SPACOMM 2020

The Twelfth International Conference on Advances in Satellite and Space Communications

February 23 - 27, 2020

Lisbon, Portugal

SPACOMM 2020 Editors

Timothy Pham, Jet Propulsion Laboratory, USA

Paul Labbé, Defence Research and Development Canada, Canada

SPACOMM 2020

Forward

The Twelfth International Conference on Advances in Satellite and Space Communications (SPACOMM 2020), held between February 23-27, 2020 in Lisbon, Portugal, continued a series of events attempting to evaluate the state of the art on academia and industry on the satellite, radar, and antennas based communications bringing together scientists and practitioners with challenging issues, achievements, and lessons learnt.

Significant efforts have been allotted to design and deploy global navigation satellite communications systems, Satellite navigation technologies, applications, and services experience still challenges related to signal processing, security, performance, and accuracy. Theories and practices on system-in-package RF design techniques, filters, passive circuits, microwaves, frequency handling, radars, antennas, and radio communications and radio waves propagation have been implemented. Services based on their use are now available, especially those for global positioning and navigation. For example, it is critical to identify the location of targets or the direction of arrival of any signal for civilians or on-purpose applications; smarts antennas and advanced active filters are playing a crucial role. Also progress has been made for transmission strategies; multiantenna systems can be used to increase the transmission speed without need for more bandwidth or power. Special techniques and strategies have been developed and implemented in electronic warfare target location systems.

We welcomed academic, research and industry contributions. The conference had the following tracks:

- Satellite and space communications
- Satellites and nano-satellites
- Satellite/space communications-based applications

We take here the opportunity to warmly thank all the members of the SPACOMM 2020 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to SPACOMM 2020. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also thank the members of the SPACOMM 2020 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that SPACOMM 2020 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the domain of satellites and space communications. We also hope that Lisbon, Portugal provided a pleasant environment during the conference and everyone saved some time to enjoy the historic charm of the city.

SPACOMM 2020 Chairs

SPACOMM Steering Committee
Timothy T. Pham, Jet Propulsion Laboratory - California Institute of Technology, USA
Stelios Papaharalabos, LN2 dB, Greece

SPACOMM Industry/Research Advisory Committee

Vittorio Dainelli, Rheinmetall Italia S.p.A., Italy

Brian Niehoefer, TÜV Informationstechnik GmbH, Germany

SPACOMM 2020

Committee

SPACOMM Steering Committee

Timothy T. Pham, Jet Propulsion Laboratory - California Institute of Technology, USA
Stelios Papaharalabos, LN2 dB, Greece

SPACOMM Industry/Research Advisory Committee

Vittorio Dainelli, Rheinmetall Italia S.p.A., Italy
Brian Niehoefer, TÜV Informationstechnik GmbH, Germany

SPACOMM 2020 Technical Program Committee

David N. Amanor, Intel Corporation, USA
Suayb S. Arslan, MEF University, Turkey
Michael Atighetchi, Raytheon BBN Technologies, USA
Xavier Bellekens, University of Strathclyde in Glasgow, Scotland, UK
Scott Burleigh, NASA-JPL, USA
Yao-Yi Chiang, University of Southern California - Spatial Sciences Institute, USA
Vittorio Dainelli, Rheinmetall Italia S.p.A., Italy
Raed S.M Daraghmah, Palestine Technical University, Palestine
Arun Das, Arizona State University, USA
Yiping Duan, Tsinghua University, China
Mohamed A. Elshafey, Military Technical College, Cairo, Egypt
Luiz Carlos Gadelha de Souza, Federal University of ABC, Brazil
Gregory Hellbourg, International Center for Radio Astronomy Research | Curtin University, Western
Australia
Sundararaja Sitharama Iyengar, Florida International University, USA
Seifallah Jardak, Bristol Research & Innovation Laboratory - Toshiba, UK
Adil Hakeem Khan, Nation College of Engineering and Technology, Guna, India
Arash Komaee, Southern Illinois University, Carbondale, USA
Fandel Lin, National Cheng Kung University, Taiwan
Pablo Madoery, Universidad Nacional de Córdoba, Argentina
Krešimir Malarić, University of Zagreb, Croatia
Sara Migliorini, Università degli Studi di Verona, Italy
Nelli Mosavi, Johns Hopkins University, USA
Brian Niehoefer, TÜV Informationstechnik GmbH, Germany
Nele Noels, Ghent University, Belgium
W. David Pan, University of Alabama in Huntsville, USA
Krishna Pande, National Chiao Tung University, Taiwan
Stelios Papaharalabos, LN2dB, LLC, Greece
Cathryn Peoples, Ulster University, UK
Timothy Pham, Jet Propulsion Laboratory, USA

Ermanno Pietrosemoli, The Abdus Salam International Centre for Theoretical Physics (ICTP), Italy
Cong Pu, Marshall University, Huntington, USA
Alexandru Rusu, University Politehnica of Bucharest, Romania
Ridha Soua, University of Luxembourg, Luxembourg
Predrag Spasojevic, Rutgers University, USA
Cristian Lucian Stanciu, University Politehnica of Bucharest, Romania
Veeru Talreja, West Virginia University, USA
Giulio Telleschi, MBDA, Italy
Stefanos Vrochidis, Information Technologies Institute, Greece
Hong Wei, University of Maryland, USA
Yunfan Gerry Zhang, Independent Researcher, USA
Yimin Daniel Zhang, Temple University, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Region-based N-cuts Polarimetric SAR Image Segmentation Algorithm

Ling-Kai Zhao
Institute of Communication Technology
Harbin Institute of Technology of China
Harbin, 150001, China
e-mail:19s005050@stu.hit.edu.cn

Ming-Chuan Yang
Institute of Communication Technology
Harbin Institute of Technology of China
Harbin, 150001, China
e-mail:mcyang@hit.edu.cn

Abstract—**Due to the imaging principle, there is a large number of noise points in polarimetric Synthetic Aperture Radar(SAR) images. This coherent noise leads to inaccurate segmentation results. To solve this problem, this paper proposes a region-based N-cuts polarimetric SAR image segmentation algorithm by combining the K-means clustering algorithm and N-cuts algorithm. Firstly, K-means clustering algorithm is used to pre-segment polarimetric SAR images to form segmented regions. Secondly, the similarity measurement matrix is constructed on the basis of pre-segmentation. Finally, the N-cuts algorithm is introduced to cluster regional nodes and to realize image segmentation. This method makes full use of the over-segmentation characteristic of K-means clustering algorithm, and it significantly reduces the computation. Combined with the global optimization of atlas segmentation algorithm, the performance of segmentation results is improved. In this paper, full polarization E-SAR(the Experimental airborne SAR System of DLR) images are used for experiments, and processing results of various polarization characteristics are compared. Experimental results show that region-based N-cuts algorithm is efficient and practical for image segmentation. The results also show that this method has higher level of precision and shorter computation time than both K-means clustering algorithm and N-cuts algorithm.**

Keywords—Polarimetric SAR; Polarization Characteristic; K-means; region-based N-cuts; Image Segmentation

I. INTRODUCTION

Synthetic aperture radar imaging is a kind of microwave remote sensing technology, which has the characteristics of working all day in all weather. It is often used in aviation and space remote sensing. Polarimetric SAR obtains the polarization scattering matrix of the imaging area by adjusting the polarization mode of the working electromagnetic wave. The matrix contains rich ground object information, which improves the ability of SAR to obtain scene target information and is of great significance for improving the accuracy of image interpretation [1].

In recent years, with the maturity of the basic theoretical system of radar polarization measurement, the segmentation and classification of polarimetric SAR images have broad application potentials in unsupervised classification, target detection and recognition.

Due to the imaging principle of SAR images, there are a large number of noise spots in polarized SAR images. This kind of coherent noise brings difficulties to SAR image segmentation. Therefore, the theory of polarization SAR image segmentation based on graph theory is still developing, and many new segmentation methods and optimization algorithms can be tried in the future.

This paper is structured as follows: Section II explains the related works about this paper; Section III introduces the characterization of polarimetric SAR images; Section IV presents the K-means Clustering algorithm and N-cuts algorithm; Section V introduces the Region-based N-cuts algorithm; Section VI presents experimental results and analysis of region-based N-cuts algorithm; and Section VII presents our overall conclusions.

II. RELATED WORK

In recent years of development, many researches combined classic segmentation algorithm and image segmentation algorithm based on specific theory. Zhu [2] used watershed over-segmentation and scattering parameters to carry out iterative classification, which effectively avoided salt and pepper noise, but failed to retain texture feature information. Xi [3] combined mean-shift algorithm and normalized cut algorithm to process ordinary images. Even though the processing effect was good, the images he used were simple-- the number of pixels in the image was small.

In this paper, we use a region-based method to process a set of polarimetric SAR images. K-means clustering is a widely used classical clustering algorithm, which has the characteristics of simple operation and local optimization. However, it is sensitive to noise and outliers. In region-based N-cuts algorithm, K-means clustering is used for over-segmentation, and the subsequent calculation units are converted from pixels to regions. Taking the obtained region into the N-cuts algorithm can effectively reduce the calculation and achieve the globally optimal image segmentation result. According to the polarization characteristics of SAR image data, we extract various polarization characteristics as data sets for image segmentation, which contain more information than ordinary images. Also, the region-based N-cuts algorithm has better processing performance and shorter calculation time than both K-means algorithm and N-cuts algorithm.

III. CHARACTERIZATION OF POLARIMETRIC SAR

Generally, the complex scattering matrix $\vec{S}$ of the target can be directly obtained from the fully polarized SAR measurement data [4]:

$$\vec{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \tag{1}$$

where: S_{HV} represents the target backscatter when horizontal *(H)* emission and vertical *(V)* polarization reception, and S_{HH}, S_{VV}, S_{VH} is similar. Considering the situation of static measurement, and the target satisfies the anisotropy $S_{HV} = S_{VH}$

Polarization scattering matrix $\vec{S}$ can represent the total scattering echo power *Span* of an object:

$$Span = |S_{HH}|^2 + |S_{HV}|^2 + |S_{VH}|^2 + |S_{VV}|^2 \qquad (2)$$

In order to extract physical information from the complex scattering matrix $\vec{S}$, it is usually constructed three-dimensional Pauli feature vector $\vec{k}$

$$\vec{k} = \frac{1}{\sqrt{2}}[S_{HH}+S_{VV} \quad S_{HH}\text{-}S_{VV} \quad 2S_{HV}]^T \qquad (3)$$

In practical situations, real objects are often considered as distributed targets. The concept of "distributed target" is derived from the fact that a non-static or unstable radar target changes over time, that is, the target exists in a changing environment, and the target scattering also changes with time and space. Therefore, the use of the scattering matrix $\vec{S}$ to characterize the object is limited, so that the polarization coherence matrix $\vec{T}$ and the covariance matrix $\vec{C}$ are introduced [5]. These two are used as new data representation methods to reflect the scattering mechanism of the distributed target. The polarization coherence matrix and the covariance matrix are obtained by time-averaging or spatial averaging processing on the $\vec{S}$ matrix.

A multi-view polarization coherent matrix can be obtained by performing non-coherent averaging of multiple single-view polarization coherent matrices. For multi-view fully polarized SAR data, the polarization coherence matrix $\vec{T}$ is defined as

$$\langle \vec{T} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i k_i^{*T} \qquad (4)$$

where N is the visual number; k_i is the scattering vector of $N{=}i$, k_i^{*T} is the conjugate transposition of k_i .

In the practical application of polarization SAR image processing, most of them are studied in the case of single station backscattering system, using the scattering matrix $\vec{S}$, the polarization coherence matrix $\vec{T_3}$ and the covariance matrix $\vec{C_3}$, the combination and operation can extract more polarization characteristics. The comprehensive utilization of ground object polarization information and texture information is of great significance for improving the accuracy of polarimetric SAR image segmentation.

IV. K-MEANS CLUSTERING ALGORITHM AND N-CUTS ALGORITHM

In this section, we introduce two existing image segmentation methods, which are combined in the region-based algorithm.

A. K-means Clustering Algorithm

As a classical clustering algorithm, K-means clustering algorithm is also an effective unsupervised classification method in the field of image processing. The algorithm divides samples into K clusters through an iterative process. It uses the mean of these K clusters to represent samples of each class in the iterative process. The iteration is terminated when the overall error is minimized.

Firstly, let N_i be the total number of samples in the number i cluster Γ_i , and m_i is the mean of samples[6]:

$$m_i = \frac{1}{N_i}\sum_{y \in \Gamma_i} y \qquad (5)$$

The sum of the squared errors between each sample y in Γ_i and the mean m_i of the class is added to each category:

$$J_e = \sum_{i=1}^{K}\sum_{j=1}^{N_i} \left\| y_{ij} - m_i \right\|^2 \qquad (6)$$

where: K is the number of categories ; y_{ij} is the j-th sample in the i-th class.

J_e represents the total squared error produced by K cluster centers $m_1, m_2, \cdots, m_k$, representing K sample subsets $\Gamma_1, \Gamma_2, \cdots, \Gamma_k$, called the sum of squared errors. So that J_e reaches the clustering result of the minimum value is the optimal result under the sum-of-squared-error criterion.

B. Image Segmentation Based on Graph Theory

a) Graph cut

With the development of image segmentation technology, applying graph theory to image segmentation is a hot topic in current research. In the field of graph theory, digital images can be mapped to a weighted undirected graph. The basic processing unit in the image is a pixel, and the basic processing unit in the figure is a vertex, which establishes a one-to-one correspondence between them: pixels in the image are mapped to vertices in the graph, adjacent relation of pixels is mapped to edge, and the similarity of pixel value between them is mapped to the weight on the edge [6].

For the graph $G = (V, E)$ is a weighted undirected graph [7], as shown in Figure 1: the circle represents the vertex, and the line segment connecting the circle represents the edge. Weight of the edge connecting vertices v_i and v_j is represented by w_{ij} , reflecting a certain property between vertices, such as the similarity relation. The degree of a vertex is defined by :

$$d_i = \sum_j w(i, j) \qquad (7)$$

The degree of a vertex represents the association of the vertex with other vertices in the graph.

If there are two graphs $G = (V, E)$ and $G_1 = (V_1, E_1)$ that satisfy the relation $V_1 \subseteq V, E_1 \subseteq E$, then graph G_1 is called a subgraph of graph G. When two subgraphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ satisfy the relation $V_1 \cup V_2 = V$, G_1 and G_2 are called complementary, see Figure 1:
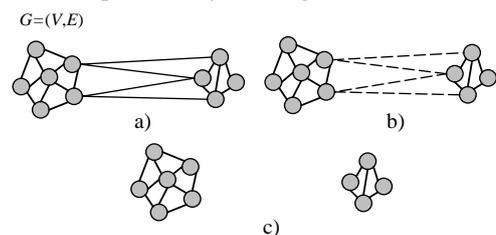

Fig. 1. Graph Segementation

b) N-cuts Criteria

For graph $G = (V, E)$, it is an undirected graph with sideband weight. Remove some edges and divide the graph into two sets of disjoint vertices to satisfy relation $A \cup B = V, A \cap B = \varnothing$. Different cuts can be obtained according to different criteria.

Shi and Malik proposed[8] the normalized cut criterion:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(B, A)}{assoc(B, V)} \qquad (8)$$

where: $assoc(A, V)$ is the sum of the weights of all the vertices in A and the vertices of the graph. In this way, the similarity between two classes is defined. The smaller the similarity between classes, the better the segmentation result.

c) Solution of the N-cuts Algorithm

In order to solve the optimal $Ncut(A, B)$ value, the N-cuts algorithm transforms the problem into the field of linear algebra research and becomes a solution to the eigenvector and eigenvalue problems of the matrix. Suppose a graph $G = (V, E)$ is divided into two parts, A and B, so that we set $N = |V|$ and x are N-dimensional indication vectors. When $x_i > 0$, node i is in A, and when $x_i < 0$, node i belongs to B.

Let D be a diagonal matrix with diagonal elements d_i. A symmetric matrix W is called a weight matrix and its element value $W(i, j) = w(i, j)$. The problem of solving the minimum value of $Ncut(A,B)$ can be transformed into the form of equation (10):

$$\min_x Ncut(x) = \min_y \frac{y^T (D - W) y}{y^T D y} \qquad (9)$$

The above formula must satisfy the two conditions of $y_i \in \{1, -b\}$ and $y^T DI = 0$ [9]. This formula is in the form of Rayleigh quotient. If the value range of y is broadened to real number, solving the minimum $Ncut$ is equivalent to solving the equation under the generalized eigenvalue system:

$$(D - W) y = \lambda D y \qquad (10)$$

The solution of the equation is:

$$y = \arg . \min_{y^T DI = 0} \frac{y^T (D - W) y}{y^T D y} \qquad (11)$$

N-cuts method obtains the eigenvector corresponding to the second small eigenvalue as the optimal solution for $Ncut(A,B)$ by solving the equation under the generalized eigenvalue system. The optimal feature vector y is an N-dimensional solution vector. Select a segmentation threshold value y^*, divide vertices with values greater than y^* into set A, and divide the rest into set B. If the image needs more detailed segmentation, the algorithm can be recursively called to realize multiple segmentation.

d) Construction of Weight Matrix in N-cuts Algorithm

The weight matrix W indicates the mapping relationship between pixels in the image and nodes in the constructed network. The smaller the weight, the smaller the degree of similarity between proved nodes. To some extent, it means the smaller probability of being divided into the same class.

We construct the weight matrix W_{ij} as follows:

$$W_{ij} = e^{\frac{-\|F_{(i)} - F_{(j)}\|_2^2}{\sigma_I^2}} * \begin{cases} e^{\frac{-\|X_{(i)} - X_{(j)}\|_2^2}{\sigma_X^2}} & \text{i,j are adjacent} \\ 0 & \text{i,j are not adjacent} \end{cases} \qquad (12)$$

where $F_{(i)}$ represents the feature space vector and $X_{(i)}$ represents the position space vector of the pixel point i. The weight matrix considers both the feature vector information and the spatial position information. σ_I and σ_X are weight coefficients, according to experience, whose values can be selected according to experience. The range of σ_I is mostly 50~80, and the range of σ_X is usually 80~100.

V. REGION-BASED N-CUTS ALGORITHM

In this paper, a new computational complexity image segmentation method is designed by combining K-means clustering algorithm and N-cuts algorithm. Firstly, we use K-means clustering algorithm to pre-segment the image. The segmented image is divided into several regions. Then, according to the K-means clustering results, the K clustering centers output by the algorithm are taken as representative points of each region to complete the construction of weighted undirected graph. The eigenvalues are used to describe the similarity relation between regional points. Finally, N-cuts algorithm is used to segment these regions.

The implementation of the algorithm and the specific steps are as follows:

(1) preprocessing the image by using a K-means clustering algorithm;

(2) The image clustered by K-means is divided into many regions, and the number of regions is greatly reduced compared with the number of pixel points in the image. Each region is extracted and labeled, and the clustering center of the region is used as a region point to replace the region. The information contained in the regional points includes feature vector information and spatial location information.

(3) According to step (2), the original image becomes an undirected graph composed of region points, which is input as N-cuts. The weight matrix W is constructed by using regional points, and then clustering segmentation is carried out by using N-cuts algorithm[11].

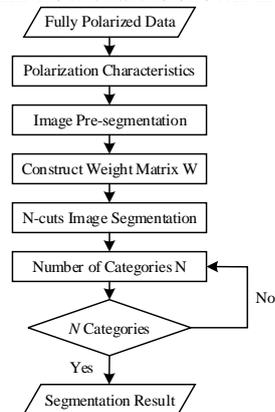The algorithm flow chart is shown in Figure 2

Fig. 2. Flow chart of region-based N-cuts algorithm

Compared with the traditional N-cuts algorithm, the region-based N-cuts image segmentation algorithm has the advantages of less computation and more accurate segmentation effect.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we introduce the selection of parameters, experimental process and analysis of experimental results.

A. Data sets description

In this experiment, a set of polarization characteristic are selected of a fully polarized SAR image. These polarization characteristic are obtained by several target decomposition methods [8], including Cloud decomposition, Yamaguchi four-component decomposition and Krogager decomposition. We also add the power spectrum image to data sets. These polarization characteristics describe ground objects from the perspective of scattering mechanism and physical structure.

A high-resolution polarimetric SAR image is used in the experiment, and the specific technical indexes are given in Table I:

TABLE I.   EXPERIENTAL DATA

Data Source	Resolution/m	Image size	Area/km^2
E-SAR	3×3	1300×1200	39.028

The image used in the experiment is shown in Figure 3(a), which is composed of fully polarized L-band data. The scene in the image is an area of Oberpfaffenhofen, Germany. Under the condition of fine division, there are about 10 kinds of ground objects in the image, including forests, buildings, farmland, bare land (obvious airport runway) and other main ground objects.

(a)   (b)   (c)   (d)   (e)   (f)   (g)   (h)   (i)
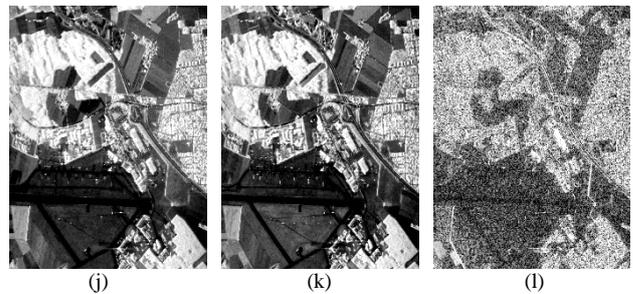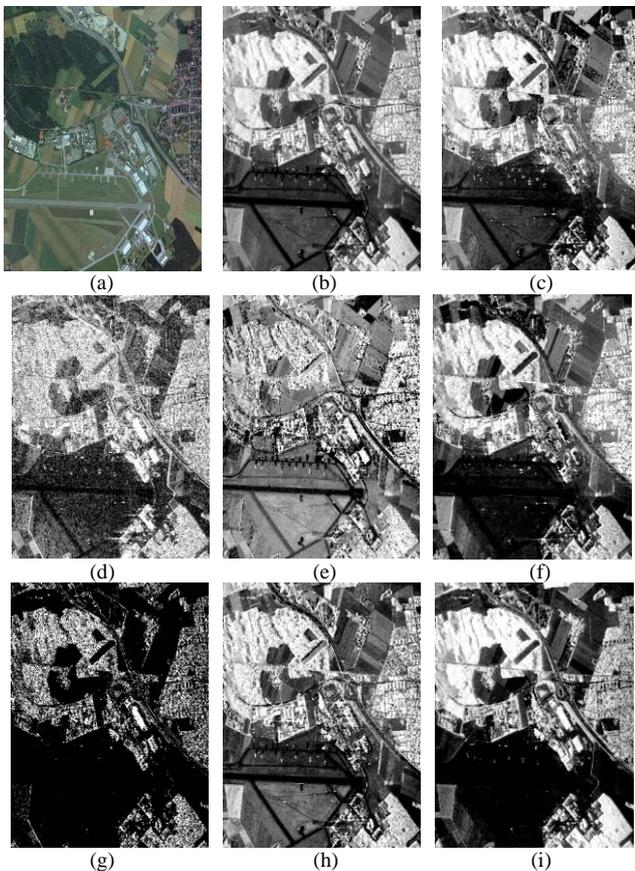

(j)   (k)   (l)

Fig. 3.   (a) Original image. (b) Power spectrum image. (c) Bragg scattering for Cloud Decomposition. (d) Dihedral scattering. (e) Volume scattering. (f) Dbl scattering component (g) Hlx scattering component (h) Odd scattering component (i) Vol scattering component (j) K_d dihedral angular scattering component (k) K_h helix scattering component (l) K_s spherical scattering component

- **Power spectrum image.** Figure 3(b) shows the power spectrum image, which use the average power represents the scattering intensity of ground objects directly.
- **Cloud Decomposition** Figure 3(c)(d)(e) shows the results of Cloud decomposition. The results represent three generation factors of the equivalent single scattering target: Bragg scattering C_{11}, dihedral scattering C_{22}, and volume scattering C_{33}, which respectively represent three mutually independent and complementary correlation scattering processes.
- **Yamaguchi Four-component Decomposition.** Figure 3(f)(g)(h)(i) shows the result of Yamaguchi four-component decomposition. Comparing the four images, we find that the urban area has the strongest double-bounce scattering, which is the main scattering mechanism of urban area. Due to the presence of relatively tall vegetation in forest area, the canopy consists of a large number of branches and leaves, mainly characterized by volume scattering. Surface features with less rough surface, such as farmland, bare land and some low vegetation, Bragg scattering is their main scattering mechanism. However, trunks and ground form a dihedral angle, so the radar tilting the scene will cause double-bounce scattering exhibited by such vegetation. The division of vegetation should be dominated by volume scattering, including Bragg scattering and double-bounce scattering. Most of the urban areas should be classified as double-bounce scattering, and should not be classified as volume scattering, which is a mistake with plants. For the helix scattering component, only the forest and urban parts are expressed, and they have certain differences. Applying this feature can effectively distinguish forests, cities and other features.
- **Krogager Decomposition.** Figure 3(j)(k)(l) shows the result of Krogager decomposition. Krogager is decomposed into three coherent components, which correspond to spherical scattering K_s, dihedral angular scattering K_d with a rotation angle θ, and helix scattering K_h. Both the spherical scattering and the dihedral angular scattering in the forest area are strong, while the urban part mainly exhibits dihedral angular scattering. Cultivated land and bare land areas are mainly characterized by spherical scattering. Similar to Yamaguchi decomposition, helix scattering features can exist mainly in forests as well as in urban areas.

B. Experimental Results of Image Segmentation

a) Environmental configuration

The hardware environment, where experiments are run is Intel Core i5-5200U CPU @ 2.20GHz 2.19GHz, 8GB RAM and 64-bit operating system. The process of image segmentation runs on Windows 10 operating system and uses Microsoft Foundation Class Library as its development tool. The program has interactive execution function.

b) Experimental Results of Polarization Feature Selection

We use a variety of polarization features to carry out experiments and select the following six better features as image segmentation results, which are segmented by region-based N-cuts algorithm. The parameter settings are shown in Table II:

TABLE II.   IMAGE SEGMENTATION PARAMETERS

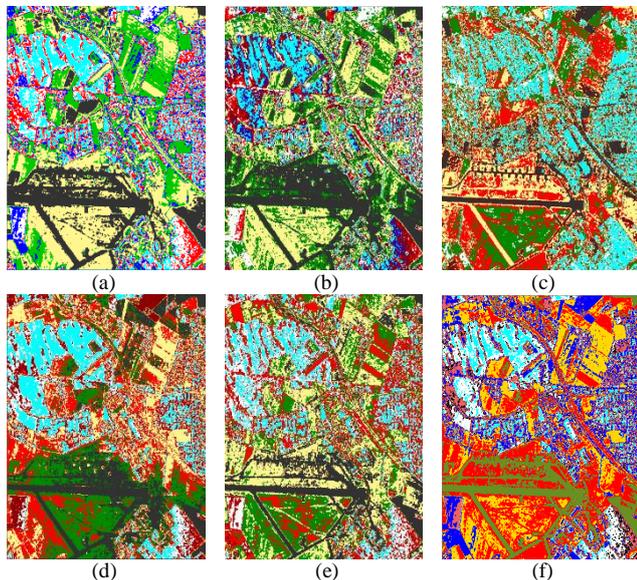| n	Max	Precision	N	σ_t^2
20	50	0.5	8	50

The results of the experiment are shown in Figure 4:

Fig. 4. (a) Krogager_ K_d scattering component (b) Krogager_ K_s scattering component (c) Cloud_ C_{33} scattering component (d) Cloud_ C_{11} scattering component (e) Yamaguchi_Odd scattering component (f) Average power λ

Reasons for selecting the above experimental data:

1) *Number of clusters n:* The original image has a higher resolution and the number of pixels is 1200×1300 . Based on actual factors, the original polarized SAR image is roughly divided into roughly 10 types of features. Therefore, the number of pre-segmented areas is at least 10. It was found through experiments that clustering numbers $k = 100$ and $k = 20$ were selected for clustering, and there was no significant difference. Moreover, the split result is used as the input of the subsequent N-cuts algorithm, and the parameter $k = 20$ is selected as the number of regions, which can reduce the calculation amount and reduce the computer memory requirement.

2) *Maximum number of iterations and precision:* Set the number of iterations to 50 and the iteration precision to 0.5 as the iteration termination condition. They can simultaneously define clusters to achieve better results.

3) *Selection of the initial cluster center:* The number of pixels in the original image is huge, and it is impossible to select a representative representative point based on experience. Therefore, the initial representative points are selected in the natural order of the images[10].

4) *Number of times algorithm was called:* Set $N = 8$, which can roughly cover the total number of features.

5) *Selection of polarization characteristics:* In the experiment, six different polarization features with large amount of information are selected, including various scattering mechanisms, such as spherical scattering, dihedral angular scattering, Bragg scattering, and volume scattering to distinguish different features. The average power can directly express the scattering intensity of the ground object, which is also a very representative feature.

According to the experimental results, it can be concluded that the average power λ is used for segmentation. Most of the other polarization features can be clearly distinguished for a particular two species, but it is impossible to distinguish between multiple types of features. For example, the Bragg scattering characteristics of Yamaguchi decomposition have a better effect on the division of cultivated land and bare land, but the effect on forest and urban area is poor. Therefore, the average power is selected for the experiment and the overall performance is better.

c) Performance evaluation metric

The experiment also selected the average power λ as the original image, changed the parameter settings in the experiment, and conducted a control experiment. For different cluster numbers K and N values in the N-cuts algorithm, the K-means clustering algorithm, N-cuts algorithm and region-based N-cuts algorithm are used to process the image separately, and the processing time is recorded. The processing results are shown in Table III. It can be seen from the experimental data that the K-means clustering is used for preprocessing, and the calculation time based on the region-based N-cuts algorithm is much shorter than the calculation time of the N-cuts algorithm. The amount of calculation depends on the number of regions. The segmentation accuracy of the region-based N-cuts algorithm is better than the K-means clustering algorithm, and the overall optimal result can be obtained.

TABLE III.   ALGORITHM PROCESING TIME

Value of N	Value of k	Time of K-means(s)	Time of N-cuts	Time of region-based N-cuts(s)
2	20	218.42	8902.19	293.17
4	20	236.48	14039.74	324.87
8	20	209.23	26839.25	388.96
8	10	54.04	26839.25	72.74
8	50	440.23	26839.25	637.28
8	100	793.29	26839.25	993.94

d) Experimental Results of Image Segmentation

The average power is selected as the original image and the N-cuts algorithm is recursively called to get the final segmentation result.

A total of 4 recursive calls are made to the N-cuts algorithm for segmentation, and 20 types of features were divided into 8 categories. Then, based on artificial experience, similar features are combined and to make further fine division of ground objects. The final segmentation results in 9 types of features. The experimental results are shown in Figure 5:
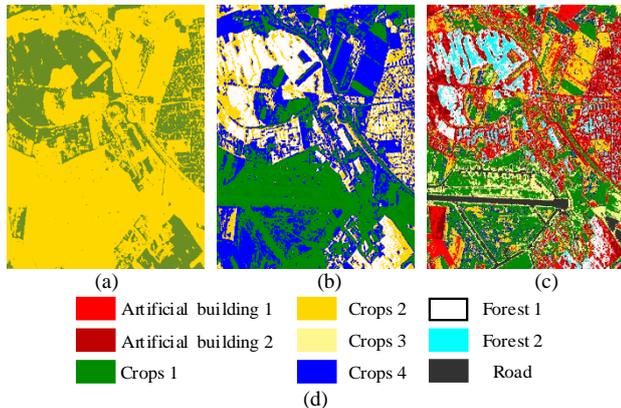
■ Artificial building 1	■ Crops 2	□ Forest 1
■ Artificial building 2	■ Crops 3	■ Forest 2
■ Crops 1	■ Crops 4	■ Road

(d)

Fig. 5. (a) Segmentation result in *N=2*. (b) Segmentation result in *N=4*. (c) Segmentation result in *N=8*. (d) Category labels

The segmentation result shows that the N-cuts algorithm is continuously called to divide the original image into several parts to achieve the purpose of classification. A better segmentation effect can be achieved by selecting the average power λ as the original image input. Moreover, the algorithm has a short processing time, occupies less computer memory, and is highly efficient.

VII. CONCLUSIONS AND FUTURE WORKS

This section presents our overall conclusions and the works needed in the future.

A. Conclusions

From the perspective of the research problem of polarimetric SAR image segmentation, this paper first analyzes the shortcomings of various existing image processing methods, and then proposes a region-based N-cuts algorithm and applies it to polarimetric SAR images segmentation. The proposed method is validated by E-SAR full polarization data, and this method is compared with other image segmentation algorithm results for evaluation. Experiments show that: experiments are carried out on multiple polarization feature images, and the influence of different features on segmentation results is compared and analyzed. The average power image is used to achieve the best segmentation effect. For polar SAR image processing, region-based N-cuts algorithm is objective and efficient compared to traditional image segmentation algorithms. Applying the mature theory in graph theory to the field of image segmentation, combined with K-means clustering, we can implement region-based image segmentation algorithm. This method can greatly improve the calculation efficiency, reduce the computer memory requirements, and achieve efficient processing. The region-based N-cuts algorithm is continuously called to divide the original image into several parts to achieve the purpose of classification. In addition, the algorithm is highly portable. In different actual scenarios, it can be improved according to specific needs and combined with different algorithms.

B. Future works

In the future experiment about region-based N-cuts image segmentation method, we consider to merge various polarization features to access some more representative new features. Also, other data sets with larger data and some other traditional image segmentation method should be carried out to compare with the proposed method. Moreover, the criteria for performance evaluation [11] need to be improved. In addition, the algorithm has good portability. So according to specific requirements, the algorithm can be improved or combined with other algorithms in different actual scenes.

REFERENCES

[1] A. Fabijanska, "Normalized cuts and watersheds for image segmentation," IET Conference on Image Processing (IPR 2012), London, 2012, pp. 1-6.

[2] T. Zhu, J. Yu and X. Li, "SAR image classification method based on superpixel and Span-Pauli decomposition" Journal of Huazhong University of Science and Technology(Natural Science Edition), 2015, 43(07): pp.77-81.

[3] Q. Xi, "An improved image segmentation algorithm base on normalized cut," 2010 2nd International Conference on Computer Engineering and Technology, Chengdu, 2010, pp. V7-293-V7-296.

[4] I. V. Bogach, D. D. Lupiak, Y. Y. Ivanov and O. V. Stukach, "Analysis and Experimental Research of Modifications of the Image Segmentation Method Using Graph Theory," 2019 International Siberian Conference on Control and Communications (SIBCON), Tomsk, Russia, 2019, pp. 1-4.

[5] W. Yang, L. Guo, T. Zhao and G. Xiao, "Improving Watersheds Image Segmentation Method with Graph Theory," 2007 2nd IEEE Conference on Industrial Electronics and Applications, Harbin, 2007, pp. 2550-2553.

[6] S. Zhu, X. Zhu and Q. Luo, "Graph theory based image segmentation," 2013 6th International Congress on Image and Signal Processing (CISP), Hangzhou, 2013, pp. 593-598.

[7] T.Jiang, F.Tan and Z.Xu, "Research on polarimetric SAR image segmentation method," wireless internet technology, 2017(12), pp. 15-16.

[8] J. Lee, "Fundamentals and Applications of Polarimetric Radar Imaging," publishing house of electronics industry, 2014, 6, pp. 23-71.

[9] T.Zhu, J.Yu, X.Li and Q.Yan, "SAR Image classification method based on superpixel and Span-Pauli decomposition," journal of huazhong university of science and technology, 2015, 43(07), pp. 77-81.

[10] X.Zhang, "Pattern Recognition (3rd Edition)," Tsinghua university press, 2010, 8, pp. 187-231.

[11] L.Zhao, E.Chen, Z.Li, Q.Feng and L.Li, "Segmentation of PolSAR Data Based on Mean-Shift and Spectral Graph Partitioning and Its Evaluation," Journal of Wuhan University Information Science Edition, 2015, 40(08), pp. 1061-1068.

A Tensor Completion-Based Selection Method of a Single Data Collection Point for Multiple Shelters in a UAV Enabled Disaster Recovery Network

Azusa Danjo* † Shinsuke Hara*, Takahiro Matsuda‡ and Fumie Ono§

*Graduate School of Engineering, Osaka City University

3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558–8585, Japan

Email: {danjo@c., hara@}info.eng.osaka-cu.ac.jp

†DAIHEN Corporation

2-1-11 Tagawa, Yodogawa-ku, Osaka 532–8512, Japan

‡Graduate School of Systems Design, Tokyo Metropolitan University

6-6 Asahiga-oka, Hino, Tokyo, 191–0065, Japan

Email: takahiro.m@tmu.ac.jp

§National Institute of Information and Communications Technology

3-4 Hikarino-Oka, Yokosuka, Kanagawa, 239–0847, Japan

Email: fumie@nict.go.jp

Abstract—In this paper, we propose a method for selecting a single data collection point for an Unmanned Aerial Vehicle (UAV)-enabled disaster recovery network where UAVs need to collect messages from refugees while hovering above multiple shelters. Just by sensing the Received Signal Strengths (RSSs) for the wireless signals transmitted from multiple shelters at *fewer randomly-selected* points in the three-dimensional (3D) region above a disaster-hit area, through a tensor completion, the proposed method reconstructs full RSS maps for the multiple shelters in the entire 3D region, and selects a single data collection point using the reconstructed RSS maps where UAVs can collect messages from them more quickly and more energy-efficiently. The experimental results reveal that by sensing RSSs in 25% of a 3D region, the proposed method can collect messages from three transmitters using a shorter route of UAV in a shorter message collection time.

Keywords–Unmanned Aerial Vehicle; Received Signal Strength; Tensor Completion.

I. INTRODUCTION

When a large-scale disaster causes catastrophic damage to infrastructure and lifeline, such as roads, buildings, water, gas, electric power and information networks, a lot of people are dead and forced to live as refugees in shelters. The primary concern of refugees is to let them know their families' safety and let their families know their safety, therefore, before the recovery of terrestrial information network, it is essential to quickly construct a temporary information network in the disaster-hit area.

Since Unmanned Aerial Vehicles (UAVs) can quickly fly anywhere in the sky, they are promised to play an essential in it, where UAVs, which are equipped with wireless communication tools, visit to collect messages from refugees while hovering above shelters [1].

In the usage of UAVs, there is one serious problem, and that is the operation time is severely limited even thanks to the current battery technology, such as typically 20-40 minutes and the energy consumption by flying and hovering is dominant. Therefore, when a region in a disaster-hit area is assigned to a UAV, if there are many shelters to visit in the region, it may be prohibitive for the UAV to visit all the shelters. In this case, it should be better for the UAV to collect messages from multiple shelters at *a single* point where the UAV can receive the wireless signals from them simultaneously in better channel conditions. This is also effective for enlarging the size of region assigned to a UAV.

Let us assume that a disaster-hit area has been divided into many three-dimensional (3D) regions. The quality of wireless channel is proportional to Received Signal Strength (RSS), so we believe it reasonable to divide the UAV operation into two stages, such as RSS sensing stage and message collection stage. In the RSS sensing stage, a UAV visits many points in a 3D region assigned to it, and it senses the RSSs for multiple shelters there. Then, we select the one out of the visited points where the RSSs can be larger for the multiple shelters. In the message collection stage, a single point has been selected for each region in the disaster-hit area, so a single or multiple UAVs can visit all the points according to the route solved as a vehicle routing problem (VRP) [2]. Note that how to apply the VRP for the UAV route planning is out of the scope of this paper.

In the RSS sensing stage, there is one more serious problem, and that is wireless channel condition is affected by the distribution of surrounding buildings in urban areas, so even for a single shelter, it is difficult to find a good point where the condition is better for the shelter. Furthermore, if the 3D region is huge, it may be also prohibitive for the UAV to fly fully in the entire 3D region. Therefore, it is mandatory to develop an efficient method to select a single message collection point even for a huge disaster-hit region.

In this paper, by dealing with the 3D-RSS spatial distribution (MAP) as a 3D tensor, we propose a tensor completion-based selection method of a single data collection point for multiple shelters. The use of UAV has been well discussed in disaster recovery scenarios; a UAV-assisted network as a substitute for terrestrial information communication network [3], image collection [4], rescue [5] and so on. However, there have been no works which try to select message collection points in a UAV-enabled disaster recovery network. VRP is
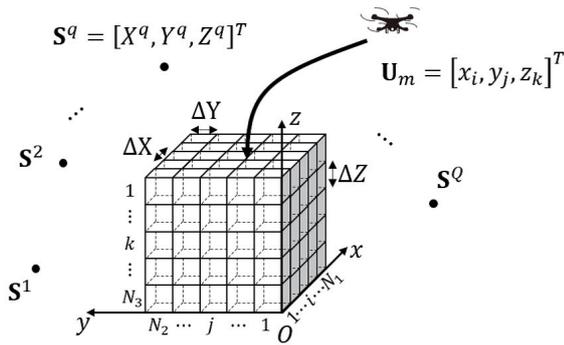
Figure 1. Layout model.

applied for a UAV delivery problem [6], but it selects delivery points in advance, although there is a lot of freedom in selecting them.

This paper is organized as follows. Section II presents the models and assumptions, and Section III mathematically states the problem. Section IV proposes three methods for selection of the RSS sensing points and collection points of messages, and Section V evaluates the performance of the proposed methods using experimental data. Finally, Section VI concludes the paper.

II. MODELS AND ASSUMPTIONS

It is assumed that the wireless communication tool between refugees and a UAV is Wi-Fi [7], since it has been widely and commonly used as a Wireless Local Area Network (WLAN) standard over the world. However, since many Wi-Fi communication links do not guarantee communication qualities, a Wi-Fi connection may take a long time, for instance, more than 10 sec [8], and this implies that if refugees independently try to access the access point installed in the UAV hovering above them, the message collection time of the UAV becomes extremely longer. Therefore, it is also assumed that in each shelter, there is a server which has already stored the messages of refugees staying it. When the UAV visits a shelter, the server tries to transmit its stored messages to the UAV by Wi-Fi. In the following, servers are referred to as "transmitters (TXs)," and their stored messages are simply referred to as "data."

A. Layout model

Figure 1 shows the layout model, where there is a 3D region of interest $\mathcal{R}$, Q stationary TXs and a flying UAV. The size of $\mathcal{R}$ is X m $\times$ Y m $\times$ Z m, and $\mathcal{R}$ is divided into $N_1 \times N_2 \times N_3 = N_v$ voxels with size of ΔX m $\times$ ΔY m $\times$ ΔZ m, where N_1, N_2 and N_3 denote the number of voxels dividing $\mathcal{R}$ in the x-axis, y-axis and z-axis directions, respectively. We label the voxels with $l = i + N_1(j-1) + N_1 N_2(k-1)$ ($l = 1, 2, \ldots, N_v$, $i = 1, 2, \ldots N_1, j = 1, 2, \ldots N_2, k = 1, 2, \ldots N_3$), define the location of the center of gravity of the $l = i + N_1(j-1) + N_1 N_2(k-1)$-th voxel as $\mathbf{V}_l = [x_i, y_j, z_k]^T$, and define the set of all $\mathbf{V}_l$s as $\mathcal{V} = \{\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{N_v}\}$. In addition, Q TXs are located at the outsides of $\mathcal{R}$ and we define the location of the q-th TX as $\mathbf{S}^q = [X^q, Y^q, Z^q]^T$ ($q = 1, 2, \ldots, Q$).

The UAV visits the centers of gravity of voxels to sense RSSs for Q TXs. We define the location of the m-th visit of

UAV as $\mathbf{U}_m$ ($m = 1, 2, \ldots, M$), the set of all $\mathbf{U}_m$s as $\mathcal{U} = \{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_M\}$ and the sensing rate as $B_{sense} = M/N_v$.

B. Route model

In the RSS sensing stage, a realistic way is to visit and sense RSSs only at fewer part of the voxels along a predetermined route Ω in $\mathcal{R}$. When the voxels to visit have been selected as $\mathcal{U}$, determining the shortest route for visiting all the voxels is a kind of the 3D Traveling Salesman Problem (TSP), and there have been a lot of solvers for the TSP [9] [10]. In this paper, therefore, we do not care about how to determine the shortest route, namely, the "route" in this paper naturally means the shortest route.

C. Wireless communication model

In the RSS sensing stage, Q TXs transmit their wireless signals for UAV, and UAV tries to receive the wireless signals from all the TXs and sense their RSSs, however, in each visit, UAV cannot always receive the wireless signals from all the Q TXs. We define the probability that UAV can receive the wireless signal from the q-th TX in the m-th visit as C_{rec}, and define the RSS as R_m^q in dBm when it is possible.

On the other hand, in the data collection stage, after a single data collection point has been selected, UAV collects the data from all the TXs while hovering at the selected point. When Wi-Fi is employed between the TXs and UAV, the data transmission rate is an increasing function of RSS [11], for instance, $f(R_{\hat{m}}^q)$ bps, where $\hat{m}$ is the voxel index for the data collection point and $R_{\hat{m}}^q$ is the RSS for the q-th TX at the data collection point. Therefore, defining the size of data for the q-th TX as D^q bits, its required data collection time is given by

$$T_{\hat{m}}^q = D^q / f(R_{\hat{m}}^q). \tag{1}$$

III. PROBLEM STATEMENT

Selection of a single data collection point is divided into the following two optimization problems.

A. RSS sensing stage

UAV sequentially visits $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_M$, so defining the route length as L_{route}, the problem is given by

$$\text{find} \quad \mathcal{U} = \{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_M\} \subseteq \mathcal{V} \tag{2}$$

$$\text{which minimizes} \quad L_{route} = \sum_{m=1}^{M-1} \|\mathbf{U}_{m+1} - \mathbf{U}_m\|. \tag{3}$$

B. Data collection stage

The data collection point, which is included in a subset of $\mathcal{V}$, should minimize the total data collection time, so the problem is given by

$$\text{find} \quad \mathbf{U}_{\hat{m}} \in \text{ a subset of } \mathcal{V} \tag{4}$$

$$\text{which minimizes} \quad T_{total} = \sum_{q=1}^{Q} T_{\hat{m}}^q. \tag{5}$$

At $\mathbf{U}_m$, the dominant factor to determine the total data collection time is the maximum among $T_m^1, T_m^2, \ldots, T_m^Q$, and from (1), the data collection time is a decreasing function

of RSS. Therefore, when $D^1 \approx D^2 \approx \ldots \approx D^Q$, defining R_m^{min} as the minimum of $R_m^1, R_m^2, \ldots, R_m^Q$, this problem is reformulated as

$$\text{find} \quad \mathbf{U}_{\hat{m}} \in \text{ a subset of } \mathcal{V} \quad (6)$$

$$\hat{m} = \arg_m \max \quad \{R_1^{min}, R_2^{min}, \ldots, R_M^{min}\}. \quad (7)$$

Once $\hat{m}$ has been selected, the corresponding voxel label can be identified. For instance, for $\mathbf{U}_{\hat{m}} = [x_{\hat{i}}, y_{\hat{j}}, z_{\hat{k}}]^T$, a distinct voxel label is given by $\hat{l} = \hat{i} + N_1(\hat{j}-1) + N_1 N_2(\hat{k}-1)$.

IV. SOLUTIONS

We propose the following three methods to solve the problem.

A. Full visits and full RSS sensing

If UAV can visit all the voxels where it senses RSSs for all the TXs, it can select the best data collection point, although the route is maximized:

$$\text{find} \quad \mathbf{U}_{\hat{m}} \in \mathcal{V} \quad (8)$$

$$\hat{m} = \arg_m \max \quad \{R_1^{min}, R_2^{min}, \ldots, R_{N_v}^{min}\}. \quad (9)$$

We refer to this method as the Full Visit/Full RSS Sensing (FV/FS) method. Note that the FV/FS method gives the performance lower-bound.

B. Random visits and random RSS sensing

It is very difficult to select M voxels, which gives the shortest route. Therefore, in this paper, we randomly select M voxels out of N_v voxels.

To formulate the problem of selecting the data collection point, we define the set of visited voxels where UAV can sense RSS for the q-th TX ($q = 1, 2, \ldots, Q$):

$$\mathcal{U}^q = \{\mathbf{U}_m \in \mathcal{U} | m = 1, 2, \ldots, M, R_m^q \text{ exists}\}. \quad (10)$$

Furthermore, we define the set of visited voxels where UAV can sense RSSs for all the Q TXs:

$$\mathcal{U}^{1 \cdot 2 \cdots Q} = \bigcap_{q=1}^{Q} \mathcal{U}^q. \quad (11)$$

Using (11), this problem is given by

$$\text{find} \quad \mathbf{U}_{\hat{m}} \in \mathcal{U}^{1 \cdot 2 \cdots Q} \quad (12)$$

$$\hat{m} = \arg_m \max \quad \{R_1^{min}, R_2^{min}, \ldots, R_M^{min}\} \quad (13)$$

where R_m^{min} is replaced by 0 when m is not included in the set of indexes of $\mathcal{U}^{1 \cdot 2 \cdots Q}$. We refer to this method as the Random Visit/Random RSS Sensing (RV/RS) method.

C. Random visits, tensor completion and random RSS sensing

Similar to the RV/RS method, we randomly select M voxels out of N_v voxels, but before selecting the data collection point, we reconstruct the RSS maps for all the TXs at all the voxels by means of a 3D tensor completion *as if UAV visits all the voxels.*

We define a 3D-RSS tensor in $\mathcal{R}$ for the q-th TX as

$$\mathbf{R}^q = \{R_{ijk}^q | \quad i = 1, 2, \cdots, N_1,$$
$$j = 1, 2, \cdots, N_2, k = 1, 2, \cdots, N_3\}. \quad (14)$$

Note that we can sense the RSSs at the voxels over the predetermined route Ω where UAV can receive the wireless signal from the q-th TX. Therefore, according to the Total Variation Low Rank Tensor Completion (TVLRTC) method [12], the reconstruction problem of a tensor $\mathbf{W}^q$ is given by ($q = 1, 2, \ldots, Q$)

$$\mathbf{W}^q = \{W_{ijk}^q | \quad i = 1, 2, \cdots, N_1,$$
$$j = 1, 2, \cdots, N_2, k = 1, 2, \cdots, N_3\} \quad (15)$$

$$\underset{\mathbf{W}^q}{\text{minimize}} \quad \lambda \sum_{n=1}^{3} \|\mathbf{F}_n \mathbf{W}_{(n)}^q\|_1 + (1-\lambda)\frac{1}{3}\sum_{n=1}^{3} \|\mathbf{W}_{(n)}^q\|_* \quad (16)$$

$$\text{subject to} \quad \mathbf{W}_{\Omega^q}^q = \mathbf{R}_{\Omega^q}^q. \quad (17)$$

In (16), $\|\mathbf{C}\|_1$ denotes the ℓ-1 norm of a matrix $\mathbf{C}$, that is, $\|\mathbf{C}\|_1 = \sum_o \sum_p |c_{op}|$, $\mathbf{F}_n = \{f_{op}\}$ is a smoothness constraint matrix, where

$$f_{op} = \begin{cases} 1 & (p = o) \\ -1 & (p = o + 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (18)$$

and $\mathbf{Q}_{(n)}$ denotes the mode-n unfolding matrix of $\mathbf{Q}$. Furthermore, $\|\mathbf{A}\|_* = \sum_m \sigma_m(\mathbf{A})$ denotes the nuclear norm of a matrix $\mathbf{A}$, where $\sigma_m(\mathbf{A})$ denotes the m-th largest singular value of $\mathbf{A}$. The first and second terms of (16) mean the total variation and low-rank tensor completion, respectively and λ is a tunable parameter to balance them. On the other hand, in (17), $(\cdot)|_{\Omega^q}$ is the operation of picking up the elements along the predetermined route Ω where UAV can receive the wireless signal from the q-th TX from $(\cdot)$. So finally, replacing the elements of $\mathbf{R}^q$ by those of $\mathbf{W}^q$, the problem is given by

$$\text{find} \quad \mathbf{U}_{\hat{m}} \in \mathcal{V} \quad (19)$$

$$\hat{m} = \arg_m \max \quad \{W_1^{min}, W_2^{min}, \ldots, W_{N_1 N_2 N_3}^{min}\}. \quad (20)$$

We refer to this method as the Random Visit/Tensor Completion/Random RSS sensing (RV/TC/RS) method.

V. EXPERIMENT AND PERFORMANCE EVALUATION

To evaluate the performance of the proposed method, we conducted an experiment in Osaka City University. Figure 2 (a) shows the photo of the experiment. We placed three TXs near the windows inside the buildings, which transmit the 2.4GHz band Wi-Fi signals, as shown in Figures 2 (b) - (d). On the other hand, the RSS sensing region had 3D sizes of $X = 30$m, $Y = 8$m and $Z = 10$m. The experimental site is located in an urban area, so flying a UAV in the air is prohibitive. Therefore, instead of UAV, we attached a receiver (RX) to the top of a pole, as shown in Figure 2 (e). Dividing the sensing region by 2m interval along the x-axis, y-axis, and z-axis, namely, $\Delta X = \Delta Y = \Delta Z = 2$m, $N_1 = 15$, $N_2 = 4$, and $N_3 = 5$, we sensed the RSSs for the TXs at each voxel in total $N_v = 300$ voxels. In addition, we set the RSS sensable probability $C_{rec} = 0.6$. TABLE I shows the specifications of experiment.

Figure 3 compares the fully sensed RSS map by the FV/FS method, the randomly selected RSS map by the RV/RS method

(a) Sensing region.

(b) TX#1.

(c) TX#2.

(d) TX#3.

(e) RX.

Figure 2. Photos of the experiment.

TABLE I. SPECIFICATIONS OF EXPERIMENT.

Location	Osaka City University
Sizes of sensing region	$X = 30\text{m} \times Y = 8\text{m} \times Z = 10\text{m}$
Sensing resolution	$\Delta X = \Delta Y = \Delta Z = 2$ m $N_1 = 15, N_2 = 4, N_3 = 4,$ and $N_v = 300$
Number of TXs (Q)	3
Wireless communication tool	2.4GHz band Wi-Fi
Antenna	Dipole
RSS sensable probability	C_{rec}=0.6

and the reconstructed RSS map by the RV/TC/RS method, for the three TXs, respectively, when setting the RSS sensing ratio B_{sense} of 0.25 (= 75/300, $M = 75$) and $\lambda = 0.07$ [13]. Since the RSS sensable probability is set as $C_{rec} = 0.6$, the RSSs for TX#1, TX#2 and TX#3 are sensed only at 48 voxels, 43 voxels and 47 voxels out of 75 voxels, respectively, but we can see that the RSS maps can be accurately reconstructed by the tensor completion. Defining the Relative Square Error (RSE) of reconstructionfor the q-th TX as

$$\text{RSE}^q = \frac{||\mathbf{R}^q - \mathbf{W}^q||_F}{||\mathbf{R}^q||_F} \qquad (21)$$

where $||(\cdot)||_F$ denotes the Frobenius norm of tensor $(\cdot)$, we obtain $\text{RSE}^1 = 0.06$, $\text{RSE}^2 = 0.03$, and $\text{RSE}^3 = 0.05$.

We have set $B_{sense} = 0.25$ and $\lambda = 0.07$ to show the RSS maps in Figure 3, but Figure 4 and Figure 5 show the

dependencies of RSE^q on B_{sense} and λ, respectively. We can see that RSE^q becomes flat when B_{sense} is more than 0.2 and also it becomes flat when λ is more than 0.07.

Figure 6 compares the total route length among the FV/FS, RV/RS, and RV/TC/RS methods when setting $B_{sense} = 0.25$ and $\lambda = 0.07$. We can see that as compared with the FV/FS method, the RV/RS and the RV/TC/RS method can reduce the total route length by around 60%.

Finally, to compare the total data collection time among the three methods, according to the data sheet of Xbee Wi-Fi, we define $f(R_{\hat{m}}^q)$ in Mbps as

$$f(R_{\hat{m}}^q) = \begin{cases} 6 & (-91 \leq R_{\hat{m}}^q < -89) \\ 9 & (-89 \leq R_{\hat{m}}^q < -88) \\ 12 & (-88 \leq R_{\hat{m}}^q < -86) \\ 18 & (-86 \leq R_{\hat{m}}^q < -83) \\ 24 & (-83 \leq R_{\hat{m}}^q < -80) \\ 36 & (-80 \leq R_{\hat{m}}^q < -76) \\ 48 & (-76 \leq R_{\hat{m}}^q < -74) \\ 54 & (-74 \leq R_{\hat{m}}^q) \end{cases} . \qquad (22)$$

Figure 7 compares the total data collection time among the FV/FS, RV/RS, and RV/TC/RS methods for the case of $D^q = 100\text{MBytes}$ (800×10^6 bits) ($q = 1, 2, 3$) when setting $B_{sense} = 0.25$ and $\lambda = 0.07$. The RV/RS method needs to select the best voxel out of fewer voxels (19 voxels) where the UAV happens to be able to sense the RSSs for all the
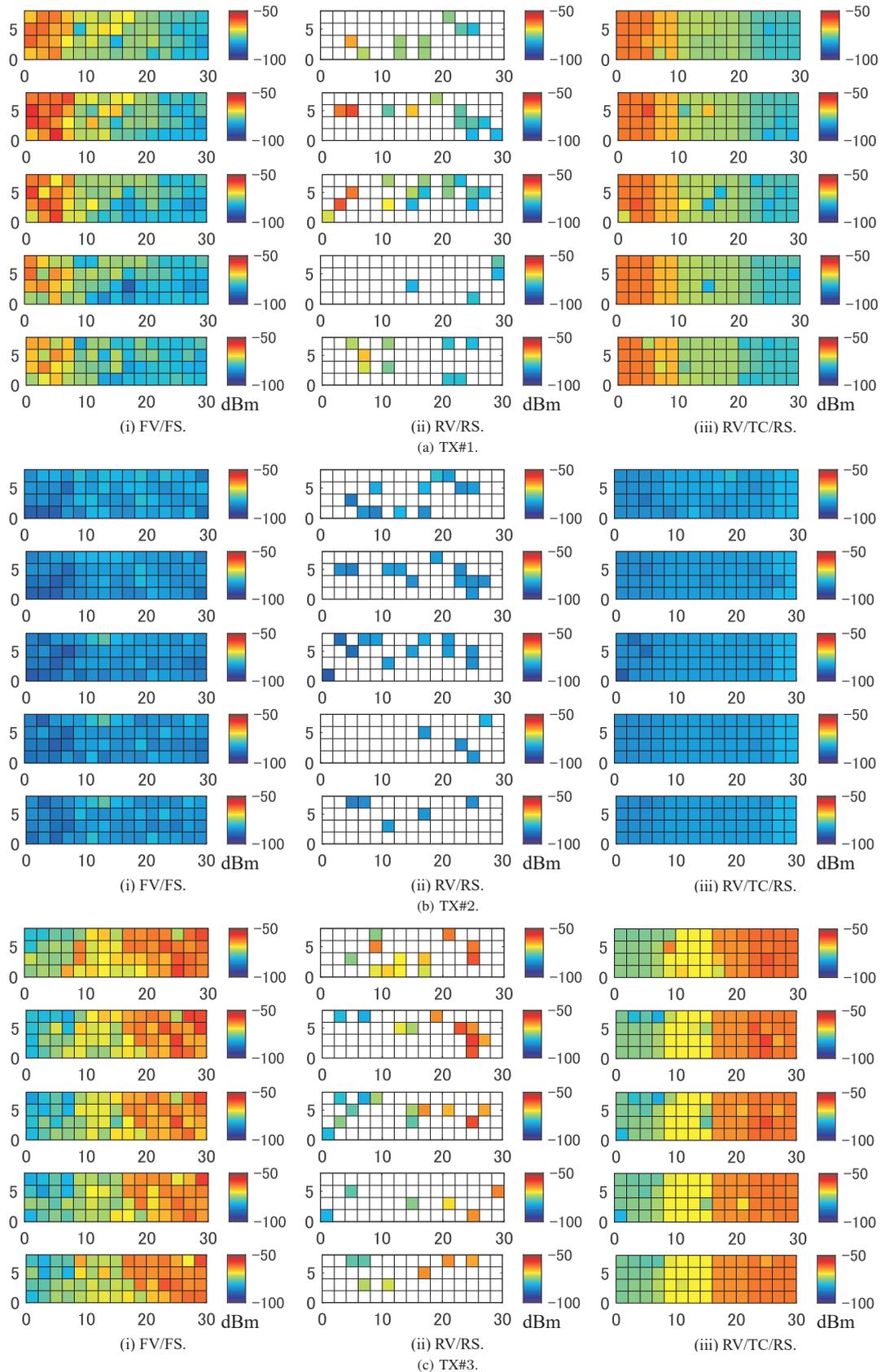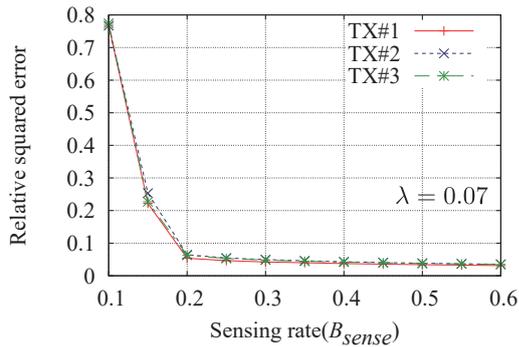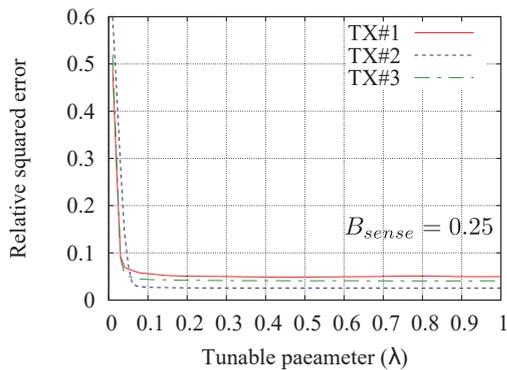
Figure 3. RSS maps ($B_{sense} = 0.25$, $\lambda = 0.07$).

Figure 4. Dependency of the RSE on the RSS sensing rate (B_{sense}).

Figure 5. Dependency of the RSE on the tunable parameter (λ).

Figure 6. Total route length.

Figure 7. Total data collection time.

three TXs, so the selected voxels cannot give higher RSSs, resulting in a much longer data collection time. On the other hand, as the clue of the RSSs for the fewer selected voxels, the RV/TC/RS method can reconstruct the RSS maps for the three TXs almost perfectly through the tensor completion as if the UAV senses the RSSs fully in the entire 3D region. We can see that there is no difference in the total data collection time between the FV/FS method and the RV/TC/RS method.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed the tensor completion-based method for selecting a single data collection point for multiple transmitters in a UAV-enabled disaster recovery network. The experimental results have shown that the RV/TC/RS method can select the truly best data collection giving the shortest data collection time by sensing the RSSs in 25% of the 3D region.

We have randomly selected a fewer RSS sensing points and simply connected them with a short route, so our future work includes the selection of RSS sensing points which can be connected with the shortest route while keeping the MAP reconstruction accuracy the highest. In addition, we have discussed a single data collection point selection, so another of our future works is to select multiple but less data collection points for multiple transmitters.
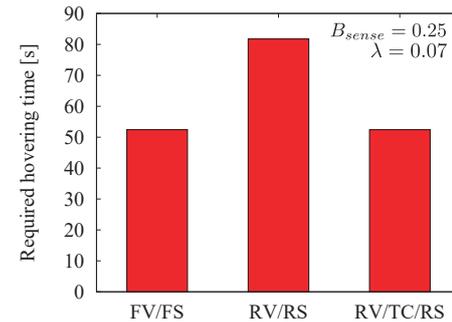
ACKNOWLEDGMENT

REFERENCES

[1] A. M. Hayajneh, S. A. R. Zaidi, D. C. McLernon, M. Di Renzo, and M. Ghogho, "Performance analysis of UAV enabled disaster recovery networks: a stochastic geometric framework based on cluster processes," *IEEE Access*, vol. 6, pp. 26215-26230, May 2018.

[2] P. Toth and D. Vigo, *The Vehicle Routing Problem*, SIAM, 2002.

[3] M. Erdelj, E. Natalizio, K. R. Chowdhury and I. F. Akyildiz, "Help from the Sky: Leveraging UAVs for Disaster Management," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 24-32, Jan.-Mar. 2017.

[4] A. Vetrivel, M. Gerke, N. Kerle, F. Nex, G. Vosselman, "Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 45-59, 2018.

[5] J.Scherer, *et al.*, "An autonomous multi-UAV system for search and rescue," *Proc. ACM, First Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use*, pp. 33-38, May 2015.

[6] K. Dorling, J. Heinrichs, G.G. Messier, and S. Magierowski, "Vehicle routing problems for drone delivery,'' *IEEE Trans. Syst., Man, Cybern. Syst.,* vol. 47, no.1, pp. 70-85, Jan. 2017.

[7] Wi-Fi Alliance, https://www.wi-fi.org/ [retrieved: Jan. 2020].

[8] C. Pei, *et al.*, "Why it takes so long to connect to a WiFi access point, " *Proc. IEEE INFOCOM 2017*, pp. 1-9, May 2017.

[9] J. Y. Potvin, "Genetic algorithms for the traveling salesman problem," Annals of Operations Research vol. 63, pp. 339-370, June 1996.

[10] M. Dorigo and L.-M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 53-66, Apr. 1997.

[11] IEEE Standard for Information technology Telecommunications and information exchange between systems Local and metropolitan area networks Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band, 2006.

[12] T.-Y. Ji, T.-Z. Huang, X.-L. Zhao, T.-H. Ma, and G. Liu, "Tensor completion using total variation and low-rank matrix factorization," Information Sciences, vol. 326, pp. 243-257, Jan. 2016.

[13] H. Nishioka *et al.*, "A compressed sensing-based 3D-RSS MAP completion for UAV routes planning," *Proc. ICST 2018*, pp. 273-277, Dec. 2018.

Cognitive Radars (CRs) Could Improve Target Engagement Success Rate

Paul Labbé

National Defence

Defence Research and Development Canada (DRDC)

Canada

Email: Paul.Labbe@drdc-rddc.gc.ca / Paul.Labbe@forces.gc.ca

Abstract— **Based on information available on Cognitive Radar (CR) technologies, one can anticipate their potential impact on target track quality that may improve target engagement success rate. CR could multiply force capabilities to respond to threats posed by nefarious groups using advanced technologies against our nations and our allies. However, so far there seems to be no CR technologies currently deployed on military platforms or for military applications but there are indications that some are in development. In the future, CR is expected to offer the capability to detect and track small targets (or stealth platforms) over much extended range than current radar under the same conditions. It could contribute to enabling a drone to disable an opposing force platform or to eliminate drones representing threats to civilian aircraft.**

Keywords-cognition; radar; target engagement; success rate.

I. INTRODUCTION

This paper is based on research done by the author [1] for the Defence Research and Development Canada (DRDC) Science and Technology Outlook function to address the following question: What are cognitive radar (CRs), their technology enablers, their applications and their implications on the Defence and Security (D&S) domains and systems? In addition, a method for assessing target engagement success rate will be summarised and used to illustrate the potential advantage of CR over traditional radar.

The concepts of CRs were developed in the 90s. Haykin's initial thoughts about radar cognition can be found in his 1990 paper [2]: "…we have coined the term radar vision. The goal here is to make radar an intelligent remote sensing device, such that it is capable of developing cognition of the surrounding environment." Later he expanded these CR concepts in several seminal papers [3][4]. Activities under the Defence Advanced Research Project Agency (DARPA) addressed the development of CRs with similar concepts starting from a Knowledge-Based (KB) system point of view with the work led by Guerci who reported in 2014 on Cognitive Fully Adaptive Radar (CoFAR), which was tested in a real environment: "A new and fully adaptive environmentally aware (cognitive) radar and signal processing architecture is introduced to meet the challenges of increasingly complex operating environments" [5]. CoFAR uses a radar centric Sense-Learn-Adapt (SLA) approach based on an 'observe, orient, predict, decide and act' (OOPDA) loop from cybernetics [6][7] well known to Command and Control (C^2) where the learn part is a Knowledge-Aided (KA) expert system with supervised training. While becoming environmentally aware, a cognitive system develops some self-awareness of its purpose (metacognitive or introspective capacity) [8]. DARPA and sibling defence organizations offer perspective extending academic works into higher level of technology readiness. They report on radar system development and publish results from trials in real environments.

Section II presents the evolution of radar technologies from adaptive to cognitive ones. Section III relates to examples of applications. Section IV provides some of the claimed performance. Section V presents a method to evaluate the trend of CRs in improving target engagement success rate. Section VI offers comments on work in progress. Section VII provides a summary of findings.

II. FROM ADAPTIVE TO COGNITIVE RADAR

Basic radar transmits a signal via an antenna to illuminate a scene. That signal bounces back from scene objects to the same antenna (monostatic) which feeds the receiver. Processing of signal echoes allows performing a variety of measurements such as location, velocity and trajectory. A Traditional Active Radar (TAR) offers some adaptive capabilities compared to basic radar. TAR was improved with adaptive receiver processing, beamforming, and constant false alarm rate. Basic radar and TAR heavily relied on the cognitive abilities of their expert operators to select, for example, waveforms and time duration on an observation area where a target was suspected to be. CRs provide some of these cognitive abilities through a learning process using statistical methods and retention of information from previous observations. A first step toward CR capabilities, besides what experienced operators could perform, was achieved by adding feedback from the receiver to the transmitter, as described in [9]. This closed-loop feedback control radar system or fully adaptive radar was labelled Fore-Active Radar (FAR) by Haykin [10]. FARs are advanced radar systems with feedback loops between fully adaptive receivers and adaptive transmitters including antenna beamforming [10] as illustrated in Figure 1.

There is no unambiguous definition of CR. Guerci [5] proposed a practical definition as follows: CR "is a system that is capable of sensing, learning, and adapting to complex situations with performance approaching or exceeding that achievable by a Subject Matter Expert (SME), especially for real time operations which demand automation."
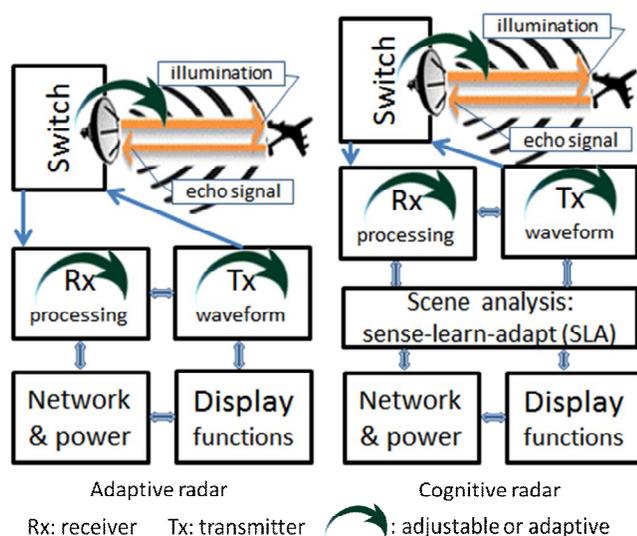
Figure 1.   Simplified diagrams of adaptive and cognitive radars.

Figure 1 provides a simplified block diagram of a CR adopting SLA as one block. Haykin [10] stated that: CR "differs from TAR as well as FAR by virtue of the following capability: the development of rules of behavior in a self-organized manner through a process called learning from experience that results from continued interactions with the environment."

"The key idea behind this new paradigm is to mimic the human brain as well as that of other mammals with echolocation capabilities (bats, dolphins, whales, etc.)" [11]. This overarching principle of a CR was inspired by the ability of bats and dolphins to track and home in on their prey. "There is much that we can learn from the echo-location system of a bat" [4]. The principle was extended keeping in mind that useful CR is expected to track multiple targets [12]. In addition, long-range detection may require different strategies or principles than biologically inspired homing in on a close target optimisation techniques.

In his 2006 CR paper [9], Haykin proposed a CR target-tracking system that continuously learned about the environment, intelligently adjusted the transmitter, and incorporated receiver-to-transmitter feedback.

In general we can say that adaptive systems react to their environment using predefined rules, but cognitive ones can develop new rules in real time with or without supervision.

Cognitive systems designed to reach users' goals use different machine learning techniques and memory retention to:
1. address immediate reaction types (parasensory and premotor);
2. plan tactics (complex and abstract information of perceptual or executive character); or
3. change strategy toward goals (dynamics of the perception–action cycle in sequential behaviour and reasoning).

A CR can learn from the observed effects from stimuli that the CR designed and generated. It can create new

algorithms based on observations of its manipulation of the environment. CRs are proactive (anticipative or predictive) while TARs are responsive—they wait until something happens. CRs probe the environment to see what happens if they transmit a signal with a given waveform or pulse shape.

A.   Enabling Technologies

CR shares technology enablers with TAR, Phased-Array Radar (PAR) systems and Digital-Array Radar (DAR) [13], these include Software Defined Radios (SDRs), SDR Sensors (SDRS) [14], software defined radar [15], field-programmable gate arrays (FPGAs), Graphics Processing Units (GPUs), cell phone Application Specific Integrated Circuits (ASICs) and Digital Signal Processors (DSPs) [16]. These common enablers include an evolution to modern phased array architectures [17] from Passive Electronically Scanned Array (PESA) to Active Electronically Scanned Array (AESA). AESAs have been deployed in missile, fighter aircraft, surveillance platforms, drones and air defence systems. AESAs' Transmit/Receive Modules (TRMs) for each antenna element, or group of antennas, enable them to accomplish multiple functions simultaneously such as radio communications, cellphones, jamming and multi pencil beams.

B.   Evolution

Interestingly, "The continued 'digitization' of radar front-ends and resultant TRM flexibility, coupled with advances in advanced KA high performance embedded computing have afforded a unique opportunity for a leap-ahead capability in a radar's ability to adapt to complex target-environment scenarios...from the nascent field of cognitive radar" [18]. One also notes impressive capabilities such as "Deep learning cognitive radar for Micro Unmanned Aircraft Systems (UAS) detection and classification," Mendis (2017) [19] showed the exceptional capabilities of this type of cognitive radar to distinguish between very small drones at relatively low signal to noise ratios. Used of a strategy to select waveforms, as for the airborne CARABAS radar, to provide good resolution to distinguish targets [20][21] at short ranges [22][23], and High Frequency (HF) Over-The-Horizon (OTH) radar by dynamically selecting frequencies and waveforms in response to sensed spectrum occupancy [23][24] and ionospheric conditions, approach cognitive radar [23][25] but lack SLA ability to design new strategies.

III.   EXAMPLES OF APPLICATION

Cognitive Radar Information Networks (CRINs) were proposed in [26] to secure large, unmanned borders such as the 3,700 km Canada/U.S. border that runs through the Great Lakes. Border security represents a major challenge given the limited sensor capabilities to cover continuously the various ships and platforms crossing in such large areas. Using cognitive sensors and radars in a network of collaborative systems could be a sensible approach to enhance risk mitigation and for reducing operator overload.

CRIN and CR should be considered as a source of inspiration in studies for updating the North Warning System

(NWS) to provide support for North American Aerospace Defense Command (NORAD) aerospace warning and control missions for insuring a continuous coverage of the air and maritime approaches to and within North America. Currently DRDC is investigating the value of adding some cognitive capabilities to our military radar systems.

IV. CR CLAIMED PERFORMANCES

In the unclassified domain one can find several CR test results from simulations and a few experiments in real environment scenes. In general the results obtained show significant improvements in all the parameters radar systems can provide about a target such as position, trajectory, velocity, acceleration, distance, altitude, jet engine modulation, and early target detection. In most operational scenarios, providing early accurate positional information is critical [27]. Reported experimental results show that CR outperformed TAR by at least one order of magnitude when using the same signal processing performance metric [10]. In a real environment using a CoFAR, [5] reports a 10 to 15 dB Signal-to-Interference-plus-Noise Ratio (SINR) for Moving Target Indicator (MTI) improvement against non-homogenous clutter. This means that the reported results show some evidence that CR significantly outperforms TAR in the situations described.

Several advantages of CR over TAR, when using the same signal processing technique, were reported in [28], e.g., to reach a Root-Mean-Square Error (RMSE) of the velocity of about 7.5 m/s, CR took 0.17 s and TAR 2.4 s, which is more than one order of magnitude faster.

The cumulated evidence from CR publications allows inferring that a substantial advantage can be gained by adopting new military CR technologies against current and emerging threats, e.g., such advanced capabilities could better defeat threats from low cost emerging technologies for Cognitive Electronic Warfare (CEW) and weaponized drones.

Assuming that CRs could rapidly change their modulation or transmitting schemes, it may make such radar signal difficult to jam with legacy jamming techniques. If advanced CRs use noise-like signal and radiation patterns with a high degree of unpredictability, they would be more difficult to jam even with CEW.

When compared to TARj with similar antenna performance, there are specific advantages that could be attributed to CRs such as:
1. Extend detection range [28].
2. Shorten the time to acquisition in target tracking [29].
3. Improve the accuracy of positional information of tracked targets [30].
4. Reduce risk in selecting an intelligent choice of decision-making mechanism in the transmitter for a prescribed goal of interest when confronted with environmental uncertainties and disturbances in real time [31].
5. Offer the agility necessary to defeat 'Digital Radio Frequency Memory' based jammer/spoofing technology which essentially captures the transmitted signal and

reradiates it towards the radar receiver, typically with some delay or modulation attached [31][32].
6. Detect smaller radar cross section (RCS) targets such as drones and stealth platforms [33].
7. Increase capabilities of passive radar and multistatic radar systems which could detect some stealth aircraft better than conventional monostatic radars, since first-generation stealth technology (such as the F-117) reflects energy away from the transmitter's line of sight, effectively increasing RCS in other directions, which multistatic passive radars can monitor [20].
8. Increase the likelihood of defeating standard electronic warfare and CEW by unexpectedly and rapidly changing waveform characteristics [31].
9. Use difficult to detect waveforms including wideband signals, limit opposing force opportunities to use or trigger their electronic countermeasures to reduce a potential range advantage a radar system may offer in targeting opposing force assets or platforms [32].
10. Reduce the susceptibility (lower likelihood) of being fooled by artificial coherent target energy, including decoys [31].
11. Use built-in shielding against misdirecting/degrading Direction-of-Arrival (DOA) measurements [31].
12. Offer enhanced geolocalization by networked CRs [31][34].
13. Deliver faster high precision information about targets [5][28].
14. Can contribute to communications when other means are jammed [34].

V. POTENTIAL IMPACT OF CR ON TARGET ENGAGEMENT SUCCESS RATE

Cybernetic models were used in analyzing coalition live and simulated exercises. These cybernetic models allowed simulating the decision-making processes made by operators at command centers including:
1. monitor the situation;
2. assess the situation and estimate adversarial intent;
3. develop alternative Courses-Of-Action (COAs);
4. predict consequences for both sides (own and opposing forces);
5. decide a COA; and
6. direct the COA execution while monitoring an evolving situation in the environment (repeat 1 to 6).

By using cybernetic models to interpret data and information collected during experiments, one can process and evaluate the stages through a set of Measures Of Performance (MOPs). Similarly, Measures Of Effectiveness (MOEs) can provide an assessment of the resulting degree of mission accomplishment in scenarios to scale MOPs relatively to MOEs, i.e., asserting both that the system performs its tasks well and that those are the correct tasks.

When running these cybernetic models over data collected from a large number of trials, one obtains a graph like the one depicted in Figure 2 relating interception success rate as a function of system time delay and track data accuracy and their timeliness. From these trials, the main delay was due to the human in the loop and imposed

transmission timing. However, track data timeliness and accuracy were also affecting target interception success rate. If legacy radars are replaced by CRs delivering more accurate data with less delay, one can infer from Figure 2 an increase in target interception success rate.
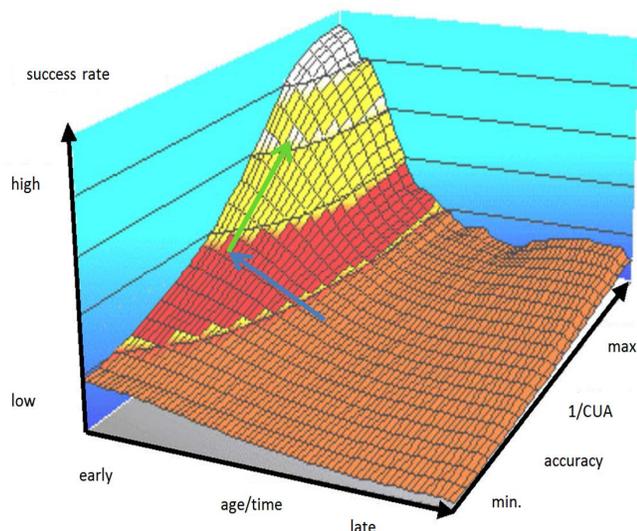

Figure 2. Potential mission success rate as function of input information age and accuracy (inverse of Circular Uncertainty Area, CUA) for a fixed effector's strategy [35].

A possible interpretation of Figure 2 is that track data in the red area and below are not appropriate for targeting with a specific effector. If a given CR track is delivered to this effector ten times earlier (blue arrow) than from a TAR and that the accuracy is several times better (green arrow), then this track data may reach the yellow/white area where it might be considered of "targeting grade". That is, this engagement has a higher likelihood of success.

In [36][37], we posited that with the advent of cognitive networks, sensors, and shooters, it seems more achievable today to accelerate and improve the sensor-to-shooter loop with a high likelihood of lower fratricide, lower collateral damage, and precise effects on intended targets or end state.

VI. MULTISTATIC MULTIBAND COMPLEXITY ADDS CAPABILITIES

Recently the author identified some new aspects of cognitive radar when adding other dimensions such as multistatic, multiband and the agility offered by the AESA approach in order to reduce the infrastructure of antennas on a platform (ship or aircraft). Platform Radio Frequency (RF) systems to provide radar, communications and Electronic Warfare (EW) functions are evolving towards an integrated system approach know as MultiFunction RF (MFRF) systems [38]. This adds some complexity to the systems but offers several advantages in trying to defeat new threats such as stealth aircraft and hypervelocity cruise missiles.

The hypothesis is how much is feasible and advantageous to build a Cognitive-Multistatic-Multiband-Radar Network (CMMRN)?

Essentially, a CMMRN is a system made of interconnected radar systems with the following capabilities:

A. Multistatic

Monostatic radar uses one antenna for transmitting a signal illuminating a scene that may include a target and uses the same antenna for receiving reflections of the signal. Multistatic radar uses at least either two transmitting or two receiving antennas, providing multistatic beam angles between the illuminating signal(s) and the reflected ones. When these beam angles are sufficiently large, such complex radar configurations are outweighed by the potential advantages of early detection of cruise missiles and stealth platforms, which increases the likelihood of successfully intercepting incoming threats [39]. Different types of multistatic radar systems exist: active and passive [40]. Passive radar uses transmitters of opportunity while active radar uses own transmitter. Currently, transmitters of opportunity are not available in Canadian Air Defense Identification Zone (CADIZ), which includes the entire Canada's Arctic Archipelago as part of the overall North American Aerospace Defense Command (NORAD) modernization, aka, Evolution of North American Defense (EvoNAD). This CADIZ is substantially extended from the previous CADIZ of the North Warning System (NWS). The modernization of the NWS not only needs to address this extended area but the range of potential threats to the continent which are more complex and increasingly difficult to detect, such as threats posed by adversarial cruise missiles and new ballistic missiles. However, with the advent of new Low-Earth-Orbit (LEO) satellite constellations, new transmitters of opportunity [41] illuminating Canadian Arctic are becoming a reality to consider.
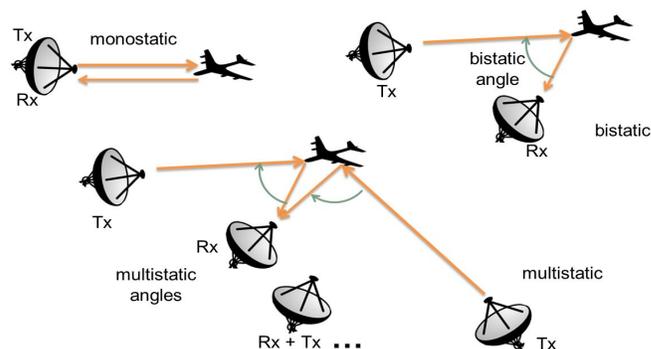

Figure 3. Monostatic, bistatic and multistatic radar systems.

B. Multiband

Radar operating in one band offers advantages and disadvantages specific to that band. A multiband radar may optimally combine advantages from operating in several bands for predefined operational goals, e.g., using the S-band for its strong immunity against weather clutter and good detection range, X-band to generate narrower beams for target tracking and improving spatial resolution, and VHF for extended range and its abilities to detect stealth targets. Multiband radar systems allow enhancing target

classification and detection, and exploiting multispectral imaging of complex targets [42].

C. AESA of New Military Radars

AESAs' Transmit/Receive Modules (TRMs) for each antenna element or group of antennas enable them to accomplish multiple functions simultaneously such as radio communications and multi-pencil beams. TRMs decrease the total energy demand, provide linearity and reduce nearby spectrum pollution from non-linearities of legacy radar main powerhouse (magnetron, klystron and travelling wave tube) [1]. Using a large number of TRMs increases reliability compared to single point of failure of legacy radars.

D. Expected Performance

According to [41], based on LEO satellites operating in the L-band, the theoretical and experimental studies show that the power budget of bistatic radar for air target detection is sufficient for some practical applications such as detecting typical air-targets against a white-noise background at distances of 30 km and more. Such radar has unique advantages from the passive mode of operation, such as not being seen as an active surveillance radar and not having to power the transmitters.

There is the challenge of the direct satellite signal competing with the faint reflected signal from the target. This interference is many orders of magnitude stronger than the reflected signal. Another challenge is the high Doppler induced by the orbiting satellites illuminating the scene and frequent hand over from one satellite to the next in the correct position for an appropriate multistatic angle. This induced Doppler may compete with the clutter Doppler shift.

The favorable results reported [41] in the power budget evaluation are encouraging for further research into air-target detection with multistatic radar based on LEO satellite signals at higher frequencies as for the new constellations under deployment.

VII. CONCLUSION

CR is expected to provide improved situational awareness with more timely and precise information. It is expected to accelerate evolution of our D&S capabilities. This evolution could be managed by progressively introducing these capabilities as in service upgrades or integrated in the development of future systems. Replacing legacy radar systems as soon as possible with AESA TARs would enable updates to CR capabilities later when CR technologies are ready and mature. Upgrading to AESA TARs and then to AESA CRs would certainly provide progressively significant advantages to coalition forces in most demanding combat and surveillance situations.

It is worth noting that using AESA TRMs decreases the total operational energy demand and reduces nearby spectrum pollution from the non-linearities of legacy radar main powerhouse, i.e., magnetron, klystron and travelling wave tube.

Other cognitive sensors of interest to D&S were noted during this research. According to several references on SOund Navigation And Ranging (SONAR), cognitive technology and associated signal processing and pattern recognition have already proven to be advantageous in underwater operations. Similarly cognitive Light Detection And Ranging (LiDAR) provides accurate representations of the environment as applied to autonomous cars.

Another aspect is that the development of CRs may have stimulated work on CEW technology since most traditional electronic warfare techniques would not be able to effectively counter the nimble and unpredictable wave patterns of agile CRs.

Overall, there is overwhelming evidence that the advantages of adopting CR technologies outweigh the risk represented by their complexity.

As illustrated in Figure 2, one can infer that CR's shorter time to deliver higher accuracy track data increases the likelihood of target engagement success rate. This is a trend that several analysts would like to confirm with field trials in the near future.

REFERENCES

[1] P. Labbé, "Cognitive radars (CRs) and their implications," Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2019-L025, January 2019.

[2] S. Haykin, "Radar vision," in *Record of the IEEE 1990 International Radar Conference*, Arlington, VA, USA, 1990, pp. 585-588: IEEE, 1990.

[3] S. Haykin, "Cognitive Radar Networks," in *IEEE Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2005: First International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Puerto Vallarta, Mexico, 2005, pp. 1-3: IEEE, 2005.

[4] S. Haykin, "Cognitive radar - A way of the future," *IEEE Signal Processing Magazine,* vol. 23, no. 1, pp. 30-40, Jan 2006.

[5] J. R. Guerci, R. M. Guerci, M. Ranagaswamy, J. S. Bergin and M. C. Wicks, "CoFAR: Cognitive Fully Adaptive Radar," in *2014 IEEE Radar Conference*, Cincinnati, OH, USA, 2014, pp. 984-989, 2014.

[6] N. Wiener, *Cybernetics: or Control and Communication in the Animal and the Machine*, 2nd ed. MIT press, 1961.

[7] J. G. Camacho *et al.*, "Open architecture as an enabler for FORCEnet Cruise Missile Defense," Monterey, California. Naval Postgraduate School, 2007.

[8] M. T. Cox, "Perpetual self-aware cognitive agents," *AI magazine,* vol. 28, no. 1, pp. 32-46, 2007.

[9] D. J. Kershaw and R. J. Evans, "Optimal waveform selection for tracking systems," *IEEE Transactions on Information Theory,* vol. 40, no. 5, pp. 1536-1550, 1994.

[10] S. Haykin, Y. B. Xue and P. Setoodeh, "Cognitive Radar: Step Toward Bridging the Gap Between Neuroscience and Engineering," *Proceedings of the IEEE,* vol. 100, no. 11, pp. 3102-3130, Nov 2012.

[11] A. D. Maio and A. Farina, "Cognitive Radar Signal Processing," North Atlantic Treaty Organization (NATO), Science and Technology Organization (STO) NATO HQ, Brussels 2015.

[12] S. Haykin and J. M. Fuster, "On cognitive dynamic systems: Cognitive neuroscience and engineering learning from each other," *Proceedings of the IEEE,* vol. 102, no. 4, pp. 608-628, 2014.

[13] W. Chappell and C. Fulton, "Digital array radar panel development," in *2010 IEEE International Symposium on Phased Array Systems and Technology (ARRAY)*, Waltham, MA, USA, 2010, pp. 50-60: IEEE, 2010.

[14] W. Wiesbeck, "SDRS: software-defined radar sensors," in *IEEE 2001 International Geoscience and Remote Sensing Symposium, IGARSS'01.* , Sydney, NSW, Australia, Australia, 2001, vol. 7, pp. 3259-3261: IEEE, 2001.

[15] M. R. Sladic and E. Fehratovic, "Software Defined Radar System: Hardware application development," Bachelor's degree in electrical engineering, Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden, 2014.

[16] U. Ramacher, "Software-Defined Radio Prospects for Multistandard Mobile Phones," *Computer,* vol. 40, no. 10, pp. 62-69, 2007.

[17] J. S. Herd and M. D. Conway, "The evolution to modern phased array architectures," *Proceedings of the IEEE,* vol. 104, no. 3, pp. 519-529, 2016.

[18] J. R. Guerci, "Cognitive radar: The next radar wave?," *Microwave Journal,* vol. 54, no. 1, pp. 22-36, January 2011.

[19] G. J. Mendis, J. Wei and A. Madanayake, "Deep learning cognitive radar for Micro UAS detection and classification," in *Cognitive Communications for Aerospace Applications Workshop (CCAA), 2017*, Cleveland, OH, USA, 2017, pp. 1-5: IEEE, 2017.

[20] A. Martone, K. Sherbondy, K. Ranney and T. Dogaru, "Passive sensing for adaptable radar bandwidth," in *2015 IEEE Radar Conference (RadarCon)*, 2015, pp. 0280-0285, 2015.

[21] R. Klemm *et al.*, *Novel Radar Techniques and Applications. Two Volume Set.* SciTech Publishing, 2017.

[22] J. Jakabosky, S. D. Blunt and B. Himed, "Optimization of "over-coded" radar waveforms," in *Radar Conference, 2014 IEEE*, Cincinnati, OH, USA, 2014, pp. 1460-1465: IEEE, 2014.

[23] NATO STO SET, "Radar Spectrum Engineering and Management," in "STO technical report TR-SET-182," April 2017.

[24] J. D. Poston and R. M. Buehrer, "Spectrum Sharing with Rotating Radar: Implications for Cognitive Radio Rendezvous," in *2014 IEEE Antennas and Propagation Society International Symposium (APURSI)*, Memphis, TN, USA, 2014, pp. 1204-1205, 2014.

[25] A. Ponsford, R. McKerracher, Z. Ding, P. Moo and D. Yee, "Towards a Cognitive Radar: Canada's Third-Generation High Frequency Surface Wave Radar (HFSWR) for Surveillance of the 200 Nautical Mile Exclusive Economic Zone," *Sensors,* vol. 17, no. 7, pp. 1-13, Jul 2017.

[26] T. J. Nohara and S. Haykin, "Cognitive radar information networks for security-Enhancing risk mitigation for reducing operator overload," in *Radar Conference (RADAR), 2013 IEEE*, Ottawa, ON, Canada, 2013, pp. 1-6: IEEE, 2013.

[27] P. Labbé and R. Proulx, "Impact of Systems and Information Quality on Mission Effectiveness," in *Proceedings of the 5th International Command and Control Research and Technology Symposium, 5th ICCRTS*, Canberra, Australia, 2000, p. 37, 2000.

[28] K. Krishnan, T. Schwering and S. Sarraf, "Cognitive Dynamic Systems: A Technical Review of Cognitive Radar," *arXiv preprint arXiv:1605.08150,* 2016.

[29] S. Haykin, A. Zia, I. Arasaratnam and Y. B. Xue, "Cognitive Tracking Radar," in *2010 IEEE Radar Conference*, Washington, DC, USA, 2010, pp. 1467-1470, 2010.

[30] R. F. Yuan, R. B. Gan and G. F. Tang, "Performance of cognitive radar detection under EMI environment," in *Proceedings of the 2015 International Conference on Mechanical Science and Engineering*, Qingdao, China, 2016, vol. 66, pp. 1-8, 2016.

[31] F. Smits, A. Huizing, W. van Rossum and P. Hiemstra, "A Cognitive Radar Network: Architecture and Application to Multiplatform Radar Management," in *2008 European Radar Conference*, Amsterdam, Netherlands, 2008, pp. 312-315, 2008.

[32] A. W. Doerry and B. Marquette, "Shaping the Spectrum of Random-Phase Radar Waveforms," Sandia National Laboratories, Albuquerque, New Mexico, , September 2012.

[33] L. Fan, J. K. Wang and B. Wang, "Time Allocation in Cognitive Radar for Multiple Target Detection," in *Proceedings of 2009 Third International Symposium on Intelligent Information Technology Application*, Shanghai, China, 2009, vol. 2, pp. 169-172: IEEE, 2009.

[34] J. Jakabosky, B. Ravenscroft, S. D. Blunt and A. Martone, "Gapped Spectrum Shaping for Tandem-Hopped Radar/Communications & Cognitive Sensing," in *2016 IEEE Radar Conference (Radarconf)*, Philadelphia, PA, USA, 2016, pp. 22-27, 2016.

[35] P. Labbé *et al.*, "Recommendations for Network-and Internet-based Synchronized E-activities for Location-and Time-dependent Information," in *7th International Command and Control Research and Technology Symposium, 7th ICCRTS*, Loews Le Concorde, Québec City, Canada, 2002, pp. 1-28: CCRP, 2002.

[36] P. Labbé, "Potential advantages of cognitive sensor-to-effector loops (CStELs)," in *23rd International Command and Control Research and Technology Symposium (23rd ICCRTS)*, Florida Institute for Human & Machine Cognition (IHMC), Pensacola, Florida, USA, 2018, vol. Topic 4: Cognitive and Socio-technical Challenges, Paper # 84, pp. 1-20: ICCRTS, International Command and Control Institute, 2018.

[37] P. Labbé, "Cognitive Networks and Systems to Improve Coalition Operations Effectiveness," in *2019 IEEE 6th Asian Conference on Defence Technology (ACDT)*, Bali, Indonesia, 2019, pp. 182-187: IEEE, 2019.

[38] P. Moo and D. DiFilippo, "Multifunction RF Systems for Naval Platforms," *Sensors,* vol. 18, no. 7, pp. 1-37, 2018.

[39] A. G. Westra, "Radar versus stealth: Passive radar and the future of US military power," *JFQ,* vol. 4th quarter, no. 55, pp. 136-143, 2009.

[40] M. Ben Kilani, "Multistatic Radar Optimization for Radar Sensor Network Applications," Ph.D., École de Technologie Supérieure Université du Québec, Montréal, 2018.

[41] M. Cherniakov, D. Nezlin and K. Kubik, "Air target detection via bistatic radar based on LEOS communication signals," *IEE Proceedings-Radar, Sonar and Navigation,* vol. 149, no. 1, pp. 33-38, 2002.

[42] F. Scotti, F. Laghezza, P. Ghelfi and A. Bogoni, "Multi-Band Software-Defined Coherent Radar Based on a Single Photonic Transceiver," *IEEE Transactions on Microwave Theory and Techniques,* vol. 63, no. 2, pp. 546-552, 2015.

LEO Satellite Constellations: An Opportunity to Improve Terrestrial Communications in the Canadian Arctic

Paul Labbé

National Defence
Defence Research and Development Canada (DRDC)
Ottawa, Canada
email: Paul.Labbe@drdc-rddc.gc.ca / Paul.Labbe@forces.gc.ca

Abstract—**This study examines communications options (data, video and voice) for operations in the Canadian Arctic accounting for adverse conditions, such as atmospheric disturbances (both natural and man-made) or adversarial attacks on satellites and terrestrial infrastructure. Potential users include Canadian Armed Forces (CAF), North American Aerospace Defense Command (NORAD), off-grid communities and Public Safety, including their respective systems requiring machine-to-machine low latency data sharing. The technologies considered include satellites, microwave relays, fiber optic links, radios like cellular phones operating in the 700 Mhz band, transceivers in the high frequency bands (20-30 Mhz), and Unmanned Aerial System (UAS) gateways.**

Keywords-communications; satellite; Arctic; latency.

I. Introduction

This Defence Research and Development Canada (DRDC) study was initiated to address Canadian Arctic communication challenges expressed in the new Canada's Defense Policy, *Strong, Secure, Engaged* (SSE) [1], which reaffirmed Canada's commitment to effective operation in the Arctic. SSE defines an extended Canadian Air Defense Identification Zone (CADIZ) that includes the entire Canada's Arctic Archipelago (see Figure 1) as part of the overall North American Aerospace Defense Command (NORAD) modernization with an improved North Warning System (NWS) requiring high-throughput low-latency communications.

In addition this study covers some aspects of current Northern Canada communication systems of the Canadian Armed Forces (CAF), off-grid communities, Search And Rescue (SAR), and Public Safety (PS). Also, the rise in commercial interest, research and tourism in this zone brings increased safety and security demands related to SAR and natural or man-made disasters to which Canada must be ready to respond.

The purpose of this paper is to examine communications options for operations in Northern Canada accounting for adverse conditions such as atmospheric disturbances (both natural and man-made) or adversarial attacks on satellites and terrestrial infrastructures. Users include CAF, NORAD, off-grid communities and PS.

Section II of this paper describes the geographic context, climate, topography and size of the area requiring assured communication in all weather and atmospheric conditions.

Section III summarizes some of the communications systems envisaged, deployed and used over the last decades in this area, such as Tropospheric Scatter, Geostationary Earth Orbit (GEO) at approximately 37000 km above the equator and Medium Earth Orbit (MEO) between 2000 to 35786 km altitudes. Section IV brings in perspective of novel options offered by several Low Earth Orbit (LEO) satellite constellations at altitudes below 2000 km. Section V introduces another option using a specific Highly Elliptical Orbit (HEO), the Three APogee (TAP with an apogee around 43000 km) [2]. Section VI proposes a possible terrestrial architecture with the addition of base stations using towers and Unmanned Aerial Systems (UASs) as gateways. Section VII compares latencies among the different satellite orbits and other communication technologies. Section VIII examines the pros and cons of some of these options. Section IX concludes with a short summary of findings and a final recommendation.
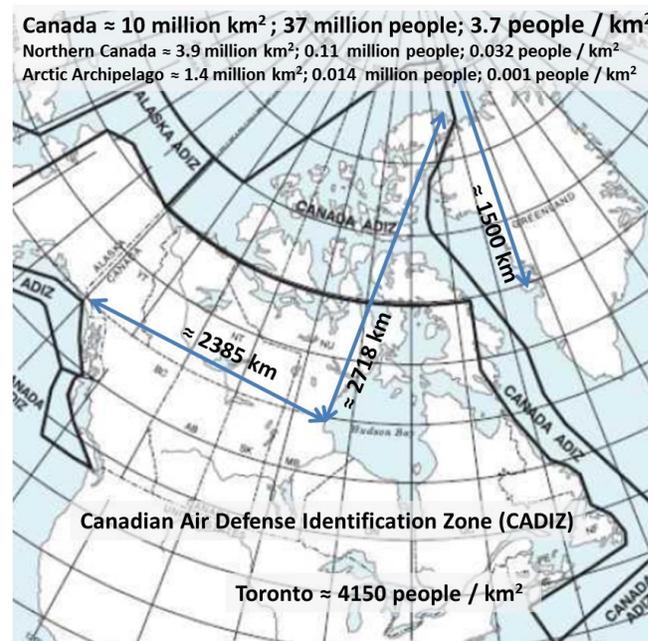

Figure 1. North America Canadian operational areas, new Canadian Air Defense Identification Zone (CADIZ), distance vectors and population densities for Canada, Northern Canada, Canadian Arctic Archipelago and Toronto.

II. CONTEXT

Northern Canada, aka the North, is the vast northernmost region of Canada variously defined by geography and politics. Politically, the term refers to three territories of Canada: Yukon, Northwest Territories, and Nunavut. This area includes the Arctic Archipelago and covers about 39% of Canada's total land area, but has less than 1% of Canada's population.

More specifically, the Arctic Archipelago, aka the Canadian Arctic Archipelago, groups together all islands lying to the north of the Canadian continental mainland excluding Greenland (an autonomous territory of Denmark). Situated in the northern extremity of North America and covering about 1424500 km^2 (550000 sq mi), this group of 36563 islands in the Arctic Sea comprises much of the territory of Northern Canada—most of Nunavut and part of the Northwest Territories. This is about 15% of Canada's geography that homes only about 0.04% of Canada's population. The Canadian Arctic Archipelago is experiencing some effects of global warming, with some computer models estimating that ice melting will contribute 3.5 cm to the rise in sea levels by 2100.

CAF's preparedness training and exercises involve several thousand participants and some observers, which temporally substantially increase the population density in Northern Canada. For example, Operation Nanook series is an annual military CAF exercise in the Arctic. It is intended to train different CAF elements (Canadian Army, Royal Canadian Air Force and the Royal Canadian Navy) along with other government organizations, such as the Canadian Coast Guard and Royal Canadian Mounted Police in disaster training and sovereignty patrols in Northern Canada. Another series conducted south of Northern Canada near CAF Base Cold Lake is MAPLE FLAG, which brings about 5000 participants. In both cases, these exercises represent significant demand of information interoperability exchange, including voice, data and video, with some requiring low latency in order to fulfill machine-to-machine requirements.

A poignant factor in deploying communication systems in Northern Canada is the low population density. With reference to Figure 1, population densities of Europe and India are respectively about 2.3 and 11.6 thousand times that of Northern Canada. It is worth noting that some aspects of evolving 5G technologies are tuned to improve network capacity in high-user density areas like cities.

A. Environmental and Logistical Considerations [3]

The North has extreme weather conditions, with temperatures ranging from -50 to +20 degrees Celsius and wind gusts of up to 150 km/h. There is very little precipitation in the Arctic, with an average total of 100 to 200 mm of rain and snow per year. The amount of daylight varies with time of year and with latitude. In Resolute Bay, for example, the sun does not rise above the horizon from early November to early February and does not set from early May to late July.

Permafrost (perpetually frozen ground) is present in most of the Arctic and although typically roughly 10 m thick, it can extend down 1 km below the surface. Construction of stable platforms required for large satellite ground stations or microwave towers is costly because support pillars must be driven below the permafrost; otherwise the platform will shift as the permafrost partially melts during the short Arctic's summers. This requires careful site selection to ensure placement where the permafrost is thin enough to enable installation. Furthermore, this type of environment is challenging for underground fibre installation.

Year-round access to all communities in the high Arctic is by plane only, increasing the overall cost of travel and accommodation when planning and installing communications equipment. Communities also have seasonal sealifts, which transport non-perishable dry cargo (e.g., construction material, household goods, vehicles, etc.) and bulk fuel to them once or twice per year. Items that are too large or costly to transport via air cargo, for example, large aperture satellite dishes or microwave towers and associated construction materials must be shipped by sea. This increases the logistic phase of any communication equipment deployment.

III. PRACTICAL ARCTIC COMMUNICATIONS SYSTEMS

This section includes the following communications systems: Tropospheric Scatter, High Frequency Ground Wave (HFGW), Point-to-Point (P2P) backhaul radio links, Fiber Optic Cable (FOC) and GEO systems. Skywave propagation modes are excluded due to their susceptibilities to radiation and atmospheric changes.

Tropospheric scatter, or Troposcatter, is a beyond-the-horizon communications solution in which microwave signals are scattered as they propagate through the troposphere. This phenomenon allows signals to be successfully received without line-of-sight between the transmitter and receiver at vast distances, even up to 500 km. Troposcatter systems have been in use for several decades and the technology has seen many advances in recent years, including improvements in throughput to 20 Mbps. However, for applications in the North it was found to be too expensive due to its power demand and infrastructure.

HFGW was documented for an alternate communications network for "nuclear-survivable means of communication for land-mobile missile systems in Europe" [4]. So, HFGW appears to be a potential communication means in case of satellite communications disruption due to solar activities or other manmade disturbances. According to International Telecommunication Union (ITU) [5] there are different uses of the terminology and the surface wave is often called the ground wave, or sometimes the Norton ground wave or Norton surface wave, after Norton who developed tractable methods for its calculation. The ground wave is the sum of direct wave, reflected wave, and surface wave. When the transmitting and receiving antennas are close to the ground, the ground reflection coefficient is -1, so the direct and ground-reflected waves act to cancel each other out, leaving the surface wave as the only important component. In such conditions, the ground wave is essentially equal to the surface wave. Empirical results [4] using broadband discone antennas with a cut-off frequency of 19 MHz operating over 20 to 30 MHz near Norway Arctic Circle and Germany

showed good link connectivity for voice and data using narrowband channels for paths over irregular terrain over distances ranging between 19 and 115 km. Based on the empirical results reported, a communication system with its signal spread over 10 MHz with code division multiplexing and sufficient coding gain would offer throughput and fading resilience for a high-reliability medium-data-rate channel.

Microwave links have been extensively used in the North providing reliable connectivity to smaller communities in the Yukon and Northwest Territories that do not have direct access to FOC backhaul. Microwave links provide high throughput with low latency, a significant advantage over GEO systems. Since microwave frequencies require line-of-sight propagation, towers and topographic features are exploited to extend the range of links beyond the limitations imposed by the curvature of Earth and terrain features.

An example of microwave technologies used in remote Arctic locations is the High Arctic Data Communication System (HADCS) of Ellesmere Island, which links Canadian Forces Station Alert (CFS Alert) to Eureka over a distance of roughly 500 km. Then the data are sent via GEO satellite from Eureka to Ottawa (4147 km). HADCS was retrofitted in 2003 to run entirely using solar power, despite prolonged darkness during winter months. Each HADCS station is powered by eight 120 W photo-voltaic (solar) panels arranged in an octagon (eight vertical panels distributed at 45° angular intervals to cover all azimuths). During the summer, with 24 hours of day light, the sun does not dip below the horizon, and all of the solar panels contribute to charging the battery banks [3]. This network comprises seven individual links ranging from 18 to 121 km in length each. The current system operates in bands near 1800 and 2100 MHz, and provides 6.3 Mbps of throughput. Helicopter is the only means of access to these repeater sites [3].

The main challenges in providing microwave backhaul throughout in the North using towers are the overall lack of infrastructure, the inaccessibility of the locations where towers would be built and powering these sites. In the North, costs of construction per kilometre for FOC and microwave links are about equal (approximately $65K/km), when serving a population of roughly 300 people. Above this number of users, fibre is less expensive [3]. These networks have at least one point of service via a satellite terrestrial terminal; the trend is two or more to increase dependability.

UAS gateways (either aerostats and high-altitude high-endurance autonomous drones) [3] provide possible communication solutions that merit some discussion. The Internet.org consortium has conducted some research into the feasibility of using UAS as communications platforms for remote and underserved locations [15]. These UAS would be deployed at an altitude of approximately 20 km. By using solar power, UAS systems would be capable of remaining above one geographic location on earth, thereby reducing complexity and cost of the ground infrastructure when compared to microwave links, or will not require tracking as for MEO and LEO satellites. As the UAS would be relatively close to Earth, cheaper low-power transmitters could be used, while still enabling high-throughput communications with low latency. A previous study [3]

assessed that this type of range extenders might have an availability no greater than 80%, but with technology advances this is different today, such UAS might be sufficiently reliable for commercial broadband.

GEO satellites are well suited to some applications in the North, given that they remain at the same point in the sky when viewed from a particular location on Earth, and therefore, do not require tracking antennas on the ground as for MEO and LEO systems. This enables the use of lower-cost stationary antennas at ground stations, which is particularly advantageous in harsh environments where moving parts are undesirable. GEO satellites provide much of the broadband Internet coverage for communities in the North; the coverage in Nunavut is currently provided by two GEO satellites, Anik F2 and Anik F3, both using C-band [3].

However, the use of GEO satellites in the North has some drawbacks. Due to geometry and the curvature of Earth, there is no line-of-sight between equatorial GEO satellites and North locations above roughly 80 degrees latitude. Furthermore, communities at latitudes higher than about 70 degrees have elevation-look angles below ten degrees to the satellite. This leads to increased absorption and scattering of radio frequency signals by the atmosphere (known as absorption) and at higher frequencies from precipitation (rain, snow and ice, known as rain fade). The five northernmost communities in Nunavut are above 70 degrees latitude. Another disadvantage of GEO satellite use for broadband applications is signal latency; due to the distance to and from the satellite, the minimum delay for a round trip is 480 ms with a median latency of 600 ms. This is an order of magnitude higher than for fibre and microwave links [3].

Tests of Mobile Satellite (MSAT) and Iridium capabilities for emergency communications in Northern Quebec showed their insufficiencies. The Canadian MSAT system uses GEO satellites, which requires, at such latitude, a high-gain antenna and higher antenna siting (implying larger installations, location constraints and thus greater costs). These conditions make emergency operations difficult and cost-ineffective. The US commercial Iridium system with its current handheld, vehicle-mounted and fixed-remote equipment demonstrated different logistic problems, such as the high cost and poor performance of the handheld telephone, which usually cannot be used inside buildings [6]. For a Canadian Arctic Underwater Sentinel Experimentation (CAUSE) project, transmission of data over a period of 7 months showed that the Iridium Pilot system (antenna about six meter above ground) did not fulfill the requirement [7]. "The Pilot data transfer rate for polar transceivers is far less than the advertised rate" [7].

The Iridium system was designed to be accessed by small handheld phones, the size of a cell phone. The omnidirectional antennas, which are small enough to be mounted on such a phone, and the low battery power, are insufficient to allow the set's radio waves to reach a GEO satellite. In order for a handheld phone to communicate with them, the Iridium satellites are at LEOs closer to Earth, at about 780 km above the surface. With an orbital period of about 100 minutes a satellite can only be in view of a handset for about 7 minutes, while the call being automatically

"handed off" to another satellite when the previous one passes beyond the local horizon. This requires a large number of satellites, carefully spaced out in polar orbits, to ensure that at least one satellite is continually in view from every point on Earth's surface. For seamless coverage at least 66 satellites are required, in 6 polar orbits containing 11 satellites each.

IV. PERSPECTIVE OF NEW LEO CONSTELLATIONS

Out of the 11 LEO satellite communications service proposals registered within the US Federal Communications Commission (FCC), the following three will be considered based on their maturity [8]: OneWeb, SpaceX and Telesat on Ku (12-18 GHz) and Ka (27-40 GHz) bands.

To ensure access to affordable high-speed Internet connectivity across rural and Northern Canada, the Government of Canada has invested $85 million and is committed to buy up to $600 million in some services over 10 years following launch in 2022 of Telesat's LEO Satellite Constellation, which is leveraging Telesat's worldwide rights to ≈ 4 GHz of Ka-band spectrum.

The conclusion of the main analysis of [8] was summarized as follows:

• The maximum total system throughput (sellable capacity) for OneWeb's, Telesat's and SpaceX's constellations are 1.56 Tbps, 2.66 Tbps and 23.7 Tbps respectively.

• A ground segment comprising of 42 ground stations will suffice to handle all of Telesat's capacity, whereas OneWeb will need at least 71 ground stations, and SpaceX more than 123.

• In terms of satellite efficiency (ratio between the achieved average data-rate per satellite and its maximum data-rate), Telesat's system performs significantly better (~59% vs. SpaceX's 25% and OneWeb's 22%). This is due to the use of dual active antennas on each satellite, and the lower minimum elevation angle required in their user links.

• OneWeb's system has a lower throughput than Telesat's, even though the number of satellites in the former is significantly larger. The main reasons for this are the lower data-rate per satellite that results from OneWeb's low-complexity satellite design, spectrum utilization strategy, orbital configuration, and payload design, as well as the lack of use of Inter Satellite Links (ISLs).

• If ISLs were to be used in OneWeb's constellation, (even with modest data-rates of 5 Gbps), the number of ground stations required could be reduced by more than half to 27 ground stations.

"To conclude, our analysis revealed different technical strategies among the three proposals. OneWeb's strategy focuses on being first-to-market, minimizing risk and employing a low-complexity space segment, thus delivering lower throughputs. In contrast, Telesat's strategy revolves around high-capable satellites and system flexibility (in diverse areas, such as deployment, targeted capacity allocation, data routing, etc.), which results in increased design complexity. Finally, SpaceX's system is distinctive in its size; although individually each satellite is not significantly more complex than Telesat's, the massive number of satellites and ground stations increases the risks and complexities of the overall system considerably" [8]. However, this offers a high level of redundancy.

V. ENVISAGED HEO/TAP CONSTELLATIONS

HEO/TAP [2], which could be considered under the Enhanced Satellite Communications Project – Polar (ESCP-P) program to provide dedicated, secure and reliable Beyond Line-Of-Sight (BLOS) communications for domestic and continental CAF operations in the Arctic. Current GEO communication satellites leave the poles uncovered; consequently, Department of National Defence (DND) is exploring options of building the capability or acquiring services from commercial providers who have future plans to cover this area. The initial operational capability is tentatively scheduled for 2029.

Quasi-geostationary coverage of Polar Regions can be achieved from HEO satellites. In accordance with Kepler's second law HEO spends most of the time in the vicinity of apogee (i.e., the farthest point from the Earth's surface). The orbit could be oriented in such a way that the apogee is over one of the two Polar Regions, so that only two HEO satellites can maintain a continuous view of an entire polar zone [9]. When satellite A leaves the optimal viewing zone and heads toward the perigee (i.e., the closest point to the Earth's surface), satellite B rises over the region to maintain the complete circumpolar region in sight. Interestingly, there are periods of several hours per day of coincident (i.e., stereo-like) imaging from the two satellites over most of a circumpolar area. Such a system could provide meteorological imaging and communication capacity, similar to GEO, focused on the polar region. The first HEO satellite system with a period of rotation equal to 12 h was called Molniya was implemented for communication purposes in 1965. It is established that a two-satellite Molniya HEO constellation can achieve continuous coverage of the polar region 58°–90°N with a Viewing Zenith Angle (VZA) less than 70°. Another HEO system—with a 24-h period—called Tundra is currently used by the satellite Sirius XM Radio service operating in North America. Both orbits, 12-h Molniya and 24-h Tundra, are launched with an orbit inclination equal to 63.4°. This value called the critical inclination, corresponds to a zero rate of apogee drift due to the second zonal harmonic of the Earth gravitational field, and insures a stable position of apogee over the polar zone. If the HEO orbit inclination differs from the critical value, then the apogee gradually drifts toward the equator. Orbital maneuvers are then required to maintain the intended orbit position. The farther the orbit inclination is from the critical value, the more resources are required to maintain the orbit. A drawback associated with the 12-h Molniya orbit is the risk linked to hazardous levels of ionizing radiation due to passing through the Van Allen belts. The highest danger originates from high-energy protons. The Molniya orbit crosses the proton radiation belts at the region of maximum concentration of energetic protons with energies up to several hundred MeV. As an alternative, the 16-h TAP HEO orbit was proposed, providing similar polar coverage as the Molniya HEO system while minimizing the proton ionizing

hazards by extending the apogee to 43000 km [2]. The TAP orbit has a ground track with three apogee points repeatable over two days. Such a constellation of two satellites in TAP orbit still revisits ground tracks every 24 h.

VI. TERRESTRIAL ARCHITECTURES

Architecture options offered by new LEO satellite constellations and terrestrial communications, such as UAS, FOC and HFGW given advances in signal processing, multi-beam antennas, spatial diversity and low cost software-defined radios could substantially improve communications availability and reliability in the North. Figure 2 illustrates such a multi-technology architecture where ISLs and Inter UAS Links (IUASLs) play important roles. Potential services provided with appropriate parameters and technologies should allow anticipating improvement in coverage, resilience, redundancy, dependability, data rate and low latency for services, such as:

- fixed installations like CFS Alert, NWS radar networks, and Forward Operating Bases (FOBs);
- mobiles near fix installations, air or tower gateways or their communication relays; and
- deployed personnel and platforms for military exercises and operations, emergency operations or off-grid communities of Northern Canada.
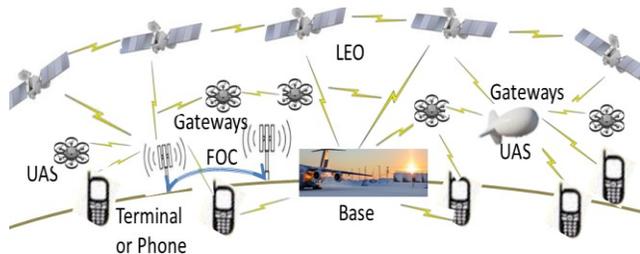

Figure 2.   Simplified Northern Communications Architecture.

VII. LATENCY

Latency is the delay, usually measured in milliseconds (ms), that occurs in a round-trip data transmission. The latency of a message is the amount of time it takes to reach its destination from its source. Results for FOC in large networks based on Round-Trip Times (RTTs) [10] using Content Delivery Network (CDN) show useful latency information, such as 18 ms for a distance of 1418 km. HADCS and HFGW offer low latency in this order of magnitude but with lower data rates. High latency typical of GEO satellite links can be very disruptive for some applications, such as video conferencing, and could increase risk in remote health applications including remote surgery. Satellite link latencies can also cause low data throughput, caused by the default behaviour of communication protocols, which are optimized for shorter distances. A GEO satellite one-way propagation delay is approximately 240 ms due to the distance between Earth and satellite. Therefore, round-trip delivery of a data packet with acknowledgement is approximately 480 ms. This does not include the network delay, which can generally add 50 to 200 ms, depending on

where the server is located. GEO satellite systems have a median latency of nearly 600 ms, which includes a median delay of 120 ms due to processing speed and networks in both directions. This makes GEO systems an unsuitable replacement for cable or fiber systems for applications requiring low latency, such as machine-to-machine interoperation. The lower orbits of LEO satellites, however, should result in latencies that are much closer to landline quality. The average orbit of the proposed constellations is around 1200 km, that is a round trip of 2400 km or delay of 31 ms. This is 93.5% improvement over a GEO round trip of 480 ms. If the processing speed of LEOs equaled that of GEOs then their total median latency time would be 151 ms. However, OneWeb tune-up, recorded an average latency of 32 ms in July 2019. As new LEO satellites are designed for high throughput, their overall processing time and network delay must be lower than those used in legacy GEO systems in order to obtain such small latency score.

For LEO latency, in this article, since no large sets of empirical results are available, a conservative approach is to use some of the simulation results from [11] for optimal Expected Latency Minimization (ELM) algorithm using Software-Defined Networking (SDN) context, which addresses more completely the entire network aspect including fading dependence on atmospherics. An interpretation of [11] for its ELM-SND hypotheses is that LEO's average latency be around 40 ms while the maximum average latency be around 90 ms.

For HEO TAP with an apogee of 43000 km this means that the round-trip time is increased by16% from GEO, i.e., 558 ms or a median latency of 778 ms. Assuming a MEO median orbit of 18893 km, that is 51% of GEO, then MEO median latency could be in the order of 306 ms.

VIII. ANALYSIS

Considering the advantages of LEOs in extending communications coverage and throughput in the North in conjunction with terrestrial network communications with appropriate gateways, this study must be extended to include experimentation. Also, there is a need to investigate how Canada could protect space and terrestrial network installation assets.

Satellites are more susceptible to radiation, jamming and atmospheric disturbances than FOC and over-the-horizon HFGW propagation. HFGW at 20 to 30 MHz is expected to provide reliable medium throughput terrestrial communications [4]. FOC offers high throughput and low latency, commonly deployed around the world and expanding in Northern Canada [12][13]. FOC and HFGW do not offer the area coverage of LEO satellites.

Two challenges of LEO constellations are the frequent handover and terrestrial base station antenna tracking. These seem achievable under low energy and cost regimes by the availability of lower-cost components, such as high-speed low-power chip sets and Active Electronically Scanned Array (AESA) integrated boards for microwave systems in Ka and Ku bands. However, Figure 3 shows the disadvantage of transmitting at higher frequencies, but such higher frequencies offer higher throughput.
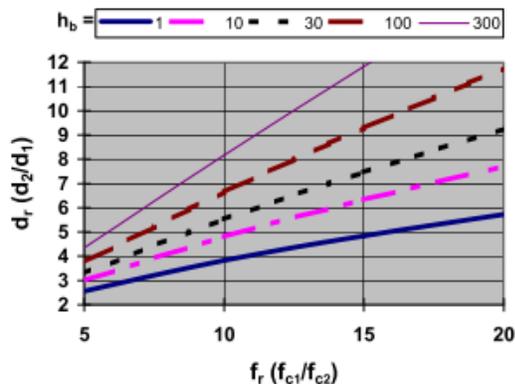
Figure 3.   Path loss distance ratio d_2/d_1 as a function of the carrier frequency ratio f_{c1}/f_{c2} for five base heights and mobile at one meter [6].

Figure 3 allows estimating the increase in radio coverage when stepping down from 1500 MHz to 150 MHz. Reducing the mobile operating frequency by a factor of 10 extends the communication range by a factor of about 5 for a base station whose effective antenna height is 30 m. If the cell size were 5 km for normal service, it might extend to 25 km for the emergency temporary service, reducing the logistic burden of covering an area affected by a disaster. Currently, the frequency of 700 MHz is allocated for emergency in Canada.

Colman *et al.* [3], present an example with microwave link systems. One is operating at 1.8 GHz and the other at 11 GHz, both with similar radiating power. The system at lower frequency offers a free space maximum range of 333 km while the other, at six times the frequency, qualifies for a free space maximum range of 30 km. That is 11 times shorter. However, the maximum effective throughput rises from 65.4 Mbps to 232 Mbps, which is 3.5 times faster.

Other considerations include the challenge of powering terrestrial systems in the North. Sources like solar and wind mill could be combined with sodium-ion batteries, which could operate at low temperature. Long endurance UAS could be using solar with hydrogen fuel-cells. However, sources requiring refueling every six months would be ideal.

IX.   CONCLUSION

This article addressed some of difficult challenges that remain after many years of research and development for communications systems in the North. Findings address DND/CAF challenges to improving capabilities required by future demands from expected developments in the Arctic and for NORAD operations.

Selected options to improve communications in off-grid areas, more specifically in the North, are expected to provide timely improved shared situational awareness in support of operations where it is currently not well provided, or not available. Such solutions would be revolutionary for our Defense and Security (D&S) capabilities and would progressively provide significant advantages to coalition forces and when CAF operates in collaboration with other Canadian departments including PS and local police in the most demanding emergency and disaster situations.

In the North, if low latency communications are critical for applications or operation objectives, GEO, MEO and HEO satellite systems are not recommended. The least expensive communications systems with low latency in the North should include UAS and LEO systems to ensure fast deployment and access to all participants. UAS offers rapid deployment capability on demand in response to PS and CAF situations. LEO/UAS hybrid systems could definitely extend capabilities of available legacy infrastructures.

The most significant finding is that the advent of low-cost high-performance LEO satellite systems will definitely improve communications in the North.

REFERENCES

[1] Canada: Department of National Defence and Canadian Armed Forces. (2017). *Strong, Secure, Engaged. Canada's Defense Policy*.

[2] A. P. Trishchenko, L. Garand and' L. D. Trichtchenko, "Three-apogee 16-h highly elliptical orbit as optimal choice for continuous meteorological imaging of polar regions," *Journal of Atmospheric and Oceanic Technology,* vol. 28, no. 11, pp. 1407-1422, 2011.

[3] G. Colman *'et al.'*, "Communications in the North," Communications Research Centre (CRC), Ottawa, 2014.

[4] J. R. Champion, "An Empirical Investigation of High-Frequency Ground Wave Propagation," *Johns Hopkins APL Technical Digest,* vol. 13, no. 4, pp. 515-525, 1992.

[5] ITU, "Handbook on Ground Wave Propagation," in "International Telecommunication Union; Radiocommunication Bureau," R-HDB-59, 2014.

[6] P. Labbé, "GPS and GIS Integration in Mobile Equipment for Improved Mobile Emergency Operations," in *Proceedings of the 12th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS-99); Session A2 - Land Systems & Public Safety*, Nashville Convention Center, Nashville, Tennessee, 1999, pp. 545-555, 1999.

[7] G. J. Heard, "Initial report of CAUSE investigated technologies for remote system deployment," Defence Research and Development Canada (DRDC) – Atlantic Research Centre, August 2019.

[8] I. del Portillo, B. G. Cameron and' E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta Astronautica,* vol. 159, pp. 123-135, 2019.

[9] A. P. Trishchenko, L. Garand, L. D. Trichtchenko and' L. V. Nikitina, "Multiple-apogee highly elliptical orbits for continuous meteorological imaging of polar regions: challenging the classical 12-h Molniya orbit concept," *Bulletin of the American Meteorological Society,* vol. 97, no. 1, pp. 19-24, 2016.

[10] I. N. Bozkurt *'et al.'*, "Dissecting Latency in the Internet's Fiber Infrastructure," *arXiv preprint arXiv:1811.10737,* 2018.

[11] W. Cho and' J. P. Choi, "Cross layer optimization of wireless control links in the software-defined LEO satellite network," *IEEE Access,* vol. 7, pp. 113534-113547, 2019.

[12] A. Fiser, "Mapping the long-term options for Canada's north: Telecommunications and broadband connectivity," 2013: Conference Board of Canada, 2013.

[13] M. Levitt, "Nation-building at home, vigilance beyond: Preparing for the coming decades in the arctic," Speaker of the House of Commons, Ottawa, 2019.

Column Norm Sorting Based Successive Interference Cancelation Algorithm For Multi-Beam Satellite Communication System

Jingchao WANG
Peng Cheng Laboratory (PCL)
Shenzhen, China
wangjc61s@163.com

Quan Yu
Peng Cheng Laboratory (PCL)
Shenzhen, China
yuquan61@qq.com

Abstract—**Multi-beam satellites have the potential to offer high capacity with frequency reuse among different beams. Recently, more radical frequency reuse strategy or even full frequency reuse has been adopted to further improve the system capacity, leading to high co-frequency interference between adjacent beams. Successive Interference Cancellation (SIC) is normally used to deal with the interference. In this paper, we propose an optimized SIC algorithm by user grouping and sorting to cancel the interference in a better way. Simulation results show that the proposed algorithm can improve the bit error rate performance compared with the classic SIC algorithm.**

Keywords—satellite communication; multi-beam satellite; successive interference cancellation algorithm.

I. INTRODUCTION

A multi-beam satellite is defined as a satellite using multi-point beam and frequency reuse techniques. With the same spectrum resources, the capacity of the satellite is several times that of the traditional satellite [1]. When the frequency reuse scheme is more radical, e.g., full frequency reuse, serious interference will be generated between the beams, and interference cancellation is needed to reduce the impact of interference on the system performance. In the literature, Multiuser Detection (MUD) techniques based on Successive interference cancellation (SIC) are usually adopted in the reverse link to combat the interference.

In [2], MUD is proved to be an effective solution to maximize the capacity of multi-beam satellite system. A good review and evaluation of various MUD techniques for broadband multi-beam satellite systems is provided in [3]. It has been shown that the capacity of current multi-beam satellite systems is greatly affected by inter-beam interference, which calls for efficient advanced signal processing techniques to overcome the interference.

In [4], the physical layer performance achievable by a specific algorithm using MUD is evaluated. The results indicate that detecting by two users may be sufficient to greatly improve the spectral efficiency of the link. The authors in [5] analyzed two basic techniques for MUD: theoretically optimal Maximum Likelihood (ML) detection and sub-optimal Minimum Mean-Squared Error (MMSE) detection. The performance of theoretically optimal ML detection is very close to the theoretical single-user limit. As the number of users increases, the performance decreases. For a small number of users, MMSE detection performs very close to ML. When the number of users increases, the performance will drop significantly.

A hybrid QR decomposition with Malgorithm Maximum Likelihood Detection (QRM-MLD) MMSE-SIC adaptive MUD algorithm is proposed in [6]. In contrast to the optimal ML method, QRM-MLD significantly reduces the complexity. A tradeoff between complexity and performance is achieved and the performance is compared with the hybrid ML-MMSE-SIC scheme.

In this paper, the traditional SIC algorithm is optimized in two aspects, which can improve the Bit Error Rate (BER) performance. First, the received signals are sorted by strength before SIC, then the detection is performed in descending order of signal strength. Second, the satellite beams are grouped. In the process of decoding, parallel and serial interference detection are mixed to obtain better performance.

The rest of this paper is organized as follows: Section II provides the system model of the multi-beam satellite system. The classical SIC algorithm is introduced and analyzed in Section III. Two improved SIC algorithms are proposed in Section IV. Simulation results are given in Section V. The conclusions follow in Section VI.

II. SYSTEM MODEL

In this paper, we consider a multi-beam satellite system and focus on the reverse link of the system. Assume the system is based on TDMA, where K users on the ground send data in the same frequency band in the same time slot. In this situation, K ground terminals can be regarded as K transmitting end antennas and the satellite receives signals from multiple users in the coverage area of multiple beams.

Denote by s_j the signal transmitted by user j and h_{ij} the channel coefficient between user j and beam i , the reveiced signal of the ith beam is:

$$y_i = \sum_{j=1}^{K} h_{ij} s_j + n_i \tag{1}$$

where n_i is the additive Gaussian white noise with variance σ^2 . Considering K users jointly, the above equation can be expressed in matrix form:

$$Y = Hs + N \tag{2}$$

Where $s = [s_1, \ldots s_K]$ is the signal vector sent by K users. $N = [n_1, \cdots, n_K]$ is the noise vector. $H = [h_1^T, \ldots h_k^T]$ is a $K \times K$ channel coefficient matrix, where the propagation loss, beam antenna gain, rain fading, small-scale fading are considered.

The reverse link channel coefficient matrix H of the multi-beam satellite communication system can be expressed as:

$$H = BH_R \Phi_d \tag{3}$$

where B is the channel gain matrix, H_R is the small-scale fading matrix, and Φ_d is the large-scale fading matrix.

We use diagonal matrix Φ_d to represent the large-scale fading factor of K users, and formula (3)-(5) can be expressed as:

$$\Phi_d = \{\varepsilon_1, ..., \varepsilon_K\} \qquad (4)$$

The small-scale fading of the channel can be modeled as Rice distribution. Therefore, channel coefficient matrix H_R is expressed as follows:

$$H_R = \sqrt{\frac{K}{K+1}}H_R^1 + \sqrt{\frac{K}{K+1}}H_R^2 \qquad (5)$$

Where K is the Rice factor, H_R^1 represents the light of sight components, and H_R^2 is the matrix containing random interference of scattering components.

III. CLASSIC SUCCESSIVE INTERFERENCE CANCELLATION ALGORITHM

The basic principle of SIC is to reconstruct the successfully decoded signal and eliminate the interference of the signal from the received signal. The SIC detector detects the signals from multiple transmitters through various algorithms. Once a signal is detected, the signal is multiplied by the corresponding column in the channel coefficient matrix, and then the value is subtracted from the original received signal. In this way, the Multiple Access Interference (MAI) caused by the signal is eliminated. Recursing this process, the MAI can be completely eliminated. The performance of the SIC is greatly improved compared with the traditional detector, and the hardware is not changed much, so it is easy to implement. The SIC algorithm is shown in Figure 1.
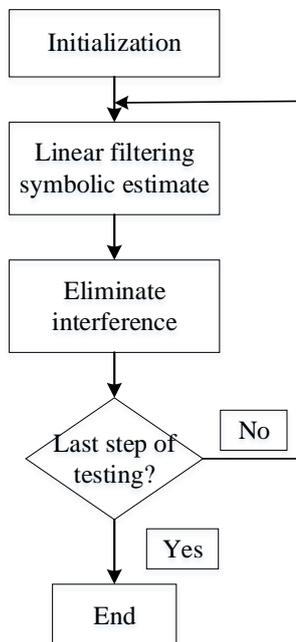

Figure 1. SIC algorithm flow chart

At the satellite receiving side, since the received signal is the sum of signals transmitted from users in multiple co-frequency beams, if the signal cannot be distinguished, the useful information cannot be identified. The SIC detector can first make decision on the strongest signal and obtain its estimated value. If there are N transmitted signals, the multiple access interference cancellation can be performed on other N-

1 users in the following stage of detection, that is, the interference caused by the decoded signal is eliminated. It can be seen that the larger the number of detection stages, the better the multiple access interference will be eliminated.

According to the detection algorithm, successive interference cancellation can be divided into ZF-SIC and MMSE-SIC. Both the ZF algorithm and the MMSE algorithm are linear detection algorithms. Their basic principle is to multiply the received signal vector by the filter matrix G to obtain an estimate from which the final solution vector can be determined. There are two different benchmarks when calculating the filter matrix, so the linear detection algorithm can be divided into (ZF) algorithm and minimum mean square error (MMSE) algorithm.

The filter matrix of the ZF algorithm and the MMSE algorithm are:

$$\mathbf{G}_{ZF} = (H^H H)^{-1} H^H \qquad (6)$$

$$G_{MMSE} = (H^H H + \sigma^2 I_M)^{-1} H^H \qquad (7)$$

The Zero-forcing (ZF) algorithm is simple to implement and has low computational complexity. However, during the detection process, the noise is amplified by multiplication with the pseudo-inverse matrix, which greatly degrades the detection performance. The MMSE algorithm takes into account the amplification effect of noise when designing the filter matrix. Therefore, its performance is improved compared to the ZF algorithm, but its computational complexity is also relatively high. The successive interference cancellation detection algorithm is performed on users one by one, so the SIC algorithm involves the problem of user detecting order. For the SIC algorithm, the detection order is worth further researching.

IV. OPTIMIZED SUCCESSIVE INTERFERENCE CANCELLATION ALGORITHM

The problem with SIC is that the interference cancellation can only be performed in the order of the transmitting antennas, and the interference intensity caused by the previous signals is not necessarily the largest. When the strongest signal is detected at the end of the detection sequence, the progressive interference cancellation effect will be very limited. Moreover, if strong interference cannot be eliminated in advance, it may also have a bad influence on the bit error rate performance of the entire system. Based on this, we propose an Optimized Successive Interference Cancellation algorithm (Optimized SIC).

A. Optimized successive interference cancellation algorithm based on column norm sorting

The sorted successive interference cancellation algorithm can improve the performance of the detection algorithm, improve the accuracy of signal detection, reduce error propagation, and does not significantly increase the complexity.

In a multi-beam satellite communication system, the proportion of the signal transmitted by the ground user i in the received signal is proportional to the column norm corresponding to the channel coefficient matrix, and thus is ordered according to the column norm size. The larger the column norm, the higher the detection priority, so we only

need to calculate the *K*-th column norm. The algorithm is as follows:

1. Define the initial variables.

$$H^{(i)} = H (i=1) \tag{8}$$

Here, the transmitted symbol of the user *i* is detected and a hard decision is made. The detection process uses ZF or MMSE detection to obtain ZF-OSIC and MMSE-OSIC detection algorithms. The filtered pseudo-inverse matrix $G^{(i)}$ is:

$$G^{(i)} = (H^{(i)^H} H^{(i)})^{-1} H^{(i)^H} \tag{9}$$

$$G^{(i)} = (H^{(i)^H} H^{(i)} + \frac{\delta_n^2}{\delta_s^2} I)^{-1} H^{(i)^H} \tag{10}$$

2. Sort the column norm for matrix $G^{(i)}$, select the column with the largest column norm value as W_i, multiply the matrix by W_i and y_i, and obtain the estimated value $\hat{x}_i$ of the transmitted signal by hard decision.

$$x_i = Q(w_i y_i) \tag{11}$$

Eliminating the interference caused by the detected signal in step 2 on the undetected signal from the received signal,

$$y_{i+1} = y_i - [H]_{:,i} \hat{x}_i \tag{12}$$

3. Set the *i*-th column of $H^{(i)}$ to 0 and remove it from the matrix $H^{(i)}$ to get $H^{(i+1)}$ and updates *i=i+1*.

B. Column Norm Sorting Group Multi-Order Successive Interference Cancellation Algorithm (GOSIC)

When the number of users is too large, the multi-beam satellite communication system needs to process multiple user data, the system complexity of the successive interference cancellation cannot be ignored. The concept of grouping is introduced on the basis of column norm sorting. Grouping based successive interference cancellation divides the entire coverage area into multiple detection groups with a certain spatial isolation by grouping multiple users according to their separated distance. Considering the uplink of the multi-beam satellite communication system, the coverage pattern of the system is illustrated in Figure 2.
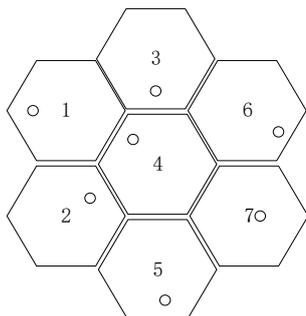

Figure 2. The coverage pattern of satellite multi-beam

Considering the proposed SIC algorithm based on column norm sorting, we introduce multi-level detection using the concept of grouping. Parallel and successive detection are combined in the decoding process to obtain better BER performance.

Due to the directivity of the multi-beam satellite antenna, the interference between the beams with large distance is

small. Taking the beam cluster in Figure 2. as an example, the received signal of beam *i* can be expressed as:

$$y_1 = [H]_{1,:} x + n_1 = \sum_{k=1}^{K} h_{1k} x_k + n_1 \tag{13}$$

When considering 7-beam coverage, beam 1 and beam 7 are very far apart, so the interference between them is small. Therefore, the interference of the user 7 to user 1 can be regarded as noise. Then, the computational complexity of the detector can be simplified. Based on the above thoughts, the seven spot beams in Figure 2. are grouped according to the detection order. The following grouping results are obtained to make the interference between groups relatively small: $G(1)=\{1,2\}$, $G(2)=\{3,4,5\}$, $G(3)=\{6,7\}$.
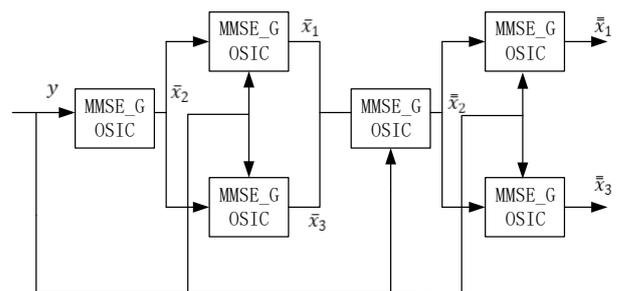

Figure 3. Schematic diagram of multi-stage MMSE-GOSIC algorithm

It can be seen from Figure 2 that users in the second group cause relatively stronger interference to users in the first and third groups. So in the simulation, we first perform interference cancellation on the second group of users, and first make decision on the second group of signals through the original signal. Then, interference cancellation is performed on the first and the third group based on the signal obtained from the hard decision results of the second group and the original signal. In this way, the estimated signals of the first and the third group without interference are obtained. Next, the signals of the second group is re-decoded using the interference-free signals of the first and the third group. Figure 3 shows the block diagram of the algorithm structure.

The estimated signals of the second group are first obtained by the original received signal y by performing column norm sorting successive interference cancellation the second group:

$$\bar{x}_2 = Q(\bar{Q}_2 H^H y) = Q(\bar{G}_2 y) \tag{14}$$

The matrix $\left[\overline{Q_2}\right]_i$ is a matrix of columns of the second set of signals in the channel matrix H. Interference cancellation according to the column norm sorting method is to eliminate the interference of the i-th group user to other users, and obtain the updated signal.

$$y_{i+1} = y_i - \left[\overline{Q_2}\right]_i \bar{x}_i \tag{15}$$

Where $\left[\overline{Q_2}\right]_i$ is a packet channel matrix composed of user columns in the i-th group of the channel coefficient matrix H, and the selected column is sequentially erased in the same manner as before. After the estimated value of $\bar{x}_2$ is obtained, the interference caused by the second group of signals can be eliminated from the received signals of the third group and the first group and updated.

$$\bar{y}_1 = y_1 - H_{12}\bar{x}_2 = H_{11}x_1 + \Delta y_1 + n_1 \qquad (16)$$

$$\bar{y}_3 = y_3 - H_{12}\bar{x}_2 = H_{11}x_1 + \Delta y_3 + n_3 \qquad (17)$$

Δy_3 and Δy_1 in the equations (16) and (17) represent mutual interference of transmitted symbols between the third group and the first group. However, since the distance between the third group and the first group is far apart, Δy_3 and Δy_1 can be regarded as noise in the respective received signals, respectively, and the receiving algorithm can be simplified. Finally, by eliminating the third group and the first group of received signals $\bar{y}_1$ and $\bar{y}_3$ of the signal interference of $\bar{x}_2$, $\bar{x}_3$ and $\bar{x}_1$ can be respectively estimated by the successive interference cancellation algorithm that performs the intra-group norm sorting.

$$\bar{x}_1 = Q(\bar{Q}_{11}H_{11}{}^{H}\bar{y}_1) = Q(\bar{G}_{11}\bar{y}_{11}) \qquad (18)$$

$$\bar{x}_3 = Q(\bar{Q}_{33}H_{33}{}^{H}\bar{y}_3) = Q(\bar{G}_{33}\bar{y}_{33}) \qquad (19)$$

Among them:

$$\bar{Q}_{11} = (H_{11}{}^{H}H_{11} + \frac{\delta_n^2}{\delta_s^2}I)^{-1} \qquad (20)$$

$$\bar{Q}_{33} = (H_{33}{}^{H}H_{33} + \frac{\delta_n^2}{\delta_s^2}I)^{-1} \qquad (21)$$

Where H_{11} represents the row corresponding to the first group of users in the channel matrix and the column corresponding to the first group of beams. Since interference caused by $\bar{x}_2$ has been eliminated, the transmitted signals of the second group of users are not considered for the beams in the first group. It has been previously specified that the transmitted signal from the third group of users is considered to be noise. So for signals 1 and 2 in group 1, the specified channel matrix $H_{7;1,2}$ will be reduced to the first group $H_{1,2;1,2}$. That is, H_{11} above.

In (16) and (17), we consider Δy_3 and Δy_1 to be noise in order to simplify the reception signals of the third and the first group. At the same time, we ignore the estimation error of $\bar{x}_2$. This will affect the performance of the algorithm, so we introduce the concept of multi-stage detection. We will use the estimated signals of the third and the first group in the first stage and the original received data to perform a second-stage detection on the estimation signals sent by the users in the second group, and update $\bar{x}_2$:

$$\bar{\bar{y}}_2 = y - H_1\bar{x}_1 - H_3\bar{x}_3 \qquad (22)$$

$$\bar{\bar{x}}_2 = Q(\bar{Q}_2 H_2{}^{H}\bar{\bar{y}}_2) = Q(\bar{\bar{g}}_2\bar{\bar{y}}_2) \qquad (23)$$

Among them:

$$\bar{Q}_2 = (H_2{}^{H}H_2 + \frac{\delta_n^2}{\delta_s^2}I)^{-1} \qquad (24)$$

Next, use the updated second-stage $\bar{\bar{x}}_2$ and the first-stage $\bar{x}_1$ and $\bar{\bar{x}}_3$ to eliminate interference from the received signals of the third and first groups again:

$$\bar{\bar{y}}_1 = y_1 - H_{12}\bar{\bar{x}}_2 - H_{13}\bar{x}_3 \qquad (25)$$

$$\bar{\bar{y}}_3 = y_3 - H_{32}\bar{\bar{x}}_2 - H_{31}\bar{x}_3 \qquad (26)$$

$\bar{\bar{y}}_1$ and $\bar{\bar{y}}_3$ are used again to detect and estimate the transmitted signals of the group, and update $\bar{x}_1$ and $\bar{x}_3$. The second order detection signals $\bar{\bar{x}}_1$ and $\bar{\bar{x}}_3$ are obtained. At this point, all transmitted symbols complete the second-order detection.

V. SIMULATION

In this paper, the successive interference cancellation technique of multi-beam satellite reverse link is studied. The multi-beam satellite system model is established according to the diagram in section 2. Considering the seven spot beams in the cluster and adopting TDMA access, the performance of the algorithm studied in this paper is evaluated by comparing simulation with the classical SIC.

TABLE 1. MULTI-BEAM SATELLITE REVERSE LINK SYSTEM PARAMETERS

Link parameter	Value	Link parameter	Value
Satellite orbital altitude (Km)	35786	Maximum satellite antenna gain (dBi)	52
Carrier frequency (GHz)	2.2	Maximum terminal antenna gain (dBi)	20
Number of beams	7	Receiver noise power (dBw)	-118
Free space loss (dB)	190	Half power beam width (°)	0.6

Therefore, the scenario of multi-beam satellite reverse link is established. The various parameters of multi-beam satellite reverse link simulated by SIC algorithm are shown in Table 1.
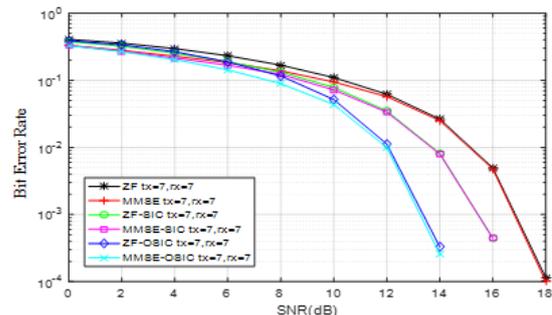

Figure 4. Bit error rate of ZF-OSIC/MMSE-OSIC detection algorithm

Figure 4. shows the bit error rate simulation of the classical linear detection ZF algorithm, MMSE algorithm, nonlinear detection ZF-SIC algorithm, MMSE-SIC algorithm and the optimized interference cancellation algorithm based on column norm sorting.

It can be seen from the Figure 4. that the performance of the MMSE algorithm is significantly better than the ZF algorithm at low SNR. This is because the noise is amplified by multiplying the pseudo-inverse matrix in the detection process in ZF algorithm, resulting in a poor performance. On the contrary, the MMSE algorithm takes the problem of noise into account. When the SNR value is high, the influence of noise becomes small, and the performance of the two linear detection algorithms is close. Combining SIC, the bit error rate of both ZF and MMSE is significantly reduced and the performance is significantly improved. This is because SIC algorithm is an iterative process. After each iteration, a signal is detected and eliminated to reduce the impact on subsequent signal detection. With the increase of SNR, the bit error rate of the six algorithms tend to decline, but the bit error rate of the OSIC algorithm drops the fastest. When the SNR value is high, the bit error rates of the six algorithms are obviously

divided into three groups. The linear detection algorithm has the highest bit error rate, the SIC algorithm has the second bit error rate, and the OSIC algorithm has the lowest bit error rate. This is because the received signals are sorted according to the column norm before the detection so that the strong signal is first detected and cancelled. Compared with the disordered SIC algorithm, the OSIC can eliminate the interference to a greater extent. Therefore, the bit error rate of the OSIC algorithm is significantly lower than the bit error rate of the SIC algorithm.

Figure 5. adopts the concept of grouping and multi-stage. Combined with the MMSE multi-user detection algorithm based on the column norm sorting, the MMSE-GOSIC algorithm is close to the bit error rate of the MMSE algorithm at low SNR. As the SNR increases, the bit error rate of the MMSE-GOSIC algorithm decreases rapidly, and its advantages gradually emerge. At high SNR, the bit error rate of the MMSE-GOSIC algorithm is the lowest compared to other algorithms, and its performance is better than other algorithms.
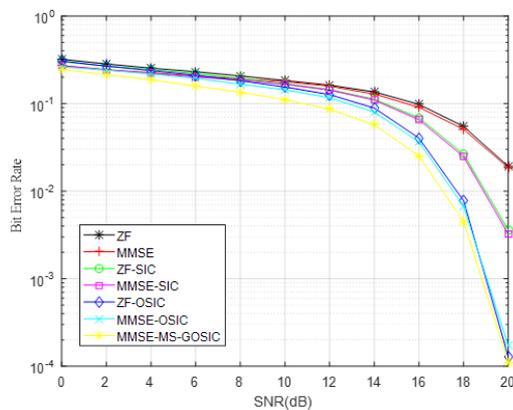

Figure 5. Bit error rate graph of MS-F-MMSE-GOSIC detection algorithm

VI. CONCLUSION

In this paper, the SIC algorithm of the multi-beam satellite system is simulated by establishing a multi-beam satellite reverse link. By investigating advantages and disadvantages of the SIC algorithm and the multi-beam satellite system, two optimization algorithms are proposed. A successive interference cancellation detection algorithm based on column norm sorting is proposed. Then, considering the high complexity as the number of user increases, the concept of grouping is proposed. The combination of grouping successive detection and multi-stage detection is used to reduce the complexity of the satellite side algorithm while ensuring accuracy. The simulation proves that the satellite MMSE multi-stage MUD algorithm based on the column norm sorting can achieve excellent performance compared with traditional SIC based detector. In the future, we will evaluate the proposed algorithm under the environments with rain fading.

REFERENCES

[1]   O. Vidal, G. Verelst, and J. Lacan. "Next generation High Throughput Satellite system", IEEE First Aess European Conference on Satellite Telecommunications. IEEE, 2012:1-7.

[2]   G. Colavolpe, A. Modenini, and A. Piemontese. "Multiuser Detection in Multibeam Satellite Systems: Theoretical Analysis and Practical Schemes", IEEE Transactions on Communications, 2017, 65(2):945-955.

[3]   F. Lombardo, E. A. Candreva, and I. Thibault, "Multi-user interference mitigation techniques for broadband multi-beam satellite systems," 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pp. 1805-1809, 2011

[4]   S. Andrenacci, M. Angelone, and E. A. Candreva., "Physical Layer Performance of Multi-User Detection in Broadband Multi-Beam Systems based on DVB-S2," 20th European Wireless Conference, pp. 1-5, 2014

[5]   C. Stallo and C. Sacchi, "Link performance analysis of multi-user detection techniques for W-band multi-beam satellites," IEEE Aerospace Conference.pp. 1-6, 2016.

[6]   U. Yeşilyurt, "Hybrid QRM-MLD-MMSE with SIC adaptive multiuser detection for SDMA-OFDM systems," 26th Signal Processing and Communications Applications Conference, pp. 1-4, 2018