. Batalama, E. Einarsson, J.P. Bird, J.M. Jornet, The TeraNova platform: an integrated testbed for ultra-broadband wireless communications at true terahertz frequencies, Comput. Netw. 179 (2020) 107370.
[3] P. Sen, J.M. Jornet, Experimental demonstration of ultra-broadband wireless communications at true Terahertz frequencies, in: IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2019.
[4] P. Sen, V. Ariyarathna, J.M. Jornet, Experimental wireless testbed for ultrabroadband Terahertz networks, in: 14th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization, 2020.
[5] S. Helal, H. Sarieddeen, H. Dahrouj, T.Y. Al-Naffouri, M.-S. Alouini, Signal processing and machine learning techniques for Terahertz sensing: an overview, IEEE Signal Process. Mag. 39 (5) (2022) 42–62.
[6] U. Nissanov, G. Singh, Antenna technology for Terahertz wireless communication. Machine Learning in Terahertz Communication, Springer, Cham, 2023.

Duschia Bodet received her BS and MS in Electrical Engineering with a concentration in Communications, Controls, and Signal Processing in 2021. During her undergraduate years, she spent six months co-oping at Raytheon in radar systems, and another six months as a research assistant for the Air Force Research Labs with the Griffiss Institute. For her master's thesis she investigated new modulation solutions for Terahertz communications under the guidance of Dr. Josep M. Jornet. She is continuing her research with Dr. Jornet as a PhD student focusing on modulation schemes, MIMO, and physical layer solutions for (sub-)THz communications.

Jacob Hall received the BS degree in electrical and computer engineering from the State University of New York Polytechnic Institute, Utica, NY, USA, in 2020, and the MS degree in electrical and computer engineering from Northeastern University, Boston, MA, USA, in 2022. His MS degree was funded by the DoD's SMART Scholarship-for-Service Program with the Air Force Research Laboratory (AFRL) Information Directorate, Rome, NY, USA, as his sponsor. He is currently with AFRL under their THz program. His research interests include deep learning and THz communications.

Ahmad Masihi is a PhD Student and Graduate Research Assistant in the Department of Electrical and Computer Engineering at Northeastern Univerity. He received the BS degree in electrical engineering from Khajeh Nasir Toosi University of Technology, Tehran, Iran, in 2019, and the MS degree in electrical and telecommunications engineering from Sharif University of Technology, Tehran, Iran, in 2022. He is currently pursuing the PhD degree in electrical engineering under the guidance of Dr. Josep Miquel Jornet in the Ultra-broadband Nanonetworking Laboratory.

Ngwe Thawdar received the BSc degree from the Binghamton University, Binghamton, NY, USA, in 2009, and the MSc and PhD degrees from the State University of New York University at Buffalo, Buffalo, NY, in 2011 and 2018, respectively, all in electrical engineering. Since 2012, she has been with Air Force Research Laboratory Information Directorate, Rome, NY. Her research includes wireless spread spectrum communications, cooperative communications, and software-defined radio implementation. Her current research focuses on communications and networking in emerging spectral bands such as mm-waves and terahertz band frequencies. Dr. Thawdar was the recipient of the 2019 AFRL Early Career Award, the 2021 AFRL Information Directorate Scientist/Engineer of the Year award, and the 2022 IEEE International Conference on Communications Best Paper Award.

Tommaso Melodia is the William Lincoln Smith Professor with the Department of Electrical and Computer Engineering at Northeastern University in Boston. He received his Laurea (integrated BS and MS) from the University of Rome - La Sapienza and his PhD in Electrical and Computer Engineering from the Georgia Institute of Technology, USA in 2007. He is an IEEE Fellow, an ACM Distinguished Member, and a recipient of the National Science Foundation CAREER award. Prof. Melodia is the Founding Director of the Institute for the Wireless Internet of Things, a research institute, think tank, and technology incubator in the areas of wireless, 5G/6G, networking, IoT, and applications of AI to systems. It counts over 150 researchers, faculty, and PhD students. Prof. Melodia is also the Director of Research for the PAWR Project Office, a $100M public-private partnership building testbeds for advanced wireless research throughout the United States. Prof. Melodia is the Director of Colosseum, the Open RAN Digital Twin and the world's largest emulator of wireless systems. He received several best paper awards, including at IEEE Infocom 2022. Prof. Melodia is a frequent Keynote Speaker at prime IEEE and ACM events. He is the Editor in Chief for Computer Networks and a co-founder of the 6G Symposium, and served as the Technical Program Committee Chair for IEEE Infocom, and General Chair for ACM MobiHoc, among others. Prof. Melodia's research on modeling, optimization, and experimental evaluation of wireless networked systems has been funded by many US government and industry entities, including US National Science Foundation, DARPA, the Office of the Undersecretary of Defense, the Air Force Research Laboratory, NTIA/Department of Commerce, Office of Naval Research, and Army Research Laboratory, among others. His current research interest include Open RAN (open, programmable, and virtualized wireless systems), AI for inference and control in wireless systems, infrastructure and spectrum sharing for wireless systems.

Francesco Restuccia is an Assistant Professor in the Department of Electrical and Computer Engineering at Northeastern University. He received his PhD in Computer Science from Missouri University of Science and Technology in 2016, and his BS and MS in Computer Engineering with highest honors from the University of Pisa, Italy in 2009 and 2011, respectively. His research interests lie in the design and experimental evaluation of next-generation edge-assisted data-driven mobile systems. Prof. Restuccia's research is funded by several grants from the US National Science Foundation and the Department of Defense. He received the Office of Naval Research Young Investigator Award, the Air Force Office of Scientific Research Young Investigator Award and the Mario Gerla Award in Computer Science, as well as best paper awards at IEEE INFOCOM and IEEE WOWMOM. Prof. Restuccia has published over 60 papers in top-tier venues in computer networking, as well as co-authoring 16+ U.S. patents and three book chapters. He regularly serves as a TPC member and reviewer for several top-tier ACM and IEEE conferences and journals. He is a Senior Member of the IEEE and ACM.

Josep M. Jornet (M'13–SM'20–F'24) is a Professor in the Department of Electrical and Computer Engineering, the director of the Ultrabroadband Nanonetworking (UN) Laboratory, and the Associate Director of the Institute for the Wireless Internet of Things at Northeastern University (NU). He received a Degree in Telecommunication Engineering and a Master of Science in Information and Communication Technologies from the Universitat Politècnica de Catalunya, Spain, in 2008. He received his PhD degree in Electrical and Computer Engineering from the Georgia Institute of Technology, Atlanta, GA, in August 2013. Between August 2013 and August 2019, he was in the Department of Electrical Engineering at the University at Buffalo (UB), The State University of New York (SUNY). He is a leading expert in terahertz communications, in addition to wireless nano-bio-communication networks and the Internet. In these areas, he has co-authored >250 peer-reviewed scientific publications, including one book, and has been granted five US patents. His work has received over 18,000 citations (h-index of 62 as of July 2024). He is serving as the lead PI on multiple grants from U.S. federal agencies including the National Science Foundation, the Air Force Office of Scientific Research, and the Air Force Research Laboratory as well as industry. He is the recipient of multiple awards, including the 2017 IEEE ComSoc Young Professional Best Innovation Award, the 2017 ACM NanoCom Outstanding Milestone Award, the NSF CAREER Award in 2019, the 2022 IEEE ComSoc RCC Early Achievement Award, and the 2022 IEEE Wireless Communications Technical Committee Outstanding Young Researcher Award, among others, as well as four best paper awards. He is a Fellow of the IEEE and an IEEE ComSoc Distinguished Lecturer (Class of 2022–2023, Extended to 2024). He is also the Editor-in-Chief of the Elsevier Nano Communication Networks journal and Editor for IEEE Transactions on Communications and Nature Scientific Reports.