Union Radio-Scientifique Internationale

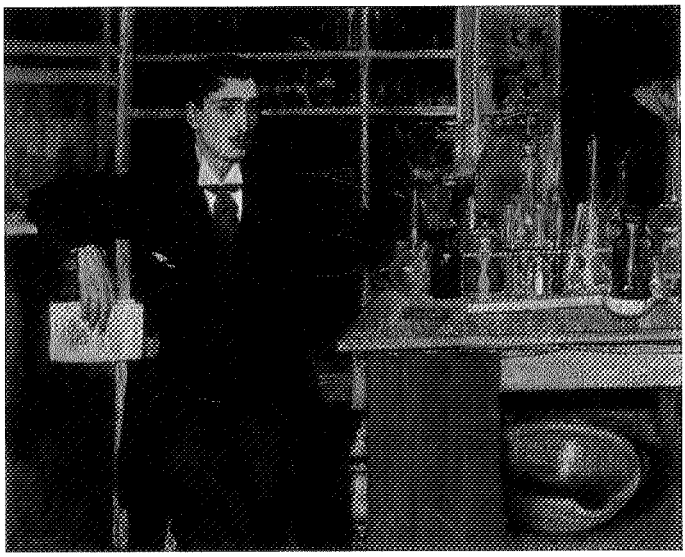
International Union of Radio Science

Space and Radio Science Symposium

Sous le Haut Patronage
de Sa Majesté le Roi Albert II

Under the high patronage of
His Majesty King Albert II

26-27 April 1995
Palais des Académies, Academy House
Brussels, Belgium



Editors:
Peter Van Daele & Paul Delogne

Front cover:

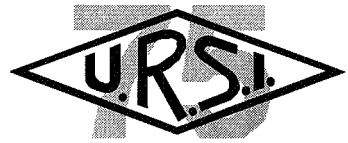
Robert Goldschmidt (°1877, †1935), the first Secretary General of URSI, shown in a painting by H. Evenepoel (Le chimiste 1896)

(Koninklijke Musea voor Schone Kunsten van België, Brussels, Belgium)

The International Union of Radio Science (URSI) is a foundation Union (1919) of the International Council of Scientific Unions as direct and immediate successor of the Commission Internationale de Télégraphie Sans Fil Scientifique (TSFS) which dates from 1913.

Union Radio-Scientifique Internationale

International Union of Radio Science



Space and Radio Science Symposium

Editors

Prof. Peter Van Daele
Universiteit Gent
Department of Information Technology
St. Pietersnieuwstraat 41
B-9000 Gent
Belgium

Prof. Paul Delogne
Université catholique de Louvain
Telecommunications and Remote Sensing
Place du Levant 3
B-1348 Louvain-la-Neuve
Belgium

URSI Secretariat
c/o University of Gent (INTEC), Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium
Tel. (32) 264 33 20, Fax (32) 264 42 88, E-mail inge.heleu@intec.rug.ac.be

Union Radio-Scientifique Internationale

International Union of Radio Science

Space and Radio Science Symposium

Under the High Patronage
His Majesty King Albert II

Organized by
The Belgian URSI Committee

and sponsored by

Koninklijke Academie
voor Wetenschappen,
Letteren en Schone Kunsten van België

•

Nationaal Fonds voor
Wetenschappelijk Onderzoek

Académie Royale
des sciences, des lettres
et des beaux-arts de Belgique

•

Fonds National de la
Recherche Scientifique

26-27 April 1995
Palais des Académies, Academy House
Brussels, Belgium

Union Radio-Scientifique Internationale

International Union of Radio Science

Space and Radio Science Symposium

NATIONAL PATRONAGE COMMITTEE

Michel Daerden
Ministre de la Politique Scientifique
et de l'Infrastructure
Minister van Wetenschapsbeleid
en Infrastructuur

•

Michel Lebrun,
Ministre de l'enseignement supérieur,
de la recherche scientifique,
de l'aide à la jeunesse,
et des relations internationales

Luc Van Den Bossche
Vlaams minister van
Onderwijs en Ambtenarenzaken

INTERNATIONAL PATRONAGE COMMITTEE

Agence spatiale canadienne - Canadian Space Agency (Canada)

•

Centre National d'Etudes Spatiales - agence française de l'espace (France)

•

Deutsche Forschungsanstalt für Luft- und Raumfahrt e.V. (Germany)

•

European Space Agency - Agence Spatiale Européenne (the Netherlands)

•

The Institute of Space and Astronautical Science (Japan)

INDUSTRIAL SUPPORT

SAIT Systems SA - Brussels

Union Radio-Scientifique Internationale

International Union of Radio Science

Space and Radio Science Symposium

Symposium Chair

Dr. P. Bauer (CESBIO-CNES, France), President of URSI

Scientific Programme Committee

J. Bach Andersen	Aalborg University Centre	Denmark
P. Bauer	CESBIO-CNES	France
R. Bonnet	ESA	France
G. Brussaard	Eindhoven University of Technology	The Netherlands
P.J.B. Clarricoats	Queen Mary & Westfield College	United Kingdom
P. Delogne	Université Catholique de Louvain	Belgium
R.D. Ekers	Australia Telescope	Australia
L. Emiliani	ESA	France
W.E. Gordon	Honorary President of URSI	USA
E.V. Jull	University of British Columbia	Canada
R.C. Kirby	ITU-R	Switzerland
P. Lagasse	University of Gent	Belgium
J. Lemaire	Institut d'Aéronomie Spatiale	Belgium
H. Matsumoto	Kyoto University	Japan
A.P. Mitra	Council Scientific & Industrial Research	India
T. Okoshi (†)	NAIR	Japan
V. Radhakrishnan	Raman Research Institute	India
R. Sagdeev	University of Maryland	USA
T.B.A. Senior	University of Michigan	USA
J. Van Bladel	University of Gent	Belgium

Local Organizing Committee

Chairman :	J. Van Bladel	Universiteit Gent
Secretary:	P. Van Daele	Universiteit Gent
Members:	A. Barel	Vrije Universiteit Brussel
	M. Blondel	Faculté Polytechnique de Mons
	L. Bossy	Université catholique de Louvain
	P. Delogne	Université catholique de Louvain
	E. Schweicher	Ecole Royale Militaire / Koninklijke Militaire School
	A. Van De Capelle	Katholieke Universiteit Leuven
	J.L. Van Eck	Université libre de Bruxelles

URSI Secretariat
c/o University of Gent (INTEC), Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium
Tel. (32) 264 33 20, Fax (32) 264 42 88, E-mail inge.heleu@intec.rug.ac.be

Union Radio-Scientifique Internationale

International Union of Radio Science

Space and Radio Science Symposium

Table of Contents

ABOUT THE PROGRAMME

PAUL DELOGNE

In early 1994 the URSI Board decided to organise a two days symposium to celebrate the 75th anniversary of URSI. The decision was also made to put the emphasis on a central scientific theme which would enlighten URSI's activities at the forefront of a science representative of all Commissions. It appeared that Space can play this central role. The preference was also expressed for a high-level symposium based on a limited number of invited lectures given by internationally renowned scientists.

A preliminary list of topics was prepared by the URSI Board, which also appointed an international scientific committee comprising more than twenty members. Its task was to identify the subjects and to propose speakers. A first run of ideas in the programme committee started in January 1994. This produced a preliminary list of 18 topics and 24 potential authors. The selection process could start. The final programme was complete by October 1994. Key elements in the choice were the interest of topics, the quality of speakers (and, of course, their availability), a good balance between the areas covered by the URSI Commissions and a wide international representation.

It is my pleasant duty to express warm thanks to the members of the international programme committee for their enthusiastic suggestions. My gratitude also goes to the authors who accepted the heavy task of preparing oral and written contributions. Their quality guarantees the success of the symposium.

Prof. Paul Delogne, chairman of the Scientific Programme Committee is with the
Université catholique de Louvain
Telecommunications and Remote Sensing
Place du Levant 3, Bâtiment Stévin
B-1348 Louvain-la-Neuve, Belgium
Tel: + 31 (10) 47 23 07 Fax: + 32 (10) 47 20 89

THE BIRTH OF URSI

JEAN VAN BLADEL

The first years of the 20th Century saw an explosion in the use of wireless communication. The Belgian kings Leopold II and Albert I immediately saw the interest of the new technique for communications in the vast, mostly unexplored territory of what was then Belgian Congo. First trials in 1902 were unsuccessful, and it was concluded that radio communications were probably impossible in equatorial regions. Leopold II did not abandon the idea, though, and asked Robert Goldschmidt to perform new experiments, and develop links, not only between Kinshasa and the then province of Katanga, but between Belgium and Congo itself. Goldschmidt, shown on the cover in an 1897 painting by Evenepoel, was a remarkable inventor and entrepreneur, who had previous experience with TSF (Télégraphie Sans Fil). He soon managed to obtain encouraging results, and established, in 1911, a first link between the two small cities of Boma and Banana. Goldschmidt, who had serious scientific credentials through his doctorate in chemistry, soon discovered that the radiopropagation problem, largely unexplored, was fundamental, and in great need of professional investigation. He discussed his ideas with various persons, in particular at the first "Conférence internationale de l'heure", held in Paris in 1912. The upshot was the creation, in october 1913, of the "Commission internationale de T.S.F". King Albert I accepted the honorary presidency of the new Commission. Dudell (UK) was president, and Wien (Germany) vice-president. A first scientific meeting was held in 1914 at the Royal Palace in Laeken, where extensive experimental facilities had been built. The First World War interrupted these promizing activities, but these were resumed in 1919. In July of that year, the Royal Belgian Academy of Sciences, Letters and Fine Arts invited the just-founded International Research Council to hold its First General Assembly in Brussels, in the very building in which the present Symposium is held. King Albert I attended the Opening Ceremony. It is during the Assembly that the "International Union of Scientific Radiotelegraphy" was founded. The first officials were General Ferrié (France) as President, Austin (U.S.A.), Eccles (U.K.) and Vanni (Italy) as Vice-presidents, and Goldschmidt (Belgium) as Secretary-general. What was later to become URSI was launched...

Prof. Jean Van Bladel, chairman of the Local Organizing Committee is with the
University of Gent,
Department of Information Technology,
St. Pietersnieuwstraat 41
B-9000 Gent, Belgium
Tel: + 32 (9) 264 33 21 *Fax: + 32 (9) 264 42 88*

THE ACTIVITIES OF URSI SINCE ITS FIRST GENERAL ASSEMBLY IN 1922

PIERRE BAUER

It is almost impossible to appreciate the progress made in Radio Science without a short glance back to the situation in 1922 when URSI held its first General Assembly. The 73 years which have elapsed since, have seen many advances in the ever expanding field of interest covered by the Union. The life of URSI is characterized by a continuous revision of its goals, and the introductory talk attempts to show the evolution of the individual Scientific Commissions, as well as the most striking results achieved. Mention is also made of the pioneering role played by URSI in other branches of science, for example astronomy and geophysics, by providing them with appropriate tools for further research. URSI meetings provide a unique forum for the exchange of ideas and results between specialists of various scientific disciplines, and also between radioscientists and engineers.

Dr. Pierre Bauer, URSI President and chairman of the "Space and Radio Science Symposium" is with the CESBIO - CNES
18, Avenue Edouard Belin
F-31055 Toulouse, France
Tel: + 31 (16) 61 55 85 25 Fax: + 31 (16) 61 55 85 00

GLOBAL SATELLITE NAVIGATION SYSTEMS : USES OF SPACE-TIME FIXES FROM GEODESY TO SAILING

JOHN E.B. PONSONBY

Abstract

There are two global satellite navigation systems, the American GPS (Global Positioning System) which consists of a complete constellation of 24 operational satellites and the Russian GLONASS, which does not yet have its full complement of 24 satellites in orbit. The two systems are very similar and operate at closely adjacent frequencies at L-band. They follow in the long tradition of earlier means of global navigation, being based on precise ephemerides for the movement of celestial bodies and the use of precision portable clocks. Unlike earlier methods however the celestial bodies are now man made, the satellites themselves, and by the use of radio technology, position is established by the measurement of range rather than of angle. Using one of these systems it is now possible to establish one's position on or near the Earth to within a few metres in all weathers and at all times. This is a remarkable development, a remarkable application of radio science, and it promises to revolutionise the navigation of all classes of vehicle form space craft to sailing dinghies. By making prolonged measurements at fixed locations it is possible to obtain position with geodetic accuracy of a few cm.

The essence of both systems is the measurement of the range of a receiver from the known positions of the satellites by measuring the one-way transmission times of radio signals. If in general four or more such measurements are made simultaneously, it is possible to obtain a fix for the receiver in 4-dimensional space-time. To do this of course the 4-D coordinates of the satellite transmission events must themselves be made know to the receiver. This is made possible by the satellites carrying knowledge of precise time in on-board atomic clocks and by them broadcasting the necessary elements of their orbits. This is achieved to an accuracy of a few nanoseconds of time and a corresponding few metres of space.

The details of the two systems will be described and compared. Their potential range of application will be outlined and the current technical and institutional obstacles to their widespread adoption for all navigational and timing purposes discussed. These obstacles are such as to have provoked proposals for the establishment of a third, wholly civilian and international Global Navigation Satellite System (GNSS) with a larger constellation of some 30 to 40 satellites.

Dr. John E.B. Ponsonby is with the
University of Manchester,
Nuffield Radio Astronomy Laboratories, Jodrell Bank,
Macclesfield, Cheshire, SK11 9DL
United Kingdom
Tel: + 44 (1477) 57 13 21 Fax: + 44 (1477) 57 16 18

ANTENNAS IN COMMUNICATIONS: MODERN DEVELOPMENTS AND CHALLENGES

Y. RAHMAT-SAMII

Abstract

Antennas are an essential segment of advanced satellite and personal communication systems, modern astronomical missions and high resolution earth observation platforms. This paper focuses on some representative examples in the area of modern synthesis and analysis of satellite and personal communication antennas. This paper contains a portion of presentation material to be delivered at URSI 75th Anniversary Symposium. In particular, novel design concepts will be highlighted using recent advances associated with diffraction synthesis/analysis of array-fed multiple reflector antennas and applying finite difference time domain (FDTD) approach for characterizing personal communications antennas including human interactions.

1. Introduction

Without antennas no wireless communications can be performed effectively. With modern manufacturing techniques and computer aided designs, the researchers, engineers and designers of modern satellite communications systems, advanced astronomical missions, sophisticated radar systems have been able to utilize ever-increasingly complex antenna configurations. The body of literature addressing the state-of-the-art in modern antennas is quite voluminous. The reader is encouraged to review several newly published antenna handbooks, textbooks on antennas and tremendous amount of research publications in various scientific journals. Antennas can be classified in various ways depending on their application, radiation pattern characteristics, bandwidth characteristics, physical configurations, etc. Modern antenna systems typically utilize reflector antennas, array antennas, various antenna elements or their combinations.

In this paper, our focus will be on modern computer aided design methodologies for array-fed multiple reflector antennas and antennas mounted on handsets for personal communications applications including human interactions. The material presented here is based on recent research activities of the author and his graduate students at UCLA. Due to the page limitations, no attempts have been made to review all the related published material in these areas, and instead,

Prof. Y. Rahmat-Samii is with the
Electrical Engineering Department
University of California - Los Angeles
Los Angeles, California 90024-1594
USA
Tel: + 1 (310) 206 3847 *Fax: + 1 (310) 206 8495*

the reader is referred to some relevant publications by the author which in turn provide substantial amount of reference material to other published data [1-26]. This paper heavily draws from the content of the author's previous publications.

2. Diffraction Synthesis of Reflector Antennas

Utilization of sophisticated synthesis/analysis techniques is required to fulfill the ever increasingly stringent requirements on reflector antenna performances for applications such as direct broadcast satellite (DBS) and satellite communications, radar systems, and radio astronomy. For this purpose, a versatile diffraction synthesis technique, which combines the optimization algorithms and the diffraction analysis techniques of Physical Optics (PO) and Physical Theory of Diffraction (PTD), is summarized in this paper. Edge diffraction, near-field effect, and blockage effect are taken into consideration automatically in this synthesis technique, which overcomes the accuracy limitation of the geometrical optics (GO) shaping algorithms. Furthermore, this synthesis technique is diversified in the sense that it can be applied to single- and dual-reflector antennas, to arbitrarily configured array feed, to reflectors with various aperture boundaries, and to a variety of radiation patterns as shown in Figure 1. The specific steps of the diffraction synthesis technique is described in the following section.

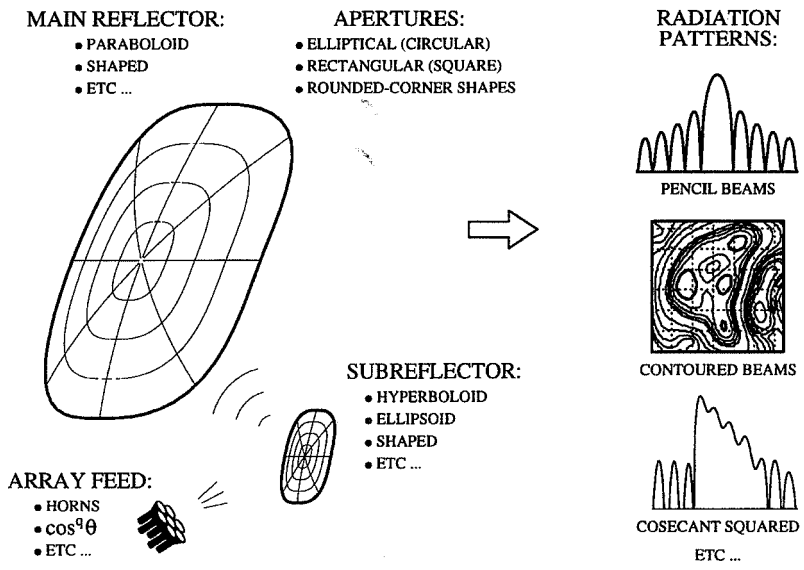


Fig. 1 - Diffraction synthesis/analysis technique applicable to the design of array-fed dual offset reflector antennas for satellite and radar applications.

2.1. Methodology

The diffraction synthesis technique consists of two steps: *parameterizing* the antenna system, and *optimizing* the parameters. Figure 2 illustrates these steps.

In the *parameterization*, the reflector surface is represented by a global Jacobi-Fourier expansion [27]. The expansion coefficients are the parameters that characterize the reflector shape; reflector shaping is achieved by adjusting (optimizing) these expansion coefficients. Due to the functional representation, the resultant reflectors are characterized by smooth surfaces, well defined

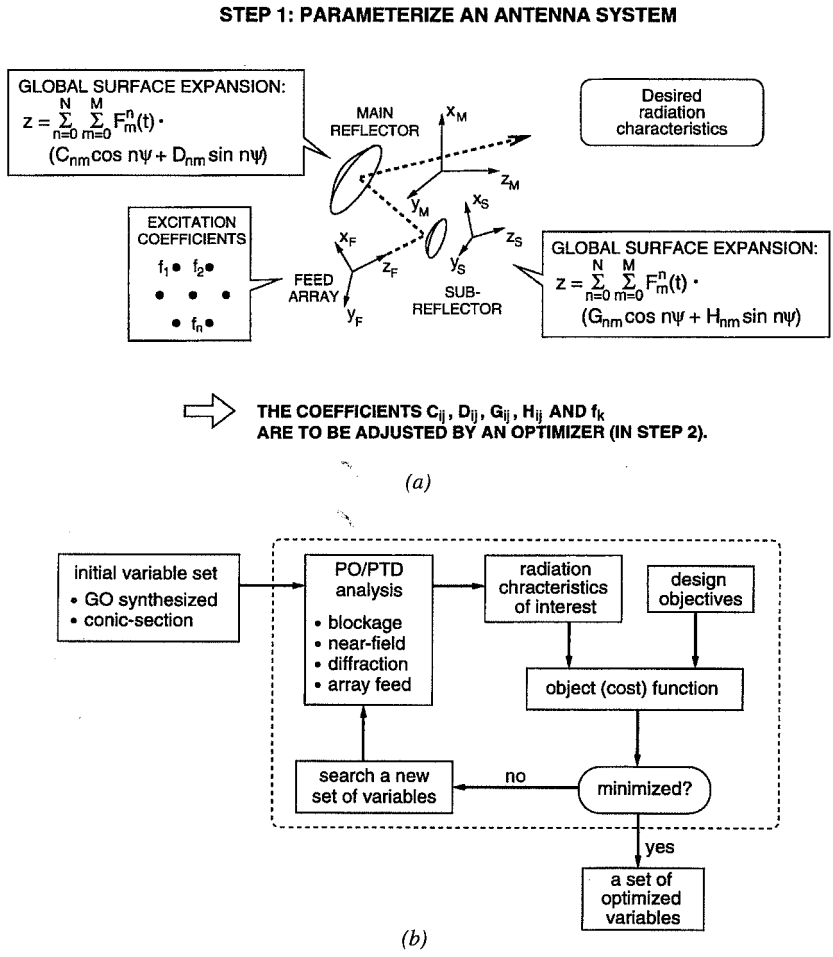


Fig. 2 - PO/PTD diffraction synthesis of array-fed reflector antenna systems.
(a) Step 1: parameterization of the antenna system. (b) Step 2: optimization of the parameters.

circumferences, and continuous surface derivatives; the shaped reflectors are ready for manufacturing (by computerized machining, for example) without requiring interpolation. In the current implementation, the boundary of the reflector aperture is allowed to be an ellipse (circle), a rectangle (square), or any intermediate rounded-corner shapes described by the superquadric function [11]. Other antenna parameters such as the excitation coefficients, positions, and orientations of the array feed elements can also be used as unknowns in the optimization.

In the *optimization*, the antenna parameters are initialized first. For example, the initial reflectors can be conic-section surfaces, or those obtained from the geometrical optics (GO) shaping algorithms, and the initial excitations of the array feed can be uniform or empirically determined. In each iteration of the optimization process, the radiation characteristics of interest are computed by the PO/PTD analysis. Then the difference between these computed values and the desired values are used to construct the object (cost, penalty) function that is to be minimized. If a termination condition (when the value or the gradient of the object function becomes smaller than a prescribed number, for example) is satisfied, the iteration ends. Otherwise, new parameter values are determined by an optimization algorithm, and the next iteration is initiated.

2.2. Design Examples

Two diverse applications of the generalized diffraction synthesis technique are illustrated in this section. In the first application, three types of contoured beam satellite antennas are designed: a shaped single-reflector antenna, a shaped dual-reflector antenna with circular main reflector aperture, and a shaped dual-reflector antenna with elliptical main reflector aperture. The challenging aspects of these designs are the use of a single feed. In the second application, two schemes for compensating reflector surface distortion are studied: compensation using an array feed, and compensation using a subreflector fed by a single feed.

These examples by no means exhaust the applications to which the diffraction synthesis technique is applicable. Instead, they are selected to illustrate the nature of the methodology such as the evolution of optimization and the coefficients-controlled surfaces, and the variety of solutions that one may assess using this technique. Other design examples may be found in the literature. For instance, simultaneous optimization of reflector and feed was demonstrated in [27], and synthesis of dual-offset reflector antenna fed by horn arrays was documented in [15].

2.2.1. Contoured Beam Satellite Antennas

Contoured beam antennas find many applications in modern satellite communication systems. These antennas radiate shaped beams that cover prescribed service regions in which the highest possible antenna gain is required.

Contoured beam reflector antennas (CBRA) can have various configurations. A typical CBRA consists of a paraboloidal reflector and a multi-horn feed. The shaped radiation is produced by adjusting the positions and excitations of the feed horns. Due to the complexity of the beam

forming network (BFN) associated with a large array feed, interest has been recently focused on reducing the number of feeds without degrading the quality of the contoured beam. One is challenged with, in the extreme case, using a single feed to illuminate a reflector. In this situation, the reflector must be shaped in order to generate the desired contoured beams. If high cross-polarization discrimination is required, one has to consider yet other CBRA configurations such as Gregorian type dual-offset reflector antennas.

For purpose of illustration, we present the design of three types of CBRA's in the following: (i) a single-reflector antenna, (ii) a dual-reflector antenna with circular main reflector aperture, and (iii) a dual-reflector antenna with elliptical main reflector aperture, all with single feed and the same coverage region. The goal is to produce a contoured beam that covers the contiguous United States (CONUS). In order to compensate for factors such as rain attenuation, gain corrections are specified in different areas or sites. This results in a weighted (or, nonuniform) plateau profile. For the current application, the desired contoured beam pattern is depicted in Figure 3, in which there are three prescribed "gain zones" in the contoured beam. A ± 0.1° pointing error has been considered in Figure 3. The locations of the observation sites (represented by dots in the figure) sample the gain regions in a uniform manner, and are separated by about half of the half-power beam width of the pencil beam radiated by a paraboloidal main reflector.

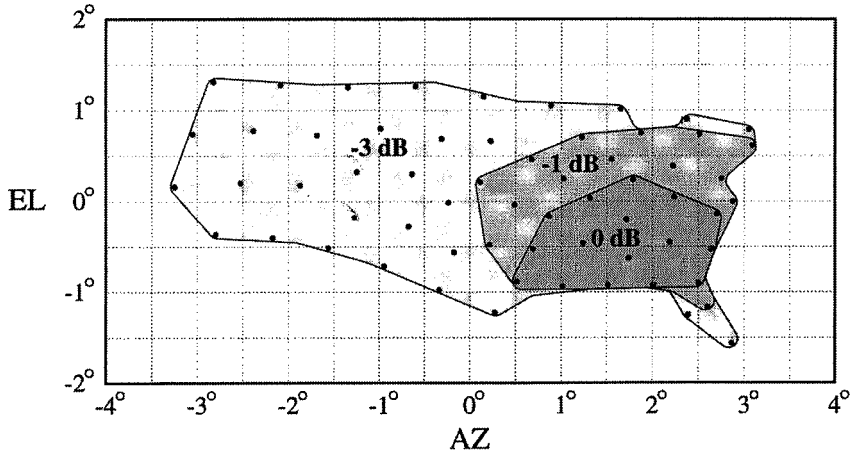


Fig. 3 - The CONUS coverage map. A ± 0.1° pointing error is considered. The three gain zones are designed for compensating rain attenuation.

Let N_c be the number of observation sites, and D_i the computed directivity at the ith site modified by gain correction. The object function F is constructed as

$$F = -D_{av} + w.\delta D \tag{1}$$

$$D_{av} = \frac{1}{N_c} \sum_{i=1}^{N_c} D_i \tag{2}$$

$$\delta D = \left[\frac{1}{N_c} \sum_{i=1}^{N_c} |D_i - D_{av}|^2 \right]^{1/2} \tag{3}$$

where D_{av} is the average directivity (including gain corrections), δD represents the root- mean-square gain ripple, and w is a weighting coefficient with a typical value of unity. The operation frequency is 11.95 GHz. The object function (1) can be further modified for cross-polarization improvements.

2.2.2. Single Reflector/Single Feed

A circular offset reflector with D = 2a = 1.524 m, F = 1.506 m, H = 1.245 m, and θ_f = 42.77° (Figure 4, with v = 1 for the superquadric boundary) is used. The feed is an x-polarized $(\cos \theta)^q$ source with -12 dB edge taper (q = 14.28). The reflector is represented by 28 terms in the Jacobi-Fourier expansion. (Clearly one could start with the Zernike polynomials as an alternative). The first (i.e. the constant) coefficient C_{00}= 0.849 (meters) is left intact in order to keep the shaped reflector in proximity to the original paraboloid, and the other 27 coefficients are optimized with the initial values derived from a paraboloid. It has been observed that in the early stages the optimizer spreads out the energy contained in the initial pencil beam, and then polishes the shaped pattern until it precisely delineates the gain regions with minimum gain ripples. Profile of the resultant shaped reflector is depicted in Figure 5. The deviation from the initial paraboloid is within (-0.4 λ, -0.6 λ).

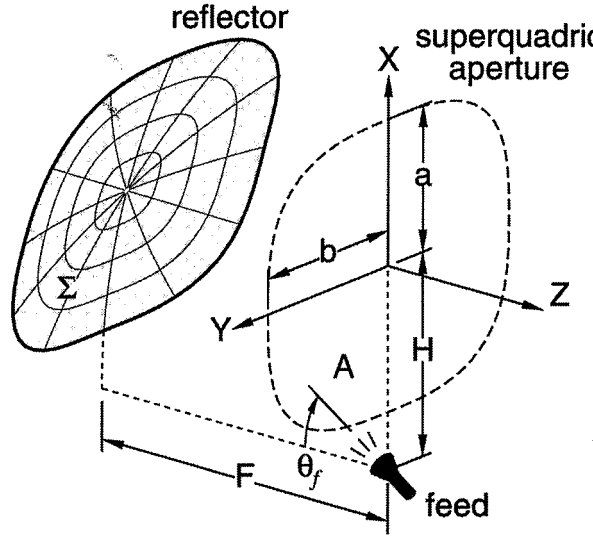


Fig. 4 - Antenna geometry of an offset single-reflector antenna

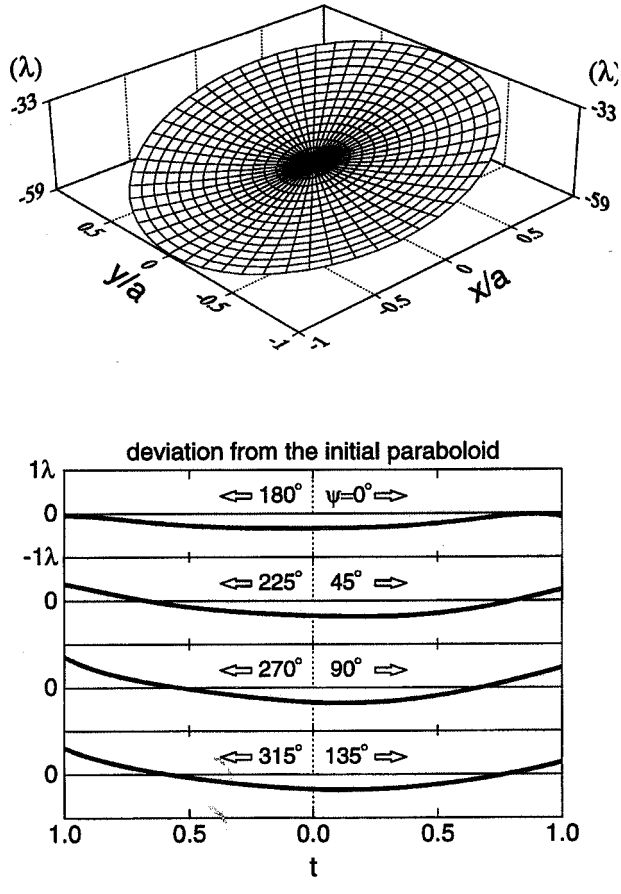


Fig. 5 - The shaped reflector for CONUS coverage and its deviation to the original paraboloid. (a = D/2 is the radius of the reflector aperture.)

2.2.3. Dual Reflector, Single Feed, Circular Aperture

The cross-polarized field of the above single-reflector antenna as shown in Figure 7(a) is excessively high for stringent requirement on the discrimination level (30 dB, for example). To reduce the cross-polarized field level, we choose a Gregorian type dual-offset reflector antenna with optimum tilted angles [21] as shown in Figure 6. The feed is an x-polarized $(\cos\theta)^q$ source with -12 dB edge taper ($q = 25.98$). Notice that for purpose of comparison, the same main reflector diameter is used, and the offset height and focal length are shortened in order to keep the total volume of the antenna similar to that of the single-reflector case.

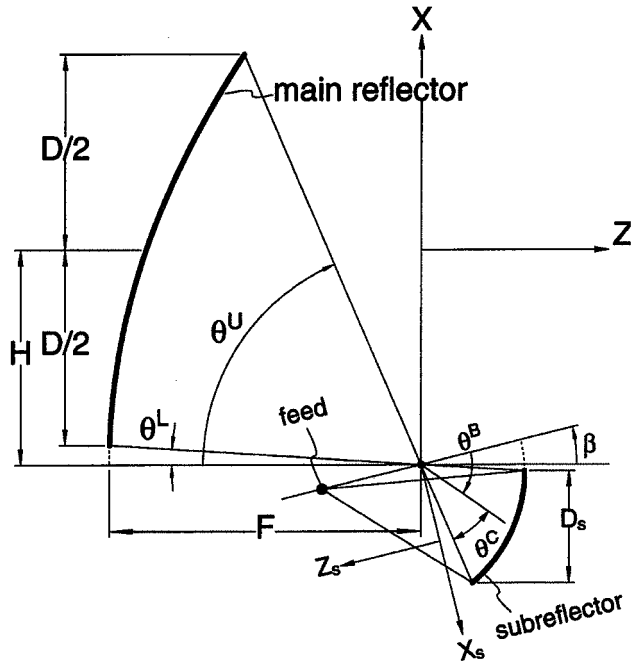


Fig. 6 - Geometry of a Gregorian type dual-offset reflector antenna (with single feed) for CONUS contoured beam application. D = 1:524 m, H = 0:838 m, F/D = 0:8, e = 0.333, β = 13:84°, inter-foci distance = 0.3995 m.

To obtain the CONUS shaped beam, the main reflector and the subreflector are shaped simultaneously. Each reflector is represented by 28 expansion terms. The optimization starts with paraboloid/ellipsoid, and there are totally 54 coefficients (27 for each reflector) to be adjusted. The resultant co-polarized and cross-polarized patterns are compared with those produced by the single-reflector antenna in Figure 7. It is seen that the cross-polarized field is significantly reduced, with the integrity of the co-polarized contours maintained.

2.2.4. Dual Reflector, Single Feed, Elliptical Aperture

Reflector antennas with elliptical apertures find many applications in modern satellite communications and radar systems. Utilization of an elliptical aperture is usually motivated by the required radiation patterns, and the reduced weight and cost. It is a challenging task though to re-direct the field from a circularly symmetric feed to reflectors with elliptical apertures. This is particularly so when a complex radiation pattern such as a contoured-beam is to be generated. The key to such design problems is reflector shaping.

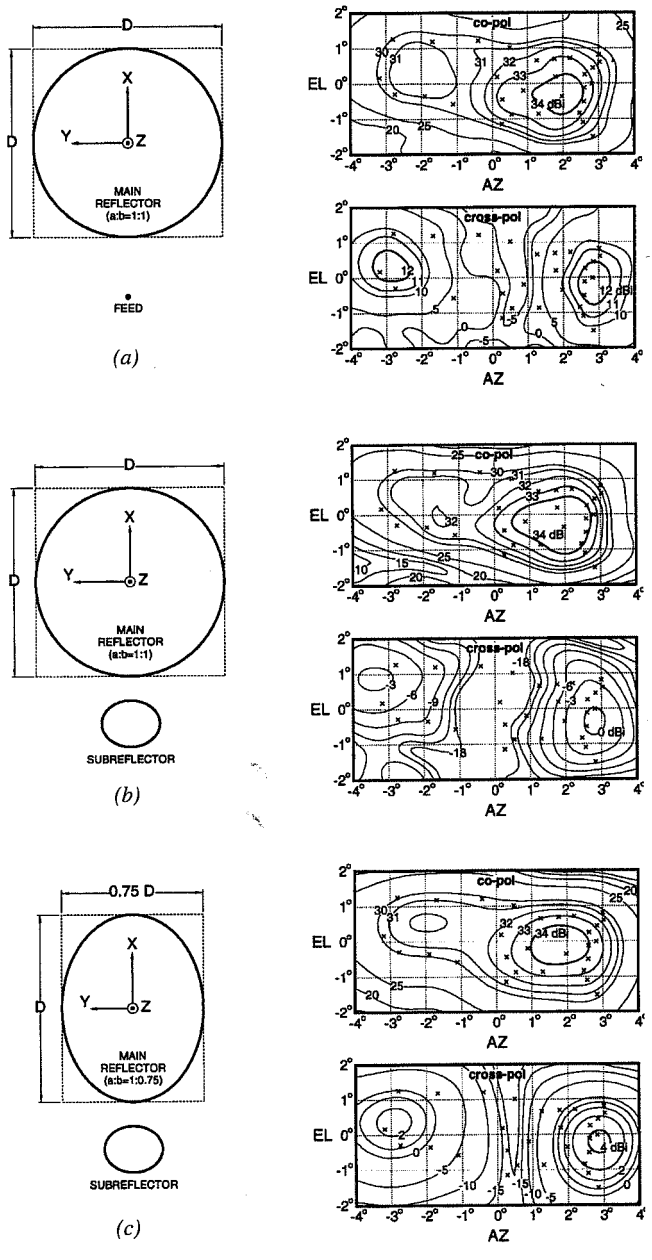


Fig. 7 - Comparison of the co- and cross-polarized CONUS patterns using (a) a single-reflector antenna, (b) a dual-reflector antenna with circular main reflector, and (c) a dual-reflector antenna with elliptical main reflector. In all cases, only a single feed is used. The frequency is 11.95 GHz.

It is observed in Figure 3 that the CONUS coverage region is elongated in the azimuthal direction. This suggests that an antenna with elongated aperture (in a direction that is perpendicular to that of the contoured pattern) may produce a similar far-field pattern if the reflectors are properly shaped. To justify this concept, we perform diffraction synthesis (shaping) on an elliptical antenna as shown in Figure 7(c), which has an aspect ratio of 1 : 0.75 but otherwise identical to the previous circular antenna. After properly shaping both reflectors, a typical resultant contoured beam is depicted in Figure 7(c). It is seen that a contoured pattern with similar average gain and gain ripple has been produced, with slightly decreased "resolution" in the azimuthal plane due to the reduced reflector size. The level of the cross-polarized field is raised about 4 dB compared to the circular case shown in Figure 7(b).

2.3. Reflector surface distortion compensation

Very large reflector antennas have been widely used in modern communications systems because these antennas produce high gain/low noise radiations and provide enhanced data transmission capacity [23]. For both ground and space antennas, however, large reflectors may suffer from systematic surface distortion due to thermal or gravitational effects. Additionally, for non-rigid reflector surfaces such as those used in unfurlable or inflatable [23, 29] antenna systems, distortion may be resulted from the mechanical construction of the reflector. The distorted reflector surface typically causes aperture phase errors and degraded antenna performance.

In this section, the diffraction synthesis technique is employed to investigate different compensating systems: (i) array feeds and (ii) a shaped subreflector fed by a single feed. For purpose of illustration, we use the example problem depicted in Figure 8, in which the distorted paraboloidal reflector is modeled as

$$z = -F + \frac{(x+H)^2 + y^2}{4F} + F_d$$

$$F_d = 0.011\left(\frac{\rho}{a}\right)^3 \cos 2\psi \qquad \text{(meters)}$$

$$\rho = \sqrt{x^2 + y^2}, \qquad \psi = \arctan\frac{y}{x}$$

where a = $D/2$ is the radius of the reflector aperture, F is the focal length of the ideal paraboloid, and H is the offset height of the reflector aperture center. The function F_d as depicted in Figure 8 represents a typical slowly varying thermal distortion. It is seen that this distortion has severely deteriorated the far-field patterns, reducing the peak directivity from 42.5 dB to 39.0 dB.

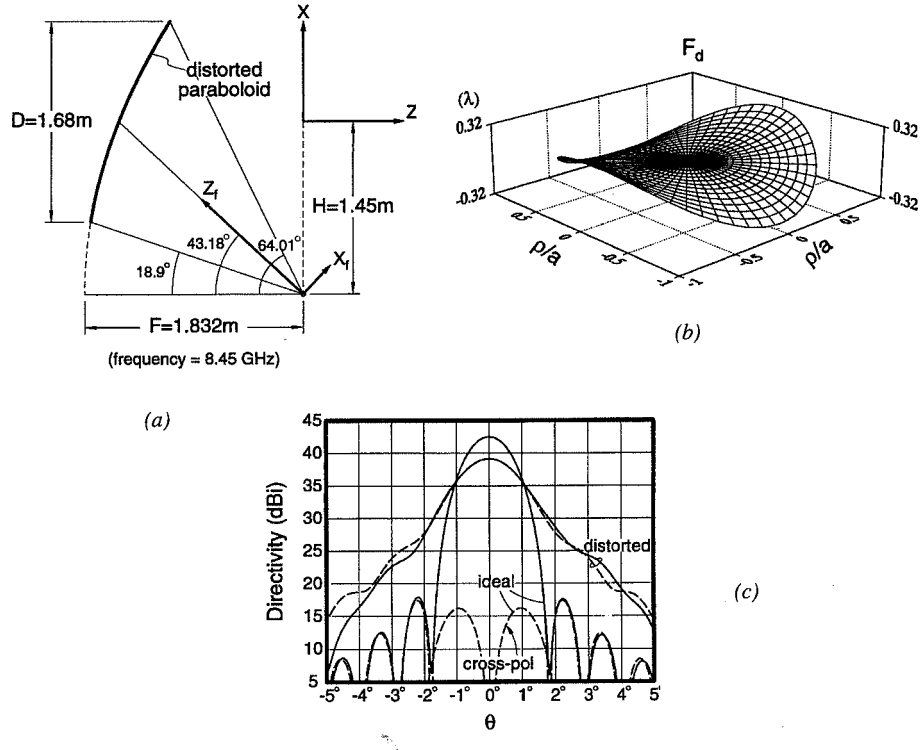


Fig. 8 - A distorted paraboloidal reflector antenna. (a) Antenna geometry. (b) The surface distortion, Fd. (c) The distorted far-field patterns. Ideal patterns refer to those produced by the paraboloid. (Solid lines: $\phi = 0°$, dashed lines: $\phi = 90°$).

2.3.1. Focal Plane Array Feeds

One way of achieving distortion compensation is to use an array feed with proper excitations, which may be obtained by the method of conjugate field matching [20]. In this paper, instead, we use the diffraction synthesis technique to determine the optimum array excitations. As an example, let us consider a representative 19-element array as depicted in Figure 9(a). In the design, the optimization variables are the real and imaginary parts of the excitation coefficients, with reference to the center element. These variables are initially set to values (1 for real parts and 0 for imaginary parts) that correspond to uniform excitations. The object function is

$$F = -D_o \tag{4}$$

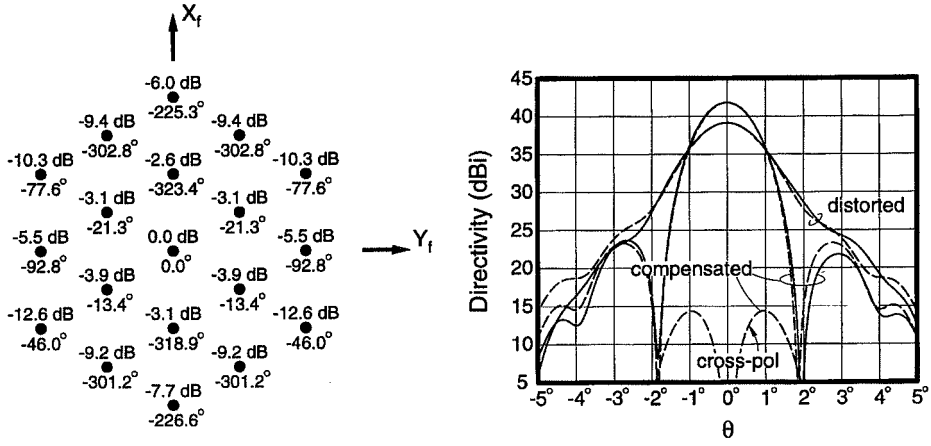


Fig. 9 - Distortion compensation using a 19-element array feed with an inter-element spacing of 1.06. (a: left) The optimum feed excitations. (b: right) Far-field patterns. (Solid lines: $\phi = 0°$, dashed lines: $\phi = 90°$).

where D_0 is the boresight directivity. The results of optimization is shown in Figure 9(b), in which it is shown that the antenna pattern has been effectively restored, with 41.8 dB boresight directivity (the actual antenna gain may be somewhat lower due to losses in the array beam forming network). The optimum excitations resulted from the diffraction synthesis technique as shown in Figure 9(a) are similar to those obtained by the method of conjugate field matching [20].

2.3.2. A Shaped Subreflector and a Single Feed

In order to avoid the increased complexity, loss, and weight of a beam forming network (BFN), another compensation scheme that employs a feed system consisting of a deformable subreflector and a single feed may be considered [30]. A design is carried out for the distorted paraboloid as shown in Figure 10(a). The feed is described by a $(\cos \theta)^q$ model with $q_1 = q_2 = 70$. The surface of the deformable subreflector is determined by the diffraction synthesis (shaping) technique with the object function:

$$F = -D_o + w.R \tag{5}$$

where w is a weighting coefficient, and R is the root-mean-square ripple of the directivities at, for example, $\theta = 1.3°$ (the -10 dB beamwidth) in the $\phi = 0°, 90°, 180°$, and $270°$ planes. The term $w.R$ is included to control asymmetry in the patterns. As a result of optimization, effective compensation is achieved as shown in Figure 10(c), with 41.3 dB boresight directivity. The deviation of shaped subreflector to the original hyperboloid is depicted in Figure 10(b). Mechanical structures that facilitate reconfigurable reflector surfaces represented by the Jacobi-Fourier expansions are recently being investigated.

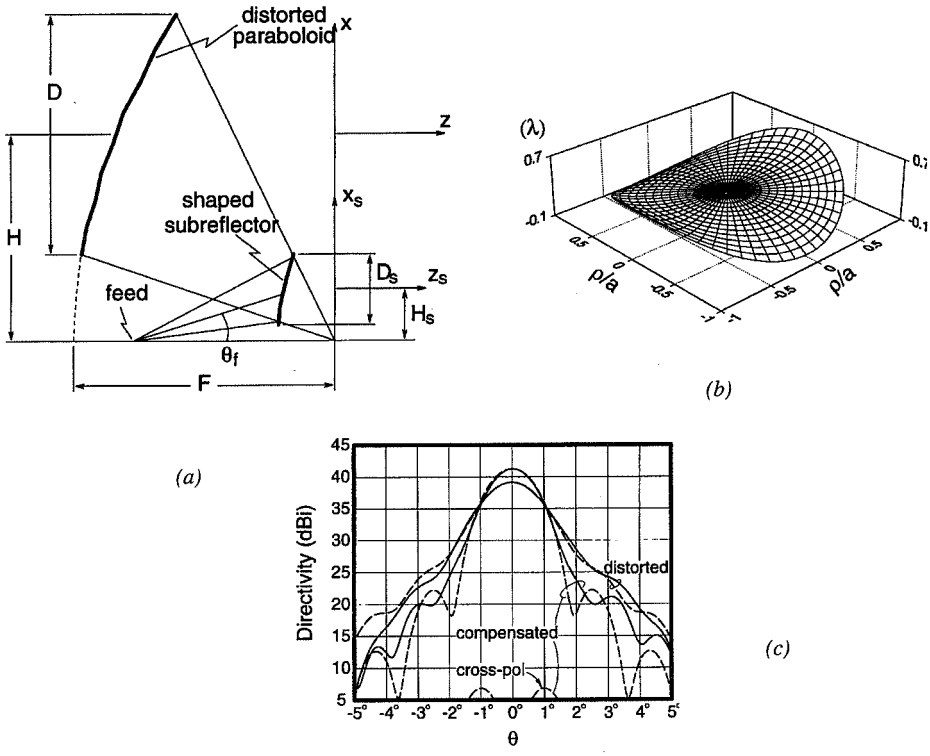


Fig. 10 - Distortion compensation using a deformable shaped subreflector and a single feed. (a) Antenna geometry. Subreflector: D_s=0.5 m, H_s=0.363 m, eccentricity = 2.3. Feed: located at 1.415 from the focal point, θ_f = 17.58°. (b) Deviation of the shaped subreflector to the original hyperboloid. (c) Far-field patterns. (Solid lines: ϕ = 0°, dashed lines: ϕ = 90°)

3. Handset Antennas Including Human Interactions

As wireless systems assume an ever-increasing role in communications networks, increasingly stringent requirements are placed upon the performance of the antennas used. The designer must be able to develop a highly efficient, low-profile antenna which can be mounted on a hand-held transceiver and operated in the proximity of human tissue [7, 8]. The design process for such a radiator is greatly facilitated by the use of sophisticated analysis methodologies which allow detailed simulations of the antenna when placed in its radiating environment. There exists a variety of candidate techniques which can be used to perform the required analysis, each with its own set of merits and disadvantages. Among these techniques, considerable attention has been focused on the direct solution of Maxwell's time-domain differential equations using the finite-difference time-domain (FDTD) [31, 32] for characterization of very complex radiating structures. It is this approach which will be discussed and applied to the simulation of integrated antennas for hand-held transceiver devices.

3.1. FDTD Approach

The underlying principle behind direct Maxwell solvers such as the FDTD methodology is the discretization of the time and spatial derivatives in Maxwell's equations into difference equations implementable on a computer. Figure 11 illustrates a typical cell with the field components spatially interleaved for convenience in implementing the finite difference approximation. We use subscripts i, j, and k to denote the spatial position ($i\Delta x$, $j\Delta y$, $k\Delta z$) and a superscript n to indicate the time such that $t = n\Delta t$, with Δt the time step and Δx, Δy, and Δz the spatial discretization step sizes. From Maxwell's curl equations in the time domain, we can formulate the following finite difference approximations:

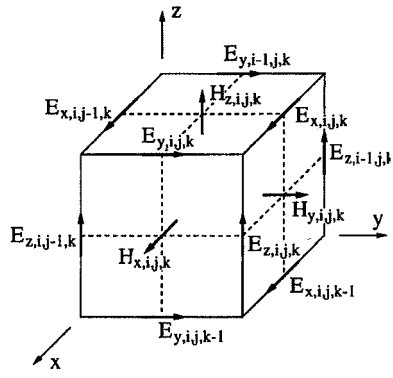


Fig. 11 - Unit cell for FDTD with position of field components.

$$H_{x,i,j,k}^{n+1/2} = H_{x,i,j,k}^{n-1/2} + \gamma\left[\frac{E_{y,i,j,k}^{n} - E_{y,i,j,k-1}^{n}}{\Delta z} - \frac{E_{z,i,j,k}^{n} - E_{z,i,j-1,k}^{n}}{\Delta y}\right] \tag{6}$$

$$E_{x,i,j,k}^{n+1} = \frac{1}{\beta}\left[\alpha E_{x,i,j,k}^{n} + \frac{H_{z,i,j+1,k}^{n+1/2} - H_{z,i,j,k}^{n+1/2}}{\Delta y} - \frac{H_{y,i,j,k+1}^{n+1/2} - H_{y,i,j,k}^{n+1/2}}{\Delta z}\right] \tag{7}$$

$$\alpha = \frac{\varepsilon}{\Delta t} - \frac{\sigma}{2},\ \beta = \frac{\varepsilon}{\Delta t} + \frac{\sigma}{2},\ \gamma = \frac{\Delta t}{\mu} \tag{8}$$

and similar expressions exist for the other four field components. In these relations, σ, ε, and μ represent the conductivity, permittivity, and permeability respectively of the medium.

To apply the FDTD algorithm to the simulation of antenna structures, the space about the radiator is filled with the unit cells of Figure 11, an arrangement which ensures continuity of tangential fields across cell interfaces. Equations (7) are used to compute the field values for each cell. Due to the "leap-frog" nature of the time-indices between E and H fields, these equations can be used in a time-stepping manner to track the time evolution of the electromagnetic fields over the spatial grid. Electric conductors within the domain are modeled by forcing tangential electric field values coincident with the conductor surface to zero at each time step. To avoid the inaccuracies caused by reflections from the outer boundaries of the computational domain, absorbing boundary conditions [33, 34] are used to truncate the computational grid. Because the FDTD method is an explicit procedure, stability is ensured if the time and space steps satisfy the condition

$$\Delta t \leq \frac{1}{c}\left(\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}+\frac{1}{\Delta z^2}\right)^{-1/2} \tag{9}$$

where c is the maximum velocity of light in the computational domain.

3.2. Design Examples: Cellular Applications

As a demonstration of the capability of the FDTD algorithm in characterizing antennas for personal communications devices, a large class of antennas have been considered including the dual planar inverted-F antennas (PIFA) situated on a conducting chassis as shown in Figure 12. As can be seen, the PIFA element consists of a conducting plate suspended above the handset chassis with a shorting wire at one end. The element is fed by a probe such as the inner conductor

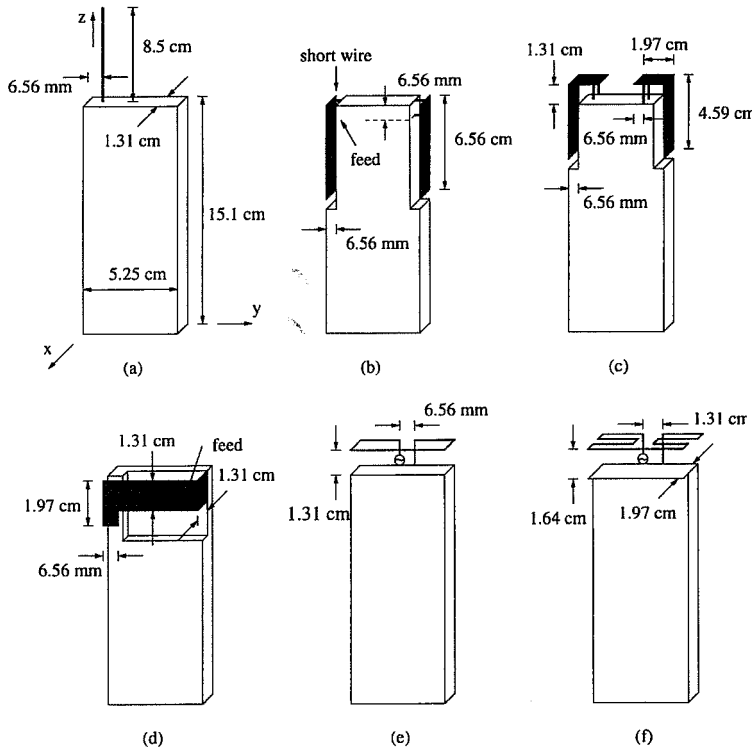


Figure 12: Geometry of the conducting handset showing various antenna element configurations considered.

of a coaxial feeding cable. The low-profile nature of this antenna topology offers the advantages that it can be conveniently packaged and protected from damage and is non-obtrusive to the user. Furthermore, two elements such as this may be used as an antenna diversity configuration to combat the effects of multipath fading.

The effects of the human operator in the vicinity of the radiator are accounted for by including models of the head and hand within the computational grid. Figure 13 illustrates the geometry for the simulated hand which consists of a layer of bone surrounded by a layer of muscle. A detailed model for the head has also been generated based upon data from magnetic resonance imaging and pictures from anatomy books. The tissue is modeled in the FDTD framework by assigning the appropriate conductivity and permittivity to each cell.

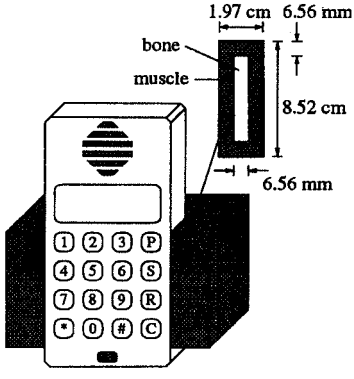


Fig. 13 - Model of the hand in the simulation of antennas on a hand-held device.

The FDTD method is used to simulate the configuration when one of the PIFA antennas is excited and the second is terminated in a matched load. A Gaussian shaped voltage pulse is applied to the feed probe of the excited PIFA, and the time-evolution of the fields is computed using the time-stepping equations. The transient voltage and current waveforms at the feed point are then Fourier transformed (via FFT) to obtain the frequency domain quantities from which the input impedance and scattering parameters can be obtained. Figure 14(a) illustrates the value of $|S_{11}|$ versus frequency for this antenna topology when the hand is absent and when the hand is present at three vertically displaced locations on the handset (distances represent the distance from the top of the antenna element to the top of the hand). Clearly, the hand has a pronounced effect on the antenna performance, especially when it begins to cover the antenna element. Measured data is shown for the case when the hand is absent and when the hand is at the middle position. Excellent agreement between the measured and computed results is obtained for this example.

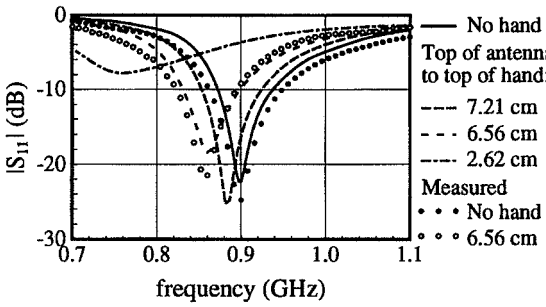


Fig. 14 - $|S_{11}|$ versus frequency for the PIFA (see Figure 12(b)) on the handset without the hand and with the hand at 3 locations. The results are compared to measured data when no hand is present and when the hand is at the middle position.

The gain patterns for the dual PIFA configuration at 915 MHz are shown in Figure 15. The plots in Figure 15(a) illustrate the patterns for the antennas on the handset when no biological tissue is present. The handset is rotated 60° from vertical

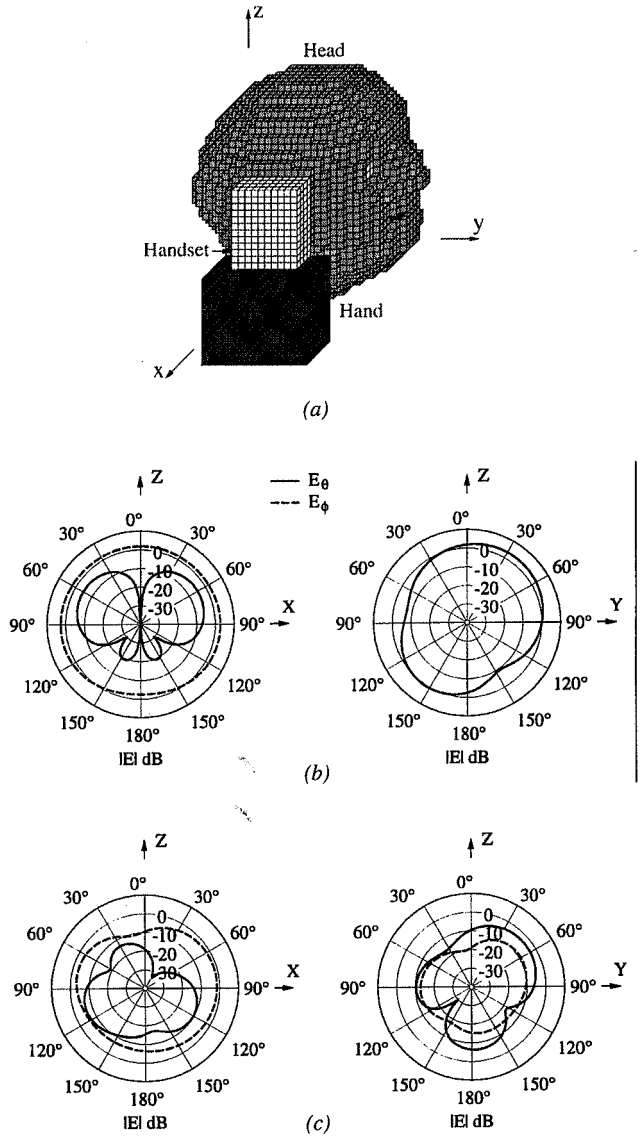


Fig. 15 - Gain patterns for the PIFA on the handset at 915 MHz: (a) geometry with head, hand, and handset; (b) patterns with no tissue present; (c) patterns with the hand at the lowest position and the head present. The handset is inclined 60° from vertical to simulate the position when speaking.

in the yz plane to simulate the scenario when the operator is speaking. The plots in Figure 15(b) represent the patterns when the hand is at its lowest position and the head is 1.3 cm away from the handset. The effect of the body in this case is very noticeable, causing alteration of the radiation, polarization, and gain characteristics of the antenna.

3.3. Design Examples: Satellite Applications

Recently there has been significant interest in developing personal satellite communications as a means of providing very large areas of wireless communications coverage. From the antenna design point of view these links offer new challenges in the way of requirements for circular polarization, slightly more directive radiating elements and higher operating frequencies (L and S band). Again the presence of the human operator must be contended with in this scenario, not only to ascertain the loss in gain due to absorbtion but also to study the possible effects the presence of the operator will have on the polarization purity of circularly polarized antennas.

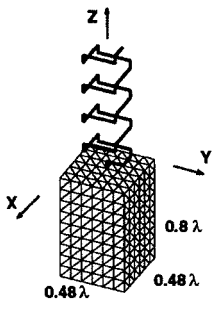


Fig. 16 - Schematics of the square helix structure mounted on a 0.48 λ x 0.48 λ x 0.8 λ conducting box.

One candidate antenna for a personal satellite transceiver handset is a square thin-wire helix. Figure 16 illustrates an example case studied, a modification to the standard helix structure mounted on a handset, where the thin-wire element is constructed of only straight wire segments. All the vertically orientated wire segments are 0.06 λ in length and all the horizontally orientated wire segments are 0.24 λ in length. This square helix structure has the potential advantages of easier construction, a geometry easier to integrate with the handset and can be simulated with a Cartesian grid based FDTD code.

Figures 17(a) and (b) are plots of the computed far-field pattern and axial ratio for the square helix structure illustrated in Figure 16 using FDTD. Figures 18(a) and (b) are plots of the FDTD computed far-field pattern and axial ratio for the square helix structure positioned next to a version of the computation head model described in the

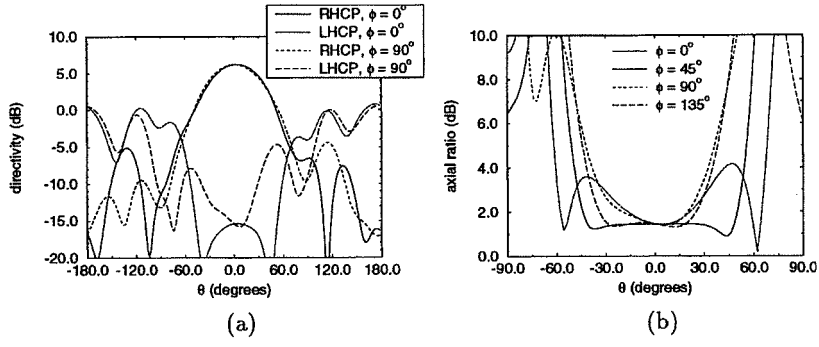


Fig. 17 - FDTD computed: (a) far-field pattern and (b) axial ratio; for the square helix structure illustrated in Figure 16.

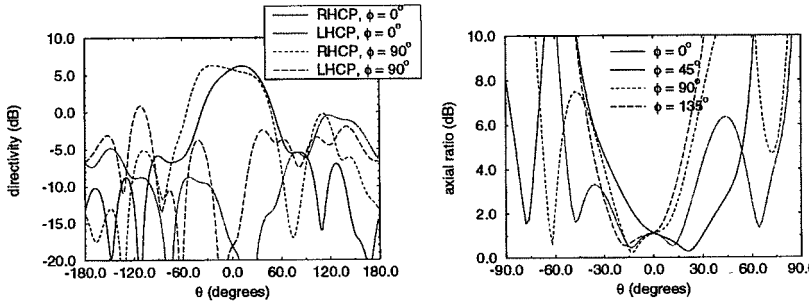


Figure 18: FDTD computed: (a) far-field pattern and (b) axial ratio; for the square helix structure illustrated in Figure 16 near the computational head model.

previous section modified for 1.8 GHz. Besides the loss in radiated power due to absorption in the head model, it can be seen in Figures 18 (a) and (b) that there are significant distortions of both the computed far-field pattern and axial ratio due to the prescene of the head model.

4. Acknowledgements

The author would like to thank his students D. Duan, M. Jensen and J. Colburn for their contributions in generating the results presented in this paper.

References

[1] Y. Rahmat-Samii, "Modern Concepts in Analysis, Synthesis and Measurements of Antennas", In Modern Radio Science 1993, edited by H. Matsumoto, pp. 75-104, Oxford University Press, 1993.

[2] Y. Rahmat-Samii, "Recent Advances in Diffraction analysis of Reflector Antennas", In Directions in Electromagnetic Wave Modeling, edited by H. L. Bertoni and L. Felsen, pp. (39-65), Plenum Publishing Corporation, New York, 1992.

[3] Y. Rahmat-Samii, "Large Antenna Experiment Aboard the Space Shuttle - Application of Nonuniform Sampling Techniques", Chapter3, in Electromagnetic Modeling and Measurements for Analysis and Synthesis Problems, edited by B. de Neumann, pp. (51-60), NATO ASI Series, Kluwer Academic Publishers, 1991.

[4] Y. Rahmat-Samii, "Antenna Diagnosis by Microwave Holographic Metrology", Chapter 2, in Electromagnetic Modeling and Measurements for Analysis and Synthesis Problems, edited by B. de Neumann, pp. (17-50), NATO ASI Series, Kluwer Academic Publishers, 1991.

[5] Y. Rahmat-Samii, "Reflector Antennas", Chapter 15, in Antenna Handbook, edited by Y. T. Lo and S. W. Lee, pp. (15-1) - (15-124), Van Nostrand Reinhold, 1988.

[6] D. W. Duan and Y. Rahmat-Samii, "A Generalized Diffraction Synthesis Technique for High Performance Reflector Antennas", IEEE Trans. Antennas and Propagation, Vol. 43, No. 1, pp.27-40, Jan. 1995.

[7] M. A. Jensen and Y. Rahmat-Samii, "EM Interaction of Handset antennas and a Human in Personal Communications", Proceedings of IEEE, Vol. 83, No. 1, pp.7-17, Jan. 1995.

[8] M. A. Jensen and Y. Rahmat-Samii, "Performance Analysis of Antennas for Hand-Held Transceivers using FDTD", IEEE Trans. Antennas and Propagation, Vol. 42, No. 8, pp. 1106-1113, August 1994.

[9] R. G. Yaccarino, Y. Rahmat-Samii and L. I. Williams, "The Bi-Polar Planar Near-filed Measurement Technique Part II: Near-field to Far-field Transformation and Holographic Imaging Methods", IEEE Trans. Antennas and Propagation, Vol. 42, No.2, pp. 196- 204,February 1994.

[10] L. I. Williams, Y. Rahmat-Samii and R. G. Yaccarino, "The Bi-Polar Planar Nearfiled Measurement Technique Part I: Implementation and Measurement Comparisons", IEEE Trans. Antennas and Propagation,

Vol. 42, No. 2, pp. 184-195, February 1994.

[11] D. W. Duan and Y. Rahmat-Samii, "Reflector antennas with superquadric aperture boundaries", IEEE Trans. Antennas Propagat., 41(8), August 1993.

[12] Y. Rahmat-Samii and A. N. Tulintseff, "Diffraction Analysis of Frequency Selective Reflector Antennas", IEEE Trans. Antennas and Propagation, Vol. 41, No. 4, pp. 476- 487, April 1993.

[13] J. P. McKay and Y. Rahmat-Samii, "An Array Feed Approach to Compact Range Reflector Design", IEEE Trans. Antennas and Propagation, Vol. 41, No. 4, pp. 448-457, April 1993.

[14] Y. Rahmat-Samii, "Advanced Synthesis and Analysis Techniques for Microwave Reflector Antennas", AEU, Vol. 46, No. 5, pp. 343-354, 1992.

[15] Y. Rahmat-Samii, D. W. Duan, D. Giri, and L. Libelo, "Canonical examples of reflector antennas for high-power microwave applications", IEEE Trans. Electromagn. Compat., 34(3):197-205, August 1992.

[16] D. W. Duan and Y. Rahmat-Samii, "A Generalized Three-Parameter (3-P) Aperture distribution for Antenna Applications", IEEE Trans. Antennas and Propagation, Vol. 40, No. 6, pp. 697-713, June, 1992.

[17] D. W. Duan, Y. Rahmat-Samii and J. P. Mahon, "Scattering from A Circular Disc: A Comparative Study of PTD and GTD Techniques", Proceedings of IEEE, vol. 79, no. 10, pp. 1472-1480, October, 1991.

[18] W. Imbriale, V. Galindo-Israel and Y. Rahmat-Samii, "On the Reflectivity of Complex Mesh Surfaces", IEEE Trans. Antennas and Propagation, vol. AP-39, no. 7, pp. 1352- 1365, September, 1991.

[19] D. W. Duan and Y. Rahmat-Samii, "Beam Squint Determination in Conic-Section Reflector Antennas with Circularly Polarized Feeds", IEEE Trans. Antennas and Propagation, vol. AP-39, no. 5, pp. 612-619, May, 1991.

[20] Y. Rahmat-Samii, "Array feeds for reflector surface distortion compensation: concepts and implementation", IEEE AP-S Magazine, 32(4):20-26, August 1990.

[21] W. V. T. Rusch, Jr. A. Prata, Y. Rahmat-Samii, and R. A. Shore, "Derivation and application of the equivalent paraboloid for classical offset Cassegrain and Gregorian antennas", IEEE Trans. Antennas Propagat., 38(8):1141-1149, August 1990.

[22] J. Huang, and Y. Rahmat-Samii, "Fan Beam Generated by a Linear-Array Fed Parabolic Reflector", IEEE Trans. Antennas and Propagation vol. AP-38, pp. 1046-1053, July, 1990.

[23] A. G. Roederer and Y. Rahmat-Samii, "Unfurlable satellite antennas: a review", Annales des Telecommunications, 44(9-10):475-488, November 1989.

[24] Y. Rahmat-Samii, "Effects of Deterministic Surface Distortions on Reflector Antenna Performance", Annales des Telecommunications, France, vol. 40, pp. 350-360 August 1985.

[25] Y. Rahmat-Samii, "Surface Diagnosis of Large Reflector Antennas Using Microwave Holographic Metrology - An Iterative Approach", Radio Science, vol. 19, pp. 1205- 1217, Sept.- Oct. 1984.

[26] Y. Rahmat-Samii, "An Efficient Computational Method for Characterizing the Effects of Random Surface Errors on the Average Power Pattern of Reflectors", IEEE Trans. Antennas and Propagation, vol. AP-31, pp. 92-98, January 1983.

[27] D. W. Duan and Y. Rahmat-Samii, "Reflector/feed(s) synthesis challenges for satellite contour beam applications", In IEEE AP-S Intl. Symp., pages 297-300, Chicago, Illinois, July 1992.

[28] R. A. Pearson, Y. Kalatidazeh, B. G. Driscoll, G. Y. Philippou, B. Claydon, and D. J. Brain, "Application of contoured beam shaped reflector antennas to mission requirements", In Intl. Conf. Antennas Propagat., pages 9-13, 1993.

[29] R. E. Freeland and G. Bilyeu, "IN-STEP inflatable antenna experiment", In 43rd Congress Intl. Astronautical Federation, 1992.

[30] S. von Hoerner, "The design of correcting secondary reflectors", IEEE Trans. Antennas Propagat., AP-24(3):336-340, May 1976.

[31] Kane Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", IEEE Trans. Antennas Propagat., 14(3):302-307, May 1966.

[32] R. Holland, "Finite-difference solution of Maxwell's equations in generalized nonorthogonal coordinates", IEEE Trans. Nuc. Sci., NS-30:4589-4591, 1983.

[33] G. Mur, "Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic field equation", IEEE Transactions on Electromagnetic Compatibility, EMC-23:377-382, 1981.

[34] B. Engquist and A. Majda, "Absorbing boundary conditions for the numerical simulation of waves", Math. Comp., 31:629-651, 1977.

[35] L. Chen, T. Uno, R. J. Luebbers, and S. Adachi, "Characteristics of a monopole antenna mounted on a conducting rectangular box", In Intl. Symp. Antennas Propagat., pages 709-712, Sapporo, Japan, September 1992; IEICE.

THE FUTURE OF DIGITAL TV AND HDTV BY SATELLITE

LEONARDO CHIARIGLIONE

1. Introduction

Many delivery media can be employed today to convey broadband information, such as required by audio-visual signals, from a source to the users. The Hertzian channel on terrestrial networks (VHF and UHF bands) has been the first to be used for delivery of television programs. This has been followed by cable, satellite, package media, etc.

The new possibilities brought about by the maturity of digital technologies are influencing the evolution of the listed existing delivery media but other broadband media that are possible candidates to carry broadband audio-visual information to the users are going to be influenced as well.

The purpose of this paper is to revisit the steps that have led to the current status of progress of digital technologies in support of audio-visual services, assess the impact on the current actions being taken to upgrade the capability to provide richer services on satellite, and to try and make some extrapolations on possible evolutions in the future.

2. How we got here

Audio-visual equipment, applications and services have come to play a very important role in the life of human beings today as the privileged means of communications, particularly of the more advanced forms of culture. They have developed over the years by utilising a variety of technologies in order to reach the intended user: printing, radio, cables, various forms of recording etc. These technologies all share the same property of being "analogue" i.e. of representing the information by means of some physical variable theoretically capable to assume a continuum of values.

Dr. Leonardo Chiariglione is with the
Multimedia and Video Services
Centro Studi e Laboratori Telecomunicazioni (CSELT)
Via G. Reiss Romoli 274,
I-10148 Torino, Italy
Tel: + 39 (11) 228 6120 Fax: + 39 (11) 228 6299

In the late '80s the investments made by the different technical and business communities had already convinced many of the laboratories that the available signal processing techniques could be extended to preserve the original quality of digital video signals after a compression factor of between 15 and 20 and to the audio (music) signals for about 1/3 of that value, while modulation techniques intended for the majority of analogue delivery media were capable of packing about 4 bit/s/Hz. That this conviction found it difficult to extend beyond the doors of laboratories is due to a number of factors that it would be too long to examine here.

The combination of signal compression and modulation had therefore the potential to provide a better exploitation of bandwidth compared to existing analogue delivery system. In fact, if the 5 MHz used by an analogue television signal is digitised a 20 Mbit/s channel is obtained. If a studio TV signal at about 166 Mbit/s is compressed, roughly 10 Mbit/s are obtained. Hence the straightforward application of digital technologies would lead to the possibility of doubling the number of television programs bringing the quality of the studio to the end user, without changing the bandwidth in use.

Slowly this awareness made it to companies' boards. In the late '80 the David Sarnoff Research Center then of the RCA had announced DVI (digital video interactive), a system to store video, audio and other data on a CD-ROM, Philips was working on FMV(full-motion video) for CD-I (compact disc interactive), a device with similar functionality, the ATV (Advanced Television) initiative of the FCC (Federal Communications Commission) in the USA was in the process of abandoning the original analogue approach to ATV for the terrestrial broadcasting network in favour of a digital solution, the European project EU 147 - DAB (digital audio broadcasting) has set out to define a new system for broadcasting of compact-disc quality music via radio, the then CCITT (now ITU-T) was well advanced in the definition of the px64 kbit/s video coding standard for visual telephony and the then CMTT (now ITUT SG 9) was struggling to find an agreement on the way to code television at 34/45 Mbit/s for contribution purposes.

MPEG was established amid these largely uncorrelated initiatives, some of them digital extensions of existing analogue systems, having in mind to respond two challenges: the first to find a mechanism to convince the different industries that there was a technology advantage in going digital together with the same solution and the second, more important, to define a single "syntax" capable to represent the audio-visual information in such a way that the common syntax could be the common platform enabling interoperability between applications.

The way in which an intrinsically analogue medium, as most of those deployed for analogue audio-visual services, is digitised also needs standardisation if equipment have to interoperate. There was, however, no attempt at development of modulation schemes looking at the broad picture as for MPEG and development took a different, more individualistic path. One reason is that delivery media have the widest spread of technical features which tend to obscure even obvious commonalties but the main reason is probably the traditional coupling of delivery media with the view of some business communities that a delivery media is their own preserve. Whatever the reason, the possibility to use the same or similar signal processing methods, of great advantage because the same or similar signal processing architecture could have been used in VLSI

implementations was lost. This meant that the reduction of development costs and the earlier attainment of a critical mass could not be achieved.

Multicarrier modulation is an example of how two apparently different delivery media, the Hertzian channel used for terrestrial broadcasting and the twisted pair used in the local telephone loop have lent themselves to two solutions based on the same approach. DMT (Discrete MultiTone), a technique being standardised by ANSI for ADSL (Asymmetric Digital Subscriber Line) uses a multiplicity of carriers to transport information. This is the same technique that is used by OFDM, that is being supported for DAB (Digital Audio Broadcasting) and DTTB (Digital Terrestrial Television Broadcasting) in Europe. The only difference is that DMT is capable of packing more bit/s/Hz thanks to a return channel, but the principle is very much the same. If DMT and OFDM had been developed in a co-ordinated way, we would be much more relaxed today in dealing with the task of converting big boxes of electronics into VLSI chips for the mass market.

3. How can existing technology be exploited for satellite services

The main result obtained by the application of up-to-date channel coding and modulation techniques to satellites is that the 8 MHz channel established by WARC, originally for satellite broadcasting of the conventional TV program, can now carry a bitrate of the order of 45 Mbit/s. To see how this can be exploited we need to know more about MPEG, and in particular the MPEG-2 technology. The following is a list of the most important features of the MPEG-2 standard:

1. the decoding process is specified not the encoding. This has a number of interesting consequences:

   a) the actual values of the (PCM) parameters are not relevant, only the input bitrate (into the decoder) and the output bitrate (from the decoder) are important. The way, say, spatial and temporal resolutions are traded in a specific video signal is not so relevant.

   b) the decoding process being described in an abstract way, there is no preferential bitrate. In general an increase in the bitrate yields a better quality and a decrease in the bitrate a lower quality, but the decision of which bitrate to choose is left to the users.

   c) there is no "MPEG-2 quality" but only "quality of an MPEG-2 encoder implementation".

2. MPEG2 is not just a "video coding" standard nor just an "audio coding" standard but an integrated audio-visual coding standard giving the possibility to a content provider or a service provider to assemble an arbitrary number of television programs with a particular bitrate and each composed of one or more video streams and of one or more (multichannel) audio streams accompanied by one or more data streams. It is even possible to exploit statistical multiplexing and have individual programs encoded at variable bitrates depending on the "activity" of each program but in such a way that the total bitrate does not exceed the assigned channel capacity (say 45 Mbit/s).

Given this level of flexibility let us examine some possible configurations of service offerings. Before doing that we have to make reference to the results obtained from a campaign of subjective tests carried out by MPEG to assess the quality of some encoder implementations.

For stereo (2 channel) audio, tests have been carried out on ten critical audio sequences and it has been verified that 256 kbit/s gives an average quality of 4.8 in the ITU-R 5-point scale with no coded sequence scoring less than 4.6. For conventional television (625/50 or 525/60 video), 6 Mbit/s gave a quality comparable to the composite studio signals (but 4 Mbit/s give a quality usually surpassing that of many signals received at home today and 1.5 Mbit/s a quality comparing favourably with that of a VHS cassette) and 9 Mbit/s a quality comparable to component studio signals. Results for HDTV can be extrapolated by multiplying the results by 4. Complete results for the multichannel audio are not available yet but the expected target is 384 kbit/s for a 5-channel audio signal.

One can then see the wealth of possible service offerings in a 45 Mbit/s stream: they stretch from a large number of programs with VHS quality (say 30) to a medium number of programs of PAL/NTSC quality (say 6-8) down to a maximum of two HDTV programs, the number of accompanying languages being left to the content of service provider in response to his/her expected audience. Ten 45 Mbit/s streams have the potential of offering a number of programs that can reach a few hundreds.

Until now we have treated all programs as independent entities, a natural thing to do at this point in time when we are leaving a world of independent suppliers of mono-program TV services. The multiplication of programs and an embedded navigation infrastructure provided by the MPEG-2 standard, however, create the conditions for a new dimension of broadcasting services.

The first observation is that there is only a small part that is specific to satellite while the major part can be extended to the use of other delivery media such as cable and terrestrial broadcasting. Before considering how broadcasting services can evolve it is appropriate to mention the role played by a new organisation called DAVIC.

DAVIC (The Audio-Visual Council) is an association established in Switzerland in 1994 with the purpose of favouring the success of emerging digital audio-visual applications and services, in the first instance of broadcast and interactive types, by the timely availability of internationally agreed specifications of open interfaces and protocols that maximise interoperability across countries and applications/services. DAVIC intends to adopt the MPEG approach of first looking at the broad picture, identifying the commonalties and then set out defining or selecting pieces of technology that have a broad application range. The expected advantages are the spreading of development costs and the increase of interoperability.

The DAVIC approach is not to standardise applications or services but to specify tools that permit the support of functionality that can be implemented in different parts of the system and which a system designer can pick up at will to build the service/application of his/her interest.

The implementation of a powerful navigation support gives the users the opportunity to enjoy an extension of services that goes beyond the sequential scanning of programs. This will make possible an electronic program guide giving information on which program can be found on what channel and Near Video on Demand services where the same program (say a movie) is broadcast on different channels at staggered times so that the user feels he/she only needs to wait a few minutes before the selected movie "starts". He/she can even be let understand that a "pause" function exists because the movie will "resume" by tuning the receiver to a different channel.

4. The expected next steps

There will be several ways in which the wide bandwidth offered by the satellite channel can be exploited to give further enhancements to the already promising new service scenarios.

The addition of a return channel is a basic requirement. Here again we are confronted with the need to achieve harmonisation, at the basic technology level and at the user interface level, in different application fields: satellite broadcasting, terrestrial broadcasting and CATV. This is again an area which is being addressed by the DAVIC initiative.

Depending on the bitrate that can be made available attractive new services can be implemented:

1. Pay per view, a service where the user is billed for the program he/she is actually seeing.
2. Some initial forms of teleshopping, where the user can decide to issue a purchase order for goods or services advertised in the downstream channel.
3. Download ability of applications, such as games.
4. Access to billing information, where the user can monitor the level of expenses accumulated so far with a particular service provider or all of them.

It goes without saying that once a bi-directional channel has been established there is no reason why it should not be used for a variety of other services that are not necessarily directly connected to broadcasting or audio-visual information, such as internet access. The initial deployments are likely to use the public switched telephone network (say by a 9.6 kbit/s modem) but higher bandwidths could become available when a return channel on the satellite itself will become possible in the near future.

PERSONAL COMMUNICATION SYSTEMS AND LOW EARTH ORBIT SATELLITES

SHUZO KATO

Abstract

To realize the ultimate goal of communications, communications with everyone at anywhere and anytime, a number of efforts have been made in various fields. The key issues are frequency utilization efficiency to accommodate as many customers as possible in the limited frequency bands and cost effective realization of nation-wide or world-wide coverage. These lead to the research and development of advanced system concepts, multiple access, modulation/demodulation, forward error correction, voice coding/decoding, channel assignment schemes and so on in addition to hardware implementation technologies which directly affect portability characteristics such as size, weight, talk time, mobility, service coverage and menus available. Cellular and satellite personal communications systems employ macro cells, from half a km to several km cell size for terrestrial cellular and at least several hundred km cell size for satellite personal communications, to realize high mobility. Since system capacity is proportional to $1/R^2$ (R: cell diameter), these systems employ low bit rate codecs to achieve high system capacity. On the other hand, personal communications systems based on digital cordless systems employ micro cells, a couple of hundreds meter in diameter and most of them have been employing a 32 kbit/s ADPCM codec. This is because micro cell systems can achieve larger system capacity and cordless phone systems must realize toll quality voice transmission for use at home or offices. Since personal communication systems are supposed to be used in a wide variety of situations and environments, a compromise must be reached between high quality voice transmission (a must for home use) and mobility (a must for mobile passengers), and integration must be carried out between terrestrial and satellite personal communications to offer cost effective mobile communication services, nation-widely and/or world-widely. Communications with everyone at anywhere and anytime by one common hand set will be seen in the third generation mobile communication era.

Dr. Shuzo Kato is with the
NTT Wireless Systems Laboratories
1-2356 Take Yokosuka 238-03
Japan
Tel: + 81 (468) 59 3470 *Fax: + 81 (468) 59 8022*

I. Introduction

There have been a number of studies and field tests on personal communication systems ranging from radio access technologies to network technologies. These have resulted in a number of different system proposals with different multiple access schemes, bit rates, modulation schemes, codecs, transmission power and so on since there is no absolute definition of personal communications systems. Among the various definitions on personal communications, the following concept is widely accepted [1].

Personal communications systems provide the capability to communicate with everyone, anywhere and anytime by featuring personal identification instead of terminal identification, person-based communications instead of location based communication, and personalized services to serve different customers' different requests.

This concept, communications with everyone, at anywhere and anytime is the ultimate goal of communications and translates to portability of services. If all services are portable, we can communicate with everyone at anywhere and anytime.

To realize this ultimate goal of communications, a number of efforts have been made in various fields. These are classified into network technologies and radio access technologies. The former is well represented by the intelligent network which provides the infrastructure needed to perform location registration, roaming, and routing to track down the called party anywhere, authentication and so on. The latter provides wireless access methods to provide last 100 m portability to customers. This comprises infrastructure, base stations and consumer products, the handset. Therefore, the key issues are frequency utilization efficiency to accommodate as many customers as possible in the limited frequency bands which lead to the research and development of advanced multiple access, modulation/demodulation, forward error correction, voice coding/decoding, channel assignment schemes and so on in addition to hardware implementation technologies which directly affect portability characteristics such as size, weight, talk time, mobility, service coverage and so on.

The important thing to remember is that when there is only one system providing communication capability, customers have no choice but to use it. However, this will not be valid in the personal communications era and only the best system will be chosen in terms of price, quality, talk time/standby time, size and weight of handset. In such circumstances, consumer products aspects rather than radio engineering aspects will dominate the success or failure of the system.

In the following chapters, the most important and often overlooked condition: customer requirements will be discussed by comparing the cordless phone and cellular phone approaches toward personal communications systems (PCS). A perspective of customer's requirements through PHS field tests in Japan will be given with survey data on acceptable tariff and what customers really want. Examples of advanced technologies will be described to meet these customers' requirements: high quality voice transmission, longer talk time, small and light weight handset.

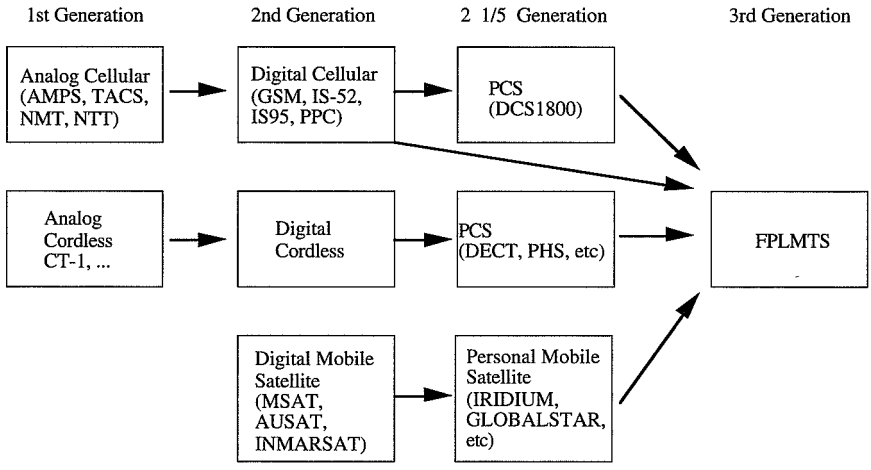


Fig. 1 - Mobile communication evolution

		JDC	ADC	GSM	NTT (Analog)		AMPS	NAMPS	CDMA (Qualcomm)
Frequency band (MHz)		800 & 1500 (New)	800	800 (New)	800		800		800
Carrier spacing (kHz)		25 Interleaving	30 Interleaving	200 Interleaving	25	12.5	30	15	1,250
Modulation		π/4 QPSK		GMSK	FM	N-FM	FM	N-FM	Forward : QPSK-SS Reverse : OQPSK-SS
Multiple access (Half rate codec)		3(6)ch TDMA		8(16)ch TDMA	FDMA		FDMA		CDMA
Carrier bit rate (kbps) [CODEC]		42 [VSELP 11.2 kbps]	48.6 [VSELP 13 kbps]	270 [RPE-LTP 22.8 kbps]	----		----		Forward: 19.2k (r=1/2FEC) Reverse: 28.8k (r=1/3FEC) [QCELP : 9.6kbps]
Equalizer		Option	Mandatory	Mandatory	Not required		Not required		Forward : 3finger RAKE Reverse : 4finger RAKE
Diversity		Option	Option	?					Spatial diversity : 2RX-4RX @base station Time diversity 20msec
Frequency hopping		----	----	Option	No		No		No
Spacial frequency utilization efficiency	Carrier spacing (kHz)	25	30	200	25		30		1,250 (60-64ch)
	kHz/ch (Full codec)	*2 33.5	*2 40	*3 100					*1 21 (with Voice activation)

Table. 1 - Major features of digital and analog cellular communication systems
**1 : 1/10 of AMPS *2 : Frequency reuse factor Analog FDMA/TDMA/CDMA : (1/7)/(1/4)/1*
**3 : as same as *2 (under estimation)*

2. Mobile communications evolution [1-5]

The evolution of mobile communications is shown in Fig. 1. Mobile communications started with analog cellular systems and analog cordless phone systems. The most important issue in analog cordless systems has been how to achieve high frequency utilization efficiency to meet traffic demands with limited frequency resources, and various analog systems have been developed in this direction as shown in Table 1. Examples include AMPS (Advanced Mobile Phone Systems) in north America, NMT (Nordic Mobile Telephone) in Europe and the NTT (Nippon Telegraph and Telephone) system in Japan. To meet higher traffic demands and advanced services, digital cellular systems were developed and have been used. Major digital cellular systems developed so far are GSM (Global Mobile Systems, developed in Europe), IS-52, IS-95 (both developed in the USA) and PDC (Personal Digital Communications, developed in Japan). Similarly, analog cordless phone systems were also digitalized and are being used. Moreover, digital mobile satellite communication systems have been developed to complement mobile communications where terrestrial mobile communications are not cost effective.

		PHS (Japan)	DECT (Europe)	DCS1800 (Europe)	CT-2 (England)	WACS (USA)
Frequency band (MHz)		1895-1918.1	1880-1900	1710-1785 1805-1880 (UK)	864.1-868.1	1850-2200
Access Scheme		TDMA-TDD	TDMA-TDD	TDMA-FDD	FDMA-TDD	TDMA-FDD
Channels/Carrier		4ch	12ch	8ch	1ch	8ch
Carrier spacing (kHz)		300	1728	200	100	300
Channel Bit Rate (kbps)		384	1152	270.833	72	400
Speech Codec		ADPCM (32kbps)	ADPCM (32kbps)	PRE-LTP (13kbps)	ADPCM (32kbps)	ADPCM (32kbps)
Modulation		$\pi/4$ QPSK	GMSK	GMSK	FSK	$\pi/4$ QPSK
Portable Transmit Power, Max/Ave		80mW/10mW	250mW/10mW	1W/125mW	10mW/5mW	200mW/20mW
Channel Control		Dedicated CC	Associated CC	?	Associated CC	Dedicated CC
Function	Termination Call	O	O	O	X	O
	Hand Over	O	O	O	X	O
	Access Network	Digital/Analog	Digital	Digital/Analog	Analog	Digital
Availability		1994	1992	1993	1989	1995?
Remarks		• Medium scale • Diversity at base station	• Too many channels for home use • Equalizer for outdoor use	• GSM air-interface	• Originating call only	• Two frequency band required • Switched antenna diversity in portable set

Table. 2 - Personal communication systems

	Iridium	Globalstar	Odyssey	Inmarsat P
Orbit inclination	90°	52°	56°	45°
Number of orbital planes	6	8	3	2
Altitude (km)	765	1,389	10,354	10,355
Number of satellites/plane	11	6	4	5+(1)
Number of satellites	66	48	12	10+(2)
Number of beams/satellite	48	6	37	121
(Total active beams)	(2,150)	(288)	(444)	(1,210)
Weight (in orbit) (kg)	700	232	1,180	1,200
Transponder	Regenerative	Non-regenerative	Non-regenerative	Non-regenerative
Life time (yrs)	5	7.5	10 - 15	10
Frequency bands				
User link (up) (MHz)	1610-1626.5	1610-1626.5	1610-1626.5	*1980-2010*
User link (down) (MHz)	1610-1626.5	2483.5-2500	2483.5-2500	*2170-2200*
Feeder link (up)	27.5-30GHz	6484-6841.5MHz	29.5-30GHz	*5100-5250MHz*
Feeder link (down)	18.8-20.2GHz	5158.5-5216MHz	19.7-20GHz	*6925-7075MHz*
Intersatellite link (GHz)	22.55-23.55	—	—	—
System capacity (voice (4.8kb/s) channels)	3,840/Sat	134,400/48 Sat	3,000/Sat	4,000-5,000/Sat
Multiple access technique	TDMA/FDMA	CDMA	CDMA	*TDMA*
Planned service				
Digital voice (kb/s)	4.8	2.4, 4.8, 9.6	4.8	*TBD*
Data (kb/s)	2.4	9.0	9.0	*TBD*
Radio determination	Yes	Yes	Yes	*TBD*
Compatibilty with cellular system	Yes (Dual mode terminal)	Yes (Dual mode terminal)	Yes (Dual mode terminal)	Yes (Dual mode terminal)
Handset outou power (mW)	390	250	250	TBD
Tariff (/minute)	$3	$0.23	<$1	$1-4
Planned operation	1998 -	1998 -	1997 -	1999 -

Table 3 - Low earth orbit satellite systems

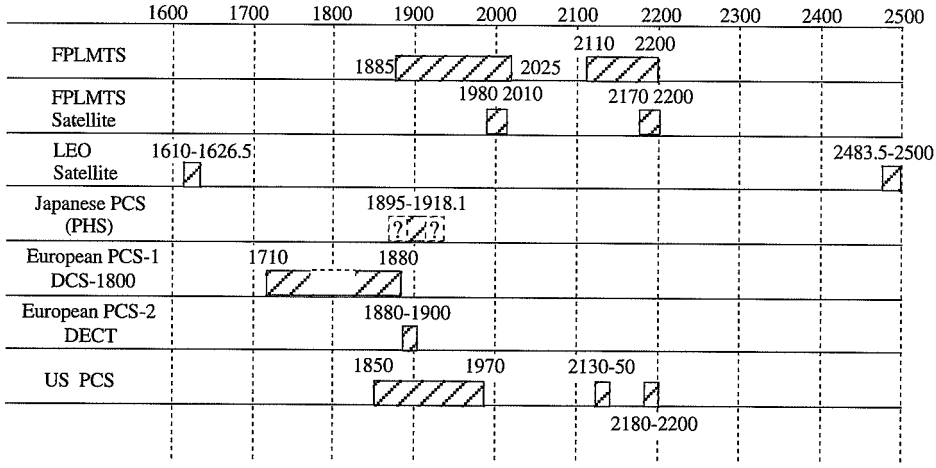


Table 4 - Frequency allocation in the 1400 - 2200 MHz band

Currently, the next (2.5) generation systems in the above mentioned three categories are being investigated, developed and tested. These are personal communication systems such as DCS 1800 based on GSM, DECT and PHS based on cordless phone systems and satellite personal communication systems such as IRIDIUM, GLOBALSTAR, Odyssey, Inmarsat P and so on as shown in Table 2 and 3. Current frequency allocations for terrestrial and satellite personal communications are listed in Table 4. As shown in this Table, the frequency bands for terrestrial and satellite personal communication systems are not much different except the 1.6 and 2. 5 GHz frequency bands for low earth orbit (LEO) satellite communication systems.

3. Terrestrial versus satellite personal communications

As shown in Table 5, cellular communications systems employ macro cells, from half a km to several km in diameter to realize high mobility for mobile passengers. Since system capacity is proportion to $1/R^2$ (R: cell diameter), these systems employ low bit rate codecs to achieve high system capacity. On the other hand, digital cordless systems employ micro cells, a couple of hundreds meter in diameter and most of them have been employing a 32 kbit/s ADPCM codec. This is because micro cell systems can achieve larger system capacity and cordless phone systems must realize toll quality voice transmission for use at home or in offices.

	Cell size	System capacity	Voice codec	Home use / Office	Unpopulated areas
Terrestrical PCS	Micro (100-500m in diameter)	Large	32kb/s ADPCM	◎	Δ
Cellular	Macro (-2km)	Medium	11-22 kb/s	○	Δ
Satellite PCS	Extra macro (-500km)	Small	4.8 kb/s	Δ	◎

Table 5 - Major features of PCS and mobile communications

One disadvantage of smaller cell size is the limited mobility management for mobile passengers. The other disadvantage is the huge number of base stations required for covering service areas. This will not cause any severe problem in urban and populated areas but may cause a big problem in rural and un-populated areas in that full coverage is not economically possible with only terrestrial personal communication systems. Thus, terrestrial personal communication systems can achieve high system capacity and give a cost effective coverage in populated areas but not in un-populated areas. Also the high capacity characteristics of personal communication systems allow us to employ higher bit rate voice codecs to offer toll quality voice transmission capability at home and offices.

On the other hand, satellite personal communication systems can provide global coverage but offer a very small capacity which can not meet all traffic demands. Although the voice codec bit rates for satellite personal communication systems are lower than current digital cellular communications as shown in Table 5, the total system capacity is very low due to the limited bandwidth and frequency reuse is very low compared with those of personal or cellular communication systems. Although the voice codec bit rates might not be high enough for home or office use, they will be good enough for outdoor, especially rural area communications where no other mobile communications can provide communications capability.

Since personal communication systems are supposed to be used in a wide variety of situations and environments, a compromise must be reached between high quality voice transmission (a must for home use), mobility (a must for mobile passengers) and better geographical coverage. Otherwise, the developed systems will be used by a few people and will not be used as it should be.

4. Two approaches to PCS [1]

Although the objective of communications is to offer the best services at the best price, customers consider the price, tariff, utilities and functions available but do not worry about system capacity, or frequency utilization efficiency which are often the most important research targets of researchers and engineers. If there is no alternative to use like life-line communication, whatever system available will be used. However, in the competitive environments which most personal communication systems will face, it is the customers' choice that determines system suitability.

Based on this basic and important fact, two approaches to PCS, cordless phone approach and cellular phone approach are compared. The condition which both systems must satisfy is to provide communication through an open-air-interface in the home, at the office and outdoors.

The cordless phone approach will be easier to deploy since a huge market can be seen for this approach by replacing currently popular analog cordless phones by digital cordless phones provided voice transmission quality is as high as toll quality and handset price is comparable to that of the analog one. Extended usage outside the house or office will give more incentives to people to buy the digital cordless phones. The drawback of this approach is limited mobility, up to walking speed for use.

The cellular phone approach provides high mobility as is well understood: however, there is a lower limit in cell size to offer this high mobility. This lower limit in cell size limits the system capacity since the traffic which can be handled is proportional to $1/R^2$. To improve system capacity, usually low bit rate voice codecs are used and unfortunately, this leads to lower than toll quality voice signal transmission. The point of this approach might be whether this voice transmission quality is acceptable in the home or office.

5. Perspective view of PCS through PHS field tests [1, 6]

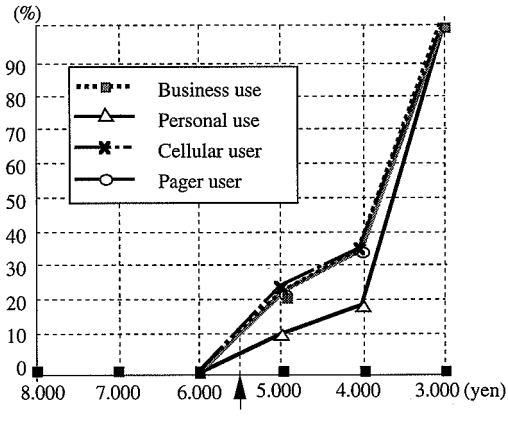


Fig. 2 - Number of people to subscribe vs. monthly charge

To see what customers really want from PCS, field tests (Phase 1) on Japanese PCS, named as PHS (Personal Handy Phone Systems) were carried out in the northern part of Japan from 1 October 1993 to March 1994 and Phase 2 PHS field tests lasted in Tokyo from last April to the end of last September. PHS employs the time division multiple access - time division duplex (TDMA-TDD), $\pi/4$ QPSK schemes and transmits up to 4 channels per carrier at a bit rate of 384 kbit/s over a 300 kHz channel. As the voice codec, a 32 kbit/s ADPCM codec is employed to realize toll quality voice transmission.

About 800 monitors tested PHS in Phase 1. Among them, the number of cellular phone users was 120 and that of pager users was 163. Major results obtained on charges are as follows:

- Acceptable monthly charge: about \$30 (100 % want to subscribe) (Fig. 2)
- Acceptable per-call charge: about \$25 (80-100 % want to subscribe)

As seen from Figure 3, survey results show that customers want higher quality, longer talk time, smaller size and lighter weight handsets.

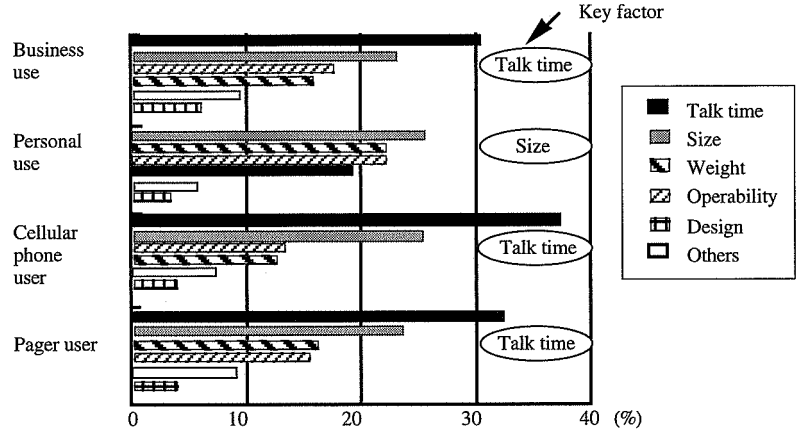


Fig. 3 - Important factors to design portable set

6. Requirements for PCS - High performance handset

For network operators, frequency utilization efficiency is a very important factor to develop personal or mobile communication systems to accommodate as many customers as possible. However, in competitive environments in the personal communications era, more important factors might be customers' satisfaction as summarized below:

(i) High-quality voice transmission

(ii) Better portability (longer talk/standby time, smaller in size and weight)

(iii) Wider coverage for seamless communications

(iv) Lower charge

To meet these requirements, a lot of effort has been taken. The following are two examples developed from this direction [7-9]. These are (1) coherent demodulation using a reverse modulation type carrier recovery scheme and (2) an error concealment technique. The proposed coherent demodulation scheme improves the frame error rate (FER) by about 3dB in terms of Eb/No — ratio of energy per bit to noise power spectral density — (Fig. 4) and the proposed error concealment technique improves voice transmission quality by 0.5 in mean opinion score (MOS) or equivalently 3 dB in terms of Eb/No. Thus, these schemes improve the receiver sensitivity by 6 dB to meet the customers requirement. To meet the other requirements, 1-chip CMOS LSIC implementation has been carried out. The developed LSIC operates at 2 volts and has the quite low power consumption of 70 mw. Moreover, a Lithium ion battery has been developed with an excellent performance of 72 Wh/kg. The handset developed by using these techniques can achieve a 10-hour talk time, and 150 hour standby time at a weight of only 120 g and a size of less than 120 cc.

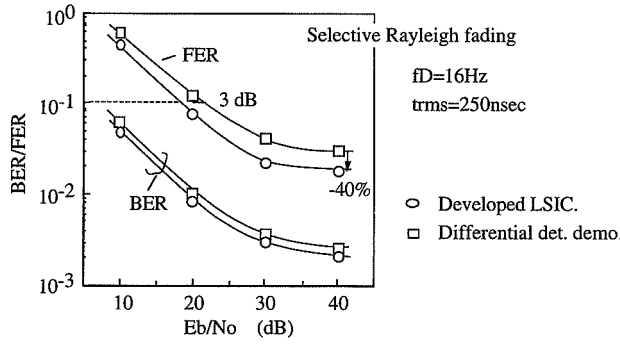


Fig. 4 - BER and FER (Rayleigh fading)

7. Technologies for Integration

- Commonalty -

A number of key technologies for terrestrial and satellite personal communications have some commonalty.

Linear modulation schemes such as quadrature phase shift keying (QPSK) and $\pi/4$ QPSK are popular for both systems to satisfy narrow band spectrum in nonlinear channels. Demodulation schemes for such systems were once considered to be differential detection to cope with the Rice or Rayleigh fading which characterize mobile satellite and personal communication channels, respectively. However, recent research results show coherent demodulation is possible in both Rice or Rayleigh fading environments resulting in significantly improved bit error rate performance [7], [10]. It is noteworthy that coherent demodulation was thought as complicated and power consuming but by employing advanced hardware and LSIC processing technologies, the penalty of better bit error rate performance is no longer conceivable as seen from the examples in the previous section. For data transmission, forward error correction (FEC) is popular for both systems, but FEC for voice signal transmission is not so popular due to slow fading and longer error bursts which degrade FEC performance quite a lot or need prohibitive interleaving length to obtain the full advantage of FEC [11].

CMOS LSI technology is promising for baseband signal processing for its high capacity and low power characteristics. The capacity of CMOS LSICs has been nearly doubled every second years and the operation voltage (V) has decreased to 3 - 2 volts. Current CMOS LSICs can accommodate up to 400 to 500k gate digital circuits or equivalent digital and analog circuits and CMOS LSICs with a quite low operation voltage of 2 volts are becoming commercially available. Since the power consumption of CMOS LSICs is proportional to V^2, digitalization of analog circuits and one-chip implementation of the baseband signal processing circuits with a low operation voltage CMOS LSIC is the obvious way to realize advanced portability of services for both systems.

RF components such as power amplifiers, synthesizers and mixers have been implemented in monolithic microwave (MM) ICs for both systems resulting in reducing size to about 1/5 every 5 to 7 year. This is very effective especially for low power applications such as PHS or DECT. MMIC implementation is also effective to reduce power consumption. GaAs MMIC s have been studied and developed for high power amplifiers and front end amplifiers for their high performances. We must accept high price and multiple supply voltages. Silicon MMICs have also been used for price sensitive but less performance-sensitive terminals. Currently, RF IC operation voltage is 3 volts for practical applications with reasonably high power-added-efficiency for power amplifiers.

- Difference -

As seen from the previous discussion, the voice codec rate is a key parameter for personal communication systems since the voice codec must provide toll quality at home and office use. On the other hand for the outside service, toll quality voice transmission is not necessary since there will be less competition and it can be considered as a kind of life-line communication.

8. Integration of terrestrial and satellite PCS

Integration of terrestrial and satellite personal communication systems is preferable for everybody: customers, operators and manufacturers. Integration of both systems is good for customers because we can provide seamless services nation-widely or globally. Since terrestrial personal communication systems are good for urban and densely populated area but not for rural areas and satellite personal communication systems can not supply high capacity to meet traffic demand in urban areas, both systems compliment each other and provide roaming capability across both systems. From the operators point of view, this integration expands service coverage cost effectively. Current cellular systems cover about 90 % of the population but their geagraphical coverage is only about 50 %. The coverage by personal communication systems is expected to be about the same (In some countries, population coverage realization of about 90 % in a fixed term is a condition for spectrum auction). Therefore, the integration of terrestrial and satellite personal communications systems is good for operators too.

- Common terminals -

There will be a number of different integration levels starting with two discrete terminals combined in one unit (lowest level) to just one set for both systems with roaming capability (highest level). As seen from the previous discussion, there is a lot of commonalty but some differences can not be resolved. A medium level integration of the terminals which employ the same RF components (satellite FPLMTS frequency bands are quite similar to that of terrestrial personal communications), the same CPU and man machine interface (MMI) components, which allow different modulation, multiple access, bit rates, bursts formats, is a good choice with upcoming technologies described in the previous section. The last components (baseband signal processing components except cpu and MMI components will consume only about 1/5 of the total power of the handset with advanced LSIC technologies, and having two baseband signal processing components increases the power consumption by a maximum of only 20 % (practically about 10 % by using common processors). This will allow customers to have high voice transmission quality at home or offices and still allow communication everywhere in the world with reasonably good voice transmission quality.

- Traffic sharing -

Since the capacity of satellite personal communications is limited and this system cannot be a major traffic carrier, satellite application areas must be carefully designed. Against this disadvantage, satellite personal communications can provide world-wide/global coverage very easily, which can not be met by terrestrial personal communications. Combining these two complementary characteristics of terrestrial and satellite personal communications systems, large system capacity and local coverage suitability, low capacity but wide coverage suitability, the idea of traffic sharing can be applied [12-13]. The integration of terrestrial and satellite communication networks in traffic sharing has been in commercial use in Japan since 1988 [12-13]. The traffic is shared in this system, as heavy traffic is carried by terrestrial digital networks and sparse and geographically distributed traffic is carried very efficiently by demand assignment time division

multiple access (TDMA) satellite communication systems. The system takes the advantage of global coverage and on demand basis channel assignment capability of TDMA satellite communication systems to complement inflexible channel assignment of terrestrial digital networks which can achieve high efficiency in carrying heavy and steady traffic. By carrying the overflow peak traffic of terrestrial digital networks intentionally caused by reducing the number of channels required for the terrestrial network to reduce installation cost, satellite channels can be used very efficiently while the total installation cost is reduced but retaining the same blocking rate for the customers.

The same idea can be applied to terrestrial and satellite personal communication systems: they complement each other except for the difference in voice codec bit rates in both systems. In traffic sharing, heavy traffic in urban areas will be carried by terrestrial personal communication systems while sparse traffic will be carried by satellite personal communication systems. The voice codec bit rate difference will not be a problem as far as both systems are supported by digital networks operating on 64 kbit/s PCM voice codecs since there must be codec translators between personal communications network and supporting digital networks such as ISDN. On the other hand, the combined system can maximize the advantages of both systems: good voice quality at homes and high capacity in populated areas, and cost effective communication ways in un-populated areas.

System integration will benefit customers by providing seamless communications at reasonable price. Moreover, network operators will benefit by not having to install terrestrial networks to meet excessive coverage requirements. Furthermore, manufacturers will benefit from this approach by producing a larger number of "common use" handsets, which will lead to lower costs. For satellite operators and handset manufacturers, the integration might be vital since the market for satellite personal communications handsets will be smaller than that of terrestrial personal communications.

Conclusion

This paper has reviewed current terrestrial and satellite personal communication systems and discussed the ideal integration of both systems. One of the important issues is what customers want and it is shown that high quality voice transmission is the first priority in PCS in its nature and also confirmed by field tests carried out on PHS in Japan. These are summarized as (i) High-quality voice transmission, (ii) Better portability (longer talk/standby time, reduced size and weight), (iii) Wider coverage for seamless communications, (iv) Lower charge. From the technological view point, the advanced technologies such as coherent demodulation can be applied to both systems to improve performances, and there will not be not much more components required for a handest (less than 10 - 20 % difference). That is to say, dual mode terminals for both terrestrial and satellite PCSs can be easily developed. The integration of both systems by traffic sharing way has been discussed to offer high voice transmission quality at home or office, and offers cost effective wide coverage for customers by complementing each other. This integration of both systems with one portable unit will be essential for both systems to glow and for customers and manufacturers.

References

[1] S. Kato: "Systems Engineering for Portability of Services" Proc. PIMRC'94, 1994

[2] J. Hutcheson: "Network Flexibility of the IRIDIUM Global Mobile Satellite System" Proc. of MDMC'94, 1994

[3] D. Rouffet: "Globalstar" Panel on Role of Mobile Satellite Communications in Personal Communications Era, IZS in Zurich, March 1994

[4] C. J. Spitzer: "Odyssey Personal Communications Satellite System" Proc. of IMSC'93, 1993

[5] R. Phillips: "Inmarsat's Global Mobile Satellite Services: Today and Tomorrow" Seminar on Trends in Satsom Access & Orbital Technologies at ALT in India, December 1994

[6] T. Kobayashi: "NTT's Personal Handy-phone System (PHS) Fields Service Tests" panel discussion, PIMRC'94

[7] Y. Matsumoto, M. Morikura and S. Kato: "$\pi/4$ shift QPSK Coherent Detection Demodulator for TDMA-TDD Systems " Proc. of the IEEE Vehicular Technology Conf., May 1993

[8] K. Enomoto, A. Dobashi, M. Suzuki, T. Hasumi, K. Seki, S. Kubota and S. Kato: "A Very Low Power Consumption ADPCM Voice Codec LSIC for Personal Communication Systems" Proc. of PIMRC'94, 1994

[9] S. Kato, S. Kubota, K. Seki, K. Enomoto, K. Kawazoe, T. Sakata, K. Kobayashi, and Y. Matsumoto: "A very Low Power and High-quality Signal Transmission Baseband LSIC for Personal Communications-Advanced Signal Processing and One-chip Integration", Proc. of Globecom'94, 1994

[10] K. Kobayashi, T. Sakai, S. Kubota, M. Morikura and S. Kato; "A New Carrier Recovery Circuit for Land Mobile Satellite Communications" IEEE Journal on Selected Areas in Communication, Vol. 10 No.8, August 1992

[11] S. Honda, S. Kubota and S. Kato; "DSD (Double Soft Decision) Concatenated Forward Error Correction" IEEE Journal on Selected Areas in Communication, Vol. 10 No. 10, October 1992

[12] M.Ohnuki, N. Nakashima, M. Umehira and S. Kato: "A New Satellite Communication System Integrated into Public Switched Networks - DYANET", IEEE Journal on Selected Areas in Communication, Vol. 10 No. 2, February 1992

[13] S. Kato, M. Morikura, S. Kubota, K. Enomoto, H. Kazama and M. Umehira: "TDMA Satellite Communication Systems for ISDN Services" IEEE Journal on Selected Areas in Communication, Vol. 10 No. 2, February 1992

MOBILE COMMUNICATION SYSTEMS AND BIOLOGICAL EFFECTS ON THEIR USERS

MARIA A. STUCHLY

Introduction

Influences of electricity and electromagnetic fields on biological systems were observed as early as the 18th century (Galvani) and the 19th century (d'Arsonval). A rigorous inquiry started after the Second World War. The interest in this field has been to a large extent stimulated by the worker and public concerns and pressures regarding safety of proliferating technologies. The main effort has concentrated in two frequency ranges: power line frequencies (50-60 Hz), and radio frequencies (RF) including microwaves. Considerable progress has been made in understanding the interactions of RF fields with living systems [1-3]. This understanding and agreement among the majority of scientists and regulators have resulted in very similar or nearly the same recommendations regarding safe exposure levels in many national standards and international guidelines [4].

Engineering contributions in dosimetry have made it possible to compare the results of biological investigations performed on various animals and to extrapolate them to the human exposure. Dosimetry is defined as theoretical and experimental evaluation of induced electric fields in tissue [5,6]. The descriptive parameter that is widely used is the specific absorption rate (SAR), i.e., the rate of energy deposition in a unit mass of tissue. The highlight of the RF dosimetry has been establishing how the average SAR depends on the field frequency and the size of the exposed body. The use of the SAR for dosimetric purposes does not in itself presume or is limited to thermal effects. The only implication it carries is that the fields induced in the body are responsible for biological interactions rather than the external exposure field itself.

With the help of dosimetric modeling, thresholds of the SAR were established for various effect in animals [1-3]. Most effects appear to be associated with thermal load due to RF exposure and are characterized by thresholds. The main limitation of the data base is a lack of well-performed chronic studies, and the paucity of data for ELF (extremely low frequency) amplitude modulated fields.

More recently, concerns have been expressed about cellular telephones and other personal communications services (PCS). There are two main issues. The first relates to the deposition of

Prof. Maria A. Stuchly is with the
Department of Electrical and Computer Engineering,
University of Victoria, Box # 3055, MS #8610
Victoria, British Columbia, V8W 3P6
Canada
Tel: + 1 (604) 721 6029 Fax: + 1 (604) 721 6052

energy (SAR) in the head. In some cases up to 50% of the device output power may be deposited in the user's head. The second issue relates to new technologies which improve the efficiency and channel utilization through digital modulations. This results in the amplitude (pulse) modulation at extremely low frequencies (ELF) of microwave signals. Because the same or similar bio-effects have been observed at RF amplitude modulated at ELF as at exposures to ELF alone, there is an uncertainty regarding health risks. The issues briefly mentioned above are discussed in more detail in this paper.

Physical Interactions

Interactions of radio frequency (RF) and microwave fields with biological tissues and bodies are complex functions of numerous parameters. Dosimetry quantifies these interactions. In free space the fields are characterized by the frequency strength of the electric and magnetic fields, their direction and polarization. The fields induced inside the body depend on the electrical properties of various tissues. The magnetic permeability ($\hat{\mu}$) of biological materials is equal to that of vacuum and therefore, can be neglected in considerations of the induced fields. The dielectric permittivity [$\hat{\varepsilon} = \varepsilon_o(\varepsilon' - j\varepsilon'')$ where ε_0 is the permittivity of vacuum] of biological tissue depends on the type of water content tissue (e.g., fat, muscle), temperature and frequency. The real part of the relative permittivity (ε') is called dielectric constant. The imaginary part of the relative permittivity (ε'') is related to the conductivity (σ) in the following way:

$$\sigma = 2\pi f \varepsilon_o \varepsilon'' \tag{1}$$

where: f is the frequency. The values of both the dielectric constant (ε') and loss factor (ε'') or conductivity (σ) vary substantially in the frequency and tissue type [7].

The dielectric permittivity and frequency also determine how far the electromagnetic wave penetrates into the body. The penetration depth varies from a small fraction of a millimeter at the upper frequencies of microwave range (above 100 GHz), to a few centimeters for high water content tissue at frequencies of a few GHz and for low water content tissue at frequencies of a few tens of GHz to over 1 m for low water content tissue at 10 MHz.

The quantification of the internal fields in complicated biological medium such as a human or animal body is, in general a difficult task due to the irregular shapes and heterogeneity of the material properties. However, only the fields inside the tissues and biological bodies can interact with them, and therefore it is necessary to determine these fields for any meaningful and general quantification of biological data obtained experimentally.

Fields inside biological bodies exposed to known external electromagnetic fields can be calculated by solving Maxwell's equations subject to given boundary conditions and taking into account the tissue electrical properties. The inhomogeneity of the dielectric properties and the complexity of the shape make a solution, or sometimes even a full formulation of the problem,

a difficult task. Only simplified models can be analyzed. The other approach is an experimental one, which is also subject to considerable limitations.

The intensity of the internal fields depends on the parameters of the external field, the frequency, strength and polarization, on the size, shape and the dielectric properties of the exposed body, on the spatial configuration between the exposure source and the exposed body, and on the presence of other objects in the vicinity. With a complex dependence on so many parameters, it is apparent that the internal fields in a mouse and a man exposed to the same external field can be dramatically different, and so will be their biological response, even not taking into account physiological differences. Conversely, different exposure conditions, i.e., at different frequencies, may induce similar fields inside such diverse shapes as a mouse and a man.

A dosimetric measure that has been widely adopted is the specific absorption rate (SAR) defined as "the time derivative of the incremental energy (dW) absorbed by, or dissipated in an incremental mass (dm) contained in a volume element (dV) of a given density (ρ)" [1].

$$\mathrm{SAR} = \frac{\mathrm{d}}{\mathrm{dt}}\left(\frac{\mathrm{dW}}{\mathrm{dm}}\right) = \frac{\mathrm{d}}{\mathrm{dt}}\left[\frac{\mathrm{sW}}{\rho(\mathrm{dV})}\right] \tag{2}$$

Using the Poynting vector theorem for sinusoidally varying electromagnetic fields, eq. (2) can be expressed as:

$$\mathrm{SAR} = \frac{\sigma}{2\rho}|\mathrm{E_i}|^2 = \frac{\omega\varepsilon_o\varepsilon''}{2\rho}|\mathrm{E_i}|^2 \tag{3}$$

where σ is the tissue conductivity in S/m, ε_o is the dielectric constant of free space $\left(\varepsilon_o = 8.85 \times 10^{-12}\ \mathrm{F/m}\right)$, ε'' is the loss factor, $\omega = 2\pi f$, f is the frequency in the Hz, $\mathrm{E_i}$ is the peak value of the internal electric fields in V/m. The SAR is expressed in watts per kilogram (W/kg) or their derivatives (i.e., mW/g). The average SAR is defined as a ratio of the total power absorbed in the exposed body to its mass. The local SAR refers to the value within a defined unit volume or unit mass, which can be arbitrarily small.

The SAR can be considered both as thermal and non-thermal dosimetric measure. For athermal biological effects that depend on the electric field strength, the SAR can be considered as an appropriate measure. The rate of temperature increase is also directly proportional to SAR, and equal to:

$$\frac{\mathrm{dT}}{\mathrm{dt}} = \frac{\mathrm{SAR}}{\mathrm{C}} \tag{4}$$

where T is the temperature, t is the time and C is the specific heat capacity.

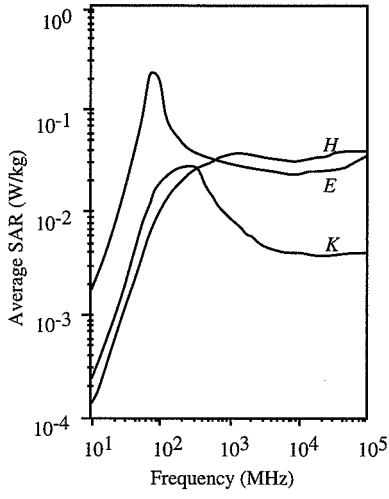


Fig. 1 - The average SAR in an average man (175 cm, 70 kg) exposed to a uniform plane wave of 1 m/Wcm^2. E polarization designates the electric field parallel to the long axis of the human body, H polarization designates the magnetic field parallel to the long axis, and K the propagation vector parallel to the long axis.

Considerable progress has been made in theoretical and experimental dosimetry. General review of theoretical and experimental methods and results are available [8,9]. The whole-body-average SAR for exposures in the far-field for a given body, is a function of frequency and polarization. Figure 1 illustrates a typical dependence, in this case for a model of an average man (175 cm, 70 kg) exposed in the far-field to 1 mW/cm^2. The E-polarization corresponds to the electric field parallel to the main body axis, H-polarization to the magnetic field parallel to the main axis of the body, and the k-polarization to the wave propagation from head-to-toe. The maximum absorption occurs at about 70 - 80 MHz for the E-polarization. This frequency is referred to as the resonant frequency for man. It should be noted that at this frequency, the power absorbed is a few times greater than that obtained by multiplying the surface area of the body cross-section by the incident power density. For a man standing in contact with RF ground, the resonant frequency shifts to about 30-40 MHz, and the SAR increases by about a factor of two. The average SAR depends also on the size and shape of the body, as illustrated in Fig. 2, which shows the SAR curves for three species for the E-polarization at 1 mW/cm^2 [8,9].

The spatial distribution of the SAR is highly non-uniform. For man at frequencies above 5 GHz the energy is deposited close to the body surface. It is at frequencies between 30 to 300 MHz, that high local SARs occur in the head and torso. At frequencies below 50 MHz for a man in contact with RF ground, very high local SARs are produced in the ankles [3]. The non-uniformities are particularly pronounced in the near-field [11,12]. Figures 3 and 4 show local SARs in the neck and the lung of a man exposed in the far-field in the E-polarization [11]. In the near-field, even greater non-uniformities in the SAR are produced[10,12].

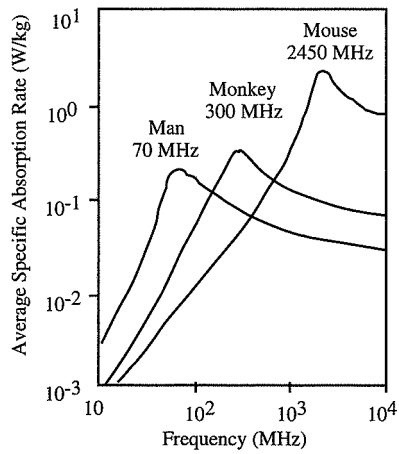


Fig. 2 - The average SAR in three species of mammals exposed to a uniform plane wave of the E polarization and a power density of 1 mW/cm^2. Approximate resonant frequencies are indicated.

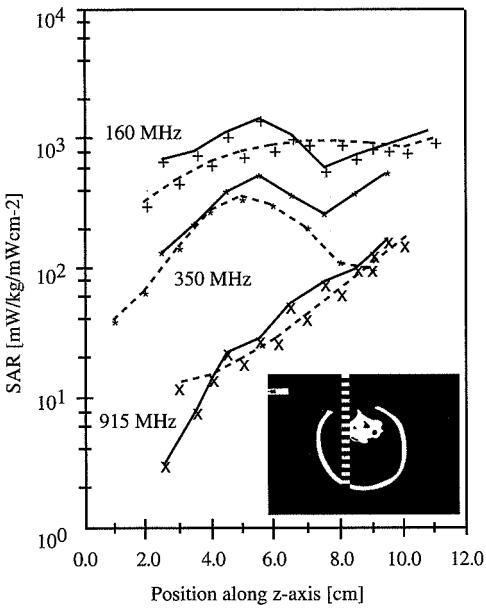


Fig. 3 - Experimentally determined local SARs along the path in the neck as marked on the insert. Exposure to 1 mW/cm², E polarized uniform plane wave.
***Solid lines** - heterogeneous model of man.*
***Dashed lines** - homogeneous model of man.*

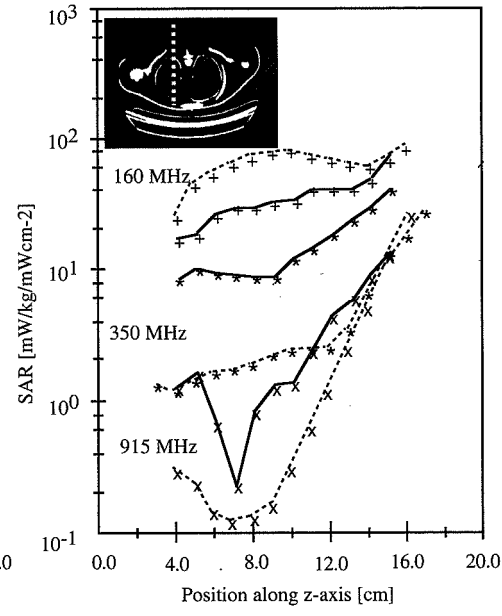


Fig. 4 - Experimentally determined local SARs along the path in the lung as marked in the insert. Exposure to 1 mW/cm², E polarized uniform plan wave.
***Solid lines** - heterogeneous model of man.*
***Dashed lines** - homogeneous model of man.*

Biological Effects

The dosimetric data on average SARs, i.e., as shown in Figures 1 and 2, are useful in extrapolating observed effects in animals to those expected in man. For example, if a mouse and a man are exposed at 1 GHz, the average SAR in the mouse is about ten times greater than that in man. Hence, if a biological effect is observed in a mouse at a given power density, a greater power density would be expected to be necessary to cause the same effect in man. Such extrapolation, however, does not take into account differences in spatial distribution of the SAR illustrated in Figures 3 and 4.

Biological effects of radio and microwave fields have been extensively investigated as documented in several reviews [1-3,13,14,15]. With the help of dosimetric modeling, thresholds of the average specific absorption rate (SAR) or rates of energy deposition in tissue in watts per kilogram have been established for various effect in animals. Some examples follow, however a comprehensive review with a full statistical analysis is not attempted here. Data quoted is from the published reviews. Radio frequency radiation is teratogenic in rodents at SARs that approach lethal levels, and a threshold for induction of birth defects is associated with the maternal core temperature of 41-42°C [1,13,15]. Chronic exposure of rats during gestation at 2.5 W/kg was reported to result in lowered fetal body weight at weaning [1]. Temporary sterility in male rats occurred at a SAR

of 5.6 W/kg, which produced a core temperature of 41°C [1]. Decreases in operant and learned-behavioral responses occur when SAR = 2.5 W/kg in the rat and at 5.0 W/kg in the rhesus monkey. Some types of behavior are affected when SARs are approximately 25-50% of the resting metabolic rate. These behavioral changes are reversible with time, after exposure ceases. Changes in the endocrine system and blood chemistry occur when SARs are greater than 1 W/kg and changes in hematologic and immunologic systems occur when SARs are equal or greater than 0.5 W/kg for prolonged exposures and appear to be associated with thermal stress[1,15]. One group of researchers reported that chronic exposure of SARs at 2-3 W/kg resulted in cancer promotion or co-carcinogenesis in mice [16]. The effect was similar to that caused by chronic stress. Neurons in the central nervous system were altered by chronic exposure at 2 W/kg [1]. All these effects appear to be associated with thermal load due to RF exposure and are characterized by thresholds [1-3, 13-15].

In summary, for whole-body exposures to RF/MW radiation, the experimental data currently available strongly suggests that biological effects in mammals occur when the average SARs are between 1 to 4 W/kg. This database is well established and consistent. This fact has been reflected in more recent exposure standards [4]. Each of these documents contain a reference either in the text or in the accompanying publication to the same database [1-3, 13-15]. The main limitation of the data base is a lack of well-performed chronic studies, and the paucity of data for ELF amplitude modulated fields. It needs to be emphasized that in the case of cellular telephones and other PCS devices, exposure is localized, and whole-body average SARs are not directly applicable.

Current Issues and Concerns

Wireless mobile communica-tions presents new challenges in terms of health effects. There are two main issues. The first is microwave power from a cellular telephone. The second issue relates to new technologies which improve the efficiency and channel utilization through digital modulations (e.g., time division multiple access - TDMA, or code division multiple access - CDMA), and use of frequency shift-keying (FSK) or phase shift keying (PSK). This results in the amplitude (pulse) modulation at extremely low frequencies (ELF) of the microwave signal. Some of the same or similar bio-effects have been observed at RF, however, there is an uncertainty regarding health risks of amplitude modulated at ELF as at exposures to ELF.

Local Energy Deposition

An evaluation of spatial distribution of SAR from cellular telephones and other mobile RF transmitters presents challenging problems both theoretical and experimental. The main difficulties are due to geometrical complexities and electrical heterogeneity of biological tissue. Several theoretical and experimental investigations of the SAR distribution in various models of the human body have been conducted [10,17-25]. Not until the 1990's have numerical techniques been developed that provide satisfactory treatment of the problem. Earlier experimental

investigations have established several important features of the energy absorption in the head of a user of a portable transmitter. The early investigations using a full-scale model of human body with heterogeneous electrical properties representing various tissues and resonant antennas indicated that:

(i) there are significant (an order of magnitude) differences in SARs in various locations between heterogeneous and homogeneous models having the same shape,

(ii) the maximum SAR is produced at the surface of the model (across from the feedpoint), and the SAR close to the surface of the model decreases exponentially with distance in the direction perpendicular to the surface, and

(iii) most of the energy is deposited in about 20% of the total body volume closest to the antenna feed point [10].

Another experimental evaluation was also performed. Notably two portable transceivers were assessed, one operating in the range of 810-820 MHz and the other operating at 850-860 MHz [20]. A heterogeneous model of the head was used with simulated skull, brain, eye, and muscle tissues. Similarly, as in [10], a maximum of the electric field on the surface and decay with distance were observed. The estimated maximum SARs on the surface of the eye were 3.2 and 1.1 W/kg per 1 W of the input power for the 810-820 MHz transmitter and 850-860 MHz transmitter, respectively. These values decreased to 1.1 and 0.7 W/kg per 1 W when the transmitter was moved 2.5 cm away from the head.

With the recent advances in the finite-difference time-domain (FDTD) technique and in computers (memory and speed), a viable tool has become available for analysis of SAR in the head due to various cellular telephones. Several groups are actively pursuing research in this area. There have been numerous reports at recent conferences. Results of some of these investigations have already been published or are in press [17-19, 21-25]. There are a few more reports accepted and submitted for publication that cannot be quoted at this time. A general picture that emerges indicates that between 10 - 50 % of the transmitter output power is absorbed in the model of a human head. The low absorption has been typically reported for antenna separations from the head of the order of 4 cm. About 50 % of even more of the antenna output power is absorbed in the head for separations of the order of 1 cm. Peak SARs reported range from 2 to 8 W/kg per 1 W of output power. Lower SARs are produced within bigger volumes of the tissue, but average SARs of over 3 W/kg in the eye have been predicted by some researchers.

There appears to be considerable confusion and disagreements with respect to the interpretation of the reported results of numerical calculations and measurements. Frequently, the data that are actually in a nearly perfect agreement is interpreted in vastly diverse ways. There are at least some apparent reasons for this phenomenon. One problem, likely a serious one, results from not sufficient attention paid to the effect of various telephone configurations. Antenna shape and feedbox, distance from and location, with respect to the head model, are a few of the most important parameters that affect the SAR in the head. Other physical differences relate to various models of heads used, and the unspecified (or not evaluated) accuracy of the results reported. The other type of problems relate to the differences in interpretation of the actual results. While not

necessarily erroneous, averaging over larger tissue volumes and time averaging, and interpretation of the results with respect to the current exposure standards, tend to obscure the data and deter from their objective comparisons. Applying in unqualified manner conditions of the current exposure standards to the data for cellular telephones is only of a limited value. The current standards have been developed on the basis of body average SARs and therefore are not formulated, neither the best suited to deal with exposures resulting in a very low average of SARs. There is a definite need to establish agreed upon methods of numerical and experimental evaluation of cellular telephones. This would permit meaningful comparisons of various designs of telephones, as well as data obtained in various laboratories. Activities of this type have recently commenced among several European centers.

Another interesting development relates to a new design of antennas for cellular telephones, particularly diversity antennas. An example of such effort has recently been reported [17]. Power absorbed in the head and other parts of the human body is lost from the communications perspective. A human body also significantly alters the radiation pattern. Therefore, modeling of the effect of the user's body on performance of cellular telephones and the minimization of the power deposited in the body illustrate good engineering sense. At the same time, the SAR in the head is decreased with such designs, even though at present the medical evidence is lacking that would indicate any real harm. Only a limited amount of work on investigations of various antenna configurations from this point of view have been published [17,19]. However, research in this area appears to have been carried out at a few laboratories.

Biological Effects of Modulated RF Fields

Biological effects have been observed at RF and MW fields amplitude modulated at ELF at SAR levels below thresholds for effects for continuous waves [1,14,26]. Many of these effects are the same or similar to effects observed for ELF electric and magnetic fields. The observed effects are usually field frequency and intensity specific. They tend to occur within relatively narrow ranges of both field parameters, and are dependent on other physical and physiological characteristics of the exposed biological system. Many of these parameters have not been fully identified and characterized. The interaction mechanisms remain unknown. The scientific database is relatively limited in this area. However, the potential importance of these effects should be critically evaluated. The scientific evidence with respect to health effects of ELF fields, while inconclusive, is suggestive of possible detrimental effects. Until the recent developments in digital communication there were hardly any situations of human exposure to RF/MW fields deeply amplitude modulated at ELF. This situation is changing rather rapidly with expansion of wireless digital communication.

The last decade witnessed considerable scientific effort as well as public concern regarding potentially harmful effects of human exposure to ELF and more specifically the power line frequency fields. The main concern is due to a number of epidemiological studies that have shown associations between exposure to low magnetic flux densities (of the order of 0.2 - 0.3 μT) and rates of childhood leukemia [e.g. 27-29]. Similarly suggestive evidence between the rates of some cancers and occupational exposures have been provided by epidemiological studies of various cohorts of workers [e.g. 30,31]. Several recent reviews outline the results and evaluate the limitations of the studies [32-34]. The limitations of some studies range from surrogate measures

of exposure to very small numbers of cases in high exposure groups limiting the statistical power of the findings. While most of these studies suffer from at least some limitations, the weight of the evidence cannot be dismissed. The main problem is a limited amount of experimental data that could support the findings of the epidemiology. There are some experimental results that may at least partly support the proposition that ELF fields affect the development of cancer, but the overall picture remains ambiguous.

A few studies of carcinogenicity have been done on animals. Large scale studies are now underway. Cancer development is considered in experimental models as a multistage process involving initiation (a change in the cell genetic material - DNA), promotion and progression. Promotion is associated with repeated exposures to an agent which may also act as an initiator (e.g. X-rays) or only as a promoter (many chemicals). Promotion involves interactions at the cell membranes. Progression is the last stage involving rapid growth of tumor and metastasis. Since electromagnetic fields have not been found to cause genetic changes, most investigations have been concentrating on promotion and co-promotion. As co-promotion, one considers modification either through an accelerated rate of development or a greater incidence of cancer produced by two chemical agents, both an initiator and a promoter. Three studies, two in Sweden [35-36] and one in Canada [37], have shown that magnetic fields do not act as promoters. Another Canadian study has shown that a relatively strong magnetic field accelerated the rate of chemically developed tumors in mice, but not the final yield in terms of the number of tumors and the number of animals affected [38]. One Swedish study [35] did not show any co-promoting effect, and the other showed a slight inhibition of the rate of tumor growth [36]. Recent Swedish studies show that intermittent exposures may be effective in cancer promotion by relatively strong magnetic fields [39]. While there appears to be some biological interaction, at least in some cases, the actual effect magnetic fields have, if any, in cancer development remains undetermined. There has been a plethora of laboratory investigations of various isolated cells in culture. These studies have clearly shown that electromagnetic fields of low frequencies can interact with biological systems at moderately low intensities. Typical responses reported were an altered cell growth rate, decreased rate of cellular respiration, altered metabolism of carbohydrates, proteins and nucleic acids, changes in gene expression and genetic regulation of cell function, and altered hormonal responses. Recently, further observations were reported regarding effects on cellular transcription, charge on the cell surface, ATP and oxygen levels in the slime mold, growth, proliferation, and functional differentiation of cells. Several studies have reported changes in calcium efflux. Another important effect of ELF fields is suppression of melatonin production and alterations in circadian rhythms. Effects on melatonin have been shown both in-vitro and in animals [26,33]. Many of these in-vitro effects are related to cancer development, but are also associated with other regulatory processes.

The main problem in evaluations of biological effects of ELF fields is the complexity of observed interactions. Effects observed are not always proportional to the field strength or the magnitude of induced currents. Sometimes, they show "window" responses in the field frequency and amplitude, and cell physiological state. The critical parameters responsible for the effect are often not well defined nor are they understood.

Conclusions

Interactions of radio and microwave energy with biological systems have been investigated vigorously for nearly half a century. A lot has been learned and thresholds for harmful effects have been established. Development of RF dosimetry, a quantification tool, has provided means for extrapolation from animal experiments into assessment of risks to human health. By the late eighties, health protection standards in various countries had converged to nearly identical limits. This reflects their reliance on the same data base and the agreement in its interpretation.

Current proliferation of devices and new services in mobile communications have raised new concerns. These concerns mostly relate to the relatively large rates of energy deposition in some parts of the body, mostly the eye and the brain, and to the data transmission resulting in the amplitude modulation of RF signals at ELF. The energy deposition is being currently evaluated for various devices using modern computational techniques. This problem is likely to be resolved in the near future, resulting perhaps in specific guidelines for device use. It is interesting to note in this context, that antenna designs that are optimal from the communications point of view also deposit the smallest amounts of energy in the device user.

The question of ELF modulation is more difficult to answer. Independent of what is the final assessment of health impact of ELF fields, it does not directly translate to the same or even similar assessment for microwave fields modulated at ELF. There is no reason to suspect any specific harmful effects related to ELF amplitude modulation of radio or microwave fields if ELF studies indicate that the detrimental impact on human health is minimal or none. On the other hand, if ELF exposures result in risks to people, that would not translate into the same risks for exposures to microwaves modulated at ELF, but would provide motivation and rationale for research into the potential problem.

Acknowledgment

This work is supported by an industrial research chair program funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), B.C. Hydro and TransAlta Utilities.

References

[1] Environmental Health Criteria 137, "Electromagnetic fields (300 Hz to 300 GHz)" World Health Organization, ISBN 92 4 1571373, Geneva, 1993.

[2] J. A. Elder, A. Czerski, M. A. Stuchly, K. H. Mild and A. R. Sheppard, "Radiofrequency radiation", Nonionizing Radiation Protection, 2nd Edition, WHO Regional Publications, ISBN 92 890 11165, 1992.

[3] M. A. Stuchly, "Biological effects of radiofrequency fields" Nonionizing Radiators, Ed: M. H. Repacholi, International Radiation Protection Association, ISBN 0 7316 23703, 1988.

[4] M. A. Stuchly, "Proposed revision of the Canadian recommendations on radiofrequency-exposure protection", Health Physics, vol. 53, pp. 649-665, 1987.

[5] M. A. Stuchly and S. S. Stuchly, "Experimental radio and microwave dosimetry" CRC Handbook of Biological Effects of Electromagnetic Fields, Ed: C. Polk and E. Postow, CRC Press, Boca Raton, Fl., 1986.

[6] J. C. Lin, "Computer methods for field intensity predictions" CRC Handbook of Biological Effects of Electromagnetic Fields, Ed: C. Polk and E. Postow, CRC Press, Boca Raton, Fl., 1986.

[7] K. R. Foster and H. P. Schwan, "Dielectric properties of tissues" Handbook of Biological Effects of Electromagnetic Fields, Ed: C. Polk and E. Postow, CRC Press, Boca Raton, Fl., pp. 27-96, 1986.

[8] C. H. Durney, et. al., "Radiofrequency radiation dosimetry handbook 3rd. Edition", Report SAM-TR-80-32. USAF School of Aerospace Medicine, Brooks Airforce Base, TX., 1980.

[9] O. P. Gandhi, "State of knowledge for electromagnetic absorbed dose in man and animals", Proc. IEEE, vol. 28, pp. 24-32, 1980.

[10] M. A. Stuchly, A. Kraszewski, S. S. Stuchly, G. W. Hartsgrove and R. J. Spiegel, "RF energy deposition in a heterogeneous model of man: near-field exposures" IEEE Transactions on Biomedical Engineering, vol. 34, pp. 951-957, Dec. 1987.

[11] S. S. Stuchly, A. Kraszewski, M. A. Stuchly, G. W. Hartsgrove and R. J. Spiegel, "RF energy deposition in a heterogeneous model of man: far-field exposures" IEEE Transactions on Biomedical Engineering, vol. 34, pp. 944-950, Dec. 1987.

[12] M. A. Stuchly, A. Kraszewski and S. S. Stuchly, "Exposure of human models in the near and far-field: a comparison", IEEE Transactions on Biomedical Engineering, vol. 32, pp. 609-616, 1985.

[13] R. D. Saunders, C. I. Kowalczuk and Z. J. Sienkiewicz, "The biological effects of exposure to non-ionizing electromagnetic fields and radiation: III Radiofrequency and microwave radiation", National Radiological Protection Board, Chilton Didcot, NRPB-R241, 1991.

[14] J. H. Bernhardt, "Non-ionizing radiation safety: radiofrequency radiation, electric and magnetic fields", Phys. Med. Biol., vol. 37, pp. 807-844, 1992.

[15] M. A. Stuchly, "Proposed revision of the Canadian recommendations on radiofrequency-exposure protection", Health Phys., vol. 53, pp. 649-665, 1987.

[16] S. Szmigielski, A. Szudzinski, A. Pierraszek, B. Bielec, M. Janiak and J. K. Wrembel, "Accelerated development of spontaneous and benzopyrene-induced skin cancer in mice exposed to 2450 MHz microwave radiation", Bioelectromagnetics, vol. 3, pp. 179-191, 1982.

[17] U. A. Jensen and Y. Rahmat-Savin, "The electromagnetic interaction between biological tissue and antennas on a transceiver handset", Digest IEEE/APS Interim. Symp., vol. 1, pp. 367-370, 1994.

[18] J. Toftgard, S. Hornsleth and J. Bach Andersen, "Effects on portable antennas of the presence of a person", IEEE Transactions on Antennas and Propogation, vol. 41, pp. 739-746, 1993.

[19] R. Mumford, Q. Bolzano and T. Taga, "Land mobile systems II: Pagers, portable phones, and safety", Mobile Antenna Systems Handbook, Ed: K. Fujimoto and J. R. James, Artech House Inc., pp 175-273, 1994.

[20] R. F. Cleveland and T. W. Athey, "Specific absorption rate (SAR) in models of the human head exposed to hand-held UHF portable radios" Bioelectromagnetics, vol. 10, pp. 173-186, 1989.

[21] P. J. Dimbylow, "FDTD calculations of SAR for a dipole closely coupled to the head at 900 MHz and 1.9 GHz", Phys. Med. Biol., vol. 38, pp. 361-368, 1993.

[22] O. P. Gandhi, "Numerical methods for dosimetry: ELF to microwave frequencies", Radio Science, vol. 30, Jan/Feb 1995.

[23] L. Martens, J. DeMoerloose, C. DeWagter and D. DeZutter, "Calculation of the electromagnetic fields induced in the head of an operator of a cordless telephone", Radio Science, vol. 30, Jan/Feb 1995.

[24] N. Kuster and Q. Balzano, "Energy absorption mechanism by biological bodies in the near field of dipole antennas above 300 MHz", IEEE Transactions on Vehicular Tech., vol. 41, pp. 17-23, 1992.

[25] P. J. Dimbylow and S. M. Mann, "SAR calculations in an anatomically realistic model of the head for mobile communication transceivers at 900 MHz and 1.8 GHz", Phys. Med. Biol., vol. 39, pp. 1537-1553, 1994.

[26] M. A. Stuchly, "Applications of time-varying magnetic fields in medicine" Crit. Rev. Biomed. Eng., vol. 28, pp. 89-124, 1990.

[27] D. A. Savitz, H. Wachtel, F. A. Barnes, E. M. John and J. G. Tvrdik, "Case-control study of childhood cancer and exposure to 60 Hz magnetic fields", American Journal of Epidemiology, vol. 128, pp. 21-38, 1988.

[28] S. J. London, D. C. Thomas, J. D. Bowman, E. Sobel, T-C. Cheng and J. M. Peters, "Exposure to residential electric and magnetic fields and risk of childhood leukemia" American Journal of Epidemiology, vol. 134, pp. 923-937, 1991.

[29] M. Feychting and A. Ahlbom, "Magnetic fields and cancer in children residing near Swedish high-voltage power lines", American Journal of Epidemiology, vol. 138, pp. 467-481, 1993.

[30] G. Theriault, M. Goldberg, A. B. Miller, B. Armstrong, P. Guenel, J. Deadman, E. Imbernon, T. To, A. Chevalier, D. Cyr and C. Wall, "Cancer risks associated with occupational exposure to magnetic fields among electric utility workers in Ontario and Quebec, Canada and France: 1970-1989. American Journal of Epidemiology, vol. 139, pp. 550-572, 1994.

[31] G. M. Matanoski, E. A. Elliott, P. N. Breysse and M. C. Lynberg, "Leukemia in telephone lineman", American Journal of Epidemiology, vol. 137, pp. 609-619, 1993.

[32] D. Savitz, N. E. Pearce, C. Poole, "Methodogical issues in the epidemiology of electromagnetic fields and cancer", Epidemiology Rev., vol. 11, pp. 59-71, 1989.

[33] "Electromagnetic Fields and the Risk of Cancer", report of an Advisory Group on Non-ionizing Radiation, National Radiological Protection Board, U.K., vol. 3, 1992.

[34] G. Theriault, "Electromagnetic fields and cancer risks" Rev. Epidem. Sante Publ., vol. 40, pp. 555-562, 1992.

[35] A. Rannug, T. Ekstrom, J. Hansson Mild, B. Holmberg, I. Gimenez-Conti and T. J. Slaga, "A study on skin tumor formation in mice with 50 Hz magnetic field exposure" Carcinogenesis, vol. 14, pp. 573-578, 1993.

[36] A. Rannug, B. Holmberg, T. Ekstrom and K. Hansson Mild, "A rat liver foci study on coexposure with 50 Hz magnetic fields and known carcinogens", Bioelectromagnetics, vol. 14, pp. 17-27, 1993.

[37] J. R. N. McLean, M. A. Stuchly, R. E. J. Mitchel, D. Wilkinson, M. Goddard, D. W. Lecuyer, M. Schunk and E. Callary, "Cancer promotion in the mouse skin model by 60 Hz magnetic fields: II Tumor development and immune response", Bioelectromagnetics, vol. 12, 1991.

[38] M. A. Stuchly, J. R. N. McLean, R. Burnett, M. Goddard, D. W. Lecuyer and R. E. J. Michel, "Modifications of tumor promotion in the mouse skin by exposure to an alternating magnetic field", Cancer Letters, vol. 65, pp. 1-7, 1992.

[39] A. Rannug, B. Holmberg, T. Ekstrom, K. Hansson Mild, I. Gimenez-Conti and T. J. Slaga, Accelerated Paper "Intermittent 50 Hz magnetic field and skin tumor promotion in SENCAR mice", Carcinogenesis, vol. 15, pp. 153-157, 1994.

REAL-APERTURE RADAR (RAR) IMAGING FROM SPACE

A.I. KALMYKOV

Introduction

The 75th URSI anniversary symposium is a brilliant testimony of a unified research community in the world and it is also a showcase of its potentialities that are not affected by the frontiers dividing the present world. And it would be worth to say that these frontiers seem to be far less tangible especially when the Earth is observed by radar right from space. I have the feeling that the very fact that we are together at this symposium makes us realise that we have gained one of the primary URSI objectives: we all enjoy the breakthroughs of individual researchers and groups of experts and we are actually proud to have them readily available to most of the countries.

In my report I will inform you about one of the developments which are actually based on the results of long-standing research work that had been previously carried out by scientists of different countries. This development concerns the side-looking radar (SLR) which had been originally tested on the Cosmos-1500 satellite and later paved the way to the elaboration of techniques for observing oceans, seas, ices and glaciers from space-borne carriers.

Now let me dwell on some background facts concerning the SLR development.

1. Historical background

As far back as in the 2nd World War, when the radar observations of targets on sea helped to locate their reflections, "sea clutters" were identified. Any attempts to synthesise and interpret this phenomenon [1] over many years ended in a failure. One of the first steps to yield certain results in gaining a deeper insight into the "clutter" mechanism was the detection of selective decametric radiowave band scattering by the sea surface [2, 3]; the decametric waves were found to be scattered by the sea waves $\Lambda=\lambda/2$. And while the computations involving the use of the small perturbation method [4, 5] helped in principle to interpret the features of the decametric band scattering by the sea surface, adequate models for the microwave band did not exist for many

Prof. A.I. Kalmykov is with the
Remote Sensing Center
National Academy of Sciences and National Space Agency of Ukraine
12 Proscura street
Kharkov, Ukraine
Tel: + 7 (0572) 44 83 97 Fax: + 7 (0572) 44 10 12

years. In the phenomenological models [1, 6, 7] which were used to make in attempts to interpret individual scattering characteristics, the emphasis was put on energy characteristics. However, with these particular models it was impossible to describe spectral, polarization and other characteristics.

In late 50s and early 60s many countries were actively engaged in the studies on the process of microwave scattering by the sea surface. The primary aim of the investigations that were carried out at the Institute of Radiophysics and Electronics of the NAS of Ukraine (Kharkov) was based on the following attemption. Inasmuch as all the elements of the surface are in a state of motion, the sea-scattered signal should have a Doppler shift ΔF whose magnitude will make it possible to determine the velocity of the scatterers V_s . In terms of the dispersion law $V_s(\Lambda)$ the nature of scatterers can be specified. To ensure that this concept could become a reality the Institute of Radiophysical and Electronics developed in 1960-61 a multifrequency coherent radar for 8 mm, 3 cm, 50 cm, 1.5 m and 4 m waves [8] and as early as in 1962 the selective scattering of microwaves [9] was identified.

In the 60s efforts were intensified to work up the procedure for computing the characteristics of the SHF radiowave scattering by the sea surface. For the case of microwave scattering a complex computational method [10] evolved into a two-scale model [11]. The basic experimental dependencies were interpreted, namely the changes in scattering cross-section s with an incidence angle $\sigma(\theta)$ and the relationship between the spectral characteristics of signals and roughness parameters. It is precisely during this period that a series of investigations aimed at updating the model was carried out in the USA [12-16]. As a result of all those studies the basic properties of scattering were found to be described in the two-scale model. The basic scatterer is the ripple $\chi_o = 2k \sin\theta$ ($\chi_o = 2\pi/Q_o$..., $k = 2\pi/\Lambda$, θ = the incidence angle) where as the large waves give rise to a modulation of ripple-scattered signals; ripples manifest themselves in the spatial-frequency structure of scattered signals.

It is precisely this phenomenon that provided the basis for early papers on oceanography, namely,those devoted to the determination of sea roughness parameters (spectra, length, distributions, periods, heights of waves) from the characteristics of scattered signals. The oceanologists were very interested in the studies of the ripple. They found that the ripple spectral density is dependent upon the wind velocity [16-18] and that σ increases with wind [19]. At any rate, this gave a stimulus to make the first attempts at conducting radar observations from space.

2. The space radar era and the first"Cosmos-1500" satellites

The main source of information on the state and structure of the seasurface in the microwave observations is the gravitational-capillar ripple.The principal characteristics of the scattering cross-section are relatedto the spectral density of the scattering ripple χ_o

Space Radar Systems

Radar Type	Program Sattellite	Date Operation life	Wavelength (cm) Polarization	Swath (km)	Spatial resolution	Data processing
Scattero-	SEASAT	1978/3m	2.1 / VV, HH	500	50 km	onground
meter	ERS-1	1991/	6 / VV	500	50 km	onground
	SEASAT	1978/3m	23 / HH	100	25 m	onground
Synthetic	SIR-A SHUTTLE	1981/5d	23 / HH	54.6	37 m	onground
Aperture	SIR-B SHUTTLE	1984/7d	23 / HH	50	35 m	onground
Radar	COSMOS-1870	1987/24m	10 / VV	20	25 m	onground
(SAR)	ALMAZ	1991/	10 / VV	20	20 m	onground
	ERS-1	1992/	6 / VV	100	30 m	onground
	JERS-1	1992/	23 / HH	75	18 m	onground
	SHUTTLE	1994	3 / VV			onground
	SIR-C / X-SAR	1994	24 / VV, HH			onground
	COSMOS-1500	1983/33m	3.1 / VV	475	0.8 x 2.5 km	onborne
Real	COSMOS-1602	1984				
Aperture	COSMOS-1766	1986				
Radar	COSMOS-1869	1987				
	OKEAN-1	1988				
	OKEAN-2	1990				
	OKEAN-3	1991				
	OKEAN-4	1994				
	OKEAN-5	1995				

Perspective Maps - Multipurpose radar rystem (spaceborn variant)

Synthetic Aperture Radar		23 / VV, HH	3 x 100	5-15 m /	onground /
with Active Phazed Array				/ 50 m	/ onborne
Real Aperture Radar		3 / VV	2 x 700	0.5 x 1.5 km	onborne
Scatterometer		3 / VV	600	(5..25) km	onground

Maps - Multipurpose radar system on the aircraft-laboratory IL-18D

Real Aperture Radar		3 / VV	2 x 45	20..50 m	onborne
Synthetic Aperture Radar		23 / VV, HH	30	2.5..25 m /	onground /
with Active Phazed Array			(120).	/ 25..50 m	/ onborne
Real Aperture Radar		0.8 / VV, HH	15	25..50 m	onborne
Synthetic Aperture Radar		180 / VV, HH	45	25 m /	onground /
with Phazed Array				/ 50..100 m	/ onborne

Table 1

$$\sigma_o \sim \alpha(\chi_o)$$

In its turn, the ripple spectraldensity is related to the speed of the near-sea surface wind W and depends on currents, surface-active films, etc. It is precisely these points that have led to the well-substantiated design of many space-borne radar systems.(Table 1).

The late 70s saw the emergence of initial space programs which envisaged the use of radar systems as the prime source to draw on necessary information. The first attempt was to develop a radar-radiometer for the"Skylab" space station.

The first practical results were obtained following the launch of the"Seasat" oceanographic satellite [21]. The SAR and a scatterometer that were originally tested in space made it possible to derive unique information which was subsequently processed and extensively used by researchers. I would not like dwell on all the peculiar features of the "Seasat" system. Their merits have received general acceptance.

When a space-borne radar was being designed in the Soviet period we allowed for the "Seasat" experience and adopted a different approach. Following the evaluated model of selective scattering, we decided on the radar parameters keeping in mind the peculiarities of radiowave scattering by the sea surface and sea ice. Initially the selection of the band $\lambda = 3$ cm resulted from the capability of observing the seasurface wind and inhomogeneities of the sea surface; it is also influenced by factors such as high sensitivity to variations in age gradation of ice, etc. A 1-3 km spatial resolution was achieved over this band using an antenna with a diameter of about 10 m. A crucial point was the use of a 50 km swath for the observations of basic mesoscale processes occurring on the ocean surface and the sea ice.

Since the processes to be studied are highly dynamic, the usefulness of data is directly dependent on how they are efficiently derived and employed. Real-time, operational observations of the Earth covers are,above all, related to all weather facilities; the band $\lambda \geq 3$ cm is particularly appropriate for this purpose. Another essential consideration in effectively gathering information by users is real-time, on-board data processing and transmission of results in a form virtually suitable for a wide network of users. This objective was achieved by forming radar data via APT (which was widely used in theoptical data transmission) and transmitting formed images in the 137 MHz wave band. The reception points for this band were more than 500.

Now let us consider some technical aspects. A side-looking radar (the radarbeam is oriented orthogonally to the satellite velocity vector) provides for sensing in a sector of incidence angles $\theta = 20° \ldots 52°$. The signal range selection is ensured by the pulse operating mode; the pulse duration was 3 μs. The transmitter used a magnetron with a radiated power (in a pulse mode) of 100 kW. The receiver was equipped with a temporal automatic gain control that equalizes the average signal level over an entire band, which is so essential for image users.

We had some serious problems with the development of the antenna system. The length of the antenna had to be approximately larger than 12 m and the antenna had to be installed on the wing

of the satellite, which was 3 m in size. In other words, the antenna should consist of 4 collapsible sections. The problems of radiation and losses on joints proved to be less significant as compared to the provision of conditions for antenna operations in outer space where the temperature drop exceeds 200°. In order to keep the antenna parameters invariable (primarily its gain), it was necessary to maintain the deviation from the plane of the 12 m long antenna within 2 mm (!). The solution to this problem turned out to be quite straightforward: the "bow" principle was employed. The antenna appears to be like a bow-string with a spring-loaded draw mechanism— the bow. This version of the antenna system has been tested on the"Cosmos-1500"/"Okean" satellites and proved to be highly reliable.

In regard to the on-board data processing it should be pointed out that, due to the integrated-data store, a high contrast-background sensitivity was achieved (nearly 1 dB). This made it possible for users to take advantage of the data thus available with no additional processing. The system is capable of forming returned signals into an image. The sweep for the range coordinate, which is orthogonal to the flight direction, is provided through the change in signal delay (line scan); the sweep for the other coordinate is produced by the satellite passage (frame scan); the scales for both coordinates were selected as being equal.

A wide spectrum of practical problems had to be tackled for analyzing the image patterns being obtained. Several applications require absolute ESA (effective scattering area) measurements. For this purpose the SLR is provided with an automatic potential calibration which is implemented by feeding a fraction of the radiated signal to the input of the receiving channel. This calibration signal is delayed as a radiated pulse and is level-variable for calibration of the dynamic range of the whole channel.The calibration signal is subject to all those transformations the returned signal has to undergo. As a result, it becomes possible to perform the σ measurements on the basis of relative measurements of a returned and calibrated signal.

The SLR polarization was chosen to be vertical so as to allow better observations of the sea surface; the selection of polarization indeed has no appreciable effect on the observations of sea ice and glaciers. With regard to ice there is a certain advantage offered by cross polarization mesurements. However, these could only be achieved by using the margin in excess of 10 dB; this accounted for some trouble at the critical stage of the SLR development.

The sector of incidence angles $\theta \geq 20°$ was chosen in such a way as to afford uniqueness of observations of sea surface and sea ice inhomogeneities against thebackground reflections by the sea. With $\theta < 20°$ one can indeed clearly see quasi-mirrors reflections whose nature is essentially due to non-diffuse scattering.

And finally, when we were actively engaged in the design of the SLR, we proceeded from the assumption that in the early 80s most of the users would not have any experience in the interpretation andutilization of radar images. Therefore, the transmitted SLR imagery was combined in different variants with optical (scanner MSU-M) and radiometric images (8 mm band scanning radiometer). The testing of the first SLR was run on the "Cosmos-1500" series of satellites. The operations of these systems over the first 3 years [22] strongly suggested that they

were highly efficient and this fact sounded quite convincing to the whole of the research community. This stimulated the development of the SLR-based space operational system "Okean", which permitted the launch of 8 satellites (Table 1)

3. Experience gained with the "Cosmos 1500"/"Okean" SLR's

Now let us consider the basic capabilities of the "Cosmos 1500"/"Okean" SLR. Although this system was designed to observe the Ocean's surface, its parameters were selected so as to enable the sea ices and glaciers to be efficiently observed and, what is more, to tackle some terrain problems.

Figure 1 presents particular observation aspects in conjunction with a number of objectives that can be achieved through the use of SLR systems. A good deal of experience has been gained in utilizing the SLR sea ices imagery [23, 24]. It will be recalled that within less than a month of the satellite launching the "Cosmos-1500" SLR got involved in the rescue of 22 cargo ships which were to deliver wintering supplies to the north-eastern area of the Arctic region. This convoy found itself trapped in pack ice during the polar night in the Long strait. Two nuclear-powered ice-breakers were unable to force their way to the distressed fleet and set about the rescue operations. The conventional methods of ice surveying had failed.The "Cosmos-1500" SLR succeeded in locating the polynya near the Wrangel Island and the channels leading to it (Figure 2). This gave an excellent chance to carry out adequate rescueoperations. Only one ship was lost. A heavy material damage estimated at 70×10^9 \$ was thus avoided, whereas the cost incurred in the SLR development amounted to 10^9 \$. Meanwhile, the most impressive was the social effect.

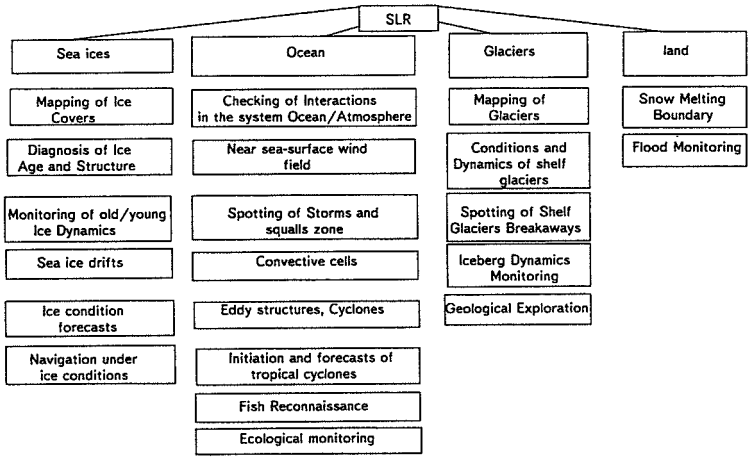


Fig. 1 - Spotting of critical changes in natural environment and warning of natural catastrophes.

Fig. 2 - "Cosmos-1500" Radar image of the Long Strait with dark channels.

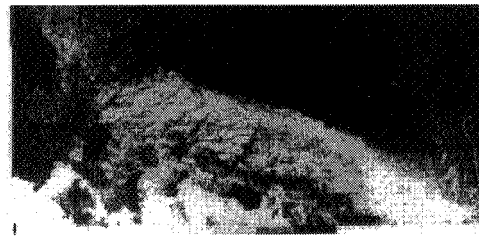


Fig. 3 - Rescue of the vessel "Mikhail Somov" in the Antartic

Two years later the research ship "Mikhail Somov" found itself in the same distress situation. It drifted for nearly 4 months in the sea ice throughout the polar night. The diesel ice-breaker "Vladivostok" was unable to make its way through the huge thickness of the sea ice to rescue that ship, which was actually in distress. The "Cosmos-1500" SLR discovered a channel in the ice mass (Figure 3), through which the ship led by the ice-breaker could travel at a speed of up 8-10 knots (as in icefree water). The whole rescue operation only lasted for more than 2 days.

An extensive use of the SLR is currently being made in Russia to provide observation data on the sea ices in the Arctic and Okhotsk seas [24]; some Scandinavian countries also benefit from the SLR data both for scientific research and a varieties of practical applications (Figure 1).

In regard with the study of oceans the SLR opens up vast fields in practical use. In this respect the most successful combination is a SLR equipped with an optical scanner. This combined system was found to be able to study numerous interaction processes in the ocean-atmosphere system, because the optical scanner is capable of monitoring the atmospheric structure under cloud cover conditions, whereas the radar is destined for observing the sea surface.

As a consequence of numerous observations of various interaction processes we have detetmined [23, 26] a high correlation of optical and radar images, that is to say, all the processes in the ocean-air system are "imprinted " on the surface and observed by the SLR.The level of radar echoes is related to the sea-surface wind speed; this enabled them to be determined by a wind velocity module (10-20% accuracy); the wind direction (accuracy of 20-50%) depends on the radar measurements (Figure 4).

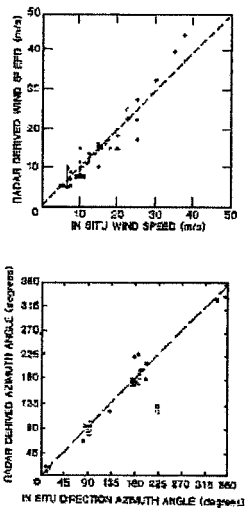


Fig. 4 - Comparison between the sea surface wind speed (a) and the sea surface wind direction (b) measured with the "Cosmos-1500" SLR and in-situ sea surface measurements. Different symbols refer to different experiments.

As can be seen, the radar is capable of determining near-hurricane winds in the range from units to tens of meters per second. While the well-developed cyclones are featured by relatively contrasted optical and radar images [23, 26] (as is clearly evident in (Figure 5 a, b)) the "young" (incipient) cyclones cannot be optically identified, especially during the polar night, although, to a large extent, they are seen to be controlling weather factors. The SLR discovers this type of cyclones at an initial stage of their generation (Figure 5 c). The SLR is also efficient in reliably detecting the storm and particularly hazardous squalls zones where in the wind is changing from the calm to a forceful hurricane (Figure 6).

During the regatta Tallinn-Riga the squall moving from west to east came upon quite abruptly; several yachts sank and many people lost their lives.

Using this sort of information in due time serves to avoid catastrophes during the navigation period in the areas affected by squalls and storms.One of the remarkable results obtained using the SLR are the "footprints"(or "imprints") of convective cells observed on the ocean surface [23, 26]. In remote sensing practice the convective cells had been observed

Fig. 5 - Radiometric (a) and radar (b) "Cosmos-1500" images of the squalls in Riga Gulf

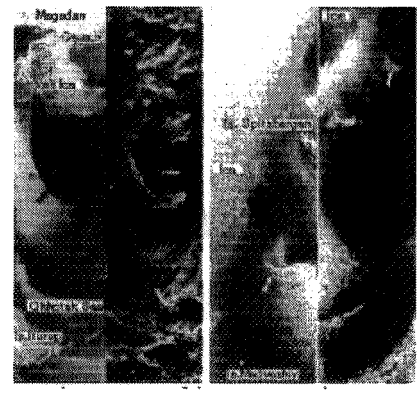


Fig. 6 - Radar (a, c) and optical (b. d) "Cosmos-1500" images of cyclones in the Polar zone, (a, b) - strong cyclone, (c, d) - "diying" cyclone

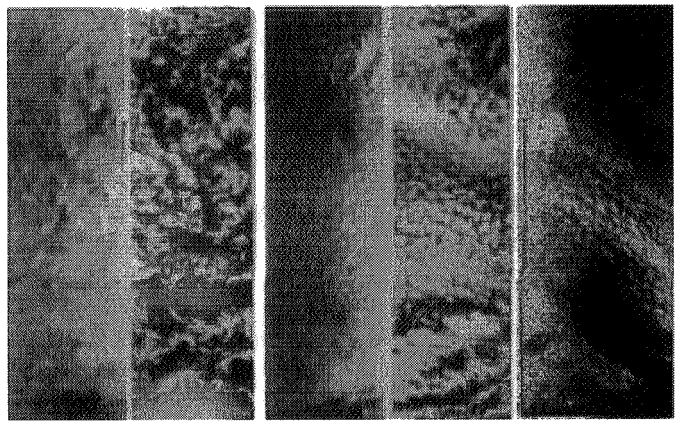


Fig. 7 - Optical (b, d) and radar (a, c, e) "Cosmos-1500" images of convective cells in different scales

in cloud patterns before. It was found that manifestations ("imprints") of convective cells on the surface could be spotted by the SLR as being more contrasting than it could be occasionally seen in the cloud structures. The convective cells were observed to have a size ranging from 1-3 km (as compared to the SLR resolution) to 150-200 km (Figure 7). A good deal of attention that was drawn to the studies of the convective cells is accounted for the fact that these phenomena are of specific importance.

Thus, the navigation at high latitudes under conditions of warm ocean-atmosphere brings about heavy icing of ships, which may subsequently result in catastrophes. According to manuals of navigation it is stated that while staying at polar latitudes it is necessary to abandon the hazardous area immediately. It is just for this reason that information derived from the SLR allows that zone to bedetected and an optimum route to be properly selected.

Another selected hazardous situation results from the merging (collapse) of convective cells into a large structure, although the energy of each cellis small; their number may be greater (Figure 7) and eventually they are likely to become a source of incipient tropical cyclones. Then a question arise: will we succeed in getting a clue to the ways of controlling the hurricane generation ?

The "Cosmos-1500"/"Okean" series SLR has proved to be an effective tool for observing tropical cyclones (Figure 8). On the one hand, the SLR parameters are conductive to this: a wide swath turns out to be useful in locating and observing the whole structure of a hurricane and this can be achieved by implementing l-3 km resolution.On the other hand, the 3 cm band provides for a high sensitivity to the wind speed which can be measured over wide ranges (Figure 4).

As a result, the data obtained from the radar measurements of the cross-section of the wind velocity field made it possible to evaluate the power of tropical cyclones [29]; thus, the power of one of the very hazardous cyclones "Diana-84" was 1.2 10^8 MW. The SLR is capable of

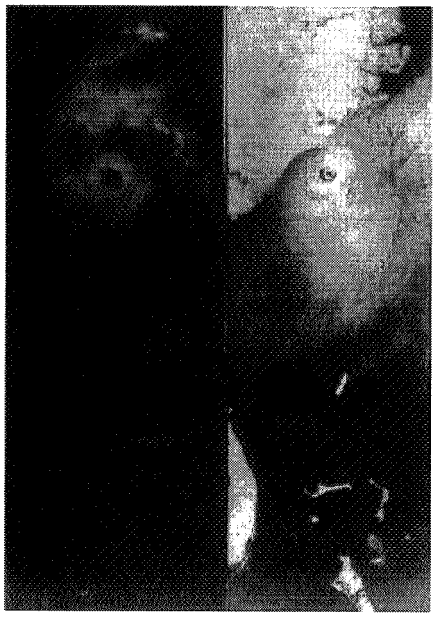


Fig. 8 - Optical (a) and radar (b) "Cosmos-1500" images of the hurricane "Diana-84"

monitoring the pattern and dynamics of hurricanes. Specifically, we were able to identify this hurricane within two weeks before it encountered land [23]; the most dangerous cases are the eye-shaped hurricane and the zone of torrential rains; they are clearly evident in the radar imagery (Figure 8).

The SLR is capable of resolving the frontal zone, the upwelling regions, and to ensures that ships move along right in the wake of a hurricane. It can help ships in discovering the areas that hold much promise for fishing. The experiments we have done yield the results that substantiate these capabilities.

The X-band SLR can be used to locate sea surface roughness variations due to simple dampening (slicks) [30]. The most hazardous of them are oil pollutions (Figure 9). We have used the SLR to perform the ecological monitoring of the maritime economic zone of the Black and Aral seas. The SLR can detect the oil slicks spread over an area ≥ 0.1 m^3 since its sensitivity makes it easy for the radar to spot the films with a thickness of more than 0.1 μm.

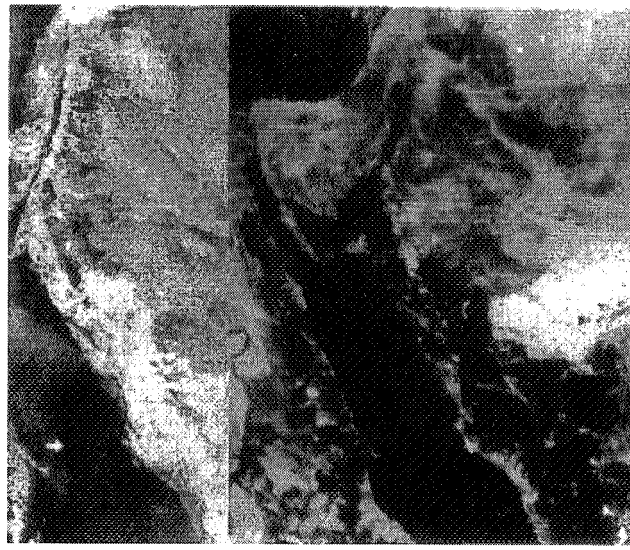


Fig. 9 - Radar (a) and optical (b) "Cosmos-1500" images of the Red Sea with oil pollutions

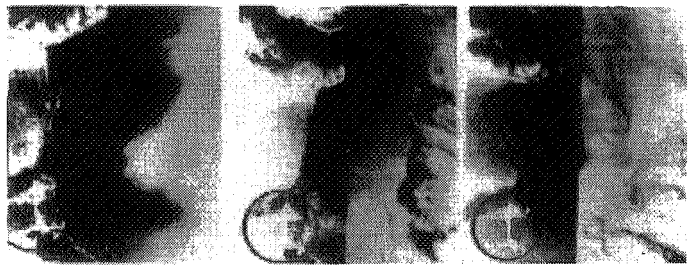


Fig. 10 - Radar (a, c, e) and optical (b, d, f) "Cosmos-1500" images of the Filchner's glacier break away. (a, b) "Cosmos-1500", 1984; (c, d) "Cosmos-1500", february 1986; (e, f) "Cosmos-1766", august 1986

The sensing of glaciers is based on the SLR capability to observe not only the surface, but also its internal structure [23, 31]. This permits mapping of the glaciers, monitoring of the dynamics of offshore glaciers, their split and iceberg formation, iceberg tracking, etc. Of special interest is geological surveying of internal glacier structure. All these features are exemplified by the fate of the Antarctic station Druzhnaya-1 (the coast of the Weddell sea) (Figure 10).

4. Use of the SLR for natural disasters reduction

One of the fields in which the SLR imagery is utilised, is the detection of critical situations and the avoidance of natural catastrophes [32]. The solution of these problems can be found through the use of SLR's capable of detecting environmental changes.Most of the issues that were taken up in connection with the SLR being applied and properly considered in the previous section of the paper clearly show that the SLR may well be utilized to the best advantage. For instance, information inferred from the SLR was, in some cases, useful in preventing huge material damages involving losses of human lives, for that matter (Figure 2-3).

Yet another example of preventing a catastrophe was the monitoring of snowmelting in Ukraine (Figure 11). The heavy snowfalls in 1988 resulted in an imminent danger of floods because of the snow melting in an area where the Chernobyl disaster had taken place two years before. The fear

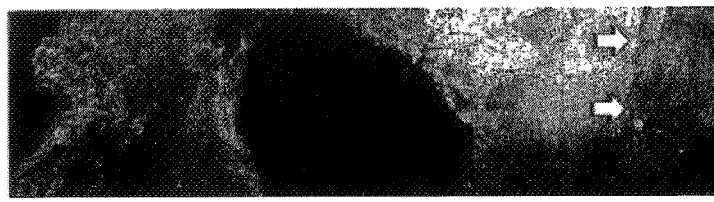


Fig. 11 - Radar "Cosmos-1500" image of the southern region of the Ukraine, during the snow melting process monitoring

was that the waters drained off in the Dnieper cascade of reservoirs could cause a heavy loss of natural resources. A team of experts were involved in supervising the rate of snow melting (according to the SLR data) and the water level in the reservoirs.

We regret to say that SLR information was not always properly used even though there did exist a good chance to do so. For instance, the squalls (Figure 6) were located by the SLR and the data obtained were not taken into account. The same is true for the disastrous situation that emerged at the Druzhnaya-1 station (Figure10). As a result, it caused great material damage and led to heavy death toll, which could have been avoided. Therefore, in this regard, the early warning is of considerable value to the countries that could be worst hit by tropical cyclones, especially by those involving death and destruction. For example, in a powerful cyclone that swept over Bangladesh in 1970, 900'000 people died and 1.3 million people were left homeless.

According to the conclusive reports by the UNESCO the timely warning of a devastating tropical cyclone reduces the number of victims by more than 20 times. In the context of the UNESCO appeal to declare the last decade of the 20th century as an International Decade for Natural Disaster Reduction (IDNDR) we believe that the potentialities of all space-borne radar systems can be successfully utilized for this noble mission.

5. Joint use of SLR and SAR

While the space-borne radar systems were initially aimed at environmental studies, recently developped radars are more and more frequently designed to deal with the problems of locating and preventing natural disasters. In this case these systems are called upon to provide observations both of land, oceans, ice and glaciers. For such applications, however, no comparison can be made between the SLR and the SAR just like it is impossible to compare and oppose a car and a truck. Nowadays efforts are being made in Ukraine to develop the space-borne radar system "Sitch" incorporating the SLR and SAR facilities. This combined system is dictated by the need to ensure that the Earth is simultaneously observed in a wide swath and with a high resolution. The wide swath is provided by the SLR whereas high resolution is achieved though the use of the SAR.

The basic parameters of the radar system"Sitch" are summarized in Table 2.

This SLR version is of the coherent type and its resolving capacity will be subsequently brought up to hundreds of meters. The SAR version is made with an active phased array antenna, which is electrically controlled in two planes. The scanning in the azimuth plane ensures that the

Radar Type	Wavelength (cm)	Polarization	Swath (km)	Resolution	Information processing
SLR	3.1	VV	2 x 750	0.5..1.5 km	onboard
SAR	23	VV HH	300	50 m	onboard
		VV HH	310	5..15 m	onground

Table 2

synthesizing process is not influenced by the SAR carrier orientation. This enables the data processing to be done onboard the radar system. Additionally, this factor also makes the requirements for carrier stabilization less stringent. Scanning at elevation angle allows the SLR and SAR images to be combined in definite zones. The SAR capabilities are clearly demonstrated in the system that had been previously tested (Table 1). The combined SLR and SAR systems open up new fields in remote sensing. For instance, the X-band SLR is efficient at measuring the sea-surface wind fields whereas the L-band SAR is only slightly sensitive to wind. However it is capable of discerning the sea roughness structure [34] (Figure 12).

A new practice of controlling oil pollution is based on the joint use of both SLR and SAR systems. The X-band SLR's are able to detect the thinnest films of oil spills whereas the L-band SAR can pick up the localised areas of films having greater thicknesses [33].

In remote sensing of sea ice the X-band is efficient in winter; in summer,the most reliable data are inferred over the L-band. Fresh studies into the Antarctic glaciers can be made by using the X-band and L-bands of the SLR andSAR respectively. Specifically the SAR could also be employed to monitor the radioactive contamination zones and earthquake precursors. Monitoring of incipient earth-quakes is based on the sensitivity of the L-band SAR to moisture, which was seen to be particularly contrasting in arid (zones).

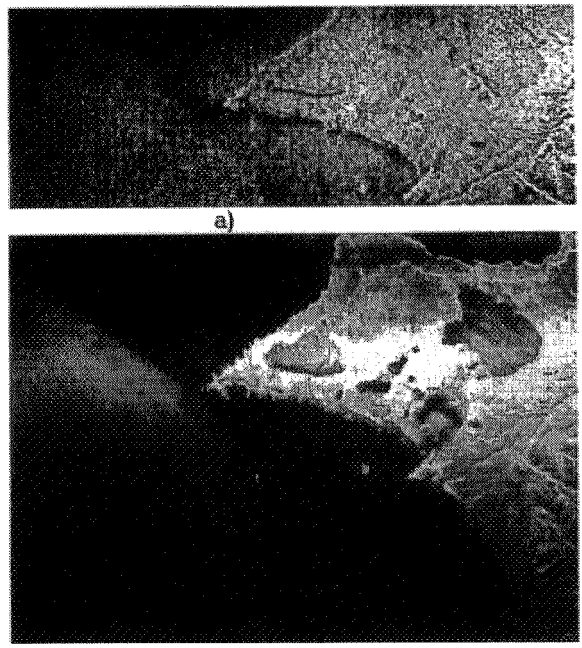


Fig. 12 - L-band SLR (a) and X-band SAR (b) image of the Pacific Ocean surface from an airplane

Conclusions

The radar observation system which iscurrently being developed in Ukraine will serve the purpose of solving national problems and promoting international collaboration and cooperation. In order to achieve this objective one of the operating modes is set in such a manner that onboard processed SLR and SAR data are transmitted to a wide network of users in the 1.7 GHz band.

We hope that the 75th URSI jubelee symposium will be a useful contribution to concerted efforts by researchers in their endeavor to seek for wider fields in which space-borne systems will be extensively employed.

References

[1] D.E. Kerr, Ed.; "Propagation of short radio waves"; v.13, MIT Rad. Lab. Series, N.Y., 1951.

[2] D.D. Crombie; "Doppler spectrum of sea echo at 13.56 Mc/s"; Nature, 1955, v. 175, p. 681

[3] "Radio-oceanographic investigations of sea surface"; Ed. S. Ya. Braude, Kiev, ASUkr, 1962 (in russian).

[4] S.O. Rice; "Reflection of electromagnetic waves by rough surfaces"; Comm. Pure and Appl. Math., 1951,No 4, p. 351.

[5] F.G. Bass, V.G.Bocharov; "On a theory of electromagnetic waves scattering from the statistically rough surface"; Radiotechnique and electronics, 1958,No 3, p. 180 (in russian).

[6] M. Katrin; "On the mechanisms of radar sea clutter"; Proc. of the IRE, 1957, v. 45, No 1, p 44.

[7] F.C. MacDonald; "The correlation of radar sea clutter on vertical and horizontal polarisation with wave height and slope"; IRE Conv. Rec., 1956,v. 4, No 21, p. 29.

[8] A.D. Rosenberg et al.; "Radar pulse coherent phasemeter on waves 50, 10, 3 and 0.8 cm"; Proc IRE ASUkrSSR, 1962,v. 10, p. 38.

[9] A.D. Rosenberg et al; "Frequency shift in radar radiowaves scattering from disturbed sea surface"; Proc IREASUkrSSR, 1962, v. 10, p. 47.

[10] B.F. Kurganov; "Sound scattering from rough surfaces with two type roughness"; Acoust. journ., 1962, No 3,p. 325 (in russian).

[11] F.G. Bass et al; "Very high frequency radiowave scattering by a disturbed sea surface"; IEEE Trans., 1968, v. AP-16, No 45, p. 554.

[12] A.K. Fung, R.K. Moore; "The correlation function in Kirkchoff's method of solution of scattering of waves from statistically rough surfaces"; J. of Geophys. Res. 23.

[13] J.W. Wright; "A new model for sea clutter"; IEEE Trans., 1968, v. AP-16, p. 217.

[14] D.E. Barrick, W.H.Peake; "A review of scattering from surfaces with different roughness scales"; Radio Science, 1968, v 3, No 8, p. 865.

[15] G.R. Valenzuela; "Scattering of electromagnetic waves from a slightly rough surface moving with uniform velocity"; Radio Science, 1968, v. 3, No 12, p. 1154.

[16] R.K. Moore, W.J. Pierson jr; "Worldwide oceanic wind and wave predictions using a satellite radar radiometer"; J. Hydronautics, 1971, v. 5, p. 52.

[17] V.I. Zeldis et al; "Investigations on amplitude characteristics of sound signals scattering by disturbed water surface"; Acoust. Journ., 1973, v. 19, No 1, p. 170.

[18] H. Mitsujasu, T.Honda; "The highfrequency spectrum of wind-generated waves"; J. Oceanogr. Soc. Jap., 1974, v. 30, No 4, p. 185.

[19] J.C. Daley; "Wind dependence of radar sea return"; J. Geophys. Res., 1973, v. 78, p. 7823.

[20] W.J. Pierson, jr et al; "The measurements of the winds near surface with a radiometer-scatterometer on Skylab"; Final Rept on EPN 550, Tech Rep., City U. on N.Y. Inst. of meteorol. sciences and Univ. of Kansas RSL,1975.

[21] "The special issues on Seasat-A"; J.Geophys Res., 1982, v. 87, NC5;1983, v. 88, NC3.

[22] "The special issue on "Cosmos-1500""; Sov. J. Remote Sens., 1985, No 4. (English. transl).

[23] "Radar observations of Earth from space. Investigation of sea surface, sea ices and glaciers using spaceside-

looking radar"; Ed. L.M. Mitnik and S.V. Victorov, 1987 (in russian).

[24] L.M. Mitnik, A.I. Kalmykov; "Structure and dynamics of sea Okhotsk marginal ice zone from the Ocean satellite radar sensing data"; J. of Geophys. Res., 1992, v. 97.

[25] A.I.Kalmykov et al.; "Radiophysical observations from space of mesoscale structure on the ocean surface"; Sov. Phys. Dokl. Acad. Sci. USSR, 1984, 29, p. 1016 (English transl.).

[26] A.I.Kalmykov et al.; "Observations of the marine environment from space-borne side-looking real aperture radars"; Remote sens. Env., 1993, No 45, p. 193.

[27] R.K. Moore, A.K. Fung; "Radar determination of winds at sea"; Proc IEEE, 1979, v. 67, p. 1504.

[28] L.C. Schroeder et al.; "SAS Smeasurements of the Ku-band radar signature of the ocean"; IEEE Journ of Ocen. Eng., 1982, v. OE-7, No 1, p. 3.

[29] V.F. Utkin et al.; "Determination of the characteristics of tropical cyclones from space-borne radar images"; Dokl. Acad. Sci. USSR, 1986, v. 286, No 2, p. 331 (in russian).

[30] A.I. Kalmykov, A.P. Pichugin; "Some features of the sea surface inhomoheneities detections by radar"; News AS USSR, Physics of atmosphere and ocean, 1981, No 7, p. 754.

[31] V.B. Yefimov et al.; "Radar studies of Antarctic glaciers conditions, Sov. J. Remote Sens, 1990, No 4, p.3.

[32] A.I. Kalmykov et al.; "Radar observations from space of critical phenomena and natural catastrophes in the world's oceans"; Prepr. Acad. Sci, UkrSSR, Inst. Radiophys. Electr., 1989, No 380.

[33] E.I. Bushujev et al.; "Multi-aim complex of national flight laboratory for remote sensing, control of environment and geological monitoring of Ukraine"; In reports Int. sci. seminar on aerospace monitoring of Earth surface and atmosphere, Kiev, 1993, p. 5.

[34] R.C. Beal; "Spatial evolution of ocean wave spectral in space-borne synthetic aperture radar for oceanography"; Johns Hopkins Univ. Press, 1981, p. 215.

MEASUREMENT OF MESOSCALE OCEANIC AND ATMOSPHERIC PHENOMENA BY ERS-1 SAR

WERNER ALPERS

Abstract

Radar images acquired over the ocean by the synthetic aperture radar (SAR) aboard the First European Remote Sensing Satellite ERS-1 delineate oceanic as well as atmospheric phenomena. The oceanic phenomena visible on ERS-1 SAR images include surface waves, internal waves, eddies, oceanic fronts, underwater bottom topography, and surface slicks. The atmospheric phenomena include katabatic wind fields, convective cells, atmospheric boundary layer rolls, internal gravity waves in the atmosphere, atmospheric vortex rows (Karman vortex streets) behind islands, and atmospheric undular bores. Examples of ERS-1 SAR images showing sea surface manifestations of mesoscale oceanic and atmospheric phenomena are presented and interpreted in terms of oceanic and atmospheric models.

1. Introduction

On July 17, 1991, the First European Remote Sensing Satellite ERS-1 was launched by the European Space Agency (ESA). It carries a C-band (5.3 GHz) synthetic aperture radar (SAR) operating at vertical polarization for transmission and reception. In the full SAR mode (geometric resolution: 25 m, swath width: 100 km) SAR data can be collected for a period of 10 minutes per orbit (period: about 100 minutes). As of February 1995, a total number of 1 million ERS-1 SAR scenes have been acquired, 426,000 SAR scenes are archived at the Processing and Archiving Facilities (PAFs) of ESA, and 2,300 SAR scenes were delivered by the PAFs to users. In addition, 25,000 ERS-1 SAR scenes were delivered to users by national foreign stations like the ones located at Tromsø (Norway), Fairbanks (Alaska), O'Higgins (Antarctica) West Freugh (Scotland), Alice Springs (Australia),.and Bangkok (Thailand).

Analyses of ERS-1 SAR data have demonstrated that ERS-1 SAR imagery of the ocean surface contains a wealth of information that can be used in oceanographic and marine meteorological studies.

Prof. Werner Alpers is with the
Institute of Oceanography
University of Hamburg
Troplowitzstraße 7,
2000 Hamburg 54, Germany
Tel: + 49 (40) 4123 5432 Fax: + 49 (40) 4123 5713

Preliminary results obtained from the analysis of ERS-1 SAR data can be found in the Proceedings of the First and Second ERS-1 Symposia held in Cannes, France, 4-6 Nov., 1992 (ESA publication SP-359) and in Hamburg, Germany, 11-14 Oct., 1993 (ESA publication SP-361), respectively, and in the Proceedings of the "ERS-1 Geophysical Validation Workshop" held in Penhors, Bretagne, France, 27-30 April, 1992 (ESA publication WPP-36).

2. Mesoscale oceanic phenomena

Mesoscale oceanic phenomena like internal waves, oceanic eddies and fronts, become visible on SAR images because they are associated with variable surface currents which modulate the surface roughness (Fu and Holt, 1982; Alpers, 1985; Johannessen et al., 1992). Figures 1 and 8 show ERS-1 SAR images of the Strait of Gibraltar on which roughness patterns associated with nonlinear internal waves propagating eastwards can be delineated. Other ERS-1 SAR images of the Strait of Gibraltar showing sea surface manifestations of internal wave forms can be found in Alpers and La Violette (1992), Brandt and Alpers (1994).

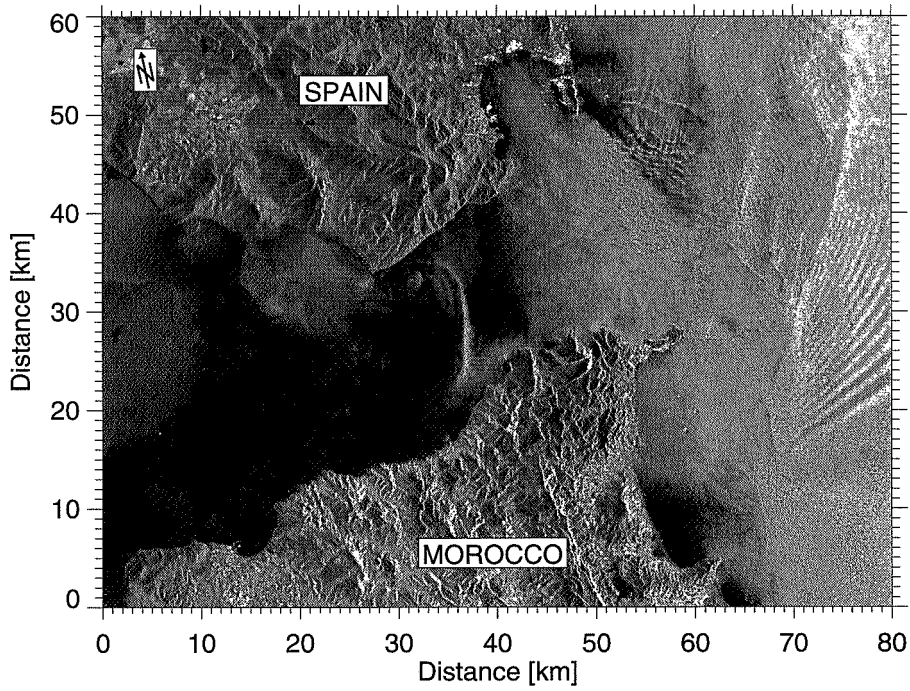


Fig. 1 - Section of an ERS-1 SAR image of the Strait of Gibraltar acquired on Jan. 20, 1994, at 11:03 UTC (orbit 13151, frames 2871/2889). It shows sea surface expressions associated with an internal soliton in the Strait (bended bright line) and an internal wave packet east of the Strait which was generated during the previous tidal cycle.

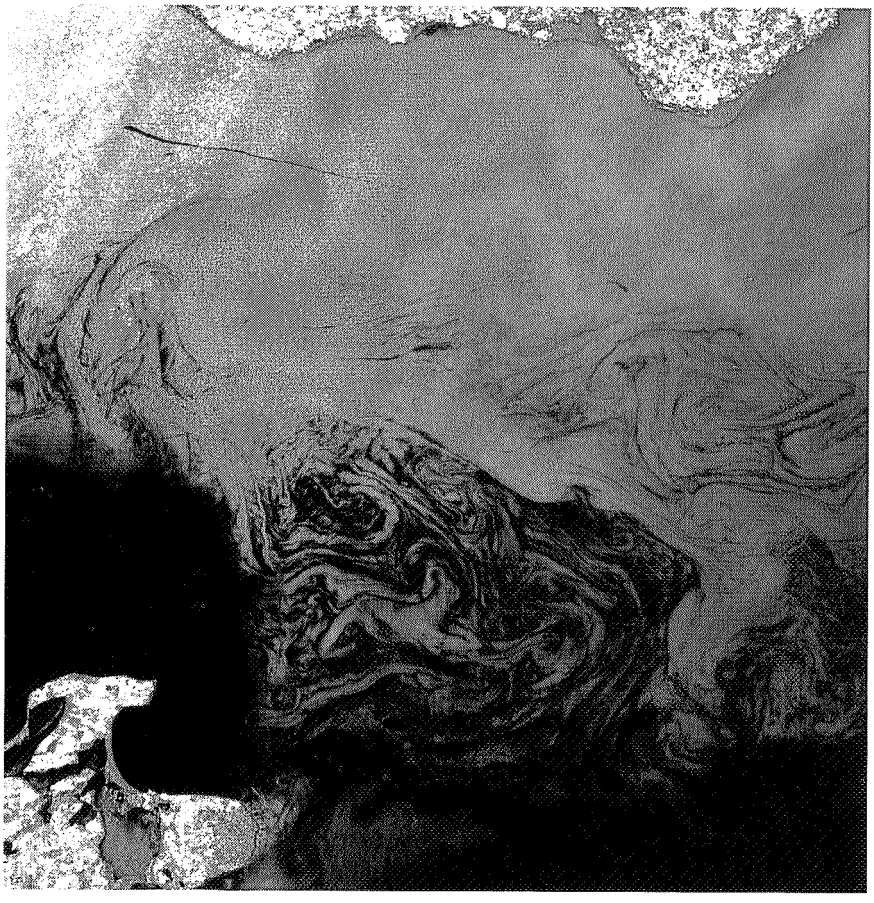


Fig. 2 - ERS-1 SAR image (100 km x 100 km) of the Baltic Sea acquired on Apr. 16, 1994, at 21:04 UTC (orbit 14390, frames 1089/1107). It shows in the center sea surface patterns caused by natural surface films. In the lower left hand corner the German island of Ruegen and on the upper edge the coast of Sweden are visible.

If slicks are floating on the sea surface, mesoscale oceanic phenomena become also visible by the spatial distribution of the slick material on the sea surface (Gower, 1993). Surface slicks are often encountered in coastal regions with high biological productivity, especially in the summer, where they follow the underlying water movements. An example of an ERS-1 SAR image delineating slick patterns is shown in Figure 2.

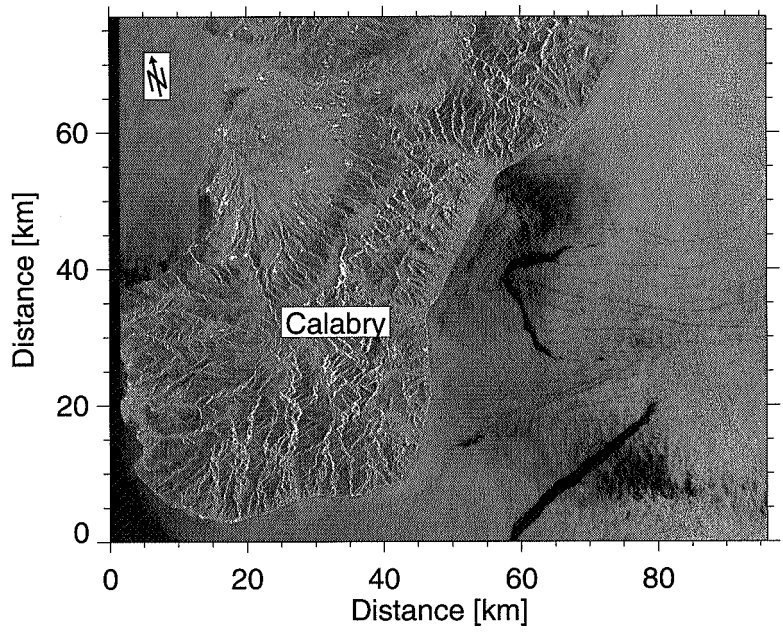


Fig. 3 - ERS-1 SAR image (100 km x 75 km) of the Mediterranean Sea (Ionian Sea) and the southern part of the Italian peninsula (Calabry) acquired on Aug. 31, 1994, at 9:38 UT (orbit 11117, frame 2835). It shows two large dark streaks off the east coast of Calabry which very likely are due to oil spills released by a ship

Mineral oil spills can also be detected on ERS-1 SAR images as shown in Figure 3. They damp the short surface waves and thus reduce the radar backscatter (Hühnerfuss et al., 1986).

Underwater sandbanks located in areas of strong tidal flow are also visible on ERS-1 SAR images. An example is shown in Figure 4. The underwater sandbanks modulate the tidal flow which in turn modulates the sea surface roughness and thus the radar backscatter (Alpers and Hennings, 1984).

3. ERS-1 SAR images of atmospheric phenomena

An imaging radar is capable of delineating mesoscale and submesoscale atmospheric phenomena because they are associated with variations of the wind stress at the sea surface. The wind stress depends on the wind speed at the sea surface and on the stability of the air-sea interface which is a function on the temperature difference between the water and the air (Keller et al., 1989). Changes in the wind stress at the sea surface disturb the small-scale sea surface roughness and thus

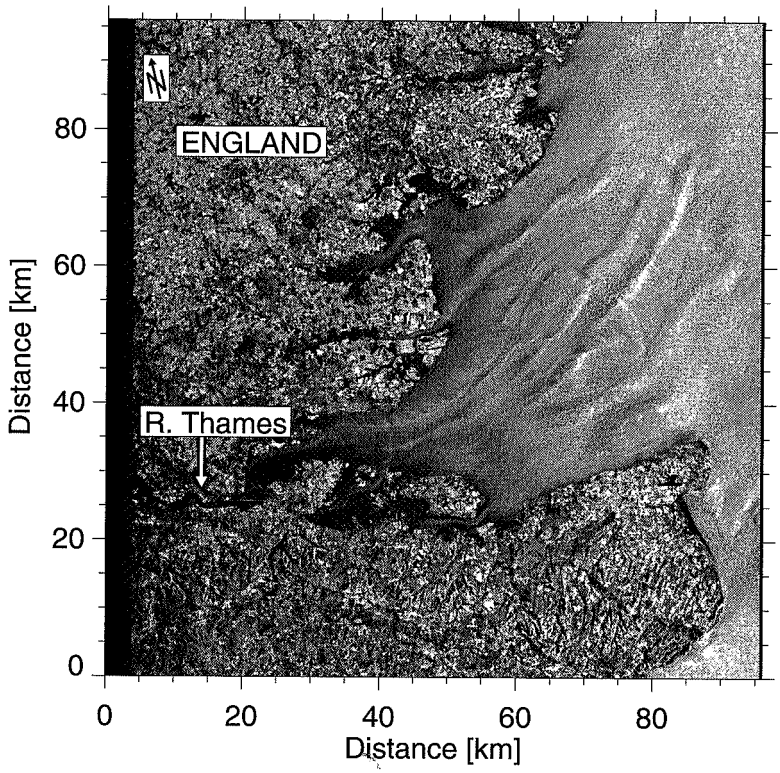


Fig. 4 - ERS-1 SAR image (100 km x 100 km) of the estuary of the river Thames (England) acquired on Jan. 21, 1993, at 10:52 UT (orbit 7940, frame 2565). It shows sea surface manifestations of sandbanks located in the estuary at a time of strong tidal flow.

give an "imprint" on the sea surface which is visible on radar images. The synthetic aperture radar (SAR) aboard the First European Remote Sensing Satellite ERS-1 is much more sensitive to changes in the wind stress than the SAR aboard the Seasat satellite because the ERS-1 SAR operates at C-band (wavelength: 5.7 cm) and not at L-band (wavelength: 24 cm) like the Seasat SAR. For ERS-1 SAR (incidence angle: 23°, VV polarization), the slope of the wind scatterometer model function is 1.1 while for Seasat SAR (incidence angle: 23°, HH polarization) it is only 0.5 (Alpers and Brümmer, 1994). Thus the ERS-1 SAR should be more suitable for studying atmospheric phenomena than the Seasat SAR. This can be seen by comparing ERS-1 and Seasat SAR images acquired over the ocean.

3.1 Katabatic wind fields

Katabatic winds are cold winds blowing in the evening and night down a sloping terrain ("gravity flow") and, at a coast, over the adjacent sea surface. These winds are generated because in the evening and night the air near the surface cools off faster over the land than over the sea. Over mountain slopes a horizontal density difference develops between the cooled air at the slope surface and the free air in the same altitude over the lower ground or the sea. This results in a "down-hill" flow of cold air. Sea surface manifestations of katabatic wind fields have often been identified in the summer on ERS-1 SAR images of coastal regions which are adjacent to

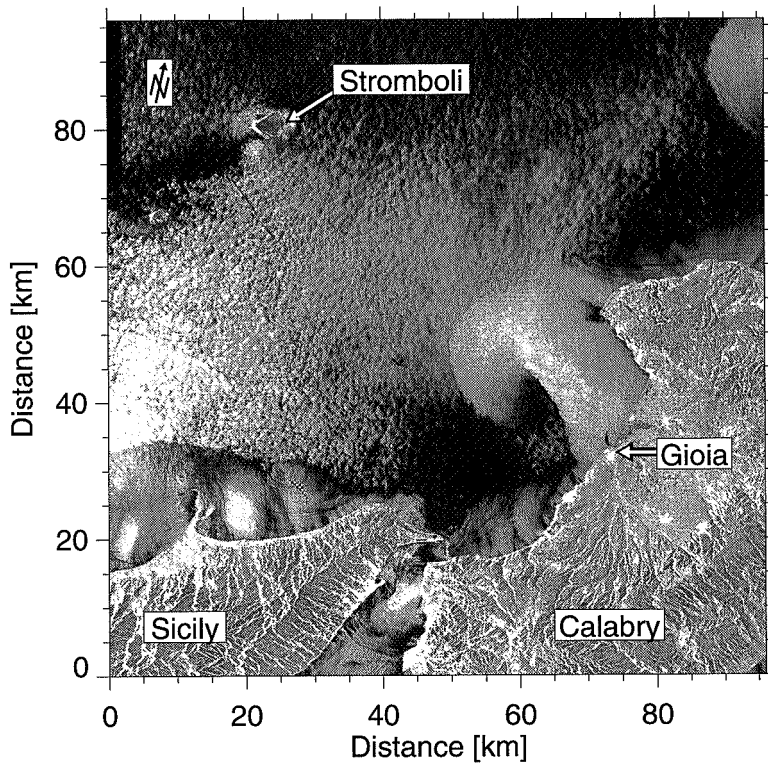


Fig. 5 - ERS-1 SAR image (100 km x 100 km) of the Mediterranean Sea (Tyrrhenian Sea) north of the Strait of Messina (Italy) acquired on Sept. 8, 1992, at 21:13 UT (orbit 6014, frame 765). It shows north-west of Gioia sea surface manifestations of a katabatic wind tongue (bright blob in the image). Furthermore, between the island of Stromboli and the Sicilian coast, a granular pattern can be delineated which we believe are sea surface "imprints" of atmospheric convective cells. This cellular structure is destroyed in the vicinity of the Sicilian coast by the katabatic wind blowing from the mountains onto the sea. In the lower section of the image an oceanic internal wave train propagating southwards in the Strait of Messina can be delineated.

mountainous areas. Figure 5 shows an example of a katabatic wind field which extends from the Calabrian coast near the Italian town of Gioia into the Tyrrhenian Sea. Also patterns arising from katabatic winds which are channeled through valleys in northern Sicily can be delineated north of the Sicilian coast.

It is also possible to extract the distribution of sea surface wind velocities from ERS-1 SAR images by using an ERS-1 wind scatterometer model function (The wind direction has to be inferred from the orientation of the valley and from the form of the katabatic tongue).

3.2 Convective cells

Convective atmospheric cells are generated over the sea when the water temperature is higher than the air temperature (unstable marine boundary layer) and when the wind speed is low. The cellular features visible in Figure 5 between the island of Stromboli and the Sicilian coast are very likely sea surface manifestations of such convective atmospheric cells. Such features can often be delineated on ERS-1 SAR images taken over the Mediterranean Sea on calm days during summer. These quasi-regular features are absent in a coastal sector adjacent to the northern Sicilian coast (see Figure 5). We suspect that there the organized cellular circulation pattern in the unstable marine boundary layer is destroyed by coastal processes associated with the katabatic wind field blowing from the northern Sicilian mountains through the valleys.

3.3 Atmospheric boundary layer rolls

Atmospheric boundary layer rolls are helical circulation patterns in the atmospheric boundary layer which are superimposed on the mean wind field, i.e., the primary flow (see Figure 6). They can be generated either by thermal instability (Rayleigh-Bénard instability) when the layer is heated from below or cooled from above, or by dynamic instabilty (inflection point instability) when the wind velocity changes with height in such a way that an inflection point occurs in the wind component normal to the roll axis. The axes of the boundary layer rolls are oriented between the directions of the mean surface wind and the geostrophic wind above the boundary layer. Usually the boundary layer is capped by an inversion so that the depth of the boundary layer and the roll layer coincide. In the case of a thermal instability, the aspect ratio, i.e., the horizontal wavelength of the roll pattern, λ, divided by the roll height, h, is 2.8 according to the linear Rayleigh-Bénard convection. However, the most frequently observed values range between 2 and 4.

If the moisture conditions are favorable, cloud streets may be formed in the upward rising branches of the roll circulation. Atmospheric boundary layer rolls contribute significantly to the vertical exchange of momentum, heat, and moisture in the atmosphere. At higher altitudes their relative contribution to the total vertical fluxes in the atmosphere can even be larger than the turbulent fluxes.

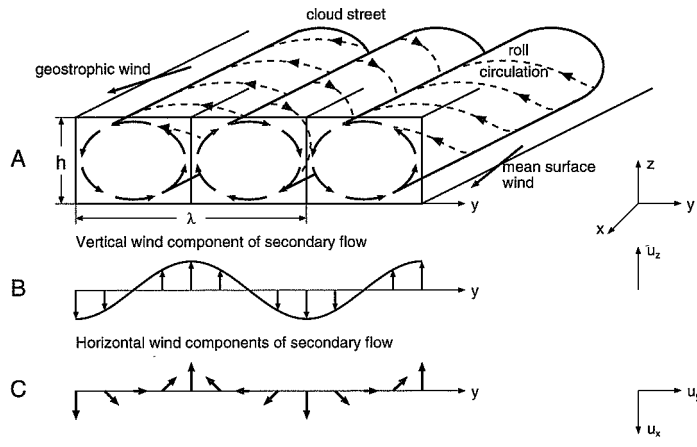


Fig. 6 - Schematic plot of secondary flow pattern associated with atmospheric boundary layer rolls. Panel A: Perspective view of the three-dimensional flow; Panel B: Variation of the vertical component uz of the wind velocity along the y direction; Panel C: Variation of the horizontal components ux and uy (in the x-y plane).

Sea surface manifestations of atmospheric boundary layer rolls have first been noted on L band SAR images which were acquired over the Atlantic ocean off the coast of Florida during the Marineland Experiment in 1975 from a NASA airplane (Thompson et al. 1983). Later such features have also been observed in L band SAR images from the Seasat satellite (Fu and Holt 1982), and from Space Shuttle Columbia during the Shuttle Imaging Radar-A (SIR-A) experiment (Ford et al. 1983), and in X band real aperture radar (RAR) images from various Russian satellites (Mitnik 1992). However, all these radars were not well calibrated, and therefore it was not possibe to extract quantitative information on sea surface wind variations from the radar images. This is, however, possible with the ERS-1 SAR since it is a calibrated instrument.

Figure 7 shows sea surface manifestations of atmospheric boundary layer rolls acquired over the Greenland Sea when a cold wind was blowing off the ice over the warmer sea surface. The streaks are aligned approximately in wind direction. The horizontal separation of the streaks increases with increasing distance from the ice edge. This increase is linked to an increase of the height of the atmospheric boundary layer with distance from the ice edge.

Other examples of sea surface manifestations of atmospheric boundary layer rolls visible on ERS-1 SAR images can be found in the paper by Alpers and Brümmer (1994).

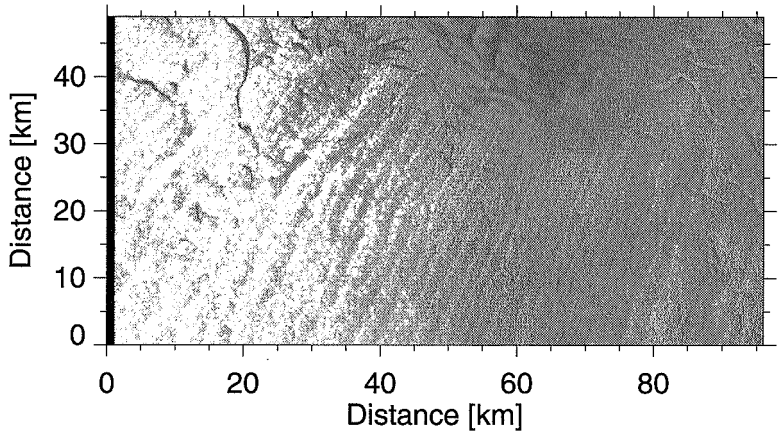


Fig. 7 - ERS-1 SAR image (100 km x 50 km) of the marginal ice zone west of Spitsbergen in the Greenland Sea acquired on March 5, 1993, at 19:41 UT (orbit 8561, frame 1629). It shows sea surface manifestations of atmospheric boundary layer rolls generated by a strong cold wind blowing off the ice over the warm sea (cold air outbreak). The ice edge is located in the upper part of the image.

3.4 Internal waves in the atmosphere

Atmospheric internal gravity waves exist in layered atmospheres. They either occur as quasi-periodic waves or as solitons. They are often generated behind mountain ranges in which case they are called lee waves. In the steady state lee waves are stationary with respect to the terrain feature, but they are propagating relative to the mean air flow above the earth surface. Lee waves are very common in visible remote sensing imagery where they manifest themselves as wave-like cloud patterns.

However, they also can manifest themselves on the sea surface since they are associated with a varying surface stress which modulates the small-scale sea surface roughness. An example of an atmospheric lee wave visible on an ERS-1 SAR image in the Mediterranean Sea is shown in Figure 8. The atmospheric lee wave is generated by the interaction of an eastwards blowing wind (approx. 7 ms^{-1}) with the Sierra del Hauz mountain range in Morocco.

Figure 9 shows an image intensity (grey level) scan normal to the eastern Morrocan coast line through the lee wave pattern. The image intensity has been converted into normalized radar cross section (NRCS) by using the conversion (calibration) factor provided by ESA. Furthermore, the normalized radar cross section can be related to the wind speed by using a C band wind scatterometer model. The wind speed scale calculated from the CMOD 4 model (Stoffelen et al. 1992) is plotted on the right-hand vertical coordinate axis.

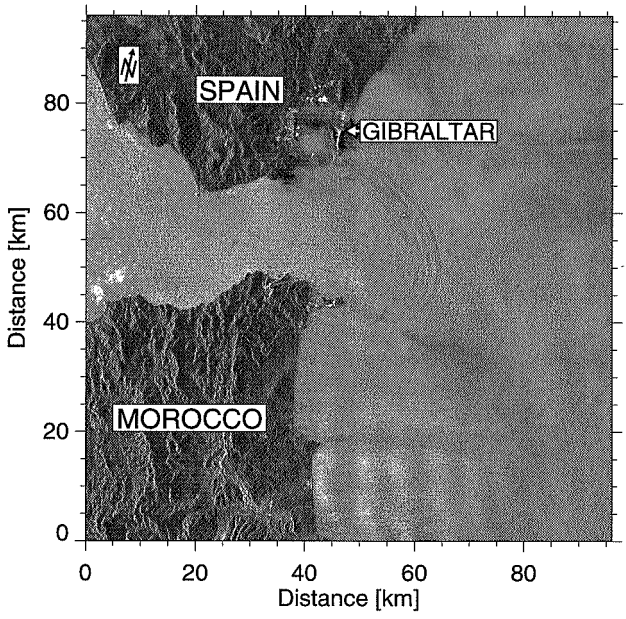


Fig. 8 - ERS-1 SAR image (100 km x 100 km) of the Strait of Gibraltar acquired on Sept. 3, 1993, at 22:39 UT (orbit 11168, frame 711). It shows in the center sea surface manifestations of an oceanic internal wave packet generated in the Strait of Gibraltar and in the lower right hand section sea surface manifestations of atmospheric internal waves (lee waves) generated by an eastward blowing wind over the 600 m high mountain range Sierra del Hauz in Morocco.

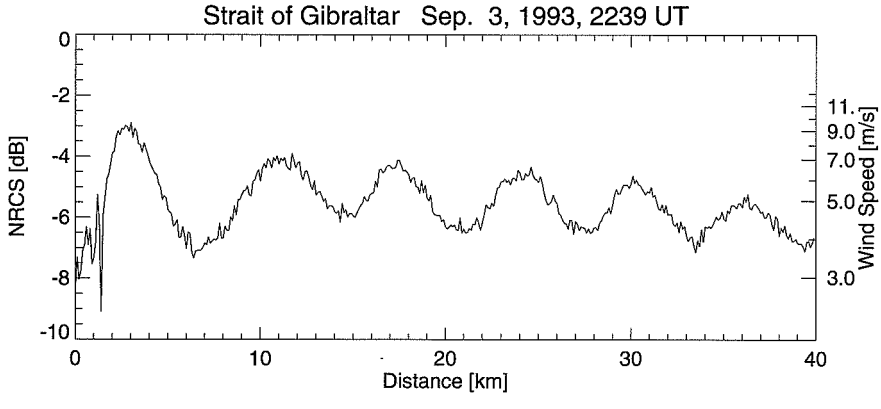


Fig. 9 - Image intensity scan from west to east through the internal lee wave pattern east of the Moroccan coast shown in Fig. 8. On the left-hand vertical coordinate axis the normalized radar cross section (NRCS) is plotted and on the right-hand vertical coordinate axis the wind speed at a height of 10 m above sea level (under neutral wind conditions).

3.5 Atmospheric undular bores

Figure 10 shows an ERS-1 SAR image of the German Bight of the North Sea which was acquired on March 18, 1992, at 10:25 UT. The wave-like pattern visible east of the island of Heligoland is not a sea surface manifestation of an oceanic internal wave, but of an atmospheric undular bore propagating eastwards. The identification of this feature as an atmospheric undular bore has been possible because simultaneous in-situ meteorological measurements were carried out at the island of Heligoland. At Heligoland the German Weather Service had erected a 80 m high meteorological mast at which wind and temperature measurements were performed every minute at height levels of 10 m, 30 m, 50 m, and 80 m. A wind with a height-averaged mean wind speed of 3.0 ms^{-1} was blowing at the time of this ERS-1 overflight from 170° N. Twenty-nine minutes before the ERS-1 SAR image was taken, a pronounced periodic wind fluctuation in east-west direction was observed at Heligoland. The period of the wind fluctuations in the reference frame moving with the wind was 9.5 minutes and the wavelength 2600 m. The wind speed fluctuations at 10 m height ranged from -1.2 ms^{-1} to 1.7 ms^{-1} (Alpers and Stilke, 1995).

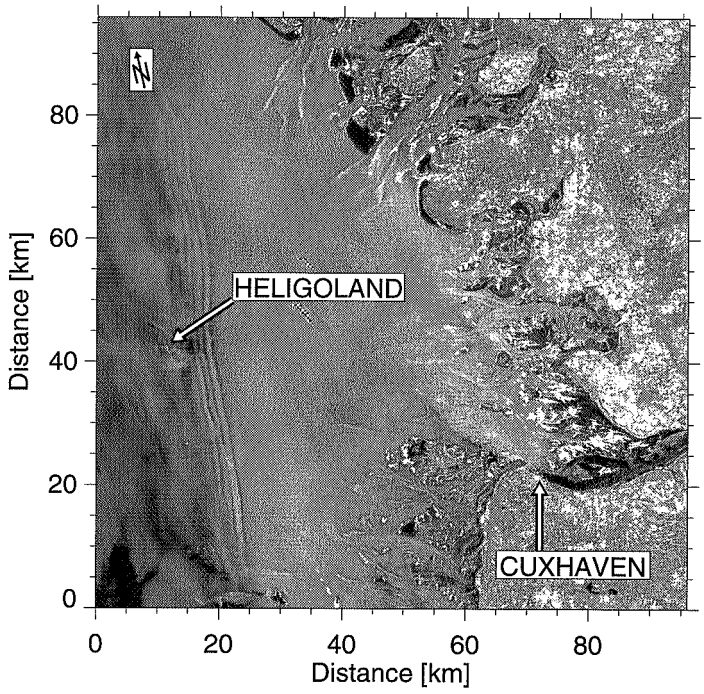


Fig. 10 - ERS-1 SAR image (100 km x 100 km) of the German Bight of the North Sea acquired on March 8, 1992, at 10:25 UT (orbit 3373, frame 2511). It shows a wave-like pattern east of the island of Heligoland which is an imprint on the sea surface of the atmospheric undular bore generated by an atmospheric cold front propagating eastwards. The bright streaks near the coast are sea surface manifestations of underwater bottom topographic features.

Atmospheric undular bores are solitary wave disturbances propagating on a stably-stratified layer in the lower atmosphere. The most spectacular examples of atmospheric undular bores are the "Morning Glories" which often occur near the southern coast of the Gulf of Carpentaria in Northern Australia (Smith 1988). The name "Morning Glory" derives from an impressive roll cloud, or series of roll clouds, which occurs early in the morning, most frequently at the end of the dry season. They are associated with surface wind gusts and with pressure jumps on the order of few hectopascals.

However, atmospheric undular bores have also been observed in other parts of the world (Doviak and Ge 1984). One example are the "Berlin Fog Waves of October 11, 1969", reported by Scherhag (1969) and Egger (1985). There exist several reports by pilots of small aircraft and crew members of small sailing boats who encountered a suddenly occurring succession of strong wind squalls or wind gusts of varying direction, the period of which is typically 5-10 minutes. These unexpectedly occurring strong wind squalls associated with atmospheric undular bores can become dangerous to small landing aircraft (Christie und Muirhead 1983).

The generation of atmospheric undular bores is usually linked with the intrusion of colder, denser air into a stably or indifferently layered atmosphere. On March 8, 1992, a cold front was moving from west to east over the German Bight. Like atmospheric lee waves, atmospheric undular bores are associated with a varying wind stress at the sea surface and thus with a varying small-scale sea surface roughness which is detectable by radar. By applying a wind scatterometer model, variations in image intensity can be converted into variations of wind speed.

Acknowledgments

The author thanks ESA for providing the ERS-1 SAR images and D. Gamarnik, J. Heybrock and Chr. Staabs for preparing the figures. This work has been supported by the Bundesministerium für Forschung und Technologie, Bonn, under contract numbers 01 QS 9014 and 50 EE 9413.

References

Alpers, W. and I. Hennings (1984). A theory of the imaging mechanism of underwater topography by real and synthetic aperture radar. J. Geophys. Res. 89: 10,529-10,546

Alpers, W. (1985). Theory of radar imaging of internal waves. Nature. 314: 245-247

Alpers, W. and P.E. La Violette (1992). Tide-generated nonlinear internal wave packets in the Strait of Gibraltar observed by the synthetic aperture radar aboard the ERS-1 satellite.Alpers, W. and B. Brümmer (1994). Atmospheric boundary layer rolls observed by the synthetic aperture radar aboard the ERS-1 satellite. In Proceedings of the First ERS-1 Symposium. ESA SP-359, 753-758

Alpers, W. and B. Brümmer (1994). Atmospheric boundary layer rolls observed by the synthetic aperture radar aboard the ERS-1 satellite. J. Geophys. Res. 99: 12,613-12,621

Alpers, W. and G. Stilke (1995). Observation of an undular bore in the marine atmosphere by the synthetic aperture radar aboard the ERS-1 satellite and by in-situ measurements, submitted to J. Geophys. Res.

Brandt, P. and W. Alpers (1994). The generation and propagation of internal wave forms in the Strait of Gibraltar studied by ERS-1 SAR imagery and a numerical model. In Proceedings of the 1994 International Geoscience and Remote Sensing Symposium (IGARSS'94), Pasadena, Cal., USA, IV: 2023-2025

Christie, D.R. and K.J. Muirhead (1983). Solitary waves: A hazard to aircraft operating at low altitudes. Aust. Meteorol. Mag. 31: 97-109.

Doviak, R.J. and R.S. Ge (1984). An atmospheric solitary gust observed with a Doppler radar, a tall tower, and a surface network. J. Atmos. Sci. 41: 2,559-2,573.

Egger, J. (1985). Die Berliner Nebelwellen. Meteorol. Rdsch. 38: 103-107.

Ford, J.P., J.B. Cimino, C. Elachi (1983). Space shuttle Columbia views the world with imaging radar: The SIR-A experiment. Jet Propul. Lab., Calif. Inst. of Technol., Pasadena, Publ. 82-95.

Fu, L.L. and B. Holt (1982). SEASAT views oceans and sea ice with synthetic aperture radar. Jet Prop. Lab., Calif. Int. of Technol., Pasadena, Publ. 81-120.

Gower, J.F.R. (1993). Mapping coastal currents with SAR, using naturally-occurring surface slick patterns. In Proceedings of the Second ERS-1 Symposium. ESA SP-361, 415-418

Hühnerfuss, H., W. Alpers, K. Richter (1986). On the discrimination between crude oil spills and monomolecular sea slicks by airborne radar and infrared radiometer - possibilities and limitations. Int. J. Remote Sensing. 8: 1,001-1,013

Johannessen, J.A., R.A. Shuchman, K. Davidson, O. Frette, R. Onstott, O.M. Johannessen (1992). Coastal ocean studies with ERS-1 SAR during NORCSEX'91. In Proceedings of the Workshop "ERS-1 Geophysical Validation". ESA WPP-36, 193-197

Keller, W.C., V. Wismann, W. Alpers (1989). Tower-based measurements of the ocean C-band radar backscattering cross section. J. Geophys. Res. 94: 924-930.

Mitnik, L.M. (1992). Mesoscale coherent structures in the surface wind field during cold air outbreaks over the Far Eastern seas from the satellite side looking radar. Extrait de la mer (Bulletin de la Société franco-japonaise d'océanographie), Tome 30 (3): 287-296.

Scherhag, R. (1969). The Berlin fog waves of Oct. 1969. Supplement to the Berlin Weather Map.

Smith, R.K. (1988). Travelling waves and bores in the lower atmosphere: The "Morning Glory" and related phenomena. Earth-Science Reviews 25: 267-290.

Stoffelen, A., D.L.T. Anderson, P.M. Woiceshyn (1992). ERS-1 scatterometer calibration and validation activities at ECMWF: B. From radar backscatter characteristics to wind vector solutions. Proceedings of the workshop "ERS-1 Geophysical Validation", held at Penhors, Betagne, France, April 27-30, 1992, published by the European Space Agency, Paris, France, ESA publication WPP-36 (Aug.92): 89-93.

Thompson, T.W., W.T. Liu, D.E. Weissman (1983). Synthetic aperture radar observations of ocean roughness from rolls in an unstable marine boundary layer. Geophys. Res., Lett. 10: 1,172-1,175.

SAR INTERFEROMETRY AND THE MONITORING OF THE EARTH SURFACE AT CENTIMETER LEVEL

DIDIER MASSONNET

Introduction

Proposed some twenty years ago for the assessment of the topography, Synthetic Aperture Radar interferometry demonstrated its huge potential only recently, when a large amount of high quality data was provided by ERS1, and, later on, by J-ERS1.

Its topographic capabilities were confirmed by these data, sometimes with an unprecedented accuracy. However, an interferometric image, obtained from two radar images taken at different times, reveals the sub-centimetre displacements between the acquisition times once the topographic contribution has been subtracted. This capability was spectacularly demonstrated by the determination of the coseismic displacements of the Landers earthquake. The technique has also been applied to other earthquakes, as well as volcanoes, landslides, artificial or natural subsidence, tectonic fault slip or expansion. Troposphere and ionosphere propagation delays were also characterised by this mean.

Being an advanced measurement tool, SAR interferometry is likely to discover previously unknown phenomena. Surface phase changes have been obtained with Seasat data and repeatedly with ERS1 data, in relation with moisture and surface occupation, indicating more potential applications in environment monitoring.

Synthetic aperture radar signal special features

Synthetic aperture radar (SAR) is often described as an observation tool for which all weather capability is obtained at the cost of unusual radiometric features. SAR is indeed an all weather tool; it is its own source of light and can therefore operate by night and the wavelengths used in the remote sensing field (3 to 25 cm) are not significantly attenuated by a cloudy atmosphere. The quality of the images, often quoted as poor, must not be judged through the criteria of optical images. The wavelength of a radar being somewhat 100000 times larger than the one of visible light, there is no wonder a radar image looks different than a conventional image. Concerning the

Dr. Didier Massonnet is with the
Centre Spatial de Toulouse, CNES
18 Avenue Edouard Belin,
F-31055 Toulouse Cedex, France
Tel: + 33 (61) 27 34 18 *Fax: + 33 (61) 27 31 67*

aesthetic aspect, one could say that a radar image looks poor when compared to an optical image (except in total cloud cover !). Why not say it looks better than an X-ray scanning ?

The real question deals with the information content. Radar is more sensitive to the physical and geometric properties of the ground, such as electrical conductivity, where optical systems reflects its chemical properties. The difference between the two systems is actually deeper than this, and concerns technology as well as physics. The aperture of a radar system may reach 10 meters at most, which is only one hundred wavelengths. The natural resolution of a radar system is therefore poor when compared to the one of an optical system. As an example, the natural along track resolution of the radar instrument of ERS-1 is about 6 km. The poor angular resolution of a radar antenna is not a problem in range, where the resolution is obtained by sorting the samples in range using their travel time. This is why imaging radar's are side looking, thus avoiding ambiguous echoes coming from each side at the same range. The situation differs in azimuth (the direction parallel to the satellite track) where the 6 km wide antenna pattern is the complex summation of, typically, 1000 end pixels if we assume that the final resolution goal is 6 metre. To achieve this goal, the radar must send pulses at a sufficient rate to ensure a ground sampling better or equal to 6 metre. It is the task of the ground processing to sort out the contribution of an individual pixel by analysing the somewhat 1000 pulses to which it participated, where it was mixed with its neighbours. This is done by taking advantage of the delay phase law experienced by a target as it closes, and then recedes, from the radar. The phase law is quadratic and characterised by a Dirac-like autocorrelation function. The final resolution is therefore obtained by complex correlation with a replica of the azimuth phase law.

The processing does not change the complex nature of the signal, which ends up as high resolution pixels made of complex numbers.

3. Signification of the phase in a radar image; interferometry

Each pixel in the image can be considered as a collection of individual targets. Each target is characterised by its amplitude and phase. The amplitude is typical of the gain of the target with respect to the direction of the radar. The phase reflects the conductivity or the internal geometric organisation of the target. The complex number which represents the pixel is the sum of the complex contributions of the individual targets. However, in this sum, each target is multiplied by a phase term which reflects the position of the target within the pixel. This phase term comes from the round trip delay experienced by the wave in reaching the target, compared to a reference delay (for instance the pixel centre). Each contribution is also modified by the combined impulse response of the instrument and the processing. This impulse response modifies the contribution of each target in amplitude and, possibly, phase depending on its position within the pixel. It is clear that a global displacement of the pixel would create the same phase delay for all the individual targets and would be observed at pixel level. We will therefore break down the phase of a radar image into three types of information :

- a macroscopic phase delay, which contains the geometric delay common to all the elementary targets of a pixel. We will consider it to be the fractional number of phase cycles obtained from the round trip delay due to the range of the centre of the pixel with respect to the instrument.

- a physical phase shift representing the electrical conductivity of the targets, assuming it varies collectively, for instance due to a global change in moisture.

- an organisation phase term, which combines the influence of the relative positions of the elementary targets within the pixel and the amplitude and phase laws applied by the instrument and the processing (i.e. the modulation transfer function).

In a radar image, exchanging the positions of two elementary targets within a pixel has much more consequences than doing the same in an optical image. This explains the fact that a same collection of targets (i.e. an homogenous landscape) may exhibit a large variation of the amplitude (the speckle noise). The third term of phase is clearly out of reach since, by definition, we have only one piece of complex information per pixel. This third term also explains why the phase of a radar image is uniformly distributed between 0° and 360° in general. This uniformity is even considered as a criterion of good image quality and is part of the standard radar image quality tests. Yet the phase is half the information provided by a radar and is it very tempting to use it. This is the purpose of radar interferometry, the principle of which is to eliminate the contribution of the third term between two radar images in order to exhibit a possible variation of the two first terms. Radar interferometry has been proposed and demonstrated 20 years ago with the data of an airborne imaging radar [1] and more recently [2] using space borne data.

To apply this idea, we simply have to subtract the phases of two radar images covering the same area in order to eliminate the third term. The challenge is only to keep the third term constant. In a simultaneous acquisition, such as the one in [1], where two antennae are mounted on a plane, the organisation of the pixel in each image can change if the processing applies a different complex weighting to each image. This point is avoided by proper image processing techniques such as a very precise registration of the complex radar images, processed with the same impulse response function. Another reason for the third term to change is a different incidence of observation in the two images. The range between two targets included in the same pixel will change according to the incidence angle of the wave. If the range differs in the two images by more than a fraction of the radar wavelength, the third phase term will build differently in the two images and the subtraction trick will not work. This places a limit on the variation of the angle of incidence which may, alternatively, be expressed by a limit on the geometric dilatation or contraction from one image to the other.

Working with radar images acquired at different times, such as in [2], also requires the third phase term to be stable in time. This implies that the elementary targets must stay at the same place, at least relative to each other, and be the same. For instance, water surface will never give the same third phase term at different times, deciduous trees will not be the same targets if observed six months apart, a snow covered landscape will not be the same as before the snow fall, etc. However,

in principle, one cannot exclude major changes in the targets which would keep the third term constant, or equivalently, to change it by the same phase in all the pixels of an image. If a field of plants blossoms and the main radar target becomes the flowers in the second image rather than the leaves in the first image, the third term could remain stable or show a constant change within the image if the geometry of the plants is such that there is always the same distance between the flower and the leaf. This phenomenon has never been evidenced but would explain some strange properties we observed, such as large variations of the amplitude between the two radar images to be compared which do not prevent the interferometric effect.

4. The geometric content of the phase and its applications

Assuming that the conditions for proper interferometric combinations have been observed and that there is no contribution from the second phase term, which is to say that the physical properties of the elementary targets did not change, the difference of phase between images contains a geometric information. The image formed by the difference of the phases of two radar images of the same area, once they have been registered, is called an interferogram. The main feature in a raw interferogram is a family of phase cycles, or fringes, parallel to the satellite track. They denote the change of range caused by the different angle of incidence in the two images, taken from slightly different points of view (in practice less than one kilometre apart for space borne images). The "condition of interferometry" of the previous paragraph can be expressed as a condition on the fringe density, which must be kept well below one

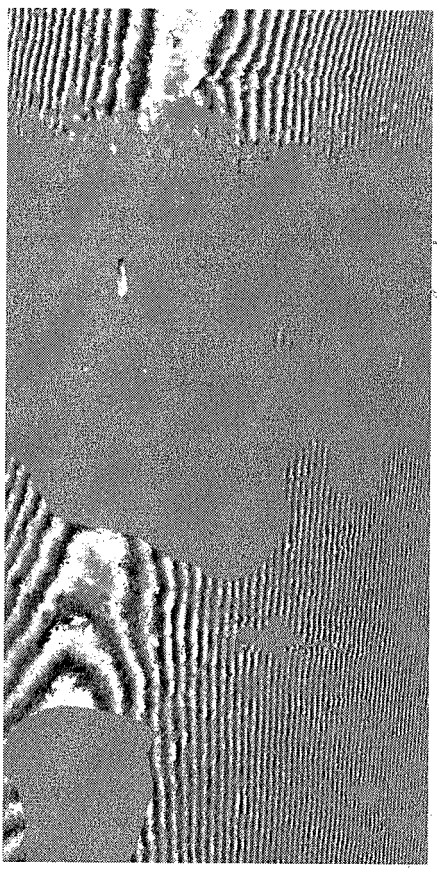


Figure 1 Example of "orbital fringes", typical of the satellite trajectories during the two data takes. The satellites travelled from bottom to top on the left side of the image. The bulk of the orbital contribution consists in a very dense fringe pattern across the image. The contribution of the one meter or so closing of the two satellite tracks along the 100 km long image is represented by fringes disappearing on the sides of the image. Each fringe corresponds to a closing of half a radar wavelength, or 28 mm. The two radar images have been acquired near Saint Petersburg on September 25 and October 1st, 1991 by the European satellite ERS1. They have been processed at CNES. The places where the fringes are not clear did not show the necessary stability. This is the case for the lakes.

fringe per pixel. These fringes (figure 1) are easily modelled and subtracted from the interferogram. If the two tracks are not exactly parallel, if they are closing for instance, the resulting phase delay reduction can be observed. A proper modelling of the orbits allows the elimination of this geometric contribution. The determination of the relative position of the two tracks with a centimetre accuracy is a by-product of SAR interferometry. A second order effect is caused by the topography, which creates a stereoscopic effect when observed from two slightly different points. This stereoscopic effect is tiny, given the proximity of the tracks demanded by interferometry. It is however observed because the accuracy is on the order of a fraction of the phase, or a few millimetres. Radar interferometry has been essentially introduced as a mean to compute topography [1,2]. Another, much more spectacular, application of SAR interferometry is the direct measurement of a displacement, provided it occurred in between the acquisition of the radar images. In this case, unlike the above "stereoscopic" effect, the displacement is directly recorded as a phase shift, with the typical accuracy of a few millimetres generally allowed by the available signal to noise ratio. This technique cannot be applied with simultaneous interferometry (i.e. not with [1]) and must, therefore, cope with the surface change possibly experienced by the ground during the time elapsed between the data take, which we may call the temporal loss of coherence of the surface. Another feature of this technique is its being unique in the sense that there are other means for precise orbit computing or digital elevation model derivation, but no other technique, so far, can map displacements with a precision of a few millimetres with such a dense spatial sampling as the one allowed by radar imaging. All these reasons made CNES undertake a research

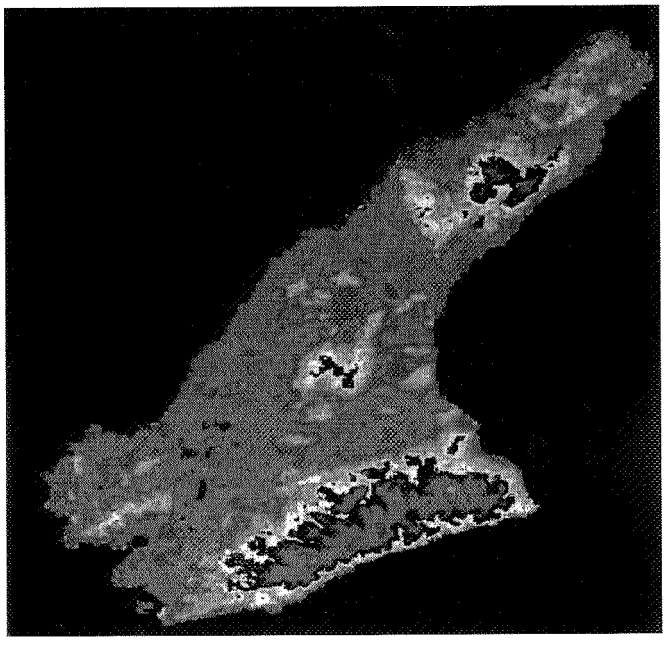


Fig. 2 - Example of topography revealed by SAR interferometry. The image shows the interferogram obtained on Awaji island [7], in Japan, using two radar images of the Japanese J-ERS radar satellite. The images have been acquired on May 31 and July 14, 1993 and were part of the investigation program proposed by CNES to NASDA (the Japanese Space Agency). Given the orbital geometry, a full cycle corresponds to a 300 m change in altitude. This site was the epicentre of the disastrous earthquake which damaged Kobe in January 1995. However, no displacement can be recorded here since the images were both acquired before this event. The island is 50 km long.

program in this field in 1985, after a paper study indicated its possibility [3]. From the beginning, it was decided that we would not mix the topographic aspect of interferometry and its displacement aspect. This is why we worked along the idea of digital elevation model elimination, where topographic fringes are simulated from the best available elevation model and the orbital estimations. Simulated fringes are then subtracted from the interferogram, which reveals displacements. This method is in contrast with the three pass method developed by JPL [4], which does not require a model of the topography, but three images meeting the conditions for

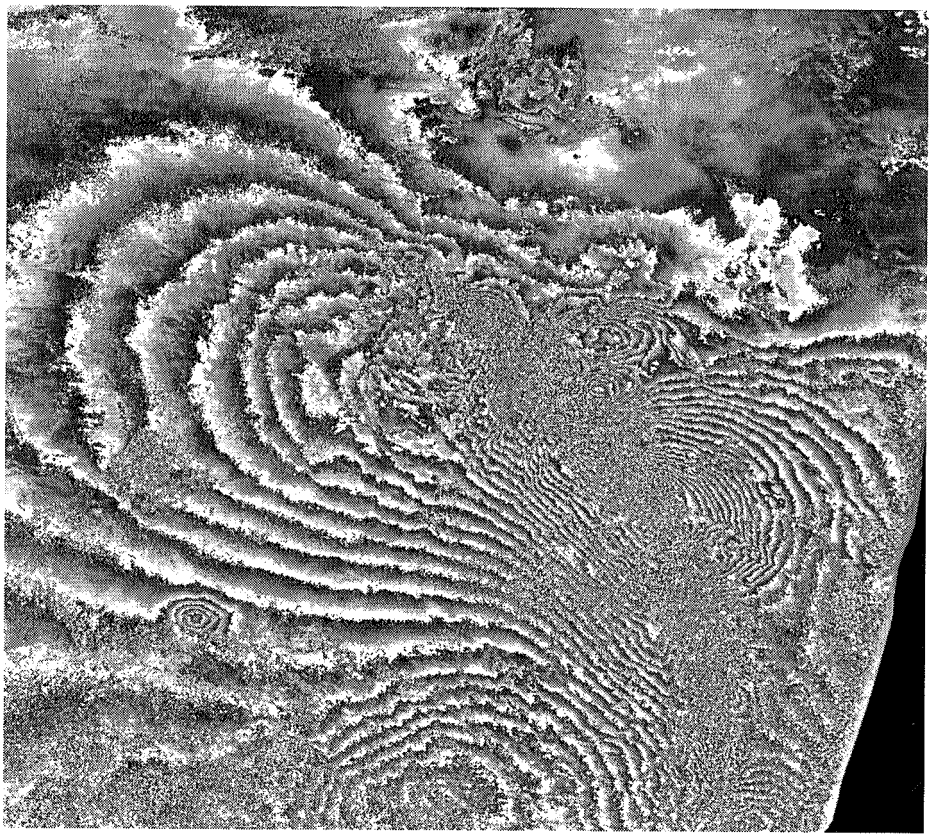
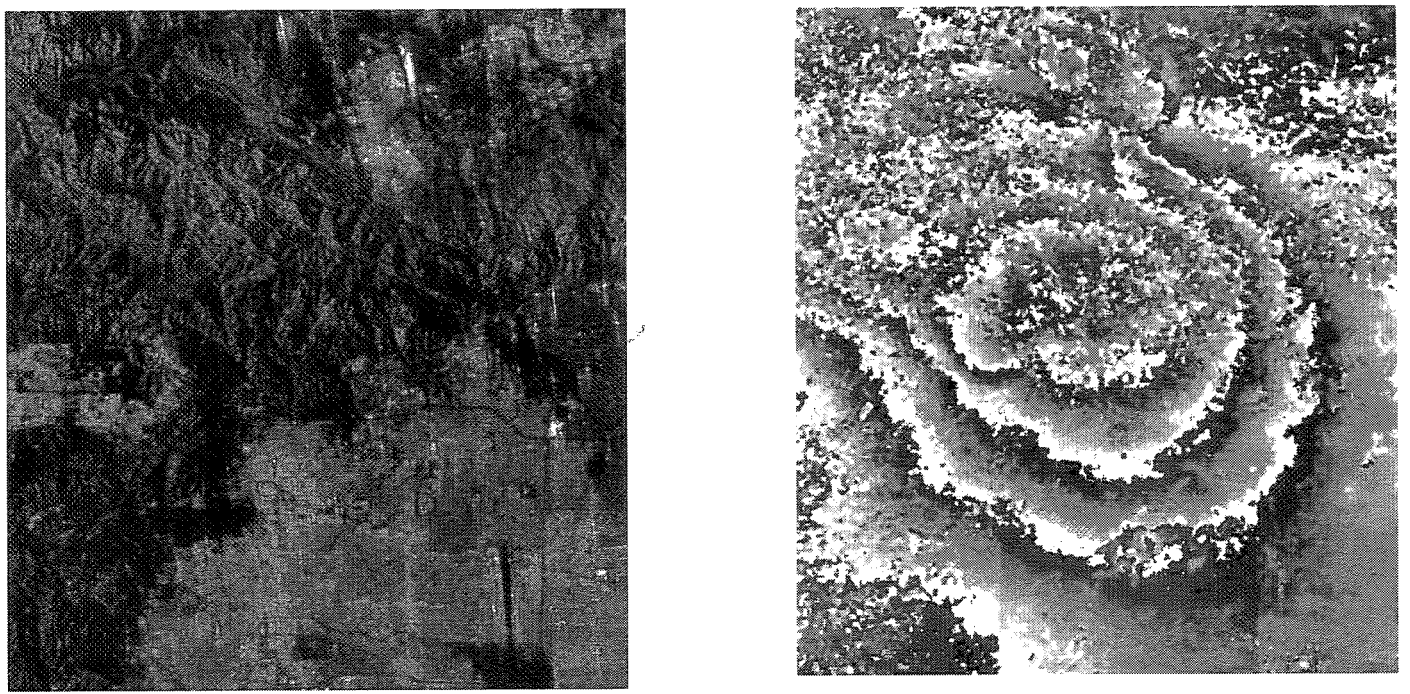


Fig. 3 - Example of a large displacement field measured by interferometry [8]. The two ERS-1 images were acquired on April 24, 1992 and June 18, 1993 and combine nicely in an interferogram despite the 14 months elapsed, which did not alter the surface in this desert environment. Each fringe corresponds to a displacement of 28 mm toward the satellite. The contribution of the topography has been removed using the digital elevation model elimination method and a USGS digital elevation model. The large scale deformation field has been caused by the (Mw=7.3) Landers earthquake, on June 28, 1992. The small concentric fringes have been caused by a (Mw=5.4) replica on December 4, 1992. The discontinuity in the fringe pattern which can be observed across the image has been caused by a slip, probably coseismic, on the Lenwood fault. The slip is half a fringe, or 14 mm, in range. The image illustrates the ambiguous nature of SAR interferometry results. Fringes must be counted to reach the final value of the displacement.

Northridge Earthquake by JERS-1



SAR Image *(NASDA data, CNES processing)* *Displacement field*

Fig. 4 - Another example of coseismic displacement, this time caught by J-ERS. The earthquake took place in Northridge, North of the city of Los Angeles, on January 17, 1994. We used two radar images provided by NASDA and combined them with an accurate 30 m grid USGS digital elevation model. The radar images were acquired on April 30, 1993 and July 14, 1994. The observed four fringes on the interferogram correspond to quite a large displacement since half the wavelength of the L-band J-ERS is 115 mm. This event has been reported in [9] using the same data. On the left, we have the radar amplitude on Northridge.

Fig. 5 - Perspective view of Mount Etna, built from a topographic model and a radar image acquired by ERS-1. Fringes obtained by an interferometric survey of the volcano have been added. They were formed using two ERS-1 images acquired one year apart [10]. The displacement ranges from zero to four fringes as we close to the top of the volcano and reaches two fringes on the volcanic edifice alone. These fringes indicate a deflation of Mount Etna during the year of observation, due to the draining of the deep magmatic chamber at the end of the eruption.

interferometry.

The technique was applied to a variety of phenomena where moderate displacements are expected, such as coseismic displacements fields [5, 8, 9], deformation of volcanoes [10], ice plate motion [6] or artificial subsidence due to mining or underground nuclear blasts (Figure 2, 3, 4 and 5). More studies are underway concerning more subtle displacements such as continental drift is rifting areas. The advantages of the technique are clearly its dense spatial sampling and its accuracy, in the sub-centimetre range. The limitations are both physical - the surfaces do not always present the required level of stability - and geometric - only the line of sight displacement is measured. This point has to be compensated by *a priori* knowledge of the three dimensional aspect of the displacement or by multiple observations along different angles.

Another limitations is the ambiguous nature of the fringes, which must be counted to restore the displacement field in full. In some cases, this is not only a lack of comfort : a field of fringes must be continuous to be "unwrapped". It happened that the deformation of an off-shore island could not be related to the deformation of the nearby mainland because the number of fringes obscured by the straight (water is not coherent) could not be observed nor guessed safely.

5. Atmospheric artefacts and new potential uses

Any change in the propagation conditions of the atmosphere will contribute to the change of range observed in SAR interferometry. The image where the problem occurred is identified because all the interferometric pairs to which it participated are affected by the artefact with a constant level. Once the faulty image has been identified, one can check, knowing the exact time of the data take, the meteorological conditions at that time and prove the case. In figure 7, the 5 to 10 km wide irregular, circular pattern, amounting to three fringes, has been linked to troposphere turbulence, caused by the formation of thunderstorm clouds. In this case, the wave is delayed. We attributed other examples of propagation artefacts to local neutralisation of the ionosphere because the wave was found to travel faster in some places than in the surrounding. We called artefacts these phenomena quite improperly, because they reflect a real physical situation. They are artefacts with regard to the measurement of ground geophysical phenomena. We detected real instrumental artefact by analysing large portion of radar data where groups of fringes cannot possibly be related to atmospheric propagation problems nor topographic or displacement contribution. These fringes are perpendicular to the track of the satellite which indicated that the observed phase changes are almost constant within a pulse. The only explanation proposed [11] is a slow drift of the carrier frequency of the radar. This drift is well within the engineering specification of the satellite which has never been designed for interferometry. We have other examples showing that this new technique, which has been developed in a quite opportunistic manner, will now generate its own specifications. The example of figure 6 suggests building radar satellites with a much more stable internal clock. The use of several frequencies may also be proposed to cope with ionosphere shortening.

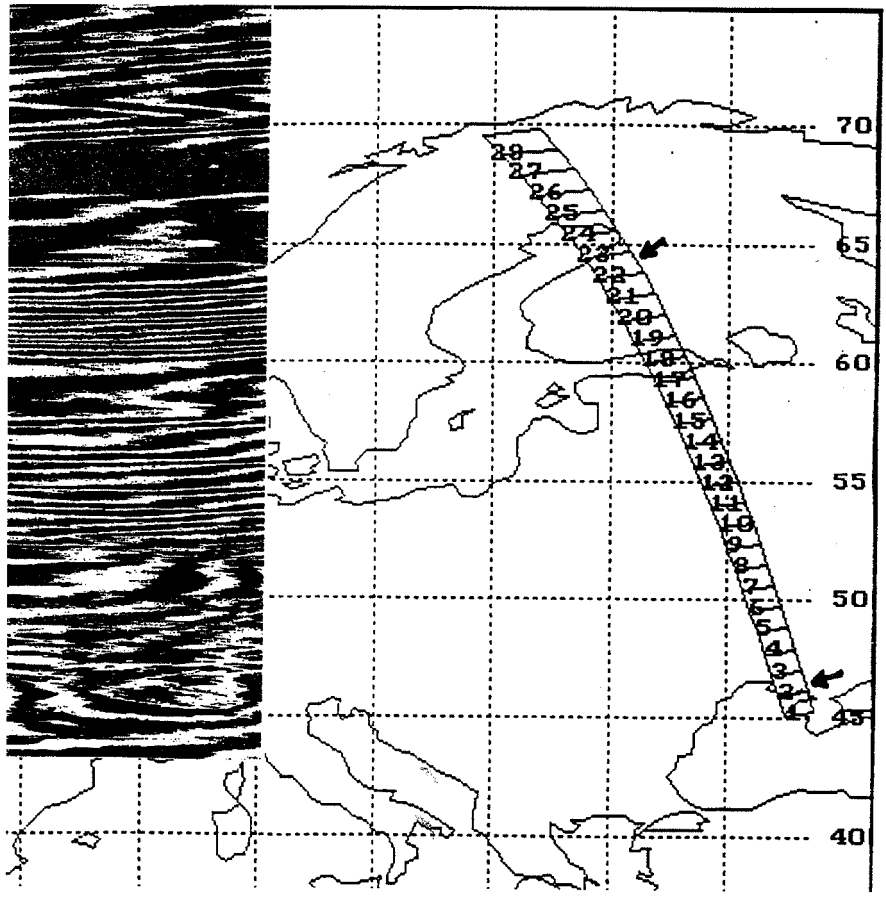


Fig. 6 - Typical interferometric artefact caused by a drift in the radar clocks [11]. SAR oscillators have been designed for keeping coherence on the typical one second of data required for image reconstruction. Radar interferometry demands a much longer stability. On this 2000 km long data segment, bracketed by arrows on the map and deformed on the left side for the sake of representation, we observe groups of fringes perpendicular to the track. These fringes cannot by attributed to topography on this very flat landscape observed by ERS-1 from two very close tracks. The 10 or 15 fringes of each group cannot be attributed to any atmospheric propagation problem due to the magnitude it would involve. Perpendicular fringes are a typical signature of perturbations associated with the spacebus or the radar itself. With only two images, It is impossible to say which image was affected by the drift or if both contributed to it.

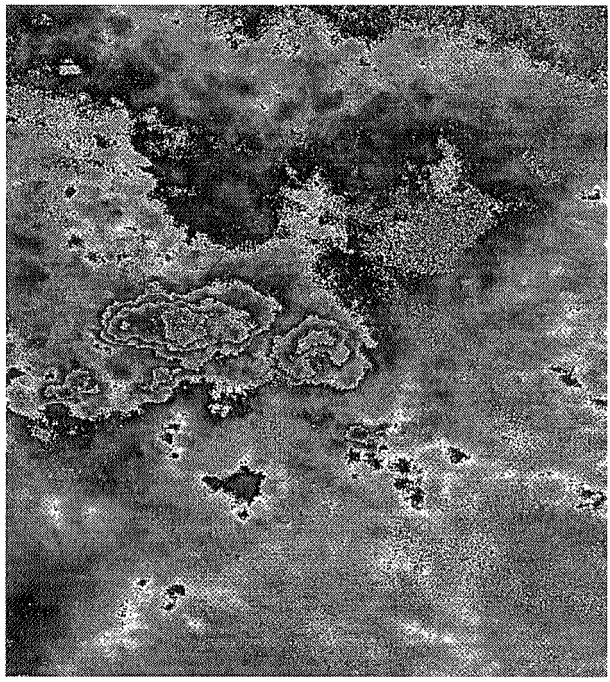


Fig. 7 - Atmospheric depth heterogeneity observed by SAR interferometry [12]. These propagation problems are linked to one particular image and appear in each interferometric combination in which it is involved. They appear with a constant magnitude regardless of the orbital configuration (as for topography) and of the time elapsed between the acquisitions (as for displacements). This particular phenomenon was caused by thunderstorm clouds forming in August 1993 over the Mojave desert (California). In this case, the turbulent atmosphere lengthened the wave path. We observed other examples where the travel was shortened with respect to surrounding areas. This shortening is attributed to a local neutralisation of ionosphere.

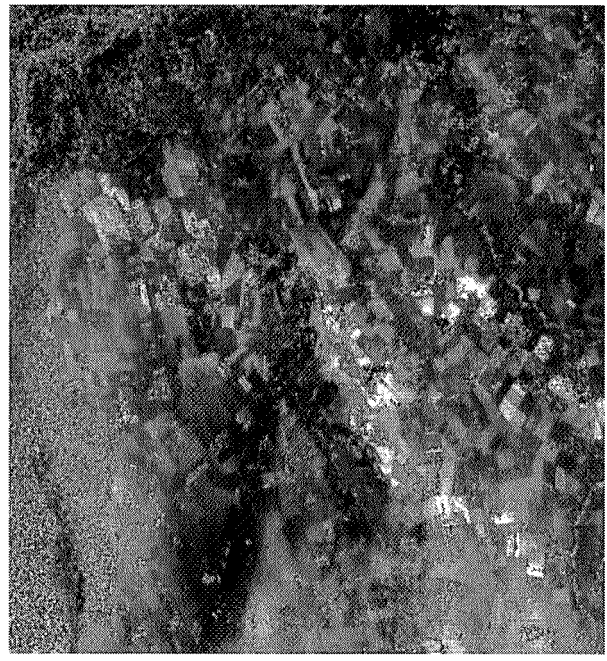


Fig. 8 - Surface phase changes associated with moisture on agricultural fields. This image was obtained from ERS-1 data, but the effect has been first observed with L-band Seasat data [4]. In the Seasat data case, the moisture was caused by irrigation. Here, it is likely to have been caused by rain. A brightness difference can be observed in the amplitudes of the images which form the interferogram over the area where surface phase change is sizeable. It is not clear whether the change is geometric (due to soil swelling under the action of moisture) or radio-electrical (due to moisture induced conductivity change).

The phase surface changes first observed in [4] using Seasat data and illustrated in figure 8, where we processed ERS-1 data, are not an artefact. We consider them as a very promising application which may allow SAR interferometry to become an investigation tool in crop monitoring. It is too early to assess the potential of SAR interferometry in this field, but the technique may end up by finding applications in most of the remote sensing domain.

6. Conclusion

SAR interferometry was first developed to compute topography, but evolved to other fields where it is really unique. The first radar satellites were not designed for this technique. They provided data in a quantity and quality which not only permitted many opportunistic and spectacular demonstrations, but allowed us to contour the specifications of a dedicated satellite system. Such a system would strengthen some points of design, such as clock accuracy, and relax others, such as radiometric precision and calibration. Making SAR interferometry enter the operational field is well within current technologies. It would allow a new dimension in Earth monitoring and create the "meteorology" of solid surface phenomena.

References

[1] L.C.Graham, "Synthetic interferometer radar for topographic mapping", Proc. IEEE vol. 62 n° 6 - June 1974

[2] H.A.Zebker & R.M.Goldstein, "Topographic mapping from interferometric SAR observations", Journal of Geophysical Research vol. 91 n° B510 April 1986

[3] D.Massonnet, "Etude de principe d'une détection de mouvements tectoniques par radar", CT/PF/TI/AS n° 326 (CNES memo.) - 16 November 1985

[4] Gabriel A.K., Godstein R.M.& Zebker H.A., "Mapping small Elevation Changes over large Areas : differential radar interferometry", Journal of Geophysical Research 94 pp. 9183-9191, 1989

[5] D.Massonnet, M.Rossi, C.Carmona, F.Adragna, G.Peltzer, K.Feigl & Th.Rabaute, "The displacement field of the Landers earthquake mapped by radar, interferometry", Nature vol. 364 n° 6433 - 8 July 1993

[6] Goldstein, RM., Engelhardt H, et al., "Satellite Radar Interferometry for Monitoring Ice Sheet Motion : application to an Antarctic Ice Stream", 1993, Science, 262, pp. 1525-1530

[7] M. Rossi, B. Rogron, D.Massonnet, "J-ERS1 : Image Quality and interferometric Capabilities", IEEE (TGARS) (Submitted 1994)

[8] D.Massonnet, K.Feigl, M.Rossi, F.Adragna, "Radar interferometric mapping of deformation in the year after the Landers earthquake", Nature vol. 369, 227-230 - 19 May 1994

[9] M.Murakami, M.Tobita, T.Saito, H.Masaharu, "Coseismic Crustal Deformation of 1994 Northridge California Earthquake Detected by Interferometric J-ERS-1 SAR", Journ. of Geophys. Research - Submitted 15 November 1994

[10] D.Massonnet, P.Briole, A. Arnaud, "New insight on Mount Etna from an 18 months interferometric monitoring", Nature (Submitted 1994)

[11] D.Massonnet, H. Vadon, "ERS-1 Internal Clock Drift Measured by Interferometry", IEEE (TGARS) vol. 33 n°2, 1-8 (03-1995)

[12] D.Massonnet, K. Feigl, "Discrimination of geophysical Phenomena in Satellite Radar interferograms", Geoph. Res. Letter (in press)

CURRENT DEVELOPMENTS IN VLBI ASTRONOMY ON THE GROUND AND IN SPACE

R.T. SCHILIZZI

1. Introduction

World-wide Very Long Baseline Interferometry (VLBI) is undergoing a major expansion in capability at the present time.

The new US Very Long Baseline Array (VLBA) is already producing eye-catching results even before its full capabilities have been brought on line. Its counterpart in Europe, the European VLBI Network (EVN) is carrying out a major upgrade of the radio frequency performance and flexibility of its member telescopes and their VLBI equipment, as well as constructing a new state-of-the-art correlator. In the southern hemisphere, the Australian VLBI array is also expanding its capabilities with a new correlator and recording terminals, and together with radio observatories in the Asia Pacific region will begin regular coordinated VLBI observations for the first time in 1995 as the Asia Pacific Telescope (APT). Millimetre-wave observatories across the globe have also banded together to form the Coordinated Millimeter-VLBI Array (CMVA) which will observe at 86 GHz three times a year starting in 1995.

And perhaps the most spectacular of all, the space VLBI era will begin in 1996 with the launch of the Japanese Muses-B satellite carrying an 8-m diameter radio telescope into Earth orbit. The mission, called VSOP (VLBI Space Observatory Programme), will combine the space borne antenna with its ground-based counterparts around the world to form radio interferometers of dimension 32000 km and maximum angular resolving power of 80 micro-arcseconds. A year or two thereafter, Russia plans to launch its 10-m diameter RadioAstron satellite into an even higher orbit than VSOP to provide a further increase of three in angular resolving power to 30 micro-arcseconds.

In addition to these instrumental developments, considerable progress has also been made in observing and data reduction techniques, so that far weaker sources can be imaged than was possible even two or three years ago. With the VLBA leading the way, images of strong sources can be made with a quality approaching that of connected element arrays like the VLA and Westerbork.

Prof. R.T. Schilizzi is with the
Joint Institute for VLBI in Europe (J.I.V.E.)
P.O. Box 2,
7990 AA Dwingeloo, The Netherlands
Tel: + 31 (5219) 7244 Fax : + 31 (5219) 7332

In this contribution to the URSI 75thAnniversary Symposium, I will review these developments in ground and space VLBI, and highlight the new capabilities in terms of recent scientific results. I will not cover the technical basis of VLBI nor the theory of interferometry in radio astronomy; readers are referred to the book by Thompson, Moran and Swenson (1986) for that information. I will also not touch on the use of VLBI in geodesy. A very comprehensive account is given in volumes 23, 24 and 25 of Contributions of Space Geodesy to Geodynamics published by the American Geophysical Union in 1993.

2. Ground-based VLBI arrays

2.1. Organisations

VLBI is both a national and international endeavour. The VLBA is run as a single instrument by the US National Radio Astronomy Observatory (NRAO), whereas the EVN is run as a coordinated instrument by the European Consortium for VLBI. The Australian VLBI array is coordinated by the Australia Telescope National Facility (ATNF), and a Consortium for the Asia-Pacific Telescope is in the process of being established. The space VLBI missions are led by the Institute for Space and Aeronautical Science (ISAS) in Japan, and by the Astro Space Center of P.N. Lebedev Physical Institute (ASC) in Russia, but both missions have International Scientific Councils to advise them on all aspects of the missions. Other organisations also are contributing to the mission hardware and operations (see section 3). Coordination of the response of the ground VLBI arrays and other major unaffiliated telescopes torequests for observing time from the space missions is being carried out by the URSI Commission J Global VLBI Working Group (GVWG).

2.2. VLBA

After many years of preparation, the US VLBA came into regular operation in 1994. The array consists of ten 25-m telescopes at sites within the United States and its territories, chosen for optimum image quality as well as ease of access for operations and maintenance (Kellermann and Thompson, 1985). Figure 1 shows the locations of the VLBA telescopes on the world map. In the US frame of reference, they are to be found in New Mexico (2), Arizona, Texas, Iowa, California, Washington, New Hampshire, Virgin Islands and Hawaii. The Array Operations Center is at Socorro, New Mexico. The locations were chosen to provide both high resolution and the most uniform aperture coverage that can be practically achieved (see Figure 2 left). The aperture plane or uv-plane coverage is one of the fundamental figures of merit of an interferometer and indicates the degree of filling of the hypothetical telescope aperture formed by the interferometers as they rotate with the Earth under the radio source. In general, the better filled the u-v plane, the better the image quality. Maximum baseline lengths of 8000 km are achieved in the VLBA.

The VLBA incorporates many features designed to yield maximum sensitivity for the observations and reduce the effects of systematic errors. A new high precision antenna was developed by the

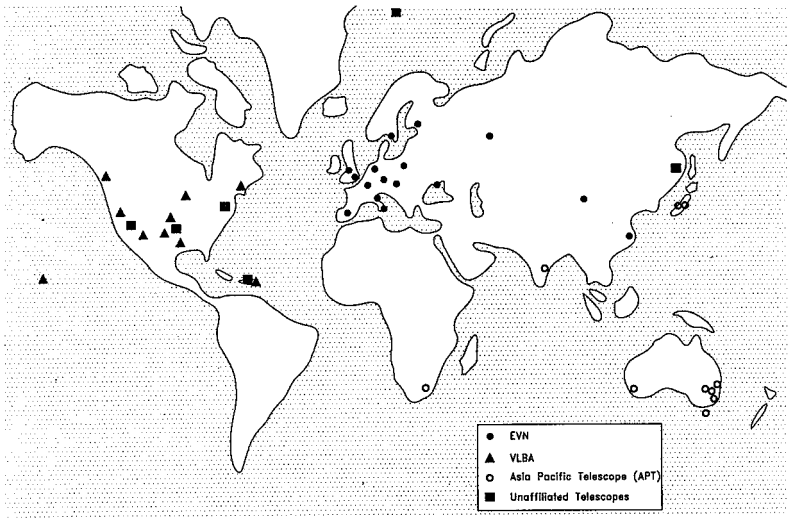


Fig. 1 - The world of VLBI. filled circles: EVN; filled triangles: VLBA; open circles: Asia-Pacific Telescope; filled squares: unaffiliated telescopes.

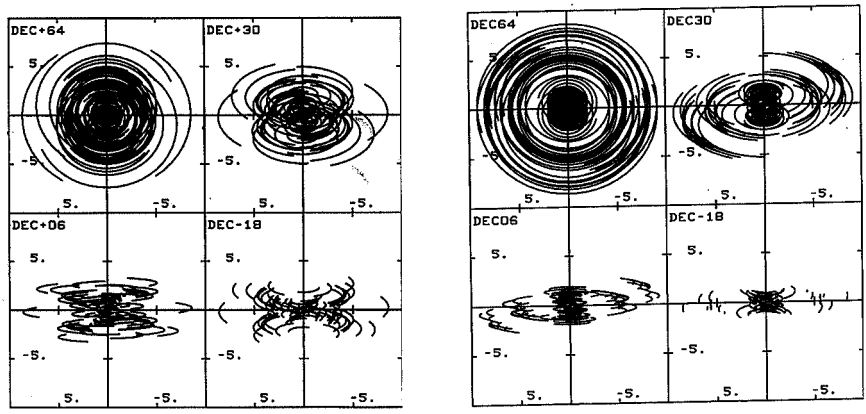


Fig. 2 - left: uv-coverage for the VLBA at declinations of 64°, 30°, 6°, and -18°. Units are 103 km. right: uv-coverage of the EVN at 22 GHz. Telescopes in the EVN are Yebes, Jodrell Bank, Cambridge, Effelsberg, Onsala, Bologna, Noto, Mets"ahovi, Torun, Urumqi, and Shanghai (see Table 1). Units are 10^3 km.

NRAO allowing efficient operation up to 43 GHz, and with reduced efficiency up to 86 GHz. Each array element is equipped with dual polarization receivers covering nine frequency bands between 327 MHz and 43 GHz; typical system noise temperatures at zenith are between 30 and 50 K for observing frequencies between 1.5 and 15 GHz, rising to 220 K at 327 MHz and 115 K at 43 GHz. Changes of receiver at the secondary focus are achieved in less than 30s by rotating

the subreflector which allows multi-frequency observations to be obtained quasi-simultaneously. Since the array elements are identical, the instrumental polarization properties of the feed/receiver systems are uniform (and low: 1-2% on axis), and the filter characteristics are also uniform. Both these properties help reduce the effects of calibration errors in the data analysis thereby increasing the effective sensitivity of the array.

COUNTRY	INSTITUTE	LOCATION	DIAMETER (m)	FREQUENCY RANGE (GHz)
Spain	National Astronomical Observatory	Yebes	14 40[1]	2.3 - 43 5 - 100
UK	Nuffield Radio Astronomy Laboratories	Jodrell Bank	76 25	0.3 - 1.6 1.4 - 43
		Cambridge	32	1.4 - 43
Netherlands	Netherlands Foundation for Research in Astronmy	Westerbork	14 x 25[3]	0.3 - 8.4
Germany	Max-Planck-Institut für Radioastronomie	Effelesberg	100	0.6 - 90
	University of Bonn	Wettsel	20	2.3 - 8.4
Sweden	Onsala Space Observatory	Onsala	25 20	1.4 - 5 2.3 - 100
Italy	Istituto do Radioastronomia	Bologna Noto	32 32	1.4 - 43 1.4 - 43
Poland	Torun Radio Astronomical Observatory	Torun	15 32[3]	0.3 - 15 0.3 - 43
Finland	Metsähovi Radio Research Station	Helsinki	14	22 - 100
Ukraine	Crimean Astrophysical Observatory	Simeiz	22	0.3 - 43
China	Shanghai Observatory Urumqi Observatory	Shanghai Urumqi	25 25	0.3 - 22[4] 0.3 - 22[4]

Table 1 - EVN telescopes with wideband recording equipment
Notes: [1]funding is being sought
[2]equivalent diameter, 93 m
[3]expected to be in operation in mid-1995
[4]upper limit to the frequency range may extend to 43 GHz

The RF signals from the receivers are mixed down to baseband and divided into as many as 16 channels with bandwidths from 16 MHz to 62.5 kHz. The filtered signals are sampled at rates ranging from 0.5 to 32 Msamples/sec with sample quantization at two or four levels. A peak data rate of 256 Mbits/sec is possible per acquisition system. The sustainable data rate is 128 Mbits/sec which fills a tape in about 10h. This rate is set by the lack of staff at the individual antennas to change tape any more frequently. With two recorders per site, unattended operation is possible for most of one day at this rate, but peak rates of 512 Mbits/sec are possible with the two recorders in operation simultaneously. Tapes can be recorded at 33 or 56 kbits/inch per track on 32 data tracks and up to 16 passes of the tape past the headstack are possible. Each tape has a capacity of 5.4 terabits.

The VLBA correlator at Socorro has 20 station inputs to accomodate global observations involving the 10 VLBA antennas with other networks, primarily the EVN. The 20 inputs will also allow correlation of 512 Mbits/sec recordin g using two recorders at the 10 VLBA sites, as well as the simultaneous correlation of two normal 10-station observations; neither of these two modes is yet operational at the correlator. The correlator first takes a fourier transform of each baseband channel from each station and then cross multiplies the resulting spectra to generate the correlation functions, a process first developed by Chikada et al. (1987) at Nobeyama in Japan, and also used in the VSOP correlator now under construction. The VLBA spectral-domain or FX correlator has 160 FFT engines each of which can operate at resolutions ranging in binary steps from 32 to 1024 spectral points. The 1024-point resolution (512 points in frequency space) is available for up to 4 channels per station in a single pass through the correlator for a 20 station array, but decreases by a factor of two for 8 channels or for dual polarization data. For a 10-station array, the maximum is 1024 spectral points.

The VLBA is pioneering a number of advances in VLBI: wide frequency coverage on one array, receiver agility essential for measurements of the continuum spectra of variable sources, ease of use, and high quality imaging (see section 4.7), while the dedicated nature of the array will make it easy to respond to Targets of Opportunity and to the demands of space VLBI. The NRAO has committed 30% of VLBA observing time to support the most highly rated VSOP and RadioAstron observing proposals, and will evaluate other space VLBI proposals in competition with purely ground-based VLBI proposals for possible additional time. It is undertaking the necessary changes to the online software in the correlator to enable it to cope with the higher fringe and delay rates encountered with orbiting interferometer elements.

2.3. The European VLBI Network

The second major VLBI facility in the northern hemisphere is the European VLBI Network (EVN). It is a part-time network operating for up to 17 weeks per year using national radio telescopes in 12 countries (see Figure 1). The EVN was formed in 1980 by a consortium of five of the major radio astronomy institutes in Europe. Since 1980, the EVN and the Consortium has grown to include 12 active institutes with 16 radio telescopes in 10 different countries from Spain in the west to China in the east. Table 1 lists the institutes and their telescopes, giving information on the diameter (or equivalent diameter for the Westerbork array) and the VLBI frequency range possible with the optics and surface accuracy. Depending on the frequency, between 7 and 10 telescopes can observe currently. The upgrade programme described later in this section includes construction of new receivers at a number of institutes so that up to a maximum of 12 telescopes will observe at any particular frequency in 1996 when the first space VLBI satellite is launched.

Considerable effort went into the optimum placing of the VLBA antennas for uv-coverage. No such licence was available for the EVN; the telescopes are where they are. Figure 2 depicts the uv-coverage for the EVN at 22 GHz in 1996. The coverage is good at northern declinations where radio sources pass overhead at the majority of telescopes, but gaps appear at lower declinations and the coverage becomes more one-dimensional the closer the source is to equatorial. A

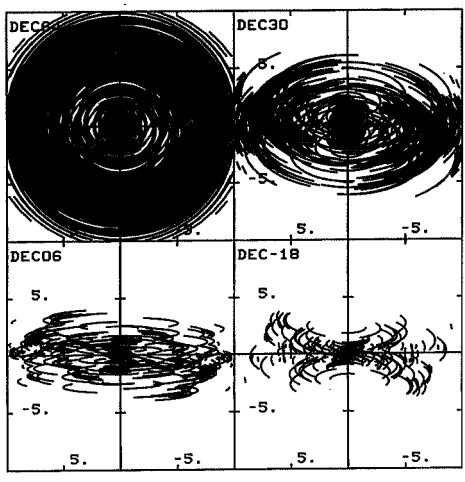


Fig. 3 - uv coverage for 20-station EVN +VLBA observations at ≤8.4 GHz. Units are 10^3 km.

telescope near the equator between western Europe and the South African telescope at Hartebeesthoek would help solve this problem. The maximum baseline length in the EVN is about 9000 km.

Making full use of the VLBA correlator (and later the new EVN correlator) allows us to construct VLBI arrays with superb uv-coverage. Figure 3 shows as an example the uv-coverage for a 20- station array composed of the EVN and the VLBA. Such uv-coverage gives us the basis on which to image compact radio sources with unprecedented quality.

A major upgrade of EVN facilities is in progress, made possible by funding from multi-national sources in Europe and from the European Union. The main aspects of the upgrade are the construction of a 16-station data processor at the newly established Joint Institute for VLBI in Europe (JIVE) in Dwingeloo (the Netherlands), upgrade of the data acquisition terminals at the individual telescopes to allow recording at 1024 Mbit/sec (the MkIV standard), and the employment of Support Scientists at JIVE and at some individual observatories to provide assistance for users of the EVN and other arrays. In addition, a number of nationally funded upgrades are being carried out at EVN member stations including receivers in new frequency bands as mentioned earlier, and replacement of older receivers with state-of-the-art HEMT-based systems.

The VLBA is pioneering a number of other instrumental features that will become standard in VLBI, such as the ability to switch receivers on timescales of less than a minute, and well-understood bandpass and polarization properties for the array. The EVN has taken on the task of achieving the same level of performance in these areas where possible.

The MkIV data acquisition system provides for up to 16 baseband channels with channel bandwidths of 62.5 kHz to 16 MHz like the VLBA. Sixty four tracks are recorded on tape simultaneously and since the tape can run at twice the speed of the VLBA acquisition system (320 inch/sec), the maximum bit rate per recorder is 1024 Mbit/sec. One tape will last about 1.25h at this rate. The increased bit rate of recording means increased sensitivity for continuum observations. The MkIV formatter will be capable of generating VLBA modes so that full backwards compatibility with the VLBA will be available for joint EVN-VLBA observations. The design of the MkIV acquisition system is due to the MIT Haystack Observatory, and the MkIV upgrade in Europe is being implemented by a team from the EVN institutes.

The EVN correlator at JIVE will have 16 station inputs and, in contrast to the VLBA, cross multiplication of the channel bitstreams will be done ahead of fourier transformation. The correlator design is a joint undertaking of JIVE supported by the EVN institutes, the Netherlands Foundation for Research in Astronomy, Haystack Observatory and NASA. The correlator is expected to handle three major types of observation: joint EVN-VLBA continuum at up to 128 Mbit/sec, EVN MkIV continuum at 1024 Mbit/sec, and EVN and EVN-VLBA spectral line at lower bandwidths. Multiple array pulsar gating and field centre correlation is alsoplanned. An array with more than 16 telescopes may require multiple passes through the correlator depending on which telescopes are in the array and the source declination. A standard mode of MkIV operation is likely to be 16-station, dual polarization, 2 bit/sample recording at 128 MHz bandwidth per polarization for a data rate per station of 1 Gbit/sec. The spectral resolution is better than 1% in full polarization mode for 16 stations, i.e. 128 complex channels per baseline for each of the four Stokes parameters. Recirculation of the data through the correlator is possible for lower sample rates than the correlator clock rate of 32 MHz, giving up to a factor of 8 higher spectral resolution.The correlator software now being developed will take account of the space VLBI requirements on fringe and delay rates. The correlator project is expected to be completed in 1997.

The EVN has agreed to reserve about 50% of its VLBI time per year (about 7 weeks per year) for the most highly rated space VLBI proposals and to evaluate other space VLBI observing proposals in competition with purely ground-based VLBI proposals for possible additional time.

2.4. The Asia Pacific Telescope (APT)

The APT has been formed to coordinate and schedule VLBI observations involving telescopes in the Asia Pacific region (Figure 1 and Tzioumis 1994a). It also provides a forum for discussion of issues of compatibility and sharing of resources for VLBI observations. The telescopes are distributed north and south of the equator giving good uv-coverage for sources in the equatorial and southern zones. At its heart is the Australian VLBI network (Jauncey et al., 1994, Tzioumis 1994b), a six-station array in Australia, plus the Hartebeesthoek telescope in South Africa, which carries out VLBI observations for 1 to 2 weeks every 4 months. The APT itself is due to have its first coordinated session in June 1995. The potential uv-coverage of the APT together with the VLBA is shown in Figure 4 for the active galaxy Centaurus A (from Tzioumis, 1994a).

A mixture of recording systems are currently available on the APT including MkII, MkIII, Canadian S-2, and the K-4. The most widespread wide bandwidth system is the S-2. The Australian correlator is S-2 based and can handle 6 stations with a maximum bandwidth of 64 MHz in continuum mode. There are many spectral line modes including 2048 complex channels per baseline for 1 MHz bandwidth for 6 stations. Working together with the VLBA and correlation in Australia will require translation of VLBA tapes to S-2, something desirable for space VLBI in any case (see section 2.6), but not yet funded.

The ATNF has guaranteed 10% of the total observing time on its telescopes to support space VLBI. The APT is likely to take on a coordinating role for the region's telescopes during the VSOP and RadioAstron space VLBI missions.

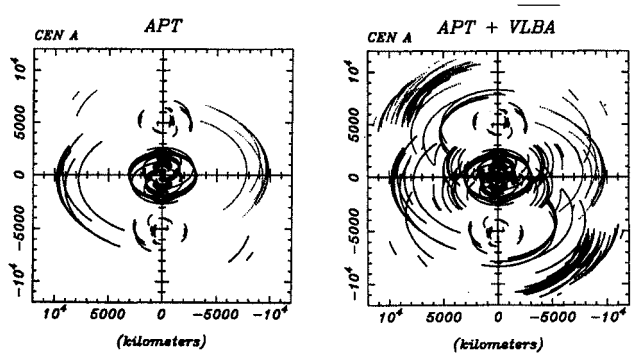


Fig. 4 - left: uv coverage of the APT for Centaurus A. right: the same for the APT+VLBA

COUNTRY	INSTITUTE	LOCATION	DIAMETER (m)	FREQUENCY RANGE (GHz)
Australia	Australia Telescope	Narrabri	22	1.4 - 115
	National Facility	Mopra	6 x 22	1.4 - 115
		Parkes	64	0.6 - 43
	NASA Tidbinbilla	DSS43	70	1.6 - 22
		DSS45	34	2.3/8.4
	Univ. Tasmania	Hobart	26	1.4 - 12
	Univ. WA/Telstra	Guangara	27.5	5
South Africa	HartRAO	Hartebeesthoek	26	1.6 - 12
Japan	CRL	Kashima	34	1.4 - 43
	Nobeyama Radio Obs.	Nobeyama	45	8.4 - 115
	ISAS	Usuda	64	1.6 - 22
China	Shanghai Observatory	Shanghai	25	0.3 - 22[1]
	Urumqi Observatory	Urumqi	25	0.3 - 22[1]
India	TIFR	GMRT Pune	30 x 45	0.3 - 1.4[2]
Russia	ISDE	Ussurijsk	70	0.3 - 22

Table 2 - The Asia-Pacific Telescope
Notes: [1]Receivers up to 9 GHz currently available, and up to 25 GHz available in late-1995.
[2]Also a member of the EVN. The frequency range may extend to 43 GHz.
[3]Available by special arrangement only. The frequency range may extend to 1.6 GHz.

2.5. The Coordinated Millimeter VLBI Array

VLBI at millimetre wavelengths has moved from observational "black art" to relative routine in the last ten years, so much so that a new organisation, the Coordinated Millimeter VLBI Array, has recently been established to coordinate observations at 86 GHz and higher frequencies in a similar way to the EVN and the old US VLBI network. Three sessions per year are envisaged with telescopes in the USA, Europe, and Chile: Haystack Observatory (MA), FCRAO (MA), NRAO Kitt Peak (AZ), OVRO (CA), Onsala (Sweden), Effelsberg (Germany), IRAM Pico Veleta

(Spain), and SEST (Chile). At these high frequencies, angular resolutions of 50 micro-arcseconds have already been achieved on strong sources. There is no overlap with the first generation space VLBI missions, but some concepts for second generation satellites include receivers at these frequencies.

2.6. Compatibility of data acquisition systems

An essential element of the cooperation necessary for VLBI is compatibility of recording system hardware and formats. Recently, however, some divergence has occurred as the result of the separate development, in different countries, of systems for wide bandwidth recording. The greatest measure of compatibility exists for the longitudinal reel to reel instrumentation recorders in use in the VLBA, EVN, and US-led geodesy networks. VLBA acquisition systems are backwards compatible with a subset of MkIII modes; MkIV systems will be backwards compatible with all MkIII and VLBA modes. However the cassette-based helical scan video recorders developed in Canada (S-2) and in Japan (VSOP/K-4) are not compatible with the VLBA/MkIII/MkIV systems or with each other. The S-2 system adopted by the Australian network stacks 8 commercial VHS recorders to provide a relatively cheap 128 Mbits/sec capacity; the VSOP/K-4 is based on the ID-1 standard and provides 256 Mbits/sec capacity. The VLBA, MkIII/MkIV and VSOP/K-4 are all of comparable (high) cost. This lack of compatibility between Europe/USA on the one hand and the APT on the other is of consequence for standard ground-based VLBI only for joint APT- VLBA observations (see section 2.4); there is very little common sky between Europe and Australia. However, it is of great consequence for space VLBI. VSOP observations in the southern sky will be recorded on S-2 at the radio telescopes in Australia, but on VLBA recorders at the tracking stations operated by NASA and by the US National Radio Astronomy Observatory, and on VSOP/K-4 at the Japanese radio telescopes and tracking station.

The solution adopted is to develop tape copying machines for VLBA to VSOP/K-4 and S-2 to VSOP/K-4 so that VSOP observations can be correlated on the VSOP correlator being built at the National Astronomical Observatory. VLBA to S-2 tape copiers are also desirable to allow correlation of (i) VLBA and APT data on the S-2 correlator in Australia, and (ii) APT and RadioAstron data recorded on VLBA recorders at the tracking stations on the S-2 corelators in Australia and in Canada. However, no funding has been identified yet for this.

An international working group is being set up to try to coordinate future developments in acquisition systems and thereby avoid such problems in the future.

3. Space VLBI

Putting a radio telescope into orbit around the Earth to increase angular resolving power is an idea that goes back 30 years to the early discussions on VLBI (Preston et al., 1983, Burke, 1991). The first paraboloid radio telescope in space was the 10 m KRT-10 which underwent deployment tests from the Salyut-6 station in 1979. The first successful radio interferometry using a space-ground

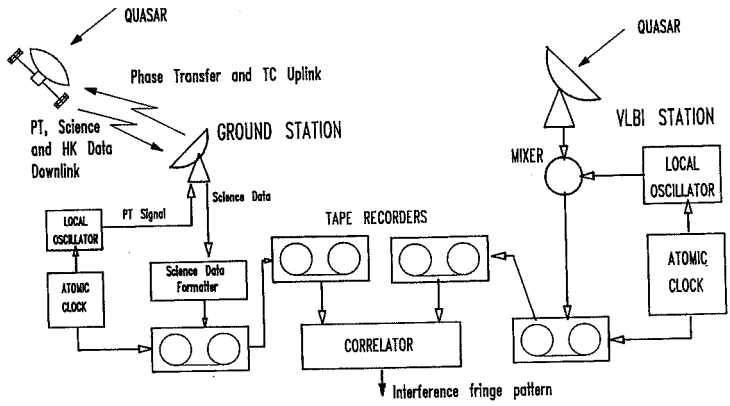


Fig. 5 - Space VLBI system

combination was achieved by the TDRSS-OVLBI experiments in 1986 and 1988 (Levy et al., 1986; Linfield et al., 1990). The 4.9 m antenna on one of NASA's Tracking and Data Relay Satellites was turned towards distant quasars to create interferometer baselines with large diameter ground- based telescopes in Australia and Japan which were twice as long as possible on Earth. During the 1980s and early 1990s, two projects were studied by international teams created by the European Space Agency: QUASAT (Schilizzi, 1988) and IVS (Pilbratt 1991). Neither project was selected for launch but, nevertheless, many of the ideas first demonstrated to be feasible in those studies have been incorporated in the first-generation missions VSOP and RadioAstron, and in plans for a second-generation mission, ARISE (Ulvestad et al., 1995).

Figure 5 shows the principles of space VLBI. The radio telescope is launched folded within the fairing of the rocket, and after reaching the desired orbit, it is deployed into its final configration. The radio signals received from cosmic radio sources are handled on board much as they are for the Earth bound telescopes, passing through various amplifiers and filters, before being digitised and formatted prior to transmission to the Earth for recording at the tracking stations. The frequency stability required for VLBI will be based on other local oscillator signals transmitted to the satellite from the tracking stations which are themselves controlled by a hydrogen maser standard, or on frequency standards carried on board the satellite. Both options will be implemented for the missions.

There are a number of substantial differences between space VLBI and ground-based VLBI (e.g. Murphy et al. 1994; Meier 1994): (i) the space antennas are of small diameter by Earth standards so interferometer sensitivity is less while the baselines are longer; (ii) spacecraft pointing constraints need to be taken into account - e.g. Sun, Moon, Earth, visibility of tracking stations by the on- board telemetry antenna, need to direct the passive cooling radiator on RadioAstron to cold sky; (iii) space VLBI uv-coverage is a strong function of the orbital elements and the spacecraft constraints; (iv) the uv-coverage changes with observation epoch due to orbital evolution, and it changes from day to day since the orbital periods are not commensurate with 24h;

and (v) the locations of the tracking stations are such that data may be lost near perigee and when the spacecraft is in the south. Whereas for ground-based VLBI with a given ground array, the uv-coverage is a function of source declination only, for space VLBI it is strongly dependent on both source coordinates (Right Ascension and declination) and the epoch of observation due to a combination of the spacecraft constraints, the precessing orbital elements, and the limited tracking network. For VSOP, the dominant orbit perturbation is due to the oblateness of the Earth and this leads to rapid precession of the orbital elements and fast changing uv-coverage; for RadioAstron, the dominant perturbations are due to lunar-solar effects which do not change very fast so that the uv-coverage is more or less fixed for a particular source once the initial orbital elements have been chosen.

	VSOP			Radioastron			
Country	Japan			Russia			
Lead Institution	ISAS			Astro Space Center			
Launch Date	Sept. 1996			1997/1998			
Launch Vehicle	M-V			Proton			
Launch Site	Kagoshima			Baikonur			
Mission Lifetime (years)	3			3			
Orbit							
Period (hours)	6.06			28.0			
Apogee Height (km)	20,000			77,000			
Perigee Height (km)	1,000			4,000			
Inclination (°)	31			51.6			
Tracking Stations							
Commanding	Kagoshima (Japan)			Ussurijsk (Russia)			
Data Acquisition + ø Link	Usuda			Ussurijsk			
Data Acquisition + ø Link	Green Bank (NRAO)			Green Bank (NRAO)			
Data Acquisition + ø Link	3 DSN sites (NASA/JPL)			3 DSN sites (NASA/JPL)			
Space Radio Telescope							
Antenna Diameter (m)	8.0			10.0			
VLBI systems	VLBA/K4			VLBA/S2			
Polarization	LCP			LCP/RCP			
Observing Bands (GHz)	1.6	4.8	22	0.33	1.6	4.8	22
System Temperatures (K)	100	120	200	100	50	50	150
Aperture Efficiencies	0.40	0.55	0.41	0.5	0.5	0.5	0.3
Integration Time (s)	100	100	100	20	100	100	100
Instanteneous Bandwidth (MHz)	32	32	32	8	32	32	32
Angular resolution (μas)	1200	400	90	2100	410	140	30
Minimal correlated flux density 8σ-level[a] (mJy)	80	65	145	455	36	36	115
Science Programmes							
AGN / QSOs	AGN Imaging AGN Monitoring			High T_6 Survey AGN Monitoring			
Interstellar Medium	-			Interstellar Scattering			
Extragalactic Masers	XGal H_20 masers			-			
Galactic Masers	Gal H_20 masers Gal OH masers			Gal H_20 masers -			

Table 3 - Characteristics of the VSOP and the RadioAstron Space VLBI systems.
[a]Co-observing with a 70m ground radiotelescope with aperture efficiencies of 0.5, 0.6, 0.6, 0.5 and system temperatures of 100, 30, 30, 50 K at 0.33, 1.6, 4.8, 22 GHz respectively.

Both missions are being developed in close coordination, and this has led to common solutions to a number of operational and data acquisition issues. One of these issues is coordination with the ground based arrays and telescopes. Space VLBI is unique amongst the space astronomy missions in that the space and ground segments are of equal value in realising the scientific aims. Many discussions over several years have led to a workable operations plan, which includes an international Science Review Committee to judge the proposals.

Table 3 summarizes the salient characteristics and performance for both VSOP and RadioAstron.

3.1 VSOP

Space VLBI observations will be carried out with an 8 m antenna aboard the MUSES-B satellite developed by the Institute for Space and Aeronautical Science (ISAS) in Japan (Hirosawa, 1994). MUSES stands for Mu Space Engineering Satellite and this, the second in the series of engineering satellites, will be launched on the first flight of ISAS's new rocket, M-V. The expected capability of the M-V rocket sets constraints on the combination of satellite mass and apogee height of the orbit. VSOP will go into an orbit with an apogee height of 22000 km, a perigee height of 1000 km, and an inclination of 31°. The satellite mass is restricted to 820 kg and the volume to 3.65 m height by 2.2 m diameter, which has necessitated some ingenious engineering solutions, particularly in the realm of antenna construction and deployment.

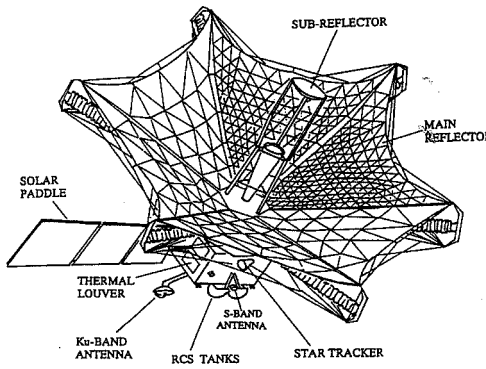


Fig. 6 - The MUSES-B satellite

In Figure 6, we can see the main elements of the MUSES-B satellite. The dominant feature is the 8 m reflector which is formed from wires and mesh stretched into parabolic form on a framework of six extendable booms. This design is based on the tension-activated truss concept developed by Miura (1986). The mass of the antenna is 225 kg and the goal for surface accuracy is 0.5 mm rms which will allow efficient observation at 22 GHz (=1.35 cm). Although the maximum structural diameter of the antenna is 10 m the effective aperture for radio astronomy is 8 m. The supporting tower for the sub-reflector is also extended in orbit to give a focal length of 3.7 m. The two major engineering challenges in developing this antenna have been the reliable deployment of the main reflector, and proving that a 0.5 mm rms surface accuracy is achievable in the zero gravity conditions in orbit while forming and tuning the antenna occurs on the ground in the 1 G environment.

The onboard radio astronomy systems comprise single-channel uncooled low noise amplifiers at 1.6, 5, and 22 GHz, down converters for these bands, an IF signal switch, two frequency synthesisers, two image rejection mixers, two A/D converters, a formatter, and calibration signal

generators. The two baseband channels can operate in two modes: 16 MHz bandwidth and 2-bit sampling or 32 MHz bandwidth and 1-bit sampling. The output bit rate of the formatter is 128 Mbps. The science data and local oscillator transfer will occur in Ku-band: 14.2 GHz for the wideband science data downlink and 15.3 GHz for the narrow band two-way LO link. The 45 cm Ku-band antenna is attached to a deployable boom to increase the field of view of the antenna (see Figure 6). The data will be recorded at the Usuda tracking station on a VSOP/K-4 data acquisition system developed in Japan, and at the NASA and Green Bank tracking stations on VLBA terminals. Tape copying machines are under development in Japan to transfer the VLBA recordings from tracking stations and ground radio telescopes to VSOP/K-4 before correlation on the VSOP correlator at the National Astronomical Observatory in Tokyo.

With the orbit apogee allowed by the launch constraints, VSOP achieves baseline lengths about 3 times that possible on Earth. Considerable effort has gone into simulating the image reconstruction capability of the VSOP mission by Murphy and collaborators (Murphy et al., 1994) since that will be its strength compared to RadioAstron (see next section) which goes into a much higher orbit with ten times the angular resolution of the ground arrays but much poorer uv-plane coverage and thus poorer imaging potential. Figure 7 (left) shows a typical uv-coverage at a particular epoch as a function of source position for VSOP co-observing with the VLBA, and in Figure 7 (right) the uv-coverage obtained at monthly intervals in a 2-year period, for a 24h observation of VSOP with the VLBA. Figure 7(left) shows very clearly that good uv-coverage cannot be obtained for any random direction in the sky - therewill be seasons for the best observations just as for other space astronomy missions and ground-based optical astronomy. One-dimensional uv-coverage is produced when the source lies in the orbital plane (equatorial sources in Figure 7(right)), and two-dimensional coverage when it lies near the orbit normals. Figure 7(right) shows that the uv-coverage varies with epoch, an effect which will have to be taken into account when analysing observations monitoring structural changes in the target sources.

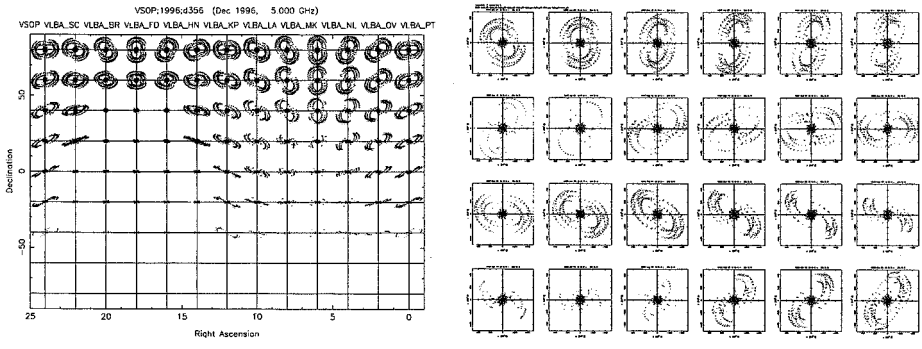


Fig. 7 - left: Typical uv-coverage as a function of source position for VSOP co-observing with the VLBA. right:uv-coverage obtained at monthly intervals in a two year period, for a 24h VSOP observation together with the VLBA of the superluminal source 1928+738. This figure illustrates how uv-coverage is a strong function of observation epoch. (Murphy et al., 1994)

3.2 RadioAstron

The RadioAstron project is led by the Astro Space Center of the P. N. Lebedev Physical Institute in Moscow, with the major construction work being undertaken by the Lavoshkin Association (Kardashev and Slysh, 1988; Kardashev, Gurvits, and Tsarevsky, 1994). The satellite itself is called Spektr-R and is one of the Spektr series of astrophysical satellites under construction by the Russia Space Agency and the Russia Academy of Sciences. RadioAstron will be launched on a Proton rocket into a high eccentricity orbit with a 28h period and will generate baselines to co-observing earth telescopes of up to 85000 km in length. The Proton is a powerful launch vehicle, so the mass budget, 5000 kg, for the RadioAstron satellite is generous compared to VSOP. The launch date is currently set for 1997, but funding slowdowns in Russia are likely to delay this date.

As we have already seen for VSOP, the dominant feature of the RadioAstron spacecraft is the 10 m diameter antenna (Figure 8) with a mass of 1500 kg. It has a central 3m section and 27 solid panels each 3.5 m long which fold up around the focal package support during launch. The petals

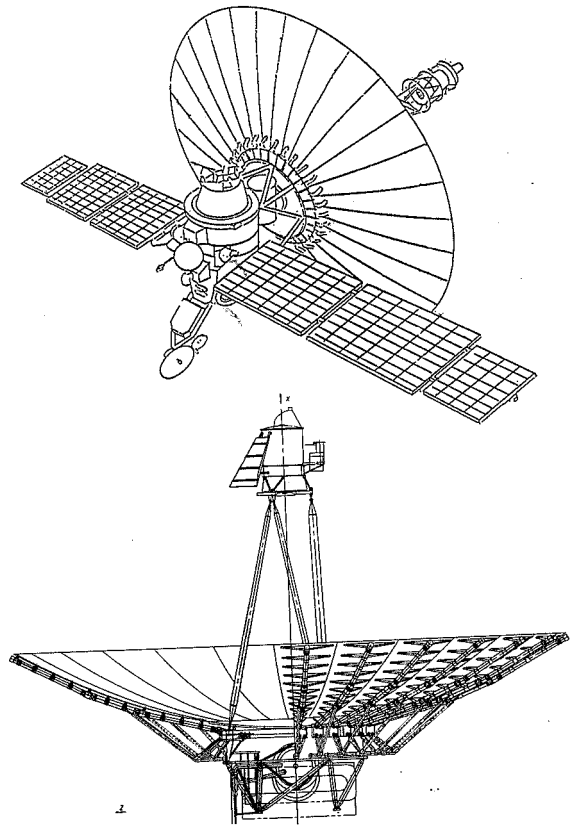


Fig. 8 - top: The RadioAstron satellite. bottom: The RadioAstron antenna

are made from a stiff carbon fibre composite material which should provide an rms surface accuracy of 0.5 mm, like VSOP, sufficient for 22 GHz operation. Testing of the thermal deformation properties of the panels has been carried out at the European Space Agency's Technology Centre (ESTEC).

RadioAstron has simpler optics than VSOP, utilizing prime focus feeds and locating the receiver electronics in the prime focus package (Figure 8). The feeds are of a novel concentric ring feed design to provide on-axis optics for all four frequencies. Each receiver is sensitive to both hands of circular polarization, which will allow VLBI polarimetry to be carried out. All receivers are cooled passively to 150 K through a 1 m2 radiator on the side of the focal package which must point to free space at all times. In addition, the three higher frequency receivers are cooled to 80 K using closed cycle cryogenic cooling technology. Institutes in Finland (22 GHz), Netherlands, Germany, Italy, Sweden, and UK (5 GHz), Australia (1.66 GHz), and India (0.327 GHz) designed and built the receivers. The IF signals are down converted to video band (channel bandwidths 2, 4, or 8 MHz), digitized and formatted in the service module below the main reflector before transmission in two 32 MHz wide streams to the tracking stations. The bandwidth available at 0.327 GHz is restricted to 2 4 MHz due to interference problems for the co-observing radio telescopes.

At the tracking stations the downlinked data will be recorded on VLBA terminals at the NASA sites, the NRAO Green Bank site, and at Ussurijsk in Russia and the S-2 system at Ussurijsk and possibly at one of the NASA DSN stations at Tidbinbilla. Correlation will take place on the VLBA, EVN and S-2 correlators in Russia and Australia. The Dominion Radio Astrophysical Observatory in Canada is providing an S-2 correlator for RadioAstron operations in Moscow.

4 Ground-based VLBI - New results and new capabilities

Of the many new results coming out in VLBI, I have selected a small number that I regard as making a significant step forward in our understanding of astrophysical phenomena and/or which advance the state of the technique substantially.

4.1 Surveys

Imaging surveys of large numbers of sources are an essential element of astronomy in making possible discoveries of new classes of source and in defining or refining source characteristics. Until recently, the time involved in carrying out VLBI observations of individual sources has limited surveys to relatively small samples. However the use of "snapshot" observations with large numbers of telescopes made possible in the first place by the 16-station MkII correlator at Caltech in Pasadena has charged attitudes to VLBI Surveys. (This correlator has now been closed down for general astronomical use since the narrow band MkII recording system has been superseded by the broader band MkIII, VLBA and MkIV systems, and the VLBA correlator has even greater capacity - 20 stations.) With many telescopes in the array (e.g. Figure 3), the uv-

coverage for three 20m observations well spread throughout the apparition of the source is sufficiently good to generate good quality images, and yet allow observation of 20 sources per day.

The Caltech-Jodrell Survey (Polatidis et al, 1995, Taylor et al, 1994) has imaged two samples of sources with declinations north of 35o and galactic latitude 10o using global arrays composed of the EVN and the VLBA. The first segment (CJ1) imaged all sources with flux densities at 5 GHz between 0.7 and 1.3 Jy (N=132), the second (CJ2) was restricted to sources with flux densities between 0.35 and 0.7 Jy but with spectral indices flatter than 0.5 at 5 GHz (N=193). Images of 276 sources have been made at 5 GHz and 132 at 1.66 GHz. The CJ1 survey together with an earlier survey (Pearson and Readhead, 1988) was primarily aimed at classifying the structures of compact radio sources, whereas CJ2 had in addition three cosmological goals: (i) to place significant limits on the mass in 106 - 109Mficondensed objects in the universe by searching for gravitational lenses in the 1 - 200 milli-arcsecond range of separations, (ii) to look for evidence of cosmological evolution in the proper motions of superluminal sources and place limits on the cosmological deceleration parameter, qo (Vermeulen and Cohen 1994), and (iii) to explore the possibility of using the angular diameter - redshift diagram for compact objects to constrain qo (Kellermann 1993, Gurvits 1994). The analysis of the CJ2 data in pursuit of these goals is in progress.

The majority of sources found in the surveys have a core-jet structure (see Figure 9 for an example) generally thought to be indicative of outflow of relativistic electrons from an active core in the host galaxy or quasar. The one-sidedness is probably due to relativistic aberration caused by bulk motion of the electrons at speeds close to that of light more or less in the direction towards the observer. This causes the illusion of faster-than-light or superluminal motion (Zensus and Pearson, 1987). Some 5 to 10% are compact symmetric objects, CSOs, (see Figure 9) which are

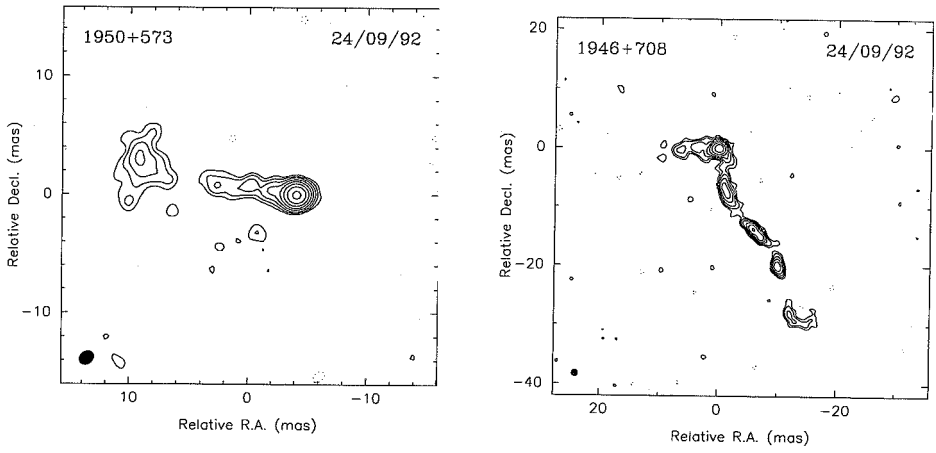


Fig. 9 - left: a typical core-jet structure. right: a Compact Symmetric Object (Taylor et al., 1994)

thought to be core-jets with the outflow more or less perpendicular to the line of sight to the source so that it is visible on both sides of the core. Their small size could be due to their being either young objects (106 years) or much older and unable to expand against an hypothesised dense interstellar medium in the host galaxy or quasar (Fanti et al, 1995, Readhead et al, in preparation).

4.2 A black hole in a nearby galaxy

Black holes are by definition impossible to detect in electromagnetic radiation, but their mass and density are such that they should be detectable through their influence on matter in their vicinity. Some stellar systems in our galaxy emit so copiously in X-rays that this is explainable only in terms of heating of material from a star as it is sucked into the accretion disk surrounding a black hole of 1 to 10 solar masses in orbit around the star. It is commonly assumed that much more massive black holes (106 to 109 solar masses) are to be found in the centres of galaxies and quasars, but evidence has proven to be elusive. The reason is that despite their great mass, the black hole is still a very small fraction (0.01 %) of the total mass of the whole galaxy and its influence is only to be seen within the inner light year of the centre. Extremely high angular resolution is required.

Recent observations of rotating gas in the centre of the nearby galaxy NGC4258 made by Miyoshi et al (1995) using the line emission from water masers, demonstrate the presence of a central mass of 4 107Mfiin a region less than 5 light months (4 1012 km) in radius. Combination of VLBA spectroscopic images in the line at 22 GHz and single dish spectroscopy revealed that the masers are found in front of the compact nucleus of the galaxy and at the tangent points of a thin warped torus surrounding the nucleus (Figure 10), both locations where maser amplification is optimum. The positions of the high velocity features define a torus with inner radius of 4 and an outer radius of 8 milli-arcseconds, and their velocities decrease with distance from the centre of rotation exactly as expected for Keplerian motion (like that of the planets around the Sun). Single dish spectroscopic observations show that the central features drift at a rate of 9.5 km s1 yr1 which is assumed due to centripetal acceleration by the enclosed mass. This allows the distance and hence the physical size of the inner radius to be calculated. Combined with the radial velocities, this then allows the mass of the object dominating the gravity field to be determined. The mass density derived is so high that the object is most likely a black hole.

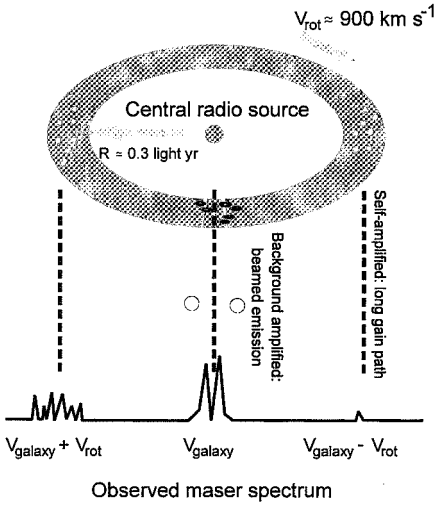


Fig. 10 - Illustration of the maser model taken from Barvainis (1995). The maser spectrum is sketched along the bottom.

4.3 The supernova in the nearby galaxy, M81

Stars explode in galaxies regularly. However no star has exploded in our Galaxy within view of our telescopes since their invention, so our knowledge of this final phase of stellar evolution comes from study of the faint remnants of explosions many centuries ago and from supernovae in nearby galaxies where several per year are detected. Details of the processes involved in the early phases of the explosion and its subsequent evolution are difficult to model theoretically since there has been very little spatial information available. The recent discovery of a radio luminous supernova in M81 (at a distance of 11.7 million light years) has provided an opportunity to investigate the early evolution of a supernova with the high angular resolution afforded by VLBI. Marcaide et al. (1995) used a global network at 8.4 GHz to image the remnant with a resolution of 0.25 milli- arcsec and showed that 8 months after the explosion a shell structure had developed, the youngest and smallest shell ever observed in a supernova event. Figure 11 shows that the shell is almost spherically symmetric in shape, but its intensity distribution is not uniform around the shell. The enhanced emission in the south-eastern part could be the result of an asymmetric distribution of pre-supernova circumstellar material. Comprehensive monitoring of SN1993J in the radio domain is continuing at a number of frequencies.

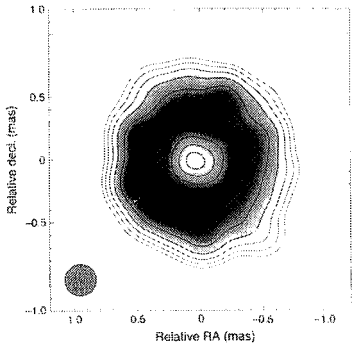


Fig. 11 - Grey-scale and contour map of SN 1993J from 8.4 GHz observations 239 days after the explosion. The gaussian beam used in the image is shown in the lower left and is 0.25 milli-arcsec in diameter. (Marcaide et al., 1995)

Using the estimate for the shell radius at the time of the observation in November 1993, and assuming zero size at explosion, the average rate of expansion for the shell radius is 2.430.15 micro arcsec/day. Making use of the maximum gas expansion speeds measured by the widths of optical absorption lines, a distance of 11.7 million lightyears is found, in good agreement with completely independent, but indirect, methods of measuring the distance to M81.

4.4 Gravitational lenses

In 1979, Walsh et al. discovered two radio sources with very similar structure close to each other on the sky, each associated with a quasar with almost identical properties. It was quickly realised that these were images of the same background object, gravitationally lensed by a foreground galaxy; predictions of this effect had been made by Einstein in the 1930's. When a sufficiently large mass concentration lies in the line of sight between the observer and a background source, space-time is distorted so much that radio and light rays are bent creating multiple images (e.g. Figure 12 top). A galactic-sized mass of 1011Mfiwill cause the images to be separated by a few arcseconds, a smaller concentration of 106-108Mfi(a naked black hole?) would cause a separation

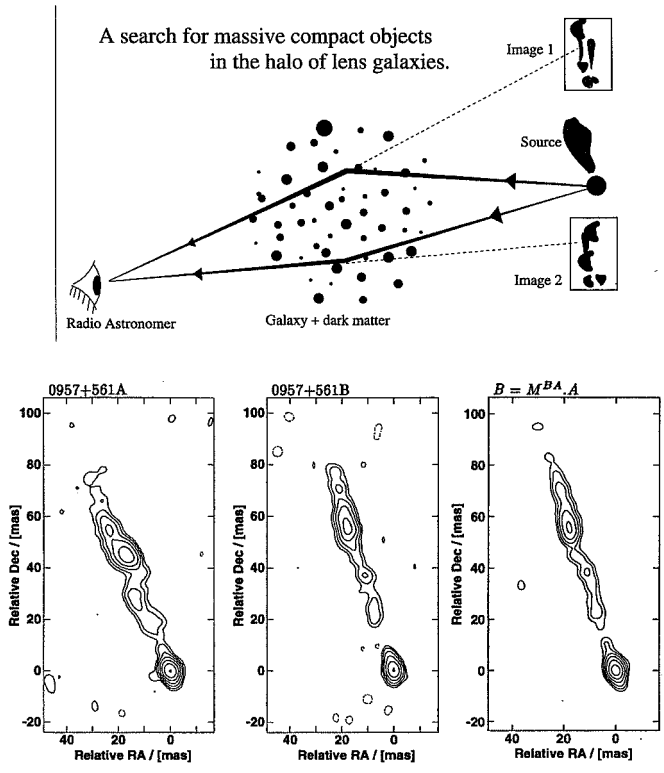


Fig. 12 - top: The geometry and principal elements required for small-scale lensing. Since the line of sight is different for each image, compact objects in the lens galaxy halo distort the lensed images in different ways. If the lens galaxy mass distribution is smooth, the images will appear very similar to each other. (Diagram from M. A. Garrett) bottom: Global VLBI observations of 0957+561 A,B. The image on the righthand side is the A image convolved with a linear transformation derived from these data. The agreement with image B (centre) is obvious. (Garrett et al., 1994)

of a few tens of milli-arcseconds. However, no confirmed small gravitational lenses were found in the Caltech-Jodrell VLBI surveys suggesting that the contribution to the total mass of the universe by a cosmologically distributed population of these lower mass objects in intergalactic space must be small. Even smaller masses of 104-106Mfiin the halos of lensing galaxies would cause distortions which would be different for each image on the milli-arcsec scale.

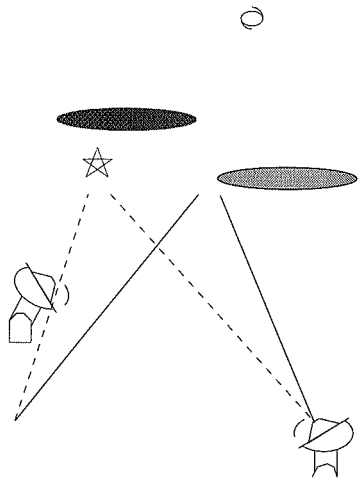
Gravitational images are recognizable as being related to each other because they have almost identical optical spectra, radio structures and radio spectra, which is very unlikely in unrelated objects. Small differences between the images and their changes in time carry information on the detailed mass distribution in the lensing galaxy; this is where VLBI observations make a unique contribution. Recent global 1.66 GHz VLBI measurements of the Walsh et al. lens, 0957+561, by Garrett et al. (1994) are shown in the lower part of Figure 12. Garrett et al. have derived a linear transformation which relates quantitatively the brightness distributions of jet features in one

image to those in the other. This transformation matrix is a function of the mass distribution in the lens and strongly constrains lens models for 0957+561. The similarity of the observed image B (centre of Figure 12) and the transformed image of A is striking, and leads to the conclusion that the mass distribution for the lens is smooth with the contribution of black holes of mass 3 106Mfibeing less than 10% of the dark matter in the halo of the lens galaxy. It is still an open question how much of the dark matter in the universe may be found in massive objects in the halos of galaxies. Perhaps VLBI measurements will provide crucial input to this (Garrett et al., 1994)

4.5 The most precise measurements of the motion and parallax of stars

The extragalactic reference frame set by the positions of several hundred compact radio sources in distant galactic nuclei is the primary celestial reference frame for astronomy. The average internal accuracy of this system is better than 1 milli-arcsecond (IERS Annual Report, 1994 published by the Observatoire de Paris). Linking the optical reference frame for our galaxy set by the Hipparcos satellite measurements to the extragalactic frame is important to unify the optical and radio coordinate systems, in particular for registration of images at both wavelengths and to calibrate the global rotation of the Hipparcos frame for dynamical studies. This link can be carried out by observing optically bright radio emitting stars in the Hipparcos catalogue with VLBI (Lestrade et al., 1992).

Since radio stars are usually weak, the technique of phase-referencing is used to increase the coherent integration time on the star so that it is well detectable after a few hours. This involves switching the telescopes between the star and a strong compact reference source within a few degrees on the sky from the star (Figure 13 every 2 or 3 minutes. The phase of the fringes detected for the reference source is used to lengthen the coherent integration time for the stellar radio emission. At the same time this allows high accuracy differential astrometry between the star and the reference source (Lestrade et al., 1990).



One of the objects in the Lestrade programme is σ^2CrB, an RS CVn close binary radio star located 0.5° away from a strong quasar. Twelve epochs of global VLBI observation over a period of nearly five years have allowed the most precise determination of the 5 astrometric parameters (2 coordinates, 2 proper motions, and parallax) ever made for any object, except the Moon. The formal uncertainties in the least squares fit were 0.08 milli-arcsec for the position relative to the quasar, 0.04 milli-arcsec/ year for the proper motion, and 0.08 milli-arcsec for

Fig. 13 - Illustration of phase referencing.

the trigonometric parallax. The lack of a sinusoidal signature in the post-fit residuals sets a limit on the mass of any planet in orbit around the star system, to somewhat more than a Jupiter. This level of precision in determining parallax should permit direct determination of distances to radio stars in our galaxy out to about 10000 light years, about half as far as the centre of our galaxy.

4.6 VLBI polarimetry

Polarization sensitive data provide information on the magnetic field structure and on the physical conditions of thermal plasma inside radio sources as well as in their neighbourhood and along the line of sight to the observer. It is one of the few diagnostic tools available for the analysis of continuum radio emission. Figure 14 shows a fine example of a 5 GHz polarization image obtained with a global array for the quasar, 3C138 (Dallacasa et al., 1995) . It shows that the polarized emission is dominated by the main jet and that the core is weakly polarized. The magnetic field geometry on scales of 10 milli-arcsec is mostly longitudinal with evidence of sub-structures at the end of the jet where the interaction with the ambient medium is most pronounced. Information of this sort for compact sources is important for models of jets mechanisms.

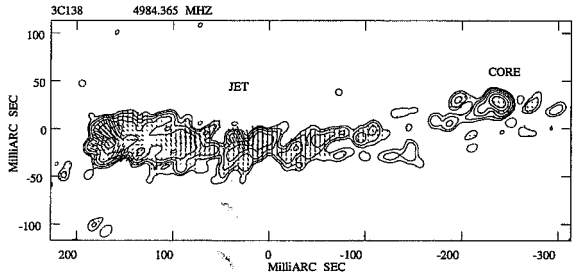


Fig. 14 - Global VLBI image of 3C138 at 5 GHz at 10 milli-arcsec resolution.The electric vectors superimposed have lengths proportional to the polarized flux density. The magnetic field directions are perpendicular to the electric vectors. (Dallacasa et al., 1995)

4.7 High quality images

For many years, VLBI image quality has been limited by poor uv-coverage, poor telescope system noise and instrumental polarization performance, and other systematic errors introduced by instrumental mismatches of one sort or another. Whereas phase-connected interferometers like the WSRT and VLA were able, with effort, to reach dynamic ranges (peak intensity in the image over rms noise) of several 100000 to 1, VLBI images seemed to be able to go no further than a few 1000 to 1. The VLBA was designed to remove or reduce the effects of the limiting factors just mentioned, so it was gratifying that test observations at 5 GHz on an almost unresolved source, DA193, have shown that a dynamic range of 120000 can be achieved using a recording bitrate of

128 Mbit/sec (Cornwell, Kemball, and Benson, VLBA Memo in preparation). Figure 15 shows this superfi cially uninteresting looking but nevertheless important result on DA193 which demonstrates that the VLBA can achieve noise-limited performance across a wide bandwidth (rms noise in the image 49 Jy/beam). Important factors in this result are the very careful editing out of bad data points, and the judicious use of deconvolution algorithms such as CLEAN and NNLS (non negative least squares matrix inversion, Briggs et al., 1994).

The European VLBI Network is inherently more sensitive to instrumental mismatches between the elements in the array because the individual telescopes are all different. Howéver, a major programme is in progress to quantify possible systematic errors and correct these either in hardware or in post-correlation software. In principle, the rms noise in an EVN image at 5 GHz should reach 10 Jy/beam for an 8h measurement in the absence of systematic errors and with state of the art receivers and MkIV recording at 1024 Mbit/sec.

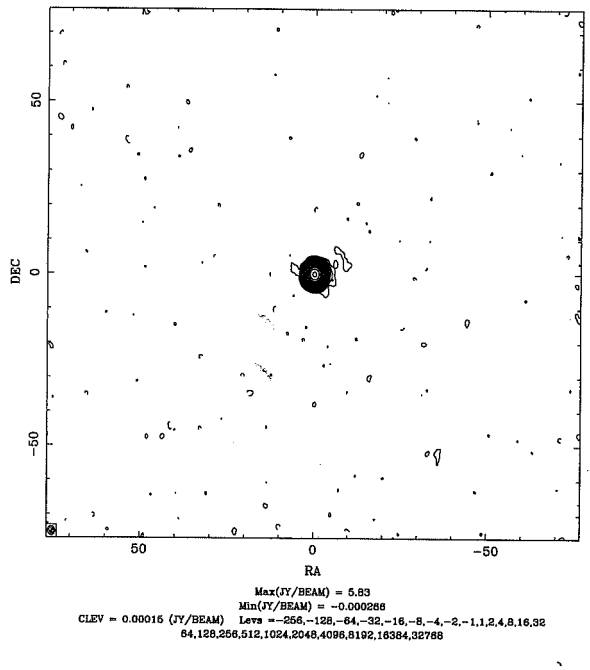


Fig. 15 - VLBA image of DA193 at 5 GHz. The dynamic range (peak intensity/rms noise in the image) is 120000, the highest ever achieved in VLBI (Cornwell et al. 1995, VLBA Memo in preparation)

5. Space VLBI - science drivers

The most exciting element of the VSOP and RadioAstron missions is the substantially increased angular resolution which allows us to explore new parameter space. One of the key projects will be a survey of compact sources. Theory leads us to predict that the maximum apparent brightness temperature in a compact radio source radiating incoherently should be about 1012degrees Kelvin, which by chance is the limit that can be set on Earth. VSOP and RadioAstron will extend this limit by one and two orders of magnitude respectively which may lead to a new understanding of the physics obtaining in these extreme conditions.

Studies of active galactic nuclei will consume much of the mission time. Questions concerning the physical origin of the unresolved core component always seen in AGNs and pinning down the region where the jets are collimated will be of prime interest. In the nearest AGN, Centaurus A, in the southern hemisphere, the physical resolution achieved by RadioAstron will be less than 21015cm or 100 times the distance from the Earth to the Sun, and this at a distance of 12 million light years! We may be able to witness events in close proximity to the central black hole.

The third major area is likely to be detailed studies of water and hydroxyl masers. Masers provide powerful tools to probe, on very small spatial scales, the physical conditions (density, velocity, turbulence, magnetic field) in flows associated with the early and late stages of stellar evolution. The measured sizes of distant hydroxyl masers in particular may be affected by scattering in the interstellar medium so that we can map the density of the medium by observing a well-distributed sample.

Polarimetry will be possible with Radioastron, and with VSOP as well although with lower accuracy. Phase referencing will be possible with both satellites when the reference and target sources are simultaneously within the primary beam of the space radio telescope.

The scientific problems to be attacked by space VLBI are governed by the receivers available on the satellite, baseband channelization (which affects the spectral resolution for spectral line work), pointing constraints (which impact the ability to monitor changes in source structure as a function of time), orbit (which governs the angular resolution and uv-coverage), and sensitivity (see Table 3 and Figure 7).

The first call for observing proposals for VSOP will be issued in May 1995. When this closes in November, there will be a clear idea of the range of science that will be carried out with the first of the first generation space VLBI satellites.

6. Concluding remarks

My aim in writing this review was to give the non-specialist reader an overview of activity worldwide in VLBI astronomy. VLBI systems are being developed rapidly at many locations around the world: regional arrays like the EVN, VLBA, and Australian VLBI array are continuing to expand their facilities while the APT and the Coordinated mm VLBI Array are about to begin operation; new observing techniques and capabilities are being exploited to make exciting scientific discoveries; and the first of the two space missions VSOP (Japan) and RadioAstron (Russia) will be launched in August 1996 heralding in the age of space radio astronomy.

Coordination of these efforts is essential both for ground and space VLBI. The organisation of ground-based VLBI has run remarkably smoothly now for many years, and there is every hope that the largest coordination task of all so far, dovetailing the operations of a satellite with those of the ground-based arrays and other radio telescopes, will also run smoothly when put to the test next year.

Acknowledgements

I am indebted to a number of colleagues for providing figures and other information for this review: T. J. Cornwell, D. Dallacasa, M. A. Garrett, L. Greenhill, L. I. Gurvits, D. L. Meier, J. M. Moran, D. W. Murphy, and R. C. Vermeulen. I wish to thank A. G. de Bruyn, T. J. Cornwell, L. I. Gurvits, D. L. Jauncey, K. I. Kellermann, H. J. van Langevelde, and R. A. Preston for comments on the manuscript, and L. I. Gurvits for considerable help in its preparation.

References

Barvainis, R. (1995) Nature 373, 103

Briggs, D. S., Davis, R. J., Conway, J. E., Walker, R. C. (1994) VLBA Memo 697A

Burke, B. F. (1991) Advances in Space Research 11/2, 349

Chikada, Y. et al. (1987) Proc. IEEE 75, 1203

Dallacasa, D. , Cotton, W. D., Fanti, C., Fanti, R., Foley, A. R., Schilizzi, R. T., Spencer, R. E. (1995) Astron. Astrophys. in press

Fanti, C., Fanti, R., Dallacasa, D., Schilizzi, R. T., Spencer, R. E., Stanghellini, C. (1995) Astron. Astrophys. in press

Garrett, M. A., Calder, R., Porcas, R. W., King, L. J. Walsh, D., Wilkinson, P. N. (1994), Mon. Not. R. astr. Soc. 270, 457

Gurvits, L. I. (1994) Astrophys. J. 425, 442

Jauncey, D. L. et al. (1994) in Very High Angular Resolution Imaging (eds. J. G. Robertson, W. J. Tango) Kluwer Academic Publishers, Dordrecht, p. 131

Hirosawa, H. (1994) in VLBI Technology - Progress and Future Observational Possibilities (eds. Sasao, T., Manabe, S., Kameya, O., Inoue, M.) Terra Scientific Publishing Company, Tokyo, p. 3

Kardashev, N. S., Slysh, V. I. (1988) in The Impact of VLBI on Astrophysics and Geophysics, Proc. IAU 129 (eds. M. J. Reid, J. M. Moran) Kluwer Academic Publishers, Dordrecht, p. 433

Kardashev, N. S., Gurvits, L. I., Tsarevsky, G. S. (1994) in Very High Angular Resolution Imaging, Proc. IAU Symposium 158 (eds. J. G. Robertson, W. J. Tango) Kluwer Academic Publishers, Dordrecht, p. 477

Kellermann, K. I., Thompson A. R. (1985) Science 229, 123

Kellermann, K. I. (1993) Nature 361, 134
Lestrade, J. F., Rogers, A. E. E., Whitney, A. R., Niell, A. E., Phillips, R. B., Preston, R. A. (1990) Astron. J. 99, 1663
Lestrade, J. F., Phillips, R. B., Preston, R. A., Gabuzda, D. C. (1992) Astron. Astrophys. 258, 112
Levy, G. S. et al. (1986) Science 234, 187
Linfield, R. P. et al. (1990) Astrophysical Journal 358, 350
Marcaide, J. M. et al. (1995) Nature 373, 44
Meier, D. L. (1994) in VLBI Technology - Progress and Future Observational Possibilities (eds. Sasao, T., Manabe, S., Kameya, O., Inoue, M.) Terra Scientific Publishing Company, Tokyo, p. 39
Miura, K. (1986) 37th Congress of the International Astronautical Federation, Insbruck, IAF-86- 206
Miyoshi, M., Moran, J. M., Herrnstein, J., Greenhill, L., Nakai, N., Diamond, P. J., Inoue, M. (1995) Nature 373, 127
Murphy, D. W., Yakimov, V., Kobayashi, H., Taylor, A. R., Fejes, I. (1994) in VLBI Technology - Progress and Future Observational Possibilities (eds. Sasao, T., Manabe, S., Kameya, O., Inoue, M.) Terra Scientific Publishing Company, Tokyo, p. 34
Pearson, T. J. and Readhead, A. C. S. (1988) Astrophys. J. 328, 114
Pilbratt, G. (1991) in Radio Interferometry: Theory, Techniques, and Applications (eds. T. J. Cornwell, R. A. Perley) Astron. Soc. Pacific Conference Series 19, 102
Polatidis, A. G., Wilkinson, P. N., Xu, W., Readhead, A. C. S., Pearson, T. J., Taylor, G. B., Vermeulen, R. C. (1995) Astrophys. J. Supp., in press
Preston, R. A., Burke, B. F., Doxsey, R., Jordan, J. F., Morgan, S. H., Roberts, D. H., Shapiro, I. I. (1983) in Very Long Baseline Interferometry Techniques (ed. F. Biraud) Cepadues Editions, Toulouse, p. 417
Schilizzi, R. T. (1988) in The Impact of VLBI on Astrophysics and Geophysics, Proc. IAU 129 (eds. M. J. Reid, J. M. Moran) Kluwer Academic Publishers, Dordrecht, p. 441
Taylor, G. B., Vermeulen, R. C., Pearson, T. J., Readhead, A. C. S., Henstock, D. R., Browne, I. W. A., Wilkinson, P. N. (1994) Astrophys. J. Supp. 95, 345
Thompson, A. R., Moran, J. M., Swenson, G. W. (1986) Interferometry and Synthesis in Radio Astronomy, John Wiley, New York
Tzioumis, A. (1994a) in VLBI Technology - Progress and Future Observational Possibilities (eds. Sasao, T., Manabe, S., Kameya, O., Inoue, M.) Terra Scientific Publishing Company, Tokyo, p. 211
Tzioumis, A. (1994b) in VLBI Technology - Progress and Future Observational Possibilities (eds. Sasao, T., Manabe, S., Kameya, O., Inoue, M.) Terra Scientific Publishing Company, Tokyo, p. 205
Ulvestad, J. S., Linfield, R. P., Smith, J. G. (1995) Proc. 33rd Aerospace Sciences Meeting, Reno, in press
Vermeulen, R. C. and Cohen, M. H. (1994) Astrophys. J 430, 467
Walsh, D., Carswell, R. F., Weymann, R. J. (1979) Nature 279, 381
Zensus, J. A. and Pearson, T. J. (1987) Eds. Superluminal Radio Sources CUP, Cambridge

ULTRACOLD ATOMS AND MICRO-GRAVITY CLOCKS

CHRISTOPHE SALOMON et al.

Abstract

Recent progress in laser cooling and trapping of neutral atoms is paving the way for the development of more accurate caesium atomic clocks. Producing micro-Kelvin atoms having an r.m.s. velocity of 1 cm/s requires no more than two diode lasers and a small glass cell.

This paper surveys the progress of ground-based fountain clocks, reporting the most recent results obtained with the prototype clock operating at LPTF. Satellite-based clocks, by taking advantage of a reduced-gravity environment, should enable the full potential of laser cooling to be realized, opening the fascinating possibility to achieve performances superior to those of their ground-based counterparts.

The expected relative stability of fountain and satellite clocks is about 3 10^{-14} at one second or ~10^{-16} per day with an accuracy in the 10^{-16} range. The intercomparison of clocks with these performances raises interesting challenges in the time and frequency transfer domain and should also allow a new generation of tests of general relativity.

1. Introduction

Today, the most stable clocks are atomic clocks, the Hydrogen Maser, trapped ion clocks and the caesium clock. The unit of time, the second, is defined using a hyperfine transition in atomic caesium. Caesium clocks have the best long term stability (~10^{-14} from 1000 s to several years) and accuracy while H masers present the best short term stability (~10^{-15} from 100 to 1000 s). These clocks are currently used on Earth in a variety of science areas ranging from fundamental tests of physics laws to Very Large Baseline Interferometry (VLBI), the timing of millisecond pulsars and the positioning of mobiles on Earth (GPS) or in deep space. A Hydrogen maser has flown in space for two hours in 1976 in the NASA Gravity Probe experiment (GPA) [18]. Several tens of caesium clocks (as well as less precise Rubidium clocks) are now in continuous operation in GPS satellites orbiting at 20'000 km above the Earth.

Authors are : C. Salomon (1), A. Clairon (2), P. Laurent (2), G. Santarelli (2), S. Ghezali (2), M. Bahoura (2), P. Lemonde (1), J. Reichel (1) and A. Michaud (1)

(1) : Ecole Normale Supérieure
Laboratoire Kastler Brossel
24 rue Lhomond,,
F-75231 Paris Cedex 05, France
Tel: + 33 (1) 44 32 33 03 Fax: + 33 (1) 45 35 00 76

(2) Laboratoire Primaire du Temps et des Fréquences
Bureau National de Métrologie
Observatoire de Paris , 61 Av. de l'observatoire,
F-75014 Paris, France

Recent progress in the field of cooling and trapping of neutral atoms or ions is having considerable impact on time and frequency standards. Ions can be trapped with magnetic or radio frequency electric fields for days and weeks leading to very long coherent interrogation times (500 seconds) [2]. Commercial systems with stability approaching that of H masers exist and space versions are being actively developed. The field of laser cooling of neutral atoms is less than ten year old and the application to the realization of better time and frequency standards is actively pursued since 1989 [3, 4].

2. The benefits of cooling

Commercial caesium clocks use a 400 Kelvin atomic beam of caesium atoms. By contrast, caesium clocks using laser cooled atoms operate with atoms at temperatures in the 1 μK range allowing a 100-fold improvement in the observation time. Nowadays, one can very easily produce a dense gas of cold caesium atoms at a temperature of 2.5 μK corresponding to atoms with an r.m.s velocity of 12 mm/s [5]. These small velocities allow interaction times between atoms and electromagnetic fields approaching one second on Earth when using an atomic fountain geometry. As the atomic resonance linewidth in caesium clocks is inversely proportional to this interaction time, an improvement of two orders of magnitude in the clock performances over conventional devices running at a temperature of 400 K is expected. It is predicted that microgravity conditions should enable a further factor of ten improvement in the interaction time with a simple and compact device [6].

Furthermore, new laser cooling techniques enabling to reach temperatures below the temperature set by the single photon recoil are presently being developed in several laboratories [7, 8, 9, 10, 11, 12]. Atoms have been cooled in one and two dimensions to much less than the recoil velocity, down to the nano Kelvin range.

2.1. Atomic Fountain

On Earth, the gravity limits the available interaction time and it was Zacharias' idea in 1953 to use an atomic fountain to obtain interrogation time in the one second range when using Ramsey's method of separated oscillatory fields. Efforts to build such an atomic fountain in that period were unsuccessful because of collisions near the oven exit. Laser cooling and the very slow

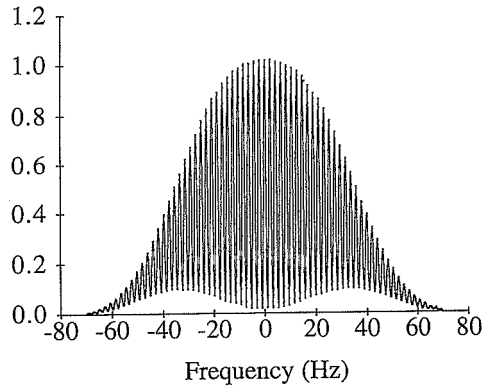


Fig. 1 - Ramsey fringes of the LPTF fountain clock for a 0.5 second flight time of the atoms above the microwave cavity [15]. The fringes are 1 Hz wide and signal to noise is 300 per point. The frequency detuning is scanned around the cesium hyperfine transition frequency at 9 192 631 770 Hz.

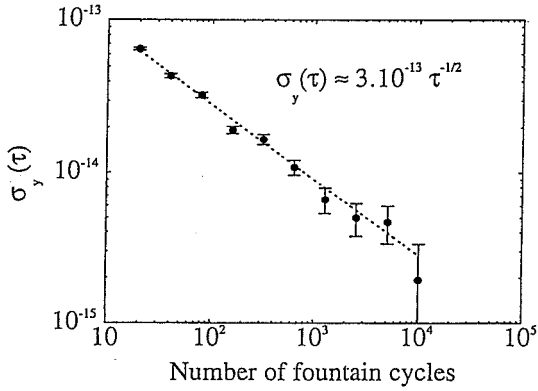


Fig. 2 - Frequency stability of the fountain clock measured against a Hydrogen Maser: Allan variance versus the number of fountain cycles [15]. Each cycle lasts 1.1 second. The stability reaches that of the H maser at about 10^4 seconds and is probably limited by the maser stability for longer measurement times.

velocities in optical molasses have enabled this idea to be brought-up to date [3, 4]. Fountain clock prototypes are being built in more than a dozen laboratories around the world and two of them (at Stanford [14] and LPTF [15]) are running as frequency standards. In the LPTF fountain, Ramsey fringes as narrow as 0.7 Hz have been obtained with a signal to noise ratio of 300 at one second integration time (Figure 1). The short term frequency stability, measured against an H-maser is $3\ 10^{-13}\ \tau^{-1/2}$ (Figure 2). While this stability is three times better than in any conventional caesium device, it is still limited by the local oscillator frequency noise (ultra-stable BVA quartz crystal) and not by the caesium fountain noise. The long term relative stability (Allan variance) as compared to a Hydrogen maser reaches $2\ 10^{-15}$ at 10'000 seconds, a value which is most probably limited by the H maser itself.

2.2. Microgravity clock

Gravity imposes an obvious limit on the resolution which can be obtained in a fountain clock: the resolution increases only as the square root of the fountain height. Going much beyond a 1 s fountain time would require an expensive and bulky apparatus and would probably be to the detriment of the accuracy. A promising alternative is to build a clock in a reduced gravity environment. Measurement time of several seconds can be envisaged using a smaller and simpler apparatus, leading to a factor of 10 improvement in resolution and accuracy. First experiments on laser cooling of atoms in microgravity were recently performed by our group in Paris using jet plane parabolic flights and very simple experimental techniques [6]. Efforts now concentrate on building a clock prototype designed for microgravity conditions. Such a system is an extension of the work developed for fountain clocks on Earth described in detail in [14, 4, 15, 16, 17]. A prototype of a micro-gravity clock is schematically described in [17] and is under study at CNES (PHARAO project). It includes a laser diode system and optics for atom manipulation and cooling,

a high stability 9.2 GHz microwave source and a vacuum tank containing a cooling region, a microwave interrogation region and a detection zone. The clock is operated fully automatically using a micro-computer.

3. Expected performances of a Space Clock

A first scientific objective of a clock experiment in space could be the demonstration of a clock running with laser cooled atoms in micro-gravity conditions and the determination of its performances. These performances are expected to be better than $10^{-13}\ \tau^{-1/2}$ where τ is the integration time in seconds. For one day of operation, the stability will be 10^{-16}, corresponding to an absolute time fluctuation of 30 picoseconds. This stability would be two orders of magnitude better than present Cs clocks. Such a demonstration would open a new era of space missions using ultra-stable clocks as space VLBI, high precision measurement of the sun gravitational potential to second order and of the Shapiro delay as proposed in the SORT mission (Solar Orbit Relativity Test) [13].

Since no other clock is known to give yet such performance for a one day measurement time, it will be necessary either to fly two identical systems in order to make frequency comparisons on board the spacecraft or to use high performance two-way time and frequency transfer systems with Earth-based clocks of similar performances. These frequency links could be in the microwave or optical domains using either X or K-band microwave signals or pulsed lasers as in the lunar ranging experiment. The required timing precision is on the order of a few ps at 10^3 s or, as said above, 30 ps at one day.

If microgravity cold atom clocks become eventually more stable than their Earth-based counterparts, two or more similar clocks on board the spacecraft are required in order to assess their performances using frequency intercomparison techniques.

4. Possible Applications

With clocks having a stability of 10^{-16} in one day, it should be possible to measure with a potential 100-fold improvement over the 1976 GPA experiment the gravitational red-shift (Einstein effect). This general relativity effect was determined at the 10^{-4} level using H masers having a stability slightly better than 10^{-15} over the 2 hour mission duration [1]. The precision on such a measurement depends directly upon the eccentricity of the orbit. A satellite with a large eccentricity (e=0.6, major axis 20'000 km and a 12 hour period) would be an ideal choice.

A second application belongs to the field of time and frequency dissemination. A geostationnary satellite is an ideal tool for dissemination toward the Earth of ultra-stable timing signals and for frequency comparisons between various Earth stations. The very low drift of the space clock

might also allow frequency comparisons between Earth-based clocks without the constraint of common view of the satellite. Positioning of mobiles on the surface of the Earth as in the GPS system could also benefit from satellite clocks with improved stability.

More ambitious space applications deal with measuring the relativistic Shapiro delay in the Sun gravitational potential with four orders of magnitude improvement in precision over current values [13]. A measurement of the second order Sun gravitational potential is also possible using these highly stable clocks.

4.1. Time and Frequency Transfer

When a stable clock is orbiting around the Earth, a crucial factor appears to be the quality of the frequency comparison between the satellite clock and one or several clocks on Earth. The picosecond time stability over a few minutes required for assessing properly the stability of the satellite clock demands that the frequency comparison be at least two-way. Since the GPA experiment, the Harvard group has performed detailed studies of several possibilities for achieving frequency comparisons below the 10^{-15} range [18]. Today, two major directions emerge, a microwave two-way link and an optical link using picosecond laser pulses [13]. By contrast to the optical link, the microwave link has the advantage of being independent of weather conditions. It is clear that both methods are very different in their approach of the time transfer. A comparison of the performances of the two methods is highly desirable and would certainly be an important issue. The synchronized two-way frequency transfers (Earth-spacecraft-Earth) and (spacecraft-Earth-spacecraft) allow complete cancellation of isolated tropospheric or ionospheric noise as well as of the first order Doppler effect [18]. Measurements of the relativistic effects (second order Doppler and redshift) requires a good knowledge of the satellite orbit and velocity.

5. Summary

The possibility to install high stability clocks in space is very attractive from the viewpoint of testing new concepts of clocks using cold atoms and for potential applications. The development of time and frequency transfer with picosecond accuracy is required in order to read these ultra-stable clocks in orbit. This would open the way for more ambitious physics missions such as space VLBI or tests of general relativity in solar orbit with unprecedented precision.

References

[1] R. Vessot et al., Phys. Rev. Lett. **45**, 2081 (1980).

[2] J. Bollinger, D. Heinzen, W. Itano, S. Gilbert and D. Wineland, IEEE Transact. on Instrum. and Measurement, **40**, 126 (1991).

[3] M. Kasevich, E. Riis, S. Chu and R. de Voe, Phys.Rev. Lett. **63**, 612 (1989).

[4] A. Clairon, C. Salomon, S. Guelatti and W.D. Phillips, Europhys. Lett. **16**, 165 (1991).

[5] C. Salomon, J. Dalibard, W. Phillips, A. Clairon and S. Guellati, Europhys. Lett. **12**, 683 (1990).

[6] B. Lounis, J. Reichel and C. Salomon C.R. Acad. Sci. Paris **316, Série 2**, 739 (1993).

[7] A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste and C. Cohen-Tannoudji, Phys. Rev. Lett. **61**, 826 (1988);

[8] F. Bardou, B. Saubamea, J. Lawall, K. Shimizu, O. Emile, C. Westbrook, A. Aspect and C. Cohen-Tannoudji, C. R. Acad. Sci. Paris, **318**, 877(1994).

[9] M. Kasevich and S. Chu, Phys. Rev. Lett. **69**, 1741 (1992).

[10] J. Lawall, F. Bardou, B. Saubamea, K. Shimizu, M. Leduc, A. Aspect, and C. Cohen-Tannoudji, Phys. Rev. Lett., **73**, 1915,(1994).

[11] N. Davidson, H.-J. Lee, M. Kasevich and S. Chu, Phys. Rev. Lett. **72**, 3158 (1994).

[12] J. Reichel, O. Morice, G. Tino and C. Salomon, Europhys. lett., **28**, 477 (1994).

[13] C. Veillet et al., TROLL proposal to ESA for M3 mission (1993).

[14] K. Gibble and S. Chu, Phys.Rev. Lett, **70**, 177 (1993).

[15] A. Clairon, Ph.Laurent, G. Santarelli, S. Ghezali, S. Lea and M. Bahoura, Proc. of CPEM 1994, (Boulder, USA).

[16] C. Salomon, B. Lounis, J. Reichel, A. Clairon, Ph. Laurent, N. Nadir, Rev. Sci. De la défense, **3**, 113 (1993).

[17] S. Lea, A. Clairon, C. Salomon, P. Laurent, B. Lounis, J. Reichel, A.Nadir and G. Santarelli, Physica Scripta, **51**, 78 (1994).

[18] R. Vessot, Proc. of the 28-e Rencontre de Moriond, (1993).

DYNAMIC INTERACTIONS BETWEEN IONOSPHERIC PLASMA AND SPACECRAFT

DAVID B. SNYDER

Abstract

In recent years, studies of the interactions between Space Station Freedom (SSF) and ionospheric plasma have led to an improved understanding of the dynamics of these interactions. Plasma currents from the ionosphere control surface potentials, but the charge stored across dielectric surfaces becomes an important consideration in predicting dynamics of arc development. Time scales for the resulting interactions can be scaled for specific circumstances. In addition, active surfaces such as antennae and switched solar array surfaces have fostered thought on the interactions of AC driven systems. These systems can, under certain conditions, give rise to radiation and enhanced sputtering of surfaces. This paper will review the work performed for the SSF program to understand the dynamics of spacecraft interactions, and will discuss implications to other spacecraft.

1. Introduction

During the previous decade, much work has been done to develop models to help understand and predict interactions between spacecraft and the plasma environment [1]. However, engineering leve codes such as NASCAP/Geo and NASCAP/LEO tend to rely on evaluating or tracing the evolution to, equilibrium conditions and drawing conclusions based on relatively constant conditions. While it is recognised that charging conditions are dynamic, the tools tend to assume rapid establishment of equilibrium conditions [2, 3]. However, with the development of new technologies for expensive spacecraft, and the need to predict and scale effects to new systems without extensive testing, it has become necessary to begin to study the dynamics of these interactions.

Early in the design of Space Station Freedom (SSF), issues of plasma were investigated with the ob- jectives of designing a plasma compatible space platform, and providing a platform suitable for ionospheric studies [1]. However, in during the several SSF redesigns and mission redefinitions, these issues were forgotten [4].

Dr. David B. Snyder is with the
NASA Lewis Research Center, MS 302-1
21000 Brookpark Rd,
Cleveland, Ohio 44135, USA
Tel: + 1 (216) 433 2217 *Fax: + 1 (216) 433 6106*

However the decision to ground SSF to the negative side of the solar arrays, Ferguson et al. raised several plasma compatibility issues [4]. This lead to the establishment of the SSF Grounding Tiger Team, which attempted to evaluate the impact of arcing of SSF [5]. While this work has raised additional questions for further research, it has also contributed to a better understanding of how spacecraft respond to various plasma interactions. As new technologies are applied to new spacecraft, in particular those performing various ionosphere investigations, some of these plasma compatibility issues may become relevant.

The purpose of this paper is to review some of the plasma compatibility issues raised in the course of the SSF investigations, and where possible discuss the dynamic characteristics of these effects, both to help spacecraft users better understand the implications of these effects on their measurements, and to suggest future directions for research in plasma-spacecraft interactions.

2. Interactions and Time Scales

If a spacecraft is not effectively 'grounded' to the ionosphere, fluctuations in spacecraft potential can occur on a variety of time scales, from DC to microseconds. There can be a DC offset driven by exposed portions of the powerpsystem, due to current density differences in positive and negative species. This is roughly driven by $\sqrt{T_e / T_i}$, where T_e refers to the electron temperature and T_i is the ion temperature. If the power generation system is sensitive to illumination (solar cells) there may be a voltage transient associated with entering and leaving eclipse. There may be transients associated with switching of systems on-board the spacecraft, for example solar array circuits, or operation of high voltage experiments. As the potential of the spacecraft changes, current is collected on the surface trying to bring the spacecraft back to equilibrium. In many cases the current collection mechanisms can be identified making it possible to estimate the times scale, and hence the frequency domain of the transients. We would like to be able to predict the magnitudes of some of these effects in order to assess their impact on measurements, or to justify requirements on spacecraft design.

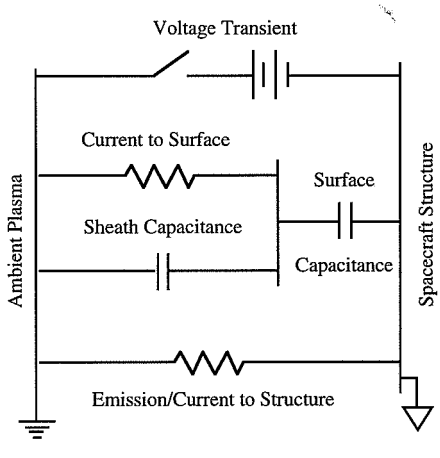


Fig. 1 - Spacecraft-Plasma interaction schematic diagram

In this work the term grounded is used in a couple of ways. The 'plasma ground' is used to describe the electrical connection between the spacecraft and the local environment, i.e. the ionospheric plasma. The 'spacecraft ground' refers to the internal process of referencing potentials to a common place on the spacecraft, usually the structure. This is comparable to a chassis ground.

To track the potential changes during transients it is helpful to look at a simplified description of the spacecraft-plasma interaction. Figure 1 shows a schematic diagram to illustrate this. The electric potential of the spacecraft can shift by changes in the energy or current of ion or electron collection, or electron emission. The capacitance of the spacecraft to plasma, or across a surface coating, plays an important role in determining the magnitude of the interaction, while the current collection and emission mechanisms then contribute to the time scale of the transient.

The capacitance of the spacecraft sheath tends to be much smaller than the capacitance of the surface. The sheath thickness tends to be fractions of a centimeter, thicker with high potentials, while dielectric materials on surfaces tend to be fractions of a millimeter. The dielectric constant of surface materials tend to be higher than that of vacuum also contributing to a higher capacitance. An interaction that occurs on a completely insulated satellite, for example a voltage shift due to auroral interactions, will involve relatively small currents even if large voltage excursions occur, because of the small capacitance. However, interactions that involve the spacecraft structure, for example arcing, can access the energy due to the relatively large dielectric capacitance of the surfaces.

3. Arcing

Transients due to arcing are of interest because (1) they may lead to damage of the affected surfaces or contamination of other spacecraft surfaces, and (2) they are probably the most severe cases of voltage change and time scale, aside from planned antenna use. Our interest has been more in how these transients may damage spacecraft than in how they effect measurements. However, these assessments illustrate some of the issues involved in evaluating the impact of other transient effects.

Arc Evolution Mechanism

There are several known mechanisms which can initiate an arc; dielectric breakdown [6], micrometeor and debris impact [7, 8], solar cell (edges or interconnects) arcs [9]. The initiation mechanism is important to understanding some arc characteristics such as arc frequency and arc threshold potentials. However as will be seen later, if a large enough initiation event occurs and the substrate is biased negative relative to the ambient plasma, the evolution of the arc appears to be independent of the initiation mechanism [10]. The key common feature appears to be a substrate biased negative relative to the ambient plasma, covered or nearly surrounded by a dielectric layer. This produces electric fields which collect electrons on the dielectric surface and focus ions back to

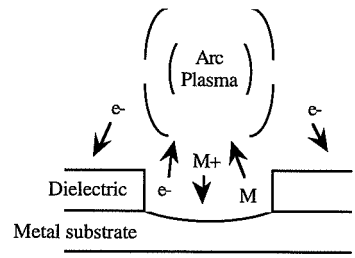


Fig. 2 - Proposed arc evolution mechanism

the metal or conductive arc surface. Three issues need to be addressed in an arc circuit mechanism; (1) Development of the arc plasma, (2) Transport of electron current to the surface, and (3) transport of current through the spacecraft. Figure 2 illustrates the hypothetical process.

In this evolution model an initiation event provides an initial ignition plasma at the arc site and, if not already exposed, exposes the underlying conductor. Due to the dielectric material surrounding the arc site, a high electric field exists at the site, which focuses ions from the plasma back to the arc site. Bombardment of the site by the attracted ions may cause sputtering or sublimation of neutrals. Electrons can be emitted from the site by a combination of thermionic and field emission. If the collision lengths are such that the electrons can collect a few ten's of eV of energy, ionization may occur when they strike the emitted neutrals, thereby creating ions and sustaining the arc plasma.

As electrons are emitted from the arc site and arc plasma, the potential of the substrate rises. The potential of the dielectric surface also rises due to capacitive coupling with the substrate. The surface now collects electrons from both the arc plasma, or the ambient plasma. Since the capacitance to space is ordinarily much smaller than the capacitance across the dielectric, most of the potential change appears across the plasma sheath, the potential of the surface also rises. Then, as charge is attracted from the plasma, discharging the dielectric. If the arc continues on the time scale of the dielectric discharge, the energy available from the dielectric capacitance can continue to drive the arc.

The current available to drive the arc is limited by the ability of the spacecraft surfaces to collect current from the arc or ambient plasma. The arc source appears to be the principal limit on arc currents, from tenths of an amp for small capacitance systems [11] (few hundred picofarads), to a thousand amps for large capacitance systems [6] (a thousand microfarads). Unless provision is made in the design, spacecraft structure does not appear to affect arc development, except perhaps in the rise times. During the Electrical Grounding Tiger Team discussions, it was argued that spacecraft inductances might prevent arc development. But during tests, the inductance of the wiring and internal inductance of capacitors were not sufficient to prevent arcs nor did they obviously effect development. Arcs appear to develop slowly enough (microseconds) that inductive effects seem to be unimportant to the arc evolution.

So far it is assumed that the structure can return the current generated by the arc plasma. But what limits are there on the current that can move to the surface? Is this current sufficient to sustain an arc? Two independent mechanisms are examined for this part of the current loop. A lower bound on the current available can be estimated by the electron thermal current to the spacecraft, i.e. for large spacecraft, the product of the electron thermal current density with the spacecraft area. This is the current available to the spacecraft due to its changing voltage, neglecting geometric and plasma sheath considerations.

For normal spacecraft voltage shifts this probe-like collection is the mechanism that governs current collection. But for arcs there is an additional source of current.

The second mechanism assumes that the current is due to the expansion of the plasma arc. Vaughn et al. [12] noted a delay in the plasma enhancement seen by a movable Langmuir probe. For their configuration (anodized aluminum biased to -240 V) they estimated a primary expansion velocity of about $3x10^4$ m/sec. This model can be used to estimate the current due to an expanding arc plasma [13] of $I = CV\, dA/dt = 2\pi CV\, v$, were A is the dielectric area covered by the arc plasma, and v is the expansion velocity. This estimate represents an upper bound as it assumes that the coating capacitance is instantaneously and completely discharged as the arc plasma moves over it. However, this assumption will break down as the arc plasma density falls due to expansion. Eventually the density may fall sufficient to terminate the arc. This mechanism suggests a current limited by the expansion velocity of the arc plasma.

A third mechanism related to the second may produce higher currents for microsecond time scales. On this time scale the ions near the 'old' plasma sheath edge have not had time to leave and form a new plasma sheath. So electrons from the arc plasma sheath are not space-charge-limited, at least in the usual sense of the term.

We have seen immediate increases in electron current on a distance scale of 1/2 meter on a sample made of several plates [14]. Figure 3 shows return currents during an arc on a set of concentric anodized aluminum rings. The total area of the sample was about 2 m^2. The arcs, instead of occuring on the center plate as intended, occured on the outside ring. In spite of the 3 to 4 meters of wire forming the electrical connection, currents arrived simultaniously at all the plate. The magnitude of the currents were approximatly proportional to the ring area.

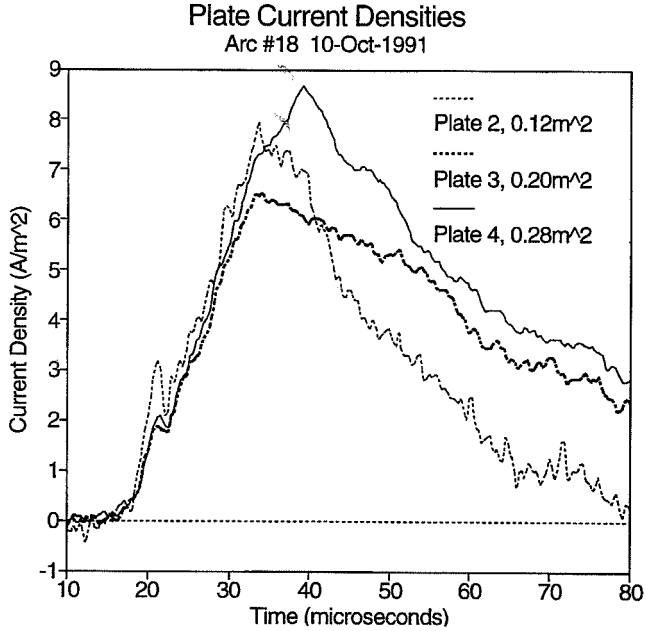


Fig. 3 - Arc return currents to a series of concentric rings

Arc Magnitude, time scales

An obvious parameter for scaling ground based studies to spacecraft is the system capacitance. The capacitance together with the potential difference to plasma at the arc site determines the charge and energy available for a discharge, and how the electric fields associated with the arc event will evolve over time. This hypothesis suggests that it may be reasonable to study arcs from large space systems by simulating them with a comparable capacitance.

Figure 4 [10] illustrates an observed correlation between peak current and system capacitance. The peak current is an interesting parameter since it puts a lower bound on the duration of an arc, and on the magnitude of the EMI generated by the arc. The data is from a variety of sources [8, 10, 15-17]. During tests, wide variations in the peak currents are seen for a particular set of conditions. Calculated standard deviations of the peak currents are on the order of the average. Points in the figure generally indicate the average of a number of arcs, but sometimes a range is indicated instead. The data from Balmain [1] 6 was for a Helium plasma at millitorr pressures.

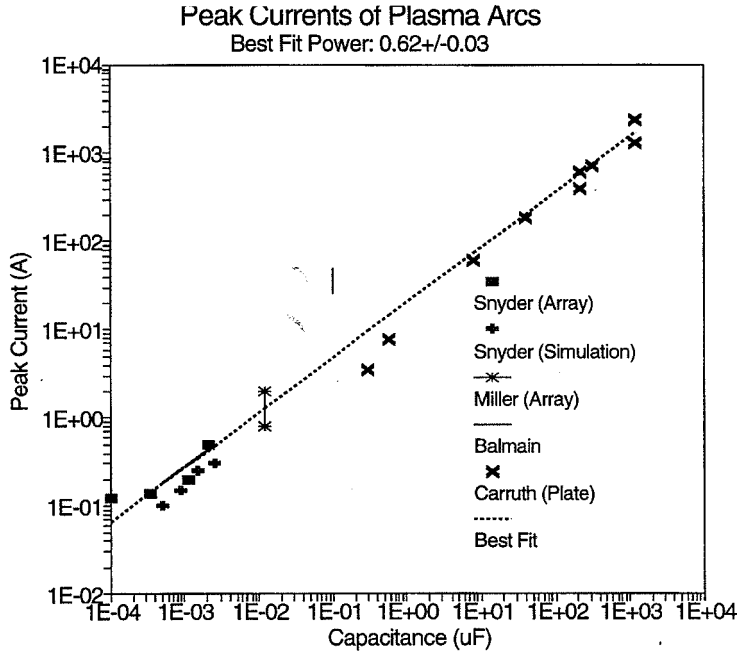


Fig. 4 - Observed peak arc currents as a function of capacitance

Figure 5 [10] shows the same sets of data plotted against energy instead of capacitance. The energy is calculated from $1/2CV_0^2$ where C is the capacitance and V_0 is the bias voltage. This introduces some dependence based on voltage. The high capacitance data was obtained in connection with space station tests at about 100 to 150 V while the low capacitance data was taken during studies of solar cell arcs with potentials closer to 1000 V. We note that the high capacitance data

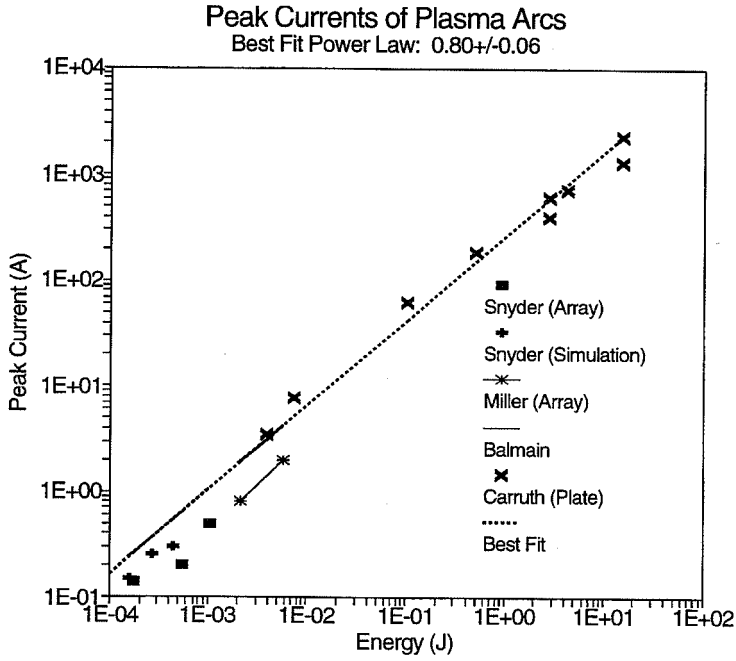


Fig. 5 - Observed peak arc currents as a function of energy

extrapolates to, and can be extrapolated from the low capacitance data. This is not quite the case when the energy data is examined. This suggests that capacitance maybe a better predictor of peak arc current than is energy. The increased voltage appears to extend the arc duration rather than increase the current.

As a working hypothesis we use the correlation between the peak arc current during an arc and the capacitance of the system to predict arc currents. But this hypothesis is not based on a rigorous arc model. Instead it is based on empirical observation. A correlation with bias voltage or energy stored in the system might be easier to understand. Development of a quantitative arc development model might suggest extrapolation procedure that could be used with more confidence.

The above technique is used as a way to estimate arc current to about a factor of two. Using time to deplete the charge stored in the system permits an estimate of the arc duration, i.e. $\Delta t = C\Delta V/I_p$, where Δt is a lower bound on thearc duration, ΔV is the change in voltage during an arc, i.e. a material dependent cutoff voltage subtracted from the bias voltage, and I_p is the estimated peak current. In practice the current dies down with the substrate voltage, but this method permits an order of magnitude estimate, enough to see in what frequency regime interference might be expected.

Mitigation

Presently we suspect that there are material dependent voltage thresholds. For silicon cells these appear to about -200V for silicon solar cells [18], and about -50V for anodic oxide coated aluminum [14], and kapton covered copper [17]. The solar cell arc threshold is an empirical observation, but coincides with the voltage where arcs appear to shut off. The copper and kapton thresholds are based on arc shutoff potentials, but simulated debris hit induced arcs have been observed at 75V on aluminum [19], and one sample arced repeatedly at 50V [14].

Two types of systems can easily develop potentials significantly different from the ambient plasma. For large spacecraft vxB induced potentials can be significant, tens of volts for ISSA (International Space Station Alpha) sized structures. More commonly the spacecraft potential will be determined by exposed biased conductive surfaces such as solar cells, or other active equipment such as high voltage experiments. It may be possible to electrically isolate these systems from the spacecraft structure and the rest of the spacecraft if they would otherwise cause excessive potentials. However, it may be difficult to provide sufficient isolation. If it is necessary to ground to the structure and both positive and negative potentials are exposed, grounding to the positive side is preferred. It is typically the exposed surfaces that provide the electrical connection to the ambient plasma. Since the positive surfaces tend to collect electrons while negative surfaces tend to collect ions, the positive surfaces have less effective resistance to plasma. It will normally be prudent to ground the positive side to structure.

It is possible to provide an electrical connection to the ambient plasma on a negative ground spacecraft using a hollow cathode plasma contactor, or some other device capable of providing relatively large currents. This is the technique being used for ISSA, where it is anticipated that about an ampere of current will be driven through ground cables to control the spacecraft potential.

4. A/C Interactions

Most studies of plasma interaction assume that a system eventually reaches some equilibrium condition, i.e. conditions stop changing. This is obviously not true of systems with a driven component. However, systems that are periodic may achieve a steady state where the changes are repetitive. This can apply to AC (Alternating Current) power distribution systems, antennae, some active experiments, and solar array power control systems. The criteria for reaching a 'steady state' is that the net charge collected during a period is zero. If both secondary emission and backscatter collisions are negligible, the electrons collected during the positive part of the cycle will be equal to the ion collection during the negative part. Except for very low frequencies, this tends to drive the system negative until the maximum negative potential on a dielectric surface is nearly twice the amplitude of the driving oscillation.

Our main concern has been that sputtering rates may be higher than expected due to ion collection at energies higher than otherwise anticipated. In fact this technique is commonly used at much higher plasma densities, voltages and frequencies to sputter dielectrics in plasma reactors. The

effect may be particularly important in low earth orbit where thin atomic oxygen resistant coats may be sputtered away exposing underlying polymers to attack. Kennedy [20] has documented this sputtering for ionosphere-like conditions. However, even under more benign circumstances the effect may be of interest to experimenters as it results in larger plasma sheaths than otherwise anticipated and may cause significant fluctuations in spacecraft potential. It is conceivable that in extreme cases the spacecraft potential could fluctuate relative to plasma, disturbing some measurements.

Mechanism: AC Surface Potentials

Figure 6 shows the circuit diagram used to calculate the potential of the surface [21]. $V_a(t)$ is the electric potential of a driving conductor measured with respect to plasma and $V_s(t)$ is the potential of the surface with respect to plasma. C_1 is the capacitance between the driving conductor and the exposed surface. This will be related to the dielectric coating of the surface but may include other artificial capacitances between the underlying conductor at the surface and the power system or other power sources. $C_p(V_s)$ is the effective capacitance between the surface and the plasma. In general this will depend on the plasma sheath thickness, and will vary with surface potential. For the case of a weak plasma this will be small and the surface potential will be near the driving voltage. However, for high density plasma and thin plasma sheath, especially when the capacitive coupling of the surface to the driving conductor is weak, C_p may be important and the surface potential will be some fraction of the driving potential. $I(V_s)$ is the current from plasma to the surface and may be obtained from probe theory.

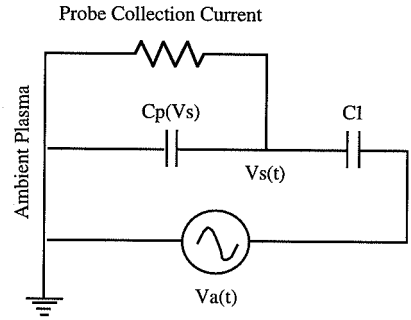


Fig. 6 - Schematic diagram of circuit to evaluate surface potentials, V_s

Since the displacement current through C_1 is equal to the displacement current plus the probe collection current through the sheath,

$$\frac{d}{dt}C_1\left(V_a(t)-V_s(t)\right)=I(V_s)+\frac{d}{dt}C_p(V_s)V_s(t) \tag{1}$$

From equation (1) the rate of change of the surface potential can be obtained,

$$\frac{dV_s}{dt}=\frac{\dfrac{dV_a}{dt}-\dfrac{I(V_s)}{C_1}}{1+\dfrac{1}{C_1}\dfrac{d}{dV_s}C_p(V_s)V_s} \tag{2}$$

Note that the factor dC_pV_s/dV_s is an effective capacitance of the plasma sheath. If C_p is independent of V_s, it reduces to C_p.

The capacitive term in the denominator is due to the voltage dividing effect of the two capacitances C_1 and Cp. Normally C_1 will be much greater than C_p, since C_1 is usually due to thin dielectric films and C_p has a minimum thickness of the plasma debye length (i.e. cm scale lengths). Thus the capacitive term will be negligible, and the surface potential, V_s will tend to track the driving potential, V_a.

The current term in the numerator serves to bring the surface toward plasma ground. If $V_a(t) = V_0 exp(i\omega t)$, the magnitude of $dV_a=dt$ will be on the order of $V_0\omega$, where V_0 is the amplitude of the driving voltage and ω is the angular frequency. If I/C_1 is much larger than this, the surface is effectively shorted to plasma ground. In practice, however, the plasma current term will be smaller than $V_0\omega$, and instead this term determines a) what the time average of the surface potential is, and b) how long it takes to get there.

Multiplying equation (2) by the denominator of the right side and integrating over from t_0 to t gives,

$$V_s(t)\left[1+\frac{C_p(V_s(t))}{C_1}\right]-V_s(t_0)\left[1+\frac{C_p(V_s(t_0))}{C_1}\right]=\left(V_a(t)-V_a(t_0)\right)-\int_{t_0}^{t}\frac{I(V_s)}{C_1}dt \qquad (3)$$

An equilibrium condition is reached, i.e. $V_s(t)$-$V_s(t_0)=0$, for periodic V_a, i.e. $V_a(t)$-$V_a(t_0)=0$, when the charge collected over a cycle, $\int_{t}^{t+\tau} I(V_s)dt$, is zero. Since electron current densities tend to be much higher than ion current densities in ionospheric plasmas, V_s will charge somewhat over each cycle resulting in a increase in the ion collecting part of the cycle at the expense of the electron collecting part, for high enough frequencies. This continues until V_s is nearly offset by $-V_0$, so that V_s varies from a small positive value to nearly $-2V_0$.

The long term behavior can be discussed by examining the change in V_s over one period, τ, of the driving voltage. Equation 3 becomes instead,

$$V_s(t+\tau)\left[1-\frac{C_p(V_s(t+\tau))}{C_1}\right]-V_s(t)\left[1-\frac{C_p(V_s(t))}{C_1}\right]=-\int_{t_0}^{t+\tau}\frac{I(V_s(t'))}{C_1}dt' \qquad (4)$$

If V_s changes so little over one period that $Cp(t_0+\tau)$ is nearly the same as $Cp(t_0)$, then

$$V_s(t_0+\tau)-V_s(t_0)=-\int_{t_0}^{t_0+\tau}\frac{I(V_s(t'))dt'}{C_1+C_p(V_s(t'))} \qquad (5)$$

Here, it can easily be seen that V_s will settle to an equilibrium condition once the charge accumulated over a cycle is zero. If $\langle I(t)\rangle$ is defined as $(q(t+\tau)-q(t))/\tau$, then

$$\frac{\Delta V_s(t)}{\tau} = \frac{\langle I(t)\rangle}{C_1 + C_p(V_s(t))} \tag{6}$$

Which suggests that for high driving frequencies, where $V_0\omega$ is much greater than the maximum probe currents, the long term behavior of V_s ignoring the driving oscillations can be described by

$$\frac{dV_s(t)}{dt} = \frac{\langle I(t)\rangle}{C_1 + C_p(V_s(t))} \tag{7}$$

Flat Plate

Some implications of this model can be examined for a simplified case of a flat plate driven by a sinusoidal voltage. For this example edge effects and details of how the plasma sheath grows will be ignored. It is assumed that for positive potentials the surface collects electrons at their thermal current density, and at negative potentials the surface collects ions at their thermal current density. Figure 7 illustrates the three applicable frequency regimes. It was generated from the results from a computer model which integrated the collected current to track the surface potential. The electron current density, J_e for this case is $-1.2\text{x}10^{-4}$ A/m^2 and the ion current density J_i is $4.9\text{x}10^{-7}$ A/m^2. The capacitance, C_1, is $4.4\text{x}10^{-8}$ F/m^2, and C_p is considered to be negligible.

The three frequency regimes are determined by a comparison of amplitude of the rate of change of the driving voltage (V_0w) with the thermal current density. At low frequencies C_1dV_d/dt is always less than the ion current density and the plasma can always supply enough current to keep the surface at 0 V. Once C_1dV_d/dt exceeds the ion current density then the surface can begin to develop negative voltages, and the average voltage begins to drop. At high frequencies, where the RMS current is saturated, the driving voltage changes rapidly enough that it is always collecting either the full electron current density or the full ion current density. For this flat plate case the steady-state charging condition dictates that ratio of time spent positive (electron collecting) to that spent negative (ion collecting) is the ratio J_i/J_e. For this case, only $4.1\text{x}10^{-3}$ of the cycle is spent with a positive surface. For the rest of the cycle the surface is negative, collecting ions, and the maximum negative potential attained is only slightly less that $-2V_0$.

At extremely high frequencies, near or above the sheath formation times, the mechanism for the transport of charge as the sheath develops becomes important and the above model does not hold.

Mitigation

The level of attention paid to addressing these issues will be mission and system dependent. Effects from some systems such as antennae will simply have to be tolerated. Hopefully antenna operate at frequencies high enough that the above analysis is not applicable. Sensitive equipment

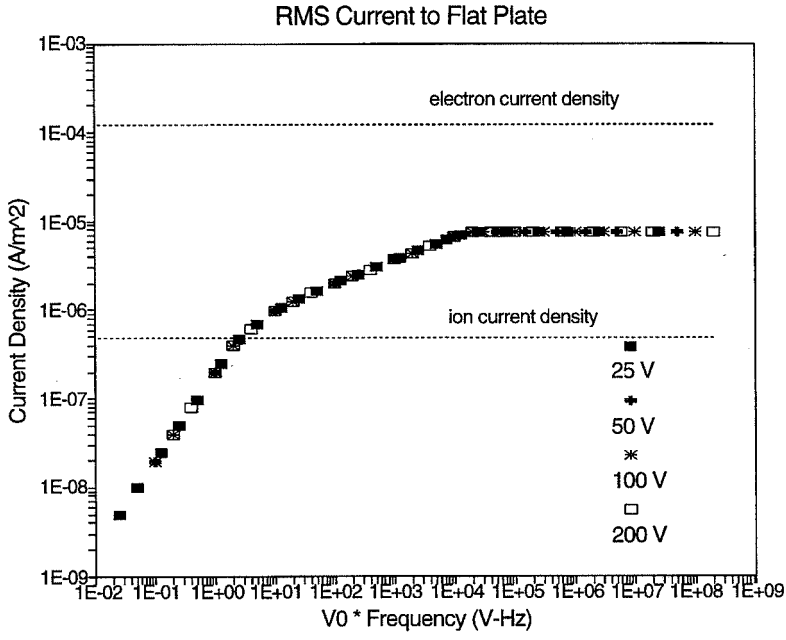


Fig. 7 - Root-Mean-Squared (RMS) currents to a flat plate

should be placed far enough away that they will not see the plasma sheath's from this equipment. Cables should be shielded, if not individually, at least collectively so the plasma does not see and react to them.

It may be possible to reduce fluctuations in spacecraft potential by including some kind of plasma contactor, i.e. a small electron emitter, to reduce the negative excursions.

Conclusions

An understanding of dynamic interactions with ionospheric plasma is beginning to be developed. The issues related to arcing are still quite controversial. This work has looked at some of the issues related to developing and sustaining arcs in ionospheric conditions. It has also presented a technique for estimating the amplitude and duration of arcs. This technique uses the capacitance of the system to estimate the peak current, and then uses the charge stored to estimate the duration of an arc.

In addition, as new technologies are implemented on spacecraft, new issues of environmental compati- bility will arise. This work has also looked at some of the issues related to driving dielectric surfaces with AC voltages. The steady-state charging criteria developed is that over an oscillation the ion charge collected is compensated for by the electron charge collected. This tends to drive the average potential negative, so that only for a small portion of the cycle is the dielectric surface positive.

The material discussed here only begins to touch on the issues related to dynamic interactions that will at least affect experiment operations and, if due care is not taken, may affect spacecraft reliability and lifetimes. Some of the work presented here is somewhat speculative, but may suggest ideas and hypotheses for future experimental and theoretical work.

References

[1] SSPIE Working Group Meeting, Cleveland, OH, April 26,27, 1987.

[2] Purvis, Carolyn K., Garrett, Henry B., Whittlesey, A. C., and Stevens, N. John, "Design Guidelines for Assessing and Controlling Spacecraft Charging Effects", NASA TP2361, Sept 1984.

[3] V.A Davis and L.W. Duncan, Spacecraft Charging Handbook, PL-TR-92-2232, Phillips Laboratory, Directorate of Geophysics, Air Force Materiel Command, Hascom Air Force Base, MA 01731-5000, 1992.

[4] Ferguson, Dale C., Snyder, David B., and Carruth, Ralph, "Findings of the Joint Workshop on Evaluation of the Impacts of Space Station Freedom Ground Configurations", NASA TM 103717, 1990.

[5] Clouber, D., and Tye, C., Space Station Freedom Program (SSFP) Electrical Grounding Tiger Team Preliminary Integrated Assessment Report,Under NASA Contract: NAS 9-18200/B, McDonnell Douglas Space Systems Company, Space Station Division, Houston, TX, 1991.

[6] Carruth, M.R., Vaughn, J.A., Bechtel, R.T, Gray,, P.A., "Experimental Studies on Spacecraft Arcing", Journal of Spacecraft and Rockets, 30(3), pp. 323-327, 1993.

[7] Schneider, E.,"Micrometeorite Impact on Solar Panels", Photovoltaic Generators in Space, European Space Agency, ESA SP-267, The Hague/Scheveningen, The Netherlands, pp. 171-174, Oct. 1986.

[8] Weishaupt, U., Kuczera, H., and Rott, M.,"Micrometeorite Exposure of Solar Arrays",Photovoltaic Generators in Space, European Space Agency, ESA SP-267, The Hague/Scheveningen, The Netherlands, pp. 175-180, Oct. 1986.

[9] Thiemann, H. Schunk, R. W. and Bogus, K., "Where Do Negatively Biased Solar Arrays Arc?", Journal of Spacecraft and Rockets, 27, pp. 563-565, 1990.

[10] Snyder, David, "Arc Current Loop", Minutes of the Electrical Grounding Tiger Team Meeting, Boeing Trade Zone, Huntsville, Alabama, May 14-17, 1991.

[11] Snyder, David B., "Characteristics of Arc Currents on a Negatively Biased Solar Cell Array in a Plasma", NASA TM 83728, IEEE Transactions on Nuclear Science, NS-31(6), pp. 1584-1586, 1984.

[12] Vaughn, J.A., Carruth, M.R., and Gray, P.A., "Current Flow in a Plasma Caused by Dielectric Breakdown", Fifth Annual Workshop on Space Operations Applications and Research, NASA CP-3127 vol 2, pp. 724-731, 1992.

[13] Snyder, D.B., "Arc Current Loop", Minutes from the Joint Meeting of the Electrical Grounding Tiger Team and the Electrical Power System Working Group, Rocketdyne Building, Fairview Park, OH, Aug 5-7, 1991.

[14] Snyder, D.B. , "Arc Return Currents, Ring Test Analysis", Plasma Testing Meeting, Memphis, 29-31 Jan 1992.

[15] Miller, W., "An Investigation of Arc Discharging on Negatively Biased Dielectric-Conductor Samples in a Plasma", Spacecraft Environmental Interactions Technology 1983, NASA CP-2359, AFGL-TR-85-0018, pp.367-377, 1985.

[16] Balmain, K.G.,"Laboratory Measurements to Characterize Anodization Arc Discharges in a Plasma", Minutes of the Electrical Grounding Tiger Team Meeting, Boeing Trade Zone, Huntsville, Alabama, May 14-17, 1991.

[17] Snyder, D.B., "Ground Testing", Minutes Space Station Plasma Interactions and Effects (SSPIE) Working Group Meeting, NASA Lewis Research Center, Cleveland, OH, June 3-5, 1987.

[18] Ferguson, D.C., "The Voltage Threshold for Arcing for Solar Cells in LEO - Flight and Ground Test Results", NASA TM 87259, NASA Lewis Research Center, Cleveland, OH, 1986.

[19] Carruth, M.R., and Vaughn, J.A.,, "Debris Induced Arc Testing", Minutes from the Joint Meeting of the Electrical Grounding Tiger Team and the Electrical Power System Working Group, Rocketdyne Building, Fairview Park, OH, Aug 5-7, 1991.

[20] Kennedy, M., Low-Energy Radio Frequency Sputtering of Copper, Anodized Aluminum, Kapton, by Argon Plasma Ions, Ph.D. Thesis, Case-Western Reserve University, Cleveland, OH, 1995.

[21] Snyder, D.B., Grier, N., Domitz, S., "AC Current Collection from a Plasma", Minutes of the Electrical Grounding Tiger Team Meetng, Rocketdyne Building, Cleveland, Ohio, Oct. 31 - Nov. 2, 1990.

SOLAR SYSTEM PLASMA WAVES

DONALD A. GURNETT

Abstract

An overview is given of spacecraft observations of plasma waves in the solar system. In situ measurements of plasma phenomena have now been obtained at all of the planets except Mercury and Pluto, and in the interplanetary medium at heliocentric radial distances ranging from 0.29 to 58 AU. To illustrate the range of phenomena involved, we will discuss plasma waves in three regions of physical interest: (1) planetary radiation belts, (2) planetary auroral acceleration regions, and (3) the solar wind. In each region we will describe examples of plasma waves that are of some importance, either due to the role they play in determining the physical properties of the plasma, or to the unique mechanism involved in their generation.

1. Introduction

This overview of solar system plasma waves is presented in celebration of the 75th anniversary of the International Scientific Radio Union (URSI).The study of naturally occurring plasma waves is an old subject that has its origins in early ground-based investigations of very-low-frequency (VLF) radio signals. The first known report of VLF radio signals of natural origin was by Preece [1894], who detected a variety of unusual audio frequency electrical signals while experimenting with a long-distance telephone line. Based on his description, the signals that he detected were probably spherics, which are lightning-generated signals that propagate in the Earth-ionosphere waveguide, and a type of VLF radio emission now known as chorus. Some years later, during World War I, Barkhausen [1919] described unusual whistling signals lasting several seconds that he detected while attempting to intercept enemy telephone communications using a rudimentary vacuum-tube amplifier. Initial progress on the understanding of these whistling signals, which soon came to be called whistlers, was slow. After several years of investigation, including detailed analyses of their frequency-time characteristics, Eckersley [1935] correctly postulated that whistlers were produced by lightning. However, the mechanism that caused the dispersion, and the long propagation paths required to achieve the several-second travel time were unknown. It was not until the 1950s that Storey [1953] was able to provide a satisfactory

Prof. Donald A. Gurnett is with the
Department of Physics and Astronomy
The University of Iowa
Iowa City, IA 52242 USA
Fax: +1 (319) 335 1753

explanation of whistlers. He showed that the dispersion occurred as the signal propagated along the magnetic field line from one hemisphere to the other in a plasma mode of propagation now known as the whistler mode.

In addition to whistlers, which are produced by lightning, a number of other unusual VLF signals were discovered in this early era that were clearly not due to lightning. The best known of these are "chorus" and "hiss." The term chorus was introduced by Storey[1953] because the signals sounded like the early morning chorus from a colony of birds. The hiss signals produced a hiss-like sound in the audio output of the receiver. At high latitudes, hiss was soon found to be associated with the aurora, and this type of emission came to be know as "auroral hiss." Among the early investigators, Ellis [1957] had the distinction of being the first to propose a theory for the generation of these emissions. He suggested that auroral hiss was produced by Cerenkov radiation from the charged particles responsible for the aurora. As will be discussed later, elements of his ideas still exist in modern theories of auroral hiss. For a more extensive review of early ground-based observations, see Helliwell [1965].

The launch of the first Earth-orbiting satellites in the late 1950s opened up an entirely new era in the study of space plasma wave phenomena. VLF receivers on Earth-orbiting spacecraft soon revealed an extremely complex variety of plasma waves in the ionosphere and in the hot magnetized plasma surrounding the Earth known as the magnetosphere. Many of these waves had never been seen in laboratory plasmas. Because of the ideal conditions that existed in space, with negligible collisions and no walls, the study of space plasma waves soon became an important element of modern plasma research. The growth of plasma waves was found to play a crucial role in establishing an equilibrium in a plasma, in much the same way that collisions in an ordinary gas play a role in achieving thermal equilibrium. Plasma waves and radio emissions also often provided a useful diagnostic tool for determining various plasma properties, such as the electron density and magnetic field strength. Because of the importance of these measurements, radio and plasma wave instruments were soon routinely included on missions to other planets. Now, nearly forty years after the launch of the first Earth-orbiting satellite, in situ measurements of plasma waves and radio emissions have been obtained at all of the planets except Mercury and Pluto, and in the interplanetary medium at heliocentric radial distances from 0.29 to 58 AU (astronomical units). Since this subject covers such a broad area of research, this overview must necessarily be limited. To restrict the scope of this review, we will concentrate on three regions of physical interest: (1) planetary radiation belts, (2) planetary auroral acceleration regions, and (3) the solar wind. In each region, we will describe certain selected examples of some importance, either due to the role they play in determining the physical properties of the plasma, or to the unique mechanism involved in their generation.

2. Planetary radiation belts

A radiation belt consists of energetic electrons and ions that are trapped in the magnetic field of a planet. The Earth's radiation belt was discovered by Van Allen [1959] using Explorer 1, which was the first U.S. satellite. Since then radiation belts have been discovered at five other planets, Mercury, Jupiter, Saturn, Uranus, and Neptune, all of which have substantial magnetic fields. The energies of the radiation belt particles vary considerably from planet to planet, but are generally in the range from a few keV to several MeV. Jupiter has by far the most intense and energetic radiation belt. In situ measurements of plasma waves and radio emissions have been made in all of the known radiation belts, except for Mercury. A summary of the various types of plasma waves that have been observed is given in Table 1. Of these, we will describe two that are of particular importance. These are (1) whistler-mode emissions, and (2) electromagnetic ion cyclotron emissions.

Type of Plasma Wave	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune
Whistlers (lightning)	X	X		X			X
Whistler-mode hiss		X		X	X	X	X(?)
Whistler-mode chorus		X		X	X	X	
Auroral hiss		X		X			
Cyclotron maser radiation		X		X	X	X	X

Table 1

A. Whistler-mode emissions

The whistler mode is an electromagnetic mode that propagates at frequencies below the electron cyclotron frequency, $f_c = eB/m_e$, where e and m_e are the charge and mass of an electron, and B is the magnetic field strength. The whistler mode is right-hand polarized with respect to the magnetic field. Two types of whistler-mode emissions, chorus and hiss, are commonly observed in planetary radiation belts. Chorus consists of numerous discrete narrow band emissions, usually rising in frequency on time scales of a few seconds. Hiss consists of a nearly steady level of band-limited noise, usually below the frequencies at which chorus is observed. A frequency-time spectrogram of chorus and hiss in the Jovian radiation belt is shown in Figure 1. These emissions were observed during the Voyager 1 flyby of Jupiter [Scarf et al., 1979], and are remarkably similar to the terrestrial chorus and hiss emissions detected during the early era of ground-based VLF studies.

Chorus and hiss are both believed to be generated by a cyclotron resonant interaction with radiation belt electrons. Cyclotron resonance occurs when the wave frequency in a frame of reference moving along the magnetic field with the particle matches the cyclotron frequency of the particle. The wave field in the moving frame of reference must also be rotating around the

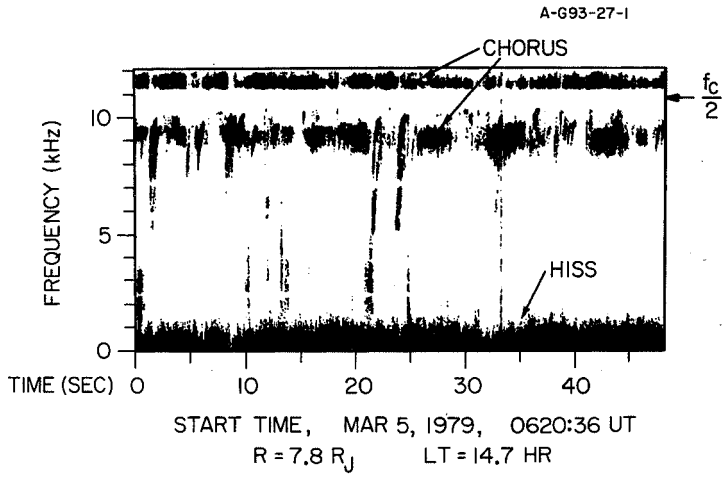


Fig. 1 - A frequency-time spectrogram showing whistler-mode chorus and hiss detected by Voyager 1 in the radiation belt of Jupiter. Chorus consists of the many discrete narrow band emissions, usually rising in frequency on time scales of a few seconds, and hiss consists of the nearly steady spectrum of band-limited noise. In this case, the chorus occurs from about 8 to 12 kHz, and the hiss occurs below about 1 kHz.

magnetic field in the same sense as the particle. Cyclotron resonance can occur for both electrons and ions. However, for the whistler mode the most important cyclotron resonance interactions occur with the electrons. In a classic paper on the subject, Kennel and Petschek [1966] showed that the parallel energy at which electrons are in cyclotron resonance with a whistler-mode wave is given by

$$W_{\parallel} = \frac{B^2}{2\mu_o N}\left(1-\frac{f}{f_c}\right)^3 \frac{f_c}{f}$$

where $B^2/2\mu_0$ is the magnetic field energy density, N is the electron number, and f is the wave frequency. Whistler-mode waves grow spontaneously if the electron velocity distribution is anisotropic in such a way that there is a deficit in the number of resonant electrons moving at small angles to the magnetic field. In a planetary radiation belt, this type of anisotropy always occurs, since any particle with a pitch angle within a cone of directions called the "loss-cone" will strike the planet and be lost from the system.

Although the whistler mode is always unstable due to the presence of the loss cone, the question of whether the wave grows to a large amplitude depends on the gains and losses along the ray path. Kennel and Petschek showed that the growth rate is proportional to the number of electrons in resonance with the wave. Therefore, the growth rate increases as the radiation belt intensity increases. Since the highest radiation belt intensities usually occur near the magnetic equator, the

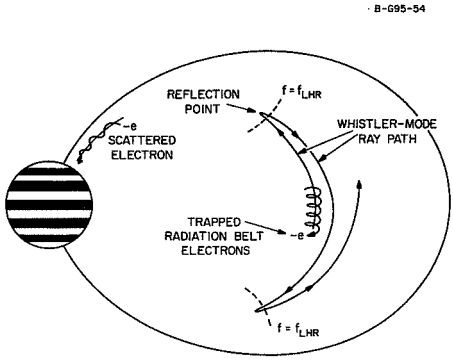


Fig. 2 - A sketch of whistler-mode ray paths in the Jovian radiation belt. For sufficiently large wave normal angles, whistler-mode waves reflect as soon as the lower-hybrid-resonance frequency, f_{LHR}, exceeds the wave frequency. The resulting ray paths are then similar to a laser, with repeated passes through the equatorial plane where wave growth occurs via cyclotron resonant interactions with energetic radiation belt electrons. These interactions then scatter the electrons into the loss cone, causing them to hit the planet.

highest growth rates tend to be near the magnetic equatorial plane. Various types of losses exist. One of the main losses is simply due to the propagation of the wave out of the system. Because of the anisotropic nature of the propagation, whistler-mode waves tend to be guided along the magnetic field lines toward the planet, where they can escape through the base of the ionosphere. However, if the initial wave normal angle is sufficiently large, the wave tends to reflect as soon as the lower-hybrid-resonance frequency, $f_{LHR} = \sqrt{m_e/m_i f_c}$, exceeds the wave frequency. A typical ray path is shown in Figure 2. This reflection process causes the waves to bounce back and forth along the magnetic field line from one hemisphere to the other. Reflections not only minimize the losses, but they also cause repeated passes through the equatorial region where the maximum growth rate occurs. From Figure 2, one can see that the growth of whistler-mode waves in a planetary radiation belt is similar to a laser, with the anisotropy in the trapped electrons providing the free energy source, and the reflections near $f = f_{LHR}$ playing the role of the mirrors.

The generation of whistler-mode waves would be relatively unimportant except for the effect these waves have on the radiation belt electrons. Since whistler-mode waves are right-hand polarized, these waves carry away right-hand angular momentum. The angular momentum must come from the resonant electrons. Since electrons rotate in the right-hand sense with respect to the magnetic field, the effect of the wave generation is to lower the pitch angle of the electrons, thereby driving them toward the loss cone. Those electrons that are scattered into the loss cone collide with the planet and are lost from the system. The growth of whistler-mode waves therefore leads directly to the loss of radiation belt electrons. This pitch-angle scattering mechanism is believed to be the primary mechanism that limits the intensity of energetic electrons in a planetary radiation belt. Whenever some process acts to increase the trapped electron intensity, whistler-mode waves grow, which increases the loss of electrons, thereby establishing a new equilibrium.

Before closing the discussion of whistler-mode emissions, it is worth commenting on the differences between chorus and hiss. Since chorus usually occurs at higher frequencies than hiss, one can see from the frequency dependence in Equation 1 that chorus resonates with lower electron energies than hiss. This difference is illustrated in Figure 3, which shows the spectrums of the chorus and hiss emissions in the top panel, and the corresponding resonance energies in the bottom panel. As can be seen, chorus resonates with electrons in the few keV range, and hiss

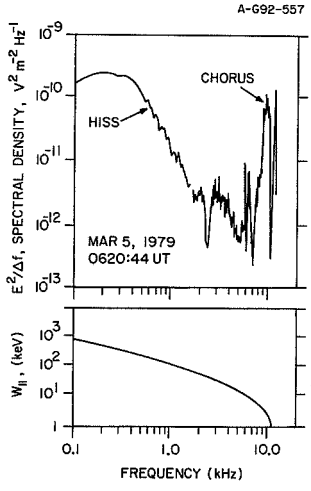


Fig. 3 - The top panel shows the spectrum of the whistler-mode chorus and hiss in Figure 1 and the bottom panel shows the parallel energy of electrons that are in cyclotron resonance with these waves. Chorus tends to resonate with relatively low energies, typically a few keV, whereas hiss resonates with much higher energies, typically several hundred keV or more.

resonates with electrons in the several hundred keV range. Since there are more resonant electrons at lower energies, chorus has higher growth rates than hiss. The higher growth rate causes nonlinear saturation effects to occur sooner, before the wave has even reached the first reflection point in Figure 2. These nonlinear effects are thought to cause the wave to evolve into the nearly monochromatic wave packets that are characteristic of chorus. Hiss on the other hand has much lower growth rates, leading to many reflections from one hemisphere to the other. The superposition of these many reflected waves and the absence of strong nonlinear effects is believed to cause the emission to evolve into the nearly steady band-limited spectrum that is characteristic of hiss. It may be worth noting that the above views on the origin of hiss are not universally accepted. Recently, Draganov et al. [1993] have proposed that the hiss in the Earth's radiation belt (called plasmaspheric hiss) may not be due to an instability at all, but rather due to the superposition of many lightning-generated whistlers that have become trapped near the equatorial plane via reflections similar to those illustrated in Figure 2.

B. Electromagnetic ion cyclotron emissions

The electromagnetic ion cyclotron mode is an electromagnetic mode that propagates at frequencies below the ion cyclotron frequency, $f_{ci} = eB/m_i$. This mode is left-hand polarized with respect to the magnetic field. Since positively charged ions rotate in the left-hand sense with respect to the magnetic field, the ion cyclotron mode interacts primarily with positively charged ions. Kennel and Petschek [1966] have shown that the ion cyclotron mode can be driven unstable by a process very similar to the whistler mode, except that the ion anisotropy is responsible for the instability rather than the electron anisotropy. Since the ion cyclotron frequency is a factor of m_e/m_i smaller than the electron cyclotron frequency, ion cyclotron emissions occur at much lower frequencies than whistler emissions. Whereas whistler-mode emissions are usually at frequencies in the few kHz range, ion cyclotron emissions are usually at frequencies of a few Hz or less. Because of their extremely low frequency, ion cyclotron waves are difficult to detect. At present, electromagnetic ion cyclotron waves have only been observed in the radiation belts of two planets, Earth and Jupiter, and possibly in the radiation belt of Neptune (see Table 1). Although electromagnetic ion cyclotron emissions are weak and difficult to detect, they are still of considerable importance. At Earth, a mid-latitude type of aurora known as a stable red arc is believed to be due to the precipitation of radiation belt ions by electromagnetic ion cyclotron waves [Cornwall, 1970].

Similar processes are also believed to cause auroral ion precipitation at Jupiter [Thorne, 1983]. Thus, electromagnetic ion cyclotron waves play a role in the loss of radiation belt ions similar to the role that whistler-mode waves play in the loss of radiation belt electrons.

3. Planetary auroral acceleration regions

The aurora consists of light produced by energetic charged particles impinging on the upper levels of a planetary atmosphere. Five planets, Earth, Jupiter, Saturn, Uranus, and Neptune, are known to have auroras. At Earth, the aurora is usually confined to a narrow region from about 65 to 75° magnetic latitude known as the auroral zone. Strong electrical currents, known as field-aligned currents, flow along the magnetic field lines linking the auroral zones with the outer regions of the magnetosphere. These currents are carried primarily by electrons. For reasons that are not completely understood, large potential differences develop along the magnetic field lines in regions of strong field-aligned currents. These potential differences accelerate some of the electrons to high energies; typically several keV, thereby forming field-aligned electron beams. Both upgoing and downgoing electron beams are observed. The auroral light is produced when a downgoing beam strikes the atmosphere. Although relatively little is known about auroral processes at planets other than Earth, there are good reasons to believe that the processes are basically similar. Many types of plasma waves and radio emissions are known to be generated in planetary auroral acceleration regions. Of these, we will focus on two in particular, (1) auroral hiss, and (2) cyclotron maser radiation.

A. Auroral hiss

Auroral hiss is a whistler-mode emission that is produced by auroral electron beams. Auroral hiss has been observed at Earth and Jupiter, and possibly Neptune (see Table 1). The absence of adequate high-latitude observations at Saturn and Uranus makes it impossible to say whether auroral hiss occurs at these planets. A frequency-time spectrogram illustrating auroral hiss observed during a high-latitude pass over the Earth's auroral zone is shown in Figure 4. The auroral hiss is the funnel-shaped emission that can be seen extending from about 100 Hz to 50 kHz. A sharp upper frequency cutoff can be seen at the electron cyclotron frequency, which is what one would expect, since the whistler mode cannot propagate at frequencies above the electron cyclotron frequency. Both upgoing and downgoing auroral hiss has been observed. The upgoing auroral hiss is associated with upgoing electron beams, and the downgoing auroral hiss is associated with downgoing electron beams. Upgoing auroral hiss is mainly observed at high altitudes, 1 to 3 R_E or more, as in Figure 4. Downgoing auroral hiss is only observed at low altitudes, 1 R_E or less. The funnel-shaped feature in Figure 4 is a propagation effect. At higher frequencies, the radiation propagates at larger angles to the magnetic field lines, thereby increasing the latitudinal region over which that radiation is observed.

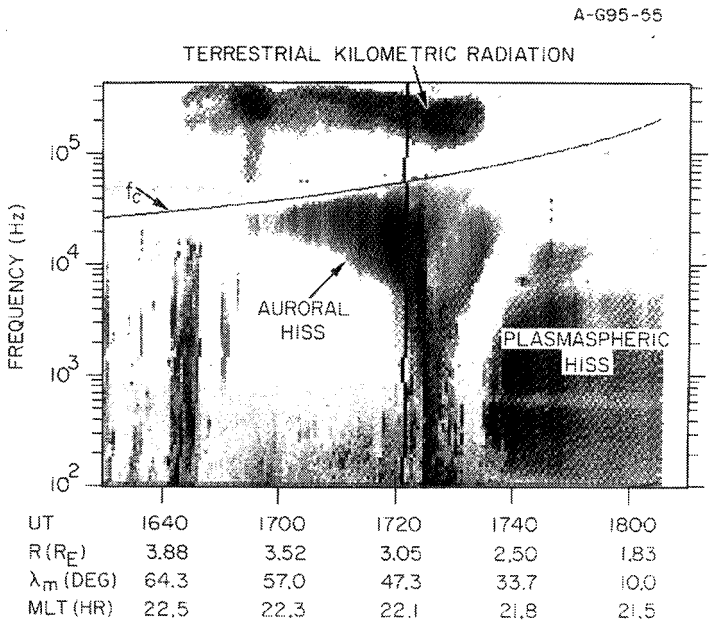


Fig. 4 - A frequency-time spectrogram from a pass of the DE-1 spacecraft over the Earth's northern polar cap. Two types of auroral radio emissions can be seen: (1) auroral hiss, and (2) terrestrial kilometric radiation. Auroral hiss is a whistler-mode emission, and terrestrial kilometric radiation is a free-space radio emission that propagates freely away from the Earth.

Whistler-mode auroral hiss is believed to be generated by a Cerenkov-like radiation process similar to that proposed by Ellis [1957]. However, the radiated power cannot be accounted for by simply summing the Cerenkov power emitted from the individual electrons in the beam. Instead, a collective process that organizes the phase of emitting electrons must be involved so that the radiated power is increased. Using modern plasma theory, it can be shown that electrons of velocity v_b interact resonantly with the whistler-mode radiation if the frequency w and wave number k satisfy the condition $v_b \equiv \omega/k_{\|}$, where the symbol || represents the component along the magnetic field. This condition is called the Landau resonance. The Landau resonance is essentially identical to the Cerenkov condition encountered in single particle radiation theory. It can also be shown that the growth rate is proportional to $\partial f/\partial v_{\|}$, where f is the electron velocity distribution. A beam always has a range of velocities where $\partial f/\partial v_{\|}$ is positive, so that wave growth will occur. This instability is often called the two-stream instability.

Auroral hiss is a very common plasma wave emission. Almost every pass over the Earth's auroral zones has intense auroral hiss. The interaction of the auroral hiss with the auroral electron beam has been extensively studied by Maggs [1976]. As the auroral hiss grows in amplitude, wave-particle interactions tend to flatten the electron velocity distribution function in the region where

$\partial f/\partial v_{\|}$ is positive, thereby driving $\partial f/\partial v_{\|}$ to zero. Whistler-mode waves therefore act to drive the electron distribution toward a stable equilibrium. The presence of this stabilization process is confirmed by the fact that well-defined "beams" are seldom observed in the auroral zones. As soon as a beam starts to develop, whistler-mode wave-particle interactions quickly spread the beam into a flat distribution.

B. Cyclotron maser radiation

During the late 1960s and early 1970s, an entirely new type of terrestrial radio emission was discovered by eccentric Earth-orbiting satellites. This radio emission was first detected by Benediktov et al. [1965] using data from the Elektron 2 and 4 satellites. A few years later, Gurnett [1974] showed that this radiation, which has its peak intensity in the frequency range from about 100 to 500 kHz, was closely correlated with the occurrence of discrete auroral arcs. Gurnett also showed that the total radiated power was very large, up to 10^9 Watts. The high power levels came as a surprise, since this radiation had not been previously detected on the ground. The reason, of course, is the radiation cannot propagate downwards through the ionosphere. The Earth was therefore found to be an intense planetary radio source, comparable in some respects to Jupiter, which had been known for many years to be an intense radio emitter [Burke and Franklin, 1955]. Since the terrestrial radiation occurs in the kilometer wavelength range, this radio emission soon became known as "terrestrial kilometric radiation" or "auroral kilometric radiation". An example of terrestrial kilometric radiation can be seen in the upper part of Figure 4, at about 100 to 400 kHz, slightly above the electron cyclotron frequency. It is now known that this same basic type of radio emission occurs at five planets, Earth, Jupiter, Saturn, Uranus, and Neptune. A comparison of the radio emission spectrums from these five planets is shown in Figure 5. The characteristic features in all cases are that the radiation is (1) very intense, (2) right-hand polarized, and (3) generated at frequencies near the electron cyclotron frequency.

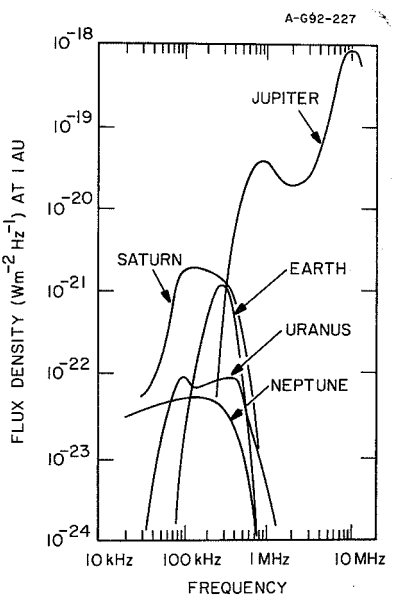


Fig. 5 - A comparison of the spectrums of cyclotron maser radiation from five planets. The spectrums are all referenced to a distance of 1 AU.

Since it is relatively easy to obtain in situ measurements of plasmas and radio emissions over the Earth's auroral zones, the discovery of the terrestrial kilometric radiation provided an unprecedented opportunity to investigate an astronomical radio emission mechanism of considerable significance. Now, after many years of study, we have a very good understanding of how this radio emission is produced. The basic mechanism is called the cyclotron maser instability.

The cyclotron maser mechanism was first discussed by Melrose [1973], and later analyzed in more detail by Wu and Lee [1979] in connection with the terrestrial kilometric radiation. The basic instability is similar to the whistler loss-cone instability in that it involves an electron cyclotron resonance. However, the mode of propagation is the free space R-X mode rather than the whistler mode. One unusual feature is that relativistic effects are fundamentally involved in the resonance condition and cannot be omitted even though the electron energies are non-relativistic (i.e., only a few keV). The free energy source that drives the instability has been the subject of considerable debate. Originally, Wu and Lee [1979] proposed that it was the loss cone in the electron distribution that provided the free energy for the instability. However, more recent studies by Louarn et al. [1989] indicated that electrons trapped in the auroral acceleration region by magnetic mirror and electrostatic forces provide the primary free energy source. Once generated, the radiation escapes freely away from the Earth, following ray paths more or less as shown in Figure 6.

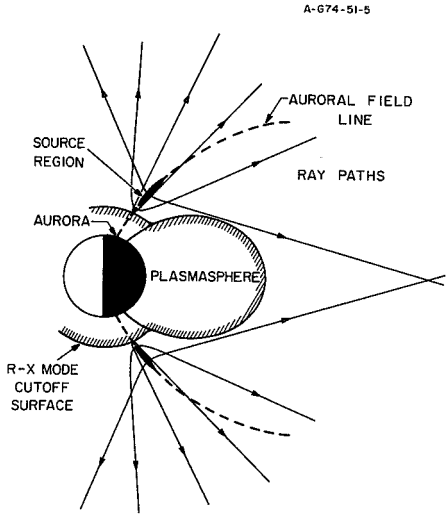


Fig. 6 - A sketch of typical ray paths of terrestrial kilometric radiation. This radiation is generated in the auroral acceleration region by an electron cyclotron maser mechanism. The radiation propagates away from the Earth in a broad conical beam, the axis of which is aligned along the magnetic field in the source region.

Although very detailed in situ measurements are available in the region where the cyclotron maser radiation is generated at Earth, comparable measurements are not available at the other planets. Even though the Voyager spacecraft flew by Jupiter, Saturn, Uranus, and Neptune, the trajectory did not pass through the source region, which in almost all cases is located at high latitudes. Thus, the only information that can be obtained about the cyclotron maser radiation mechanism at these planets is what can be gleaned from the radio emission spectrum. The Jovian cyclotron maser radiation (called decametric radiation) has one unusual feature that is worth noting. The intensity of the Jovian decametric radiation has been shown by Bigg [1964] to be strongly controlled by Jupiter's moon, Io. As Io moves through the Jovian magnetosphere, it induces a field-aligned current loop that closes in the auroral zone. This current system then causes electron acceleration, auroral light emission, and other effects comparable to the aurora at Earth. Recent pictures from the Hubble Space Telescope show that Jupiter also has an auroral zone that appears to be driven by a magnetospheric current system. Thus, at least two energy sources are probably involved in the generation of cyclotron maser radiation at Jupiter.

4. The solar wind

The solar wind is a hot, fully ionized gas that flows outward from the Sun at a supersonic speed. At the orbit of Earth, the solar wind density is approximately 5 cm^{-3}, and the speed is approximately 400 km/s. In situ measurements of plasma waves and radio emissions have been made in the solar wind as close to the Sun as 0.29 AU, and as far from the Sun as 58 AU. To illustrate the range of plasma wave and radio emission processes that can occur in the solar wind, we will focus on two examples: (1) Langmuir waves associated with type III solar radio bursts, and (2) heliospheric 2-3 kHz radio emissions.

A. Langmuir waves associated with type III solar radio bursts

Langmuir waves are electrostatic oscillations that occur in a plasma at the electron plasma frequency, $f_p = 9\sqrt{N}$ kHz, where N is the electron density in cm^{-3}. Langmuir waves are excited by electron beams and are of considerable importance in the theory of certain types of solar radio emissions. In a classic paper, Ginzburg and Zheleznyakov [1958] proposed that type III solar radio bursts are produced by a two-step process in which (1) electrons from a solar flare excite Langmuir waves at f_p via a two-stream instability, and (2) the Langmuir then decay into electromagnetic radiation at f_p and $2f_p$ via nonlinear wave-wave interactions. The two-step type III generation

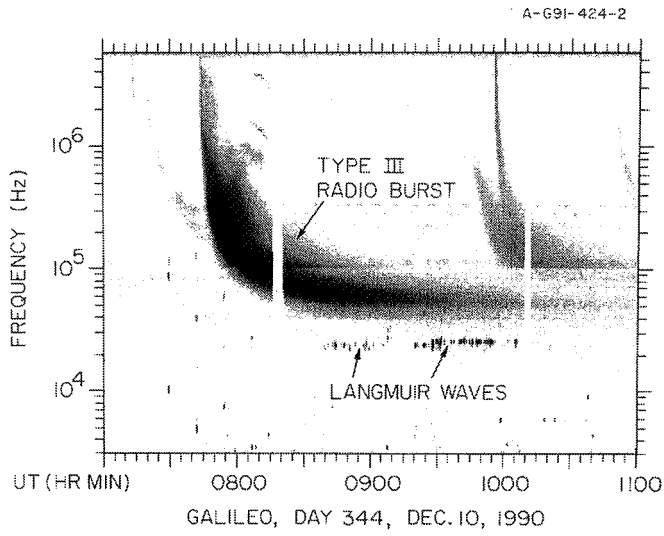


Fig. 7 - An example of a strong solar type III radio burst detected by the Galileo spacecraft at a heliocentric radial distance of about 0.98 AU. The Langmuir waves responsible for the radio emission can be seen at the local electron plasma frequency, which in this case is about 23 kHz. These waves are driven by electrons arriving from a solar flare that occurred at about 0730 UT.

process has now been confirmed by a variety of space plasma wave measurements [see Gurnett and Anderson, 1976; and Lin et al., 1981]. An example of a type III radio burst detected in the solar wind near 1 AU is shown in Figure 7. Type III radio bursts are characterized by an emission frequency that decreases with increasing time. The downward frequency drift is caused by the decreasing electron plasma frequency encountered by the solar flare electrons as they move outward from the Sun. This process is illustrated in Figure 8, which shows a typical radial variation of the plasma frequency in the solar wind. Although the type III radiation can propagate great distances, the Langmuir waves, which are a locally generated oscillation, cannot be detected until the electron beam reaches the spacecraft.

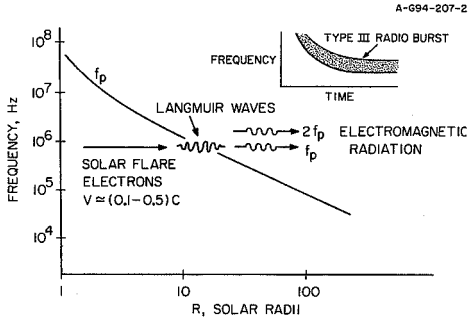


Fig. 8 - In the two-step mechanism believed to be responsible for type III radio bursts, electrons arriving from a solar flare excite Langmuir waves at the local electron plasma frequency, f_p. The Langmuir waves then generate radiation at f_p and $2f_p$ via nonlinear wave-wave interactions. The downward frequency drift of the radio burst is caused by the decreasing electron plasma frequency encountered by the solar flare electrons as they move outward from the Sun.

In Figure 7, the type III solar radio burst is associated with a solar flare that occurred at about 0730 UT. The Langmuir waves excited by the energetic electrons arriving from this solar flare can be seen at the local electron plasma frequency ($f_p \approx 23$ kHz) starting at about 0835 UT and continuing to about 1020 UT. In this case the type III radio emission is generated at the harmonic $2f_p$. Harmonic radiation is believed to result from the interaction of two Langmuir waves, L and L', such that the emitted frequency is $f_{pL} + f_{pL'} = 2f_p$. Since the emitted transverse electromagnetic wave, T, has a wave number much smaller than the Langmuir waves, conservation of momentum $(\vec{k}_L + \vec{k}_{L'} = \vec{k}_T)$ requires that the Langmuir waves, L and L', must be propagating in opposite directions. The origin of the oppositely propagating Langmuir wave is still a subject of debate. The current view is that this wave is produced by parametric decay from the original beam-driven Langmuir wave. Radiation at the fundamental, which is rare at these low frequencies, is believed to occur when a beam-driven Langmuir wave interacts with another low-frequency wave, such as a sound wave, to produce emission at $f = f_p + f_s$. Since $f_s << f_p$, the radiation occurs slightly above the plasma frequency.

B. Heliospheric 2-3 kHz radio emissions

In the early 1980s, as the Voyager 1 and 2 spacecraft were moving outward from the Sun beyond the orbit or Saturn, they began to detect an unusual radio emission in the frequency range from about 2 to 3 kHz. In the approximately twelve years since this radio emission was first detected, two particularly strong events have occurred, the first in 1983-84 [Kurth et al., 1984] and the

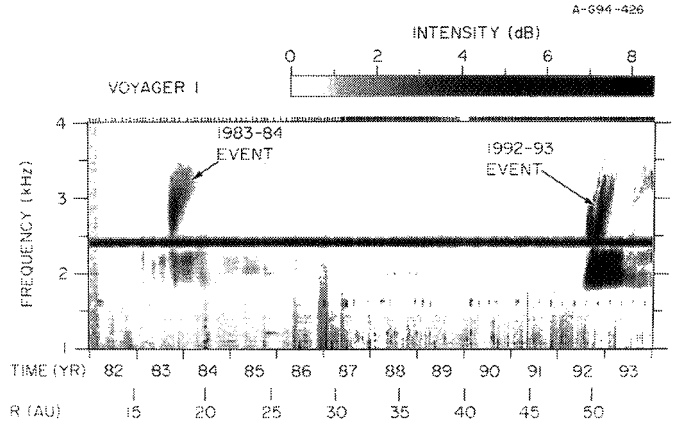


Fig. 9 - A 12-year frequency-time spectrogram showing the two intense heliospheric 2-3 kHz radio emission events detected by the Voyager 1 and 2 spacecraft in the outer regions of the solar system. These two events each occurred about 400 days after intense periods of solar activity, the first in July 1982 and the second in May-June 1991.

second in 1992-93 [Gurnett et al., 1993]. A twelve-year frequency time spectrogram from Voyager 1 illustrating these two events is shown in Figure 9. Since the solar wind electron plasma frequency varies roughly as $f_p = 20$ (1/R) kHz where R is the heliocentric radial in AU, the source of these radio emissions must be located far from the Sun, at least $R \gtrsim 10$ AU. Initially, several possible sources were considered, including (1) planetary, (2) heliospheric, and (3) stellar. Based on the most recent 1992-93 event, Gurnett et al. [1993] have estimated that the total radiated power is at least 10^{13} W, which effectively rules out planetary sources (also see Gurnett and Kurth [1994]). Because of the great distance to the nearby stars, stellar sources are also considered unlikely, since they would require extremely high power levels ($>10^{20}$ Watts) to account for the observed intensities. A heliospheric source has recently been given strong support by the fact that the 1983-84 and 1992-93 events each followed a period of intense solar activity, the first in July 1982 and the second in May-June 1991. The delay time between the peak of the solar activity and the onset of the radio emission in both cases was approximately 400 days.

The best current explanation of the 2-3 kHz radio bursts is that they are produced in the outer regions of the heliosphere by an interaction involving a shock wave or system or shock waves propagating outward from the Sun. There are two obvious boundaries where this interaction could occur, (1) the termination shock, where the solar wind undergoes a transition from a supersonic to a subsonic flow, and (2) the heliopause, where the solar wind is held off by the pressure of the interstellar medium. Because the Sun is moving through the local interstellar medium at a speed of about 26 km/s, the heliopause is expected to form a bullet-shaped boundary around the Sun [Axford, 1990]. Since the pressure in the local interstellar medium is poorly known, the distances to the terminal shock and heliopause are difficult to estimate. Pioneers 10 and 11, and Voyagers 1 and 2 are currently at 61.3, 42.2, 58.1, and 44.7 AU (as of January 1, 1995), and none have yet reach either the termination shock or the heliopause.

The most likely mechanism for generating the heliospheric 2-3 kHz radio emissions is thought to be similar to the two-step Langmuir wave mechanism, except that the electron beam is produced by an interplanetary shock. This is the mechanism by which type II radio bursts are believed to be produced. The radiation would then be generated at f_p or $2f_p$. If a two-step Langmuir wave mechanism is responsible for the radiation, then it is unlikely that the radio emission is produced at the termination shock, since the plasma frequency is too low. At Voyager 1 the plasma frequency in the solar wind is already down to about 350 Hz, so it must be even lower at the termination shock. At the heliopause, the situation is much better. Since the heliopause is a contact discontinuity, the plasma density can increase by whatever factor is necessary to maintain pressure balance with the interstellar medium. Current best estimates of the electron density in the interstellar medium [Lallement et al., 1993], indicate that the plasma frequency is in the range from 2.2 to 2.8 kHz. Thus, the plasma densities in the vicinity of the heliopause are in a suitable range to account for the 2-3 kHz radio emissions. A representative plasma frequency profile through the outer regions of the heliosphere is shown in Figure 10. If the radio emission is generated by the interaction of an interplanetary shock with the heliopause, then the distance to the heliopause can then be estimated from the travel time and speed of the shock. From the 400-day travel time, and the speeds of the interplanetary shocks (550 to 800 km/s), the distance to the heliopause can be computed, and is in the range from about 106 to 177 AU [Gurnett et al., 1995].

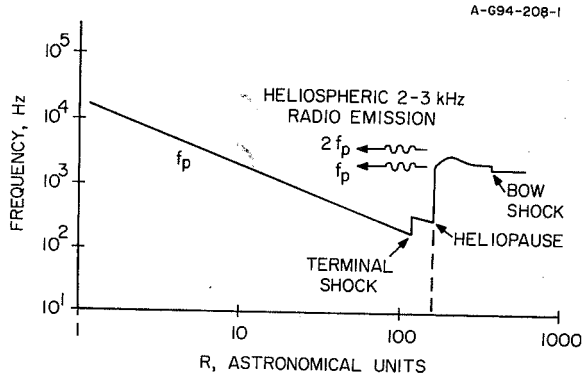


Fig. 10 - The heliospheric 2-3 kHz radio emissions are believed to be produced in the vicinity of the heliopause by an interplanetary shock wave moving outward from the Sun. The radiation is believed to be produced by a two-step process involving Langmuir waves generated by an electron beam accelerated by the shock.

5. Conclusions

In the nearly forty years since the launch of the first Earth-orbiting satellites, considerable progress has been made in the understanding of solar system plasma processes. Measurements of space plasma waves and radio emissions have played an essential role in achieving this understanding. However, much remains to be done. Although our knowledge of the plasma environment of the Earth is very good, our knowledge of plasma processes at other planets is very limited. There is a strong need to obtain plasma and plasma wave measurements in the auroral acceleration regions at Jupiter, where strong radio emissions are generated over the high-latitude polar regions, most likely in association with the aurora. There is also a strong need to explore plasma wave processes much closer to the Sun, in the region where strong radio emissions are produced in response to flares and other energetic solar processes. In the meantime, most future space plasma wave research will probably focus on measurements obtained in the vicinity of Earth. At Earth there are still many avenues of research that remain to be explored. Although the linear growth phase of most plasma wave instabilities is well understood, the nonlinear mechanisms that limit the growth and saturate the instability are poorly understood. The continued pursuit of these and other areas of space plasma wave research is likely to continue well into the 21st century.

Acknowledgement.

The research described in this paper was supported by NASA via contracts 959193 and 958779 with the Jet Propulsion Laboratory.

References

Axford, W. I., "Introductory Lecture- The Heliosphere"; ed. by S. Grzedzielski and D. E. Page, Physics of the Outer Heliosphere, COSPAR Colloquia Series, 1, Pergamon Press, Oxford, 7-15, 1990.

Barkhausen, H.; "Zwei mit Hilfe der neuen Verstärker entdeckte Erscheinungen"; Phys. Z.20, 401, 1919.

Benediktov, E. A., G. G. Getmantsev, Yu. A. Sazonov, and A. F. Tarasov; "Preliminary results of measurements of the intensity of distributed extraterrestrial radio-frequency emission at 725 and 1525 kHz frequencies by the satellite Electron-2"; Kosm. Issled., 3, 614-617, 1965.

Bigg, E. K.; "Influence of the satellite Io on Jupiter's decametric emission"; Nature, 203, 1008-1010, 1964.

Burke B. F., and K. L. Franklin; "Observations of a variable radio source associated with the planet Jupiter"; J. Geophys. Res.,60, 213-217, 1955.

Cornwall J. M., F. V. Coroniti, and R. M. Thorne; "Turbulent loss of ring current protons"; J. Geophys. Res., 75, 4699-2709, 1970.

Draganov, A. B., U. S. Inan, V.S. Sonwalkar, and T. F. Bell; "Whistlers and plasma spheric hiss: Wave directions and three-dimensional propagation"; J. Geophys. Res., 78, 11,401-11, 410, 1993

Eckersley, T. L.; "Musical atmospherics"; Nature, 135, 104-105, 1935.

Ellis, G. R.; "Low-frequency radio emission auroral"; J. Atmos. Terrestr. Phys., 10, 302-306, 1957.

Ginzburg, V. L., and V. V. Zheleznyakov; "On the possible mechanism of sporadic radio emission (radiation in an isotropic plasma)"; Sov. Astron., AJ2, 653, 1958.

Gurnett, D. A.; "The Earth as a radio source: Terrestrial kilometric radiation"; J. Geophys. Res., 79, 4227-4238, 1974.

Gurnett, D. A., and R. R. Anderson; "Electron plasma oscillations associated with type III radio bursts"; Science, 194, 1159-1162, 1976.

Gurnett, D. A., W. S. Kurth, S. C. Allendorf, and R. L. Poynter; "Radioemission from the heliopause triggered by an interplanetary shock"; Science, 262, 199-203,1993.

Gurnett, D. A., and W. S. Kurth; "Evidence that Jupiter is not the source ofthe 2-3 kHz heliospheric radio emission"; Geophys. Res.Lett., 21, 1571-1574, 1994.

Helliwell, R. A.; "Whistlers and related ionospheric phenomena"; StanfordUniversity Press, Stanford, CA, 1965.

Kennel, C. F., and H. E. Petschek; "Limit on stably trapped particle fluxes";J. Geophys. Res., 71, 1-28, 1966.

Kurth, W. S., D. A. Gurnett, F. L. Scarf, and R. L. Poynter; "Detection of a radio emission at 3 kHz in the outer heliosphere"; Nature, 312, 27-31, 1984.

Lallement, R., J.-L. Bertaux, and J. T. Clark; "Deceleration of interstellar hydrogen at the heliospheric interface"; Science 260, 1095-1098, 1993.

Lin, R. P., D. W. Potter, D. A. Gurnett, and F. L. Scarf; "Energetic electrons and plasma waves associated with a solar type III radio burst"; Astrophys. J., 251, 364, 1981.

Louarn, P., A. Roux, H. deFeraudy, and D. LeQueau; "Trapped electrons as a free energy source for the AKR"; J. Geophys. Res., 95, 5983-5995, 1989.

Maggs, J. E.; "Coherent generation of VLF hiss"; J. Geophys. Res., 81,1707-1724, 1976.

Melrose, D. B.; "Coherent gyromagnetic emission as a radiation mechanism"; Aust. J.Phys., 26, 229, 1973.

Preece, W. H.; "Earth currents"; Nature, 49(1276), 554, 1894.

Scarf, F. L., D. A. Gurnett,and W. S. Kurth; "Jupiter plasma wave observations: An initial Voyager 1 overview"; Science, 204, 991-995, 1979.

Storey, L. R. O.; "An investigation of whistling atmospherics"; Phil. Trans. Roy. Soc., London, A 246, 113-141,1953.

Thorne, R. M.; "Microscopic plasma processes in the Jovian magnetosphere"; Physics ofthe Jovian Magnetosphere, ed. by A. J. Dessler, Cambridge Univ. Press, Cambridge,454-497, 1983.

Van Allen, J. A.; "The geomagnetically trapped corpuscular radiation"; J. Geophys.Res., 64, 1683-1689, 1959.

Wu, C. S., and L. C. Lee; "A theory of terrestrial kilometric radiation"; Astrophys. J., 230, 621-626, 1979.

MICROWAVE POWER TRANSMISSION FROM SPACE AND RELATED NONLINEAR PLASMA EFFECTS

HIROSHI MATSUMOTO

Abstract

We first present a brief historical review of the development of technology and scientific research related to the transmission of electrical energy via radio waves. The idea of radio power transmission was first conceived by Tesla about a century ago. However, the first practical use of radio waves was for transmitting intelligence and information, and not for transmitting electrical power per se. At the close of World War II, engineers and scientists re-examined the original Tesla idea of transmitting electric power to a distant place via radio, as high-power microwave technology became available. These efforts in 1960's resulted in the idea of the Solar Power Satellite (SPS) which was proposed by P. Glaser in 1968. The NASA/DOE concept of the SPS was extensively developed in the late 1970's. After reviewing the history of microwave power transmission and related theoretical/experimental studies from the beginning of this century up to 1980, we will discuss recent research on microwave power transmission after 1980. Our focus will be on related experiments conducted in the 1980's and 1990's, including those on ground-to-ground microwave energy transmission, ground-to-aircraft power transmission, and rocket-to-rocket power transmission. The rocket experiment we discuss was conducted to examine a possible nonlinear resonant interaction of intense microwaves with the ionospheric plasma. The result of the rocket experiment is further studied in detail by particle model computer simulations, and the results are explained in terms of nonlinear plasma effects. Such problems of interaction between the microwave power beam and the ionosphere must be resolved before space-to-ground and space-to-space power transmission can be realistically developed.

1. Introduction

On the occasion of the 75th Anniversary of URSI, it is appropriate to re-examine the historical traces of radio utilization for transmitting electric power without wires to a distant destination. Today, radio waves are mainly used for transmitting intelligence and information. However, the threat of the lack of energy resources, especially for electrical energy, is increasing as a result of

Prof. Hiroshi Matsumoto is with the
Radio Atmospheric Science Center
Kyoto University, Uji,
Kyoto 611, Japan
Tel: + 81 (774) 33 2532 *Fax: + 81 (774) 31 8463*

the population explosion and rapid industrialization over the globe. Therefore, considering that the energy problem on our mother planet Earth, and the crisis of the Earth's environment have become urgent issues for mankind, we need to re-examine the use of radio waves for transmission of clean electrical energy from one place to another, especially from space to the ground, without wires.

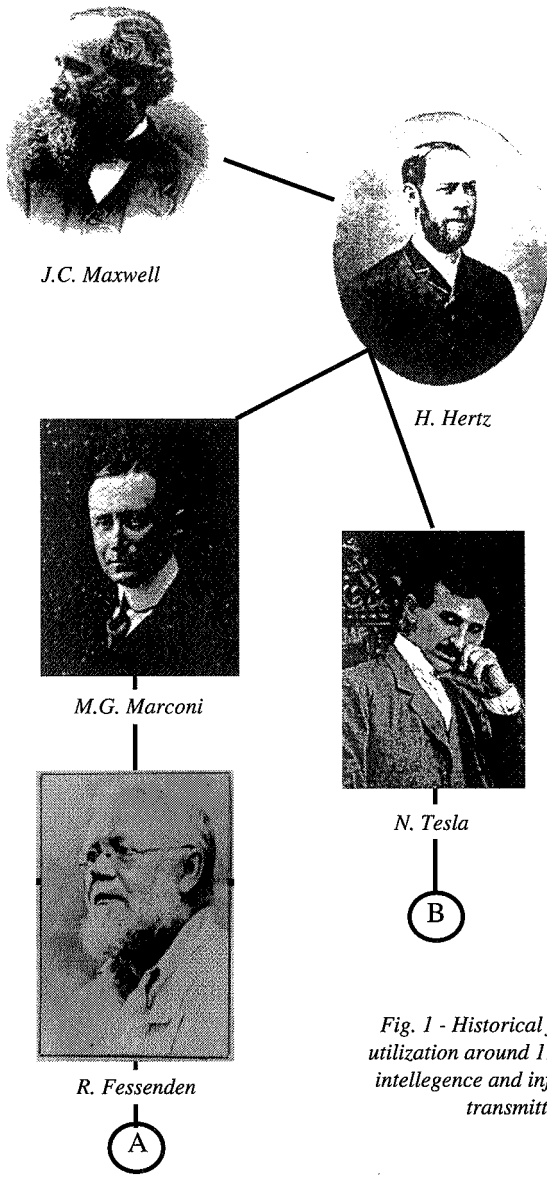
There exists a good review paper by W. C. Brown [1] on the history up to 1980 of power transmission by radio waves. We briefly describe, in Section 2, the historical footprints of radio power transmission from a century ago to 1980. In the late 1970's, the NASA/DOE sponsored extensive studies on the Solar Power Satellite (SPS). The NASA/DOE SPS studies program contained an evaluation of the impact of a microwave power beam on the plasma environment of the ionosphere. Section 3 reviews the theoretical studies on Ohmic heating of the ionosphere, the thermal self focusing instabilities caused by the SPS microwave power beam and the related ionospheric heating experiments by ground-based heating facilities. Following these studies, the present author conducted a further study of microwave action on the ionospheric plasma, focusing on the nonlinear resonant scattering of the microwave power beam by magnetized ionospheric plasma. Section 4 presents a theoretical study of the nonlinear resonant interaction of a high-power microwave beam with ionospheric plasma, and a rocket experiment called MINIX (Microwave-Ionosphere Nonlinear Interaction eXperiment) which was conducted to test the theoretical estimate of nonlinear resonant interactions. Extensive computer simulations of nonlinear resonant interactions were carried out by the present author and his colleagues to interpret the MINIX result in terms of nonlinear wave-wave-particle interactions. Section 5 describes the computer simulation and its theoretical interpretation.

In Section 6 we outline two recent microwave-driven airplane experiments: SHARP in Canada and MILAX in Japan. The recent experiment on microwave power beam steering using an active phased array system developed in Japan is described as well. In Section 7 a brief account of a recent rocket experiment and recent ground-to-ground power transmission is given. In Section 8 we conclude with a summary and discussion for future plans of research and development on microwave power transmission (MPT).

2. History of microwave power transmission before 1980

In, 1864, James Clerk Maxwell [2] predicted the existence of radio waves by means of mathematical model. Twenty four years later, in 1888, bolstered by Maxwell's theory, Heinrich Hertz [3] first succeeded in showing experimental evidence of radio waves by his spark-gap radio transmitter. This experiment stimulated Marchese Guglielmo Marconi [4], who first achieved signal transmission by means of radio waves over 10 m in 1895, and over the Atlantic Ocean in 1901. It was Reginald Fessenden [5] who first succeeded in transmitting continuous wave (CW) for voice telecommunications [6]. Thus, the road to modern radio telecommunication was opened up around the turn of the century. Modern radio utilization has been directed into the area of radio telecommunications for transmission of "intelligence and information" over rather weak radio

waves. This is one main stream of radio utilization stemming from the Maxwell-Hertz-Marconi-Fessenden work. However, another stream of work was directed toward a different radio wave application. The second stream of radio utilization was an effort to transmit electrical energy by radio to a distant place. These two streams are illustrated in Figure 1.



The idea of radio power transmission was first conceived and experimented on in 1899 by Nikola Tesla [7, 8]. He attempted to distribute ten thousand horse-power under a tension of one hundred million volts. He said "This energy will be collected all over the globe preferably in small amounts, ranging from a fraction of one to a few horse-power. One of its chief uses will be the illumination of isolated homes". He actually built a gigantic coil which was connected to a high mast of 200-ft with a 3 ft-diameter ball at its top (see Figure 2). He fed 300 kW power to the Tesla coil resonated at 150 kHz. The RF potential at the top sphere reached 100 MV.

From the turning point of the century on, however, radio has been used mainly for transmitting intelligence and information, and very few attempts have been made to transmit electrical energy over radio following Tesla's work.

The reason for a lack of interest in radio power transmission in the first half of this century is clear. People were waiting for the invention of a

Fig. 1 - Historical figures along two stream lines of radio utilization around 1900. (A) is a line toward transmission of intellegence and information over radio, and (B) is toward transmitting electric power over radio

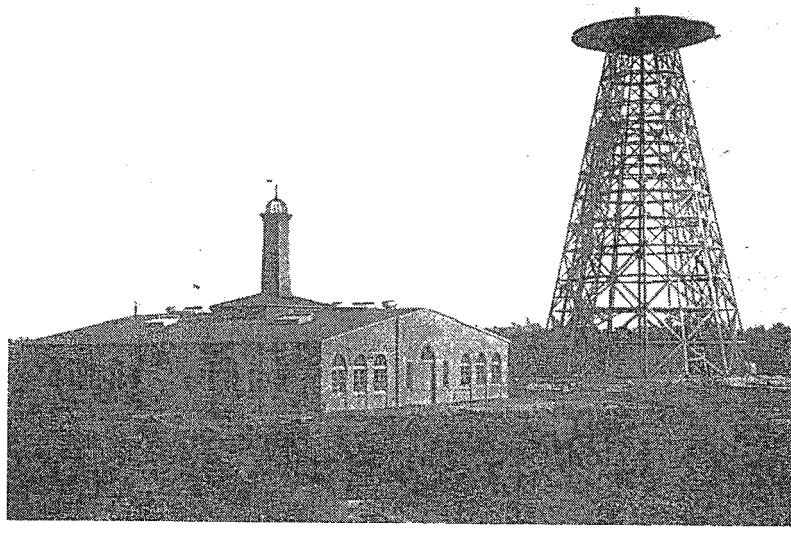


Fig. 2 - Tesla's experimental laboratory in Colorado Springs with power plant and transmitting tower [from N. Tesla, 1905]

high-power microwave device to generate electromagnetic energy of reasonably short wavelength, since efficient focusing toward the power receiving destination is strongly dependent on the use of technology of narrow-beam formation by small-size antennas and reflectors. In 1930's much progress in generating high-power microwaves was achieved by invention of the magnetron and the klystron. Though the magnetron was invented by A. W. Hull [9] in 1921, the practical and efficient magnetron tube gathered world interest only after Kinjiro Okabe [10] proposed the divided anode-type magnetron in 1928. It is interesting to note that H. Yagi and S. Uda[11], who are famous for their invention of Yagi-Uda Antenna, stressed a possibility of power transmission by radio waves in 1926, thereby displaying profound insight into the coming microwave tube era in Japan. Microwave generation by the klystron was achieved by the Varian brothers [12] in 1937 based on the first idea by the Heil brothers in Germany in 1935.

During World War II, development of radar technology accelerated the production of high-power microwave generators and antennas. A CW high power transmission over a microwave beam was investigated in secrecy in Japan. The project , the "Z-project" , was aimed at shooting down air-bombers by a high-power microwave beam from the ground, and involved two Nobel winners H. Yukawa and S. Tomonaga [13]. Figure 3 shows a 100 kW magnetron developed at that time, and an introduction of the Japanese Magnetron appeared in "Electronics" of USA immediately after World War II. However, the technology of the high-power microwave tube was still not developed sufficiently for the practical continuous transmission of electric power. Further, no power device was available to convert a microwave energy beam back to DC power until the 1960's.

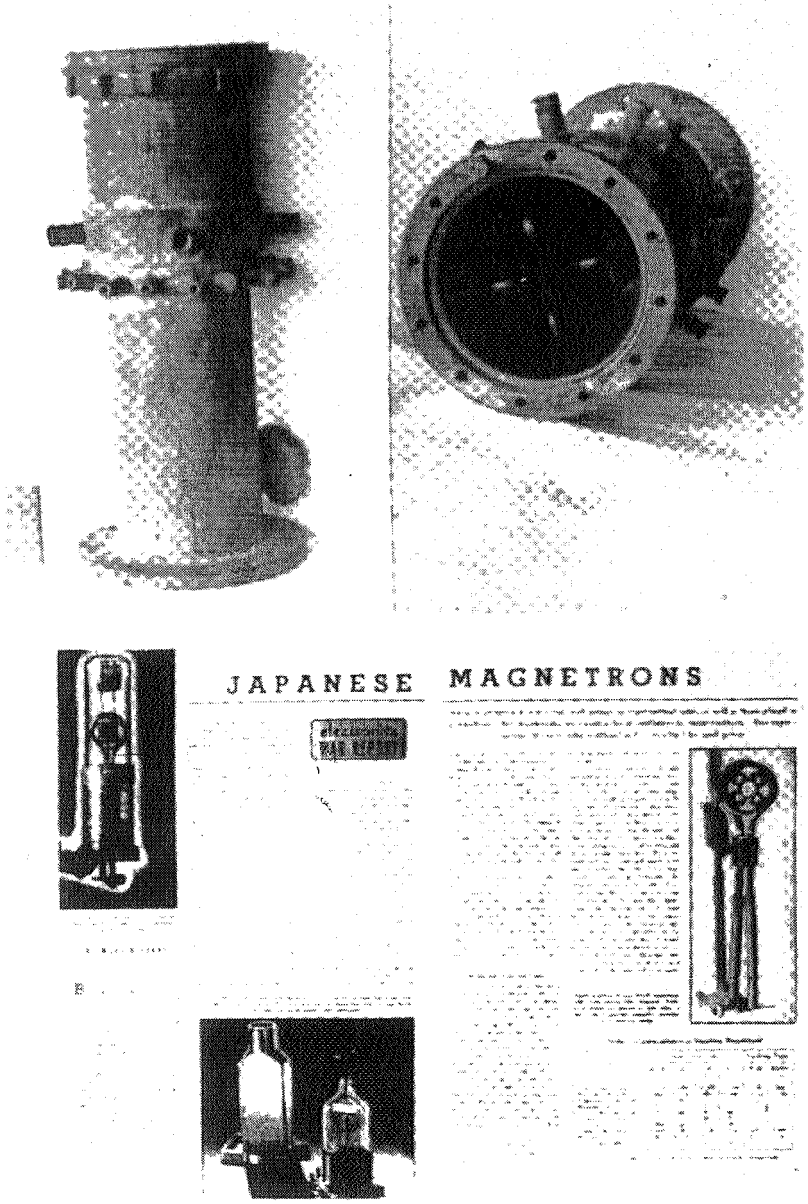


Fig. 3 - A 100 kW magnetron developed during World War II in Jaspan and a copy of an article appeared in US "Electronics" based on the onformation collected by US GHQ after the end of World War II

The post-war history of research on free-space power transmission is well documented by William C. Brown [1], who was a pioneer of practical microwave power transmission (see references in [1]). It was he [14] who first succeeded in demonstrating a microwave-powered helicopter in 1964, using 2.45 GHz in the frequency range of 2.4 - 2.5 GHz reserved for the ISM (Industrial, Scientific and Medical) applications of radio waves (see Figure 5 in [1]). A power conversion device from microwave to DC, called a rectenna, was invented [15, 16, 17] and used for the microwave-powered helicopter. The first rectenna (shown in Figure 3 in [1]) was composed of 28 half-wave dipoles terminated in a bridge rectifier using point-contact semiconductor diodes. Later, the point contact semiconductor diodes were replaced by silicon Schottky-barrier diodes which raised the microwave-to-DC conversion efficiency from 40 % to 84 %[1], the efficiency being defined as the ratio of DC output to microwave power absorbed by the rectenna. The highest record of 84% efficiency was attained in the demonstration of microwave power transmission in 1975 at the JPL Goldstone Facility [18]. Power was successfully transferred from the transmitting large parabolic antenna dish to the distant rectenna site over a distance of 1.6 km. The DC output was 30 kW [18] (see Figure 9 in [10]).

An important milestone in the history of microwave power transmission was the three-year study program called the DOE/NASA Satellite Power System Concept Development and Evaluation Program, started in 1977. This program was conducted for the study of the Solar Power Satellite (SPS), which is designed to beam down the electrical power of 5 to 10 GW from one SPS toward the rectenna site on the ground.

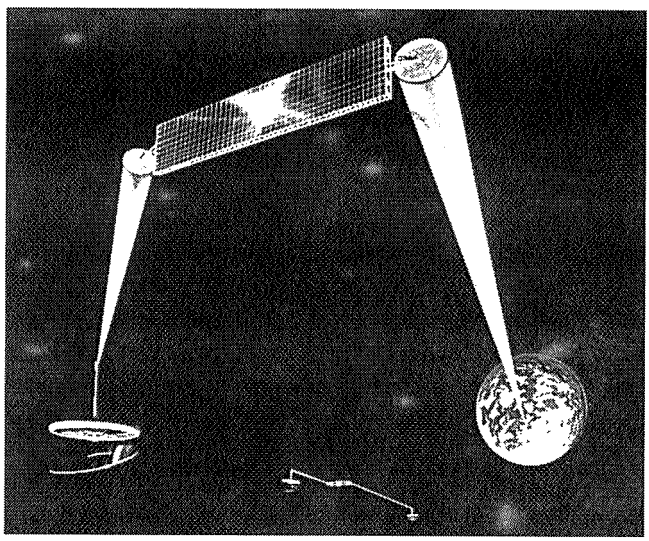


Fig. 4 - Artist concept of Solar Power Satellite. The power station will transmit electric power to the Earth and possibly to Space Factory, Space Farms and Space Cities in addition

The extensive study of the SPS ended in 1980, producing a 670 page summary document [19]. The concept of the SPS was first proposed by P. E. Glaser [20] in 1968 to meet both space-based and earth-based power needs. An artist's SPS concept is shown in Figure 4. The SPS will generate electric power of the order of several hundreds to thousands of megawatts using photo-voltaic cells of sizable area, and will transmit the generated power via a microwave beam to the receiving rectenna site. Among many technological

key issues which must be overcome before the SPS realization, microwave power transmission (MPT) is one of the most important key research issues. The problem contains not only the technological development of microwave power transmission with high efficiency and high safety, but also scientific analysis of microwave impact onto the space plasma environment. We discuss this in the following three Sections.

3. Review of ohmic heating of the ionospere and thermal self-focusing instability by SPS and a related ionospere heating experiment

The SPS studies program, carried out in the latter half of 1970's under the sponsorship of the US NASA/DOE, contained research on the effects associated with the propagation of intense microwave beams through the ionosphere. Two main effects were pointed out.

The first is the resistive (Ohmic) heating effect due to collisional damping of microwaves. Though the fraction of wave energy absorbed by the ionospheric plasma is very small, the resultant Ohmic heating can significantly modify the local ionospheric thermal balance [21, 22]. The electron temperature is determined by a balance between heating and cooling processes. The heating by microwaves is proportional to square of the electric field of the microwave E^2, to square of the ratio of the local plasma frequency and the microwave frequency, $(f_p < f)^2$and to the collision frequencies ν_{ei} and ν_{en} where ν_{ei} and ν_{en} are the collision frequencies of electrons with ions and neutral particles, respectively. The collision frequency ν_{en} increases as the electron temperature increases. Thus Ohmic heating by intense microwaves can be self-amplifying, and thereby result in thermal runaway [22, 23]. Calculation of the balance between the enhanced heating and cooling losses through vibrational excitation of N_2 and O_2 shows that the electron temperature will be raised several-fold for a microwave power density of 23 mW/cm^2. It is also found that the electron density in the E-layer will be increased by 10 - 20 % due to a decrease in the temperature-dependent recombination rate of O_2^+ and NO_2^+, while in the D-layer increase in the attachment rate to O_2^+ will cause up to a 50 % reduction in the electron number density [21, 23]. As the ionospheric heating efficiency varies inversely as the square of the radio frequency, ionospheric heating equivalent to that by the SPS microwave beam can be achieved at lower radiated power by heating at a lower frequency. With this idea, experimental tests of the enhanced electron heating theory were carried out by the Rice University Group lead by W. E. Gordon [24], using the 430 MHz radar system at the Arecibo Observatory. A series of underdense ($f_{heater} < f_p$) HF ionospheric modification experiments using the Platteville high power HF (5-10 MHz) heating facility at Colorado were conducted to simulate the effects of the SPS microwave beam, while monitoring potential impact upon telecommunication system performance [25, 26]. It was concluded that there is no significant difference between the telecommunication system performance of the OMEGA system (VLF), the LORAN system (LF) and the AM broadcast system (MF) between the times when the heating facility was operating and when it was not. Thus, the impact of the SPS intense microwave beam on the performance of VLF, LF and MF telecommunication systems would be minimal [25].

The second potential effect of the SPS microwave beam onto the ionosphere, studied extensively in the late 1970's, is the phenomenon of thermal self-focusing [27, 28, 29]. Thermal self-focusing takes place as a result of a positive feedback loop. Small natural density fluctuations in the ionosphere cause a spatial variation of the refractive index thereby giving rise to a slight focusing and defocusing of the microwave. This slight inhomogeneous (differential) heating of the ionospheric plasma results in a temperature gradient driving the plasma from the focused region and thereby amplifying the initial density fluctuations. The self-focusing instability will eventually reach a hydrodynamic equilibrium creating large-scale ionospheric irregularities. A self-focusing experiment was conducted at the Arecibo Observatory using intense HF electromagnetic wave under the overdense condition ($f_{heater} < f_p$) [30]. The experimental result showed clear self-focusing striations and large-scale structuring of the ionosphere. However, it is noted that these experiments were all conducted under the overdense condition and not the underdense condition.

The two main effects described above are caused basically by the non-resonant heating of the plasma by the intense electric field of the SPS microwave. Resonant interactions of the microwave beam with the ionospheric plasma are another interesting research area. Parametric excitation of ionospheric plasma waves was studied for multiple-frequency electromagnetic radiation. The interaction between two high frequency microwaves and a multiple of the ionospheric resonant frequency (such as the electron plasma frequency) was studied [31, 32] utilizing a model of an up-going pilot microwave signal operating at a frequency slightly separated from the downcoming power beam, and beat waves generated within the finite width of the main downcoming energy beam.

The US government suspended the NASA/DOE program study partly due to budget problems and partly because of the apparent recovery from the oil shock in the 1970's, although the NASA/DOE final report concluded that "no factors that would preclude the SPS research and development are found in the light of the highly potential future energy crisis".

4. Theory and rocket experiment (MINIX) on nonlinear plasma wave excitation by microwave power beam in the ionosphere

Most of the experiments conducted in the 1970's on the potential impact of the SPS microwave onto the ionosphere were conducted by the ground-based heating facility using much lower frequency than the SPS microwave frequency of 2.45 GHz. Therefore, the realistic resonant interaction of the SPS microwave with natural resonance frequency bands of the order of several MHz to several kHz may not have been adequately estimated. The resonant interaction naturally involves electrostatic waves which can be detected much more easily by in-situ measurement than by ground-based radar diagnostics. Based on such idea, an in-situ rocket experiment of radiating an intense 2.45 GHz microwave into the ionospheric plasma was proposed by the present author in the early 1980's in Japan. The project was named MINIX which stands for Microwave Ionosphere Nonlinear Interaction eXperiment.

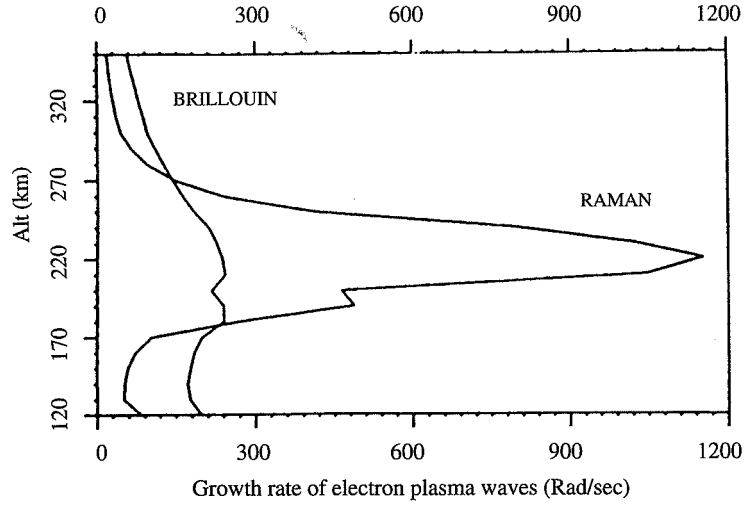
The plasma response to the injected microwave power beam was monitored by diagnostic sensors and receivers onboard the rocket. Preceding the MINIX experiment, Matsumoto [33, 34] numerically evaluated the growth rate of the resonant instabilities of electron plasma waves and ion acoustic waves as a result of Raman and Brillouin scattering of the SPS intense microwave under model ionospheric plasma parameters. The growth rate of Langmuir waves as a result of the Raman scattering of the microwave is given [34] by

$$\gamma_e = -\frac{1}{2}\left[\Gamma_e + \Gamma_2 - \sqrt{(\Gamma_e + \Gamma_2)^2 + \frac{\Pi_e^2}{\omega^2} k^2 v_0 \sin^2 \varphi}\right] \tag{1}$$

where Π_e is the electron plasma frequency and Γ_e is the damping rate of Langmuir waves; the latter includes two terms resulting from the Landau damping and collisional damping. The quantity Γ_2 represents a collisional damping of the back-scattered microwave with a frequency of

$$\omega_2 = \sqrt{\Pi_e^2 + c^2 (k - k_0)^2}$$

$v_0 = eE_0 / (m_e \omega_0)$ is the sloshing velocity of electrons by the SPS microwave, and φ is an angle between k and E_0. The suffix o denotes the quantity associated with the SPS microwave. A similar expression for the growth rate Γ_i for ion acoustic waves was also obtained [34] (not shown). Numerical calculation of Γ_e and Γ_i versus the ionospheric altitude (see Figure 5) showed that Langmuir waves are easily excited by the SPS microwave, while the growing time of ion acoustic



F ig. 5 - Growth rates of Langmuir waves excited by the Roman scattering of the SPS microwave, and fo ion acoustic waves excited by the Brillouin scattering of the SPS microwave. The microwave field E_0 is assumed to be 220 V/m (from [33]).

waves is much slower. It is noted that the growth rate expression (1) gives only the initial growth rates as a result of nonlinear interaction between the three waves: the incident intense microwave, the back-scattered microwave and the excited Langmuir wave. The nonlinear coupling of the three waves should satisfy both energy and momentum conservation in the form of

$$\omega_0 = \omega_1 + \omega_2 \,,\; \boldsymbol{k}_0 = \boldsymbol{k}_1 + \boldsymbol{k}_2 \tag{2}$$

where ω_j and k_j (j=0, 1, 2) are angular frequency and wave number respectively, and the subscripts 0, 1 and 2 represent the incident microwave, the back-scattered microwave, and the excited Langmuir wave, respectively.

The MINIX rocket experiment was carried out by a Japanese sounding rocket S-520-6 of ISAS (Institute of Space and Astronomical Sciences) on August 29 in 1983 [35, 36, 37, 38]. The experiment was conducted with a mother-and-daughter rocket system. Figure 6 shows an artist's concept of the experiment, while Figure 7 represents the ground testing of the mother-daughter system with the real flight model payloads. Two sets of high power (~ 830 watts) magnetrons (reinforced versions of Toshiba Magnetron 2M172 for home use oven) were installed on the mother section of the rocket payload and were connected to the truncated waveguide antenna (see

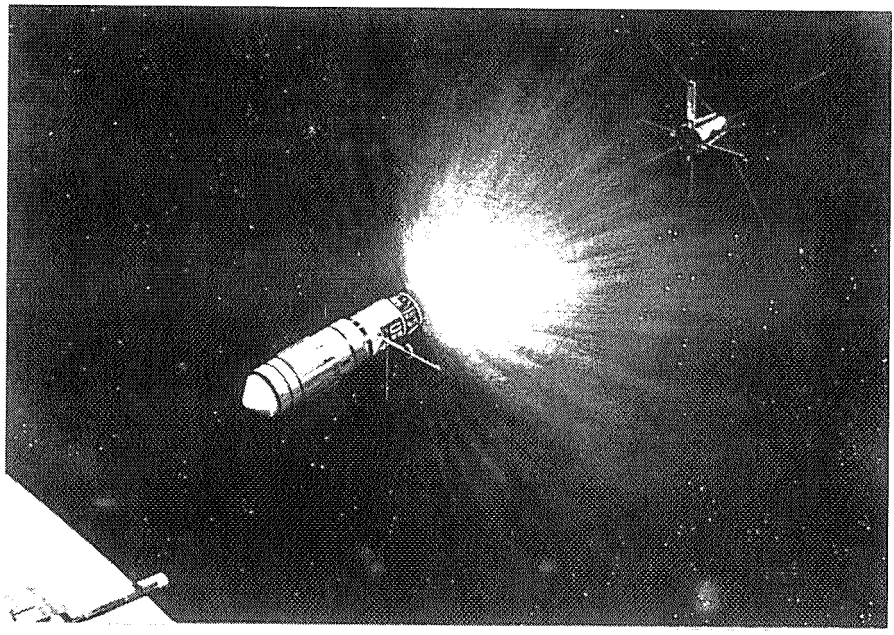


Fig. 6 - An artist concept of the MINIX rocket experiment. A high power microwave (~ 830 watts) was radtiated from the truncated wave guide antenna toward the daughter rocket section by which the nonlinear responses of the ionospheric plasma were measured.

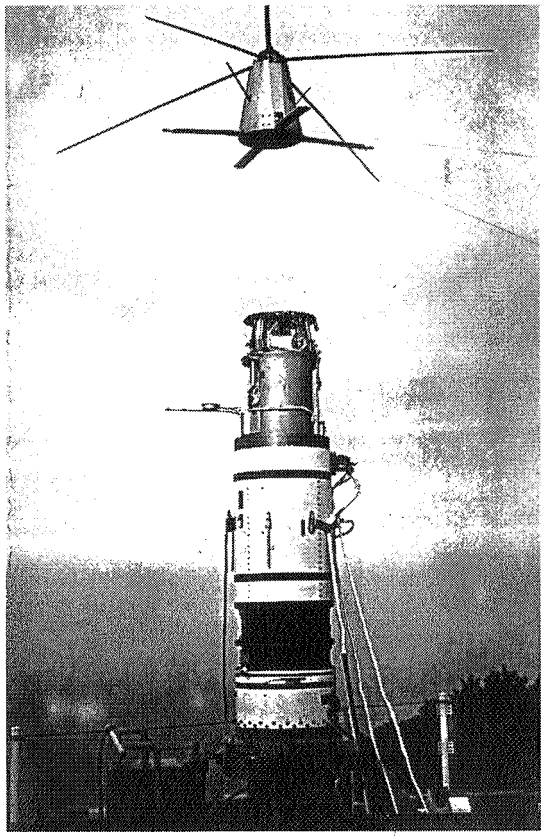


Fig. 7 - A photograph of pre-flight test scene of the MINIX payload at launching site KSC in Japan. The daughter unit with various sensors and diagnostic packages was separated by a crane from the mother section. A truncated wave guide antenna used for the side ward transmission is seen in the mother section

Figure 8) to radiate intense microwave with a frequency of 2.45 GHz. The DC power supply to the magnetrons was given by onboard batteries. A plasma diagnostic package was installed on the daughter unit of the rocket. It was composed of a VLF wide band receiver, an HF sweep frequency receiver, a geomagnetic aspect sensor, electron density and temperature meter, and a microwave receiver. Four rod antennas with a length of 2 m were extended out from the daughter rocket in the top plane of the daughter unit to detect plasma waves which are expected to be nonlinearly excited by the injected intense microwave. The HF sweep frequency receiver covered a frequency range from 100 kHz up to 18 MHz with a sweeping time of 250 msec. The VLF receiver covered a frequency range from 60 Hz up to 25 kHz. Four paddles were extended out at the bottom level of the daughter unit and were used for rectennas. The

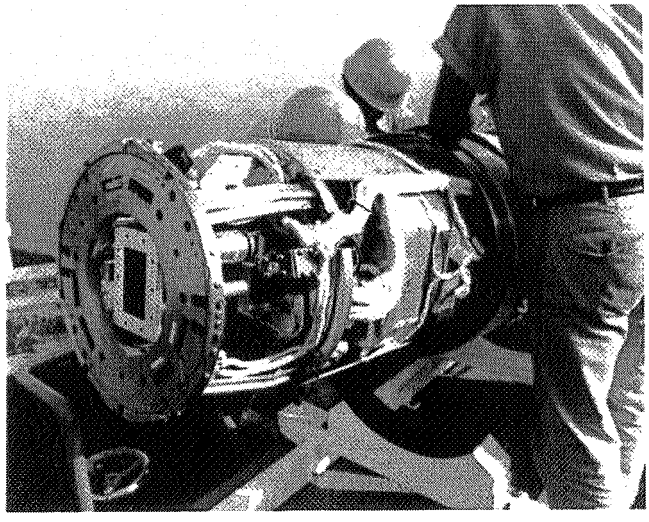


Fig. 8 - The truncated wave guide antenna aperture used for the forward transmission of 2.45 GHz microwave onboard the MINIX payload.

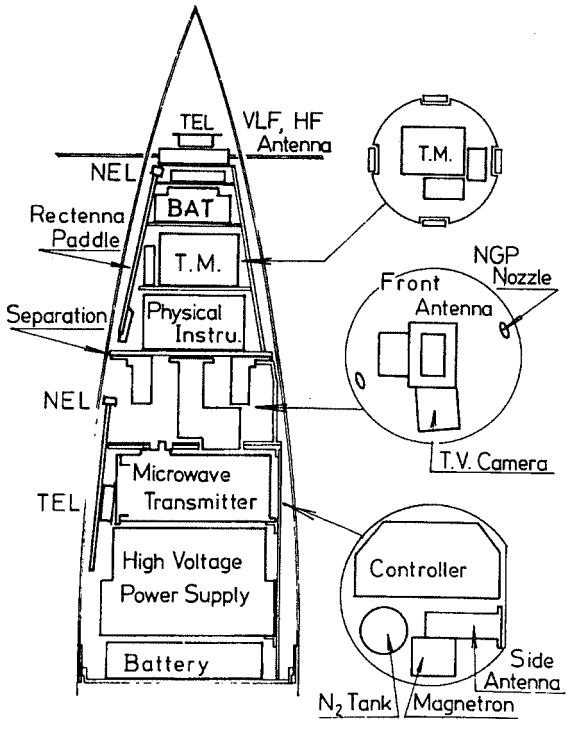


Fig. 8 - Payload configuration of MINIX rocket experiment. Mother and daughter sections are separated by spring coil during the flight in the ionosphere

arrangement of these antennas and paddles is shown in Figure 7. Figure 9 shows the configuration of the payload instrumentation on board S-520-6 rocket. The mother unit, which is the section below the level of separation plane in Figure 9, carried the power supply composed of the DC-battery and DC-DC converter, the microwave transmitter with a time sequencer, two sets of truncated wave guide antennas, a Langmuir probe, a wide-band telemetry set, a neutral gas plume of N_2 gas, and a TV monitor camera for monitoring the in-flight separation of the daughter unit. The neutral gas plume was prepared to create an artificially simulated D-layer with high collision frequency at the height of the lower F layer near the apex of the rocket orbit.

Figure 10 shows the trajectory of the S-520-6 rocket. The MINIX experimental time sequence along the orbit is also shown in the figure. The thick line parts along the orbit indicate the time interval of microwave transmission. The experiment was conducted under three different modes: The first mode (I) was devoted to the study of Ohmic heating in the ionospheric D-layer and lower E-layer. In this experimental mode, the microwave was transmitted continuously for a long duration of 10 seconds. Above the altitude of 100 km, the second mode (II) was in operation where the microwave was transmitted intermittently with a 5 sec transmission followed by a 5 sec pause-of-transmission period for the measurement of the plasma response. Under the mode (II), the microwave was radiated radially, i.e., in the direction perpendicular to the spin axis of the rocket. After the separation of the daughter unit near the apex of the orbit around 220 km altitude, the experimental mode was switched to the mode (III) where the

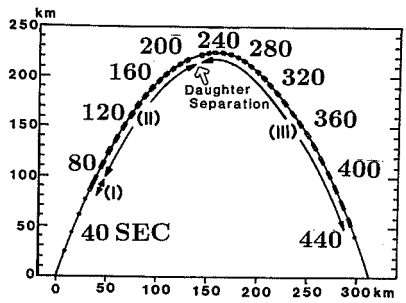


Fig. 10 - The trajectory of S-520-6 rocket and the time sequence of the MINIX along the orbit

microwave was transmitted forward in the axial direction from the truncated wave guide antenna facing the leaving daughter unit (see Figure 6), with the same ON-OFF time sequence as that in the mode II. The modes (II) and (III) were mainly prepared to detect the theoretically predicted excitation of plasma waves through nonlinear resonant interaction of the transmitted intense microwave.

All of the instrumentation on board the MINIX rocket worked perfectly and provided useful data. The measurement of the variation of electron temperature showed no temperature difference between ON and OFF periods of the microwave transmission during the mode (I). It turned out later with the use of pre-launch plasma chamber data on plasma heating by MINIX transmitter [39] that the estimated maximum temperature increase due to the Ohmic heating for the MINIX situation is below 100 K which is lower than the detectable limit by the Langmuir probe used in the MINIX. It is also noted that the effective time of microwave exposure to the ionospheric plasma particles was too short compared to the characteristic time for the Ohmic heating. The plasma volume illuminated by the intense microwave with a power density which is comparable to or higher than that of the SPS microwave is limited to within the distance of 2 m from the center of the truncated wave guide antenna. The power density and the electric field intensity of the microwave radiated by the magnetron plus wave guide antenna system are shown in Figure 11.

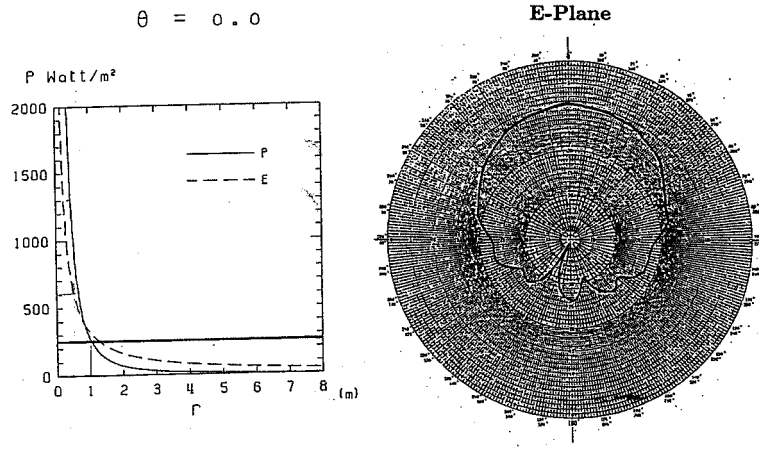


Fig. 11 - Power density and electric field intensity of the radiated microwave as a function of distance from the MINIX truncated wave guide antenna. The right panel shows the antenna pattern of the transmitting antenna.

The sweep frequency analyzer (SFA) in the HF range measured the spectrum of ionospheric plasma waves and broadcasting waves reaching the rocket altitude from the ground. The SFA detected strong plasma waves at certain frequency bands when and only when the microwave was transmitted. The plasma wave spectra from 100 kHz to 18 MHz are shown in Figure 12. The upper

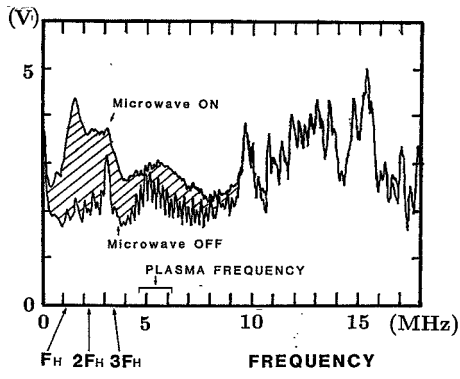


Fig. 12 - Plasma wave spectra observed by HF wave receiver onboard the daughter unit of the MINIX rocket experiment. The upper spectral line is for the time when the intense 2.45 GHz microwave is transmitted into the ionospheric plasma. The lower spectral line is for the time when the microwave is not transmitted. The spectral peaks above 10 MHz are due to the broadcasting waves reaching from the ground.

spectral curve is for the period of the microwave transmission, while the lower is for the no-transmission period. The spectra above 10 MHz are not different from each other, but those below 10 MHz show a clear difference. The shaded part shows the enhancement of the spectral components due to additional excitation of the plasma waves by the intense microwave. In Figure 12, the local electron cyclotron frequency F_H and its harmonics and the local electron plasma frequency are indicated on the horizontal frequency axis. The error bar in the electron plasma frequency was due to the ambiguity in determining the number density and temperature from the Langmuir probe. The enhancement in the spectral intensity is seen in two different characteristic frequency ranges. One is seen at odd half harmonics of the local electron cyclotron frequency in the range from 1.5 MHz to 3.5 MHz. The frequency range of these waves did not change with altitude as the local electron cyclotron frequency is almost constant in the orbit range of the rocket. Taking into account the observed frequency range and discrete peaks at odd half cyclotron frequencies, we explained these waves in terms of electron cyclotron harmonic (ECH) waves. The other enhancement is seen above the local electron plasma frequency ranging from 5 MHz to 8 MHz. The frequency range shifts upward with the altitude of the rocket, and accordingly with the increasing plasma density in the E- to F-layers of the ionosphere. We explained the enhancement of this frequency above the local electron plasma frequency in terms of Langmuir waves due to the excitation by the intense microwave through Raman scattering.

The VLF wide band receiver could not pick up the expected ion acoustic wave through Brillouin scattering. This was not surprising because the theoretically calculated [34] growth rate of the ion acoustic waves is not large enough to make these waves grow to the observable amplitude within a short time illuminated by the microwave at a fixed point of the ionosphere. The illuminated time was about 1 msec, which is calculated by the size of the plasma volume illuminated by the intense ($E > 200$ V/m) microwave, while the theoretically calculated growing time is longer than 10 msec.

The result of the MINIX on nonlinear plasma wave excitation due to intense 2.45 GHz microwaves was not much different from what had been predicted by the theoretical calculation. A similar but more complicated expression of the growth rate of the electron cyclotron waves including upper hybrid waves was obtained based on the nonlinear kinetic theory of resonant three-wave coupling[40, 41]. It is expressed as

$$\gamma_{ECH} = \sqrt{|\beta_1 \beta_2|} E_0 \tag{3}$$

where E_0 is the intensity of the incident intense microwave, and β_1 and β_2 are the coupling coefficients defined by

$$\frac{dE_1}{dt} = i\beta_1 E_2^* E_0 \tag{4}$$

$$\frac{dE_2}{dt} = i\beta_2 E_0 E_1^* \tag{5}$$

where E_1 and E_2 are the electric field of the backscattered microwave and the excited electron cyclotron wave. With a lengthy calculation and by use of the resonant condition (2), we obtain the final complicated expression of β_1 and β_2 (eqs. (80) - (93) in [41]) which contains both tensor elements of the linear dispersion relation and the integral containing resonant terms at higher harmonics of electron cyclotron frequencies. Numerical calculation of the growth rate shows that the electron cyclotron harmonic waves are also excited by the intense microwave under ionospheric conditions.

Despite the qualitative agreement with these theoretical predictions, the MINIX result showed the following features which are not consistent with the above theory. The first point is that the observed spectrum of the excited Langmuir waves are not monochromatic, nor quasi-monochromatic, but of broad-band nature in contrast to the monochromaticity predicted by theory based on the resonance condition (2). The second point is that the intensity of the electron cyclotron waves is higher than the Langmuir wave intensity. The second point clearly contradicts the theoretical prediction. The saturation level of the nonlinearly excited plasma waves (both Langmuir and ECH waves) through the nonlinear resonant three wave-coupling is proportional to the growth rate divided by the product of the coupling coefficients $\beta_0 \beta_1$. The theory shows that the saturation level of the Langmuir waves should be larger than that of the ECH waves. However, the amplitude of the Langmuir waves excited by the microwave in the MINIX was smaller than that of the ECH. Thus, these two experimental features are not well explained by the conventional nonlinear resonant three-wave coupling theory. In the next section, we will describe our attempt to overcome this contradiction by the help of computer simulations.

5. Computer Simulation of Nonlinear Interaction of Microwave Power Beam with Space Plasma

Traditionally, methods of scientific research have involved a mutual interplay between experiment and theory. Experiment attempts to collect "factual" information through repeated or controlled measurements. Theory, on the other hand, tries to order accumulated factual knowledge and thereby propose a new paradigm of description of physical processes. New theoretical descriptions

and new experiments challenge each other in turn. Such feedback between theory and experiment is generally on-going. However, as in the case of MINIX, theory and experiment sometimes show a gap which cannot be easily overcome because of difficulties of repeated experiments and/or a limit to the applicability of theory for highly nonlinear and complex processes. In order to fill such a gap, a third new approach become available with the advent of modern high speed computers. This third approach is called computer simulation or computer experiment. The basic idea of computer simulation is to simulate the physical behavior of complicated natural systems by solving an appropriate set of mathematical equations based on an accepted and fundamental physical mathematical model. As one can easily change the set of the mathematical equations as well as the boundary and initial conditions, the computer simulation can be a perfectly controlled experiment. The main advantage of computer simulations is that complicated physical systems including nonlinearity and/or strong inhomogeneity can be dealt with as easily as simpler, linear and homogeneous systems can be treated. The International Union of Radio Science (URSI) has played a significant role of accelerating the establishment of this third research tool in Radio Science, especially in the field space plasma wave studies through its activities in Commission H [43].

In order to understand the MINIX result and the nonlinear interaction of the intense microwave power beam with the ionospheric plasma (which contradicts the nonlinear resonant three wave coupling theory) we performed a series of computer simulations. The computer code used for this purpose is a particle-model simulation code called KEMPO [43]. KEMPO solves Maxwell's equations and simultaneous equations of motion of several tens of thousands to several millions of super-particles. The one-dimensional version of the KEMPO code is now available in the public domain [42]. As the MINIX result showed that there exist two different electrostatic (ES) plasma waves with two different propagation angles relative to the geomagnetic field, we have set up two different simulation models. One is the case where all of the incident intense microwave (or the "pump" electromagnetic (EM) wave in terms of nonlinear resonant three-wave coupling), the back-scattered microwave (or the back-scattered "idler" EM wave) and the excited ES plasma wave are assumed to propagate along the external magnetic field B_0. The other model assumes that they propagate in a perpendicular direction to B_0. The former and the latter cases are referred to as the "parallel case" and the "perpendicular case" hereafter. The energy and momentum conservation relation expressed by Eq.(2) can be graphically shown by a parallelogram in the ω-k diagram. Figure 13 is one example of the parallelogram in the perpendicular case. The common simulation parameters for both cases are listed in Table 1. Figure 14 shows an example of a temporal variation of the excited ES wave intensity, and that of the kinetic energy of electrons for both the parallel and perpendicular cases. The upper and lower panels correspond to the parallel and perpendicular cases, respectively. In the parallel case, an L-mode EM wave is adopted as the pump. The L-mode wave couples with a back-scattered L-mode EM wave and a Langmuir wave (LW). This coupling is referred to as L-L-LW coupling. A similar result is obtained (not shown) for R-R-LW wave coupling with the R-mode pump wave. In the perpendicular case, an X-mode EM wave is used as the pump. The X-mode pump wave is scattered producing a back-scattered X-mode EM and the ES ECH wave through the X-X-ECH coupling. A similar result is obtained for the O-O-ECH coupling (not shown). The ECH wave excited by the X-mode EM

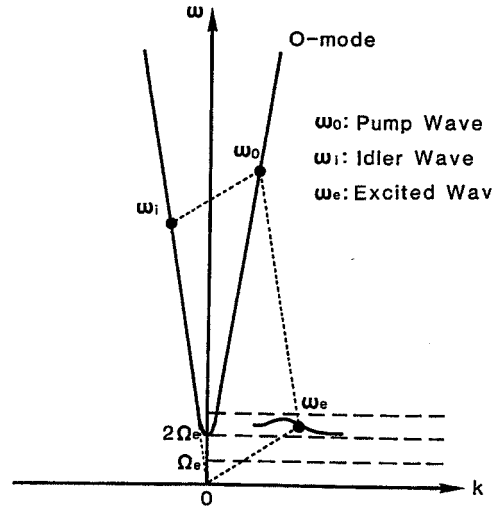


Fig. 13 - A parallelogram in the ω-k diagram showing the resonant interaction between two electromagnetic waves (microwaves) and electron cyclotron harmonic (ECH) wave.

wave, shown in the lower panel, is one of the multiple harmonic modes of ECH waves. It is the upper hybrid mode that shrinks to the upper hybrid oscillation when its wave number tends to zero. As seen in Figure 14, both the LW and ECHW grow exponentially in the early phase. The numerically measured growth rates of the LW and ECHW are $\gamma_{LW} = 0.07$ and $\gamma_{LW} = 0.04$, respectively. These growth rates agree well with the theoretically predicted value of the growth rate based on the nonlinear resonant three-wave coupling theory.

In contrast to a good agreement of the growth rates of both LW and ECHW, the saturation levels of the LW and of the ECHW contradict the theoretical prediction as seen in Figure 14. The LW saturation level is almost half of that of the ECHW. The simulation result does not agree with theory, but agrees well with the result of the MINIX rocket experiment. Thus, the computer simulation could reproduce the inconsistency between the rocket experiment and theory. One of the merits of computer simulations is that one can make the diagnostics as detailed as one wishes from the information stored in memory. In particular, compared with rocket experiments, one can

<Plasma Parameters>

Speed of Light	c	50
Electron Plasma Angular Frequency	Π_e	2.0
Electron Cyclotron Angular Frequency	Ω_e	1.0
External Magnetic Field Strength	B_{ext}	1.0
Charge to Mass Ratio of Electrons	q_e/m_e	-1.0
Total Number of Electrons	N_p	32768
Parallel Thermal Speed of Electrons	$v_{th\|\|}$	1.0
Perpendicular Thermal Speed of Electrons	$v_{th\perp}$	1.0
Dielectric Constant	ε_0	1.0

<System Parameters>

Time Step	Δt	0.01
Grid Spacing	Δx	1.0
Number of Grids	N_x	2048

<Parameters of Injected EM Wave>

Angular Frequency	ω_0	18.5~23.4
Wave Number (Mode Number)	k_0	15~19
Wave Magnetic Field Strength	B_ω	0.5

Table 1 - Numerical parameters for simulations presented in Figure 14

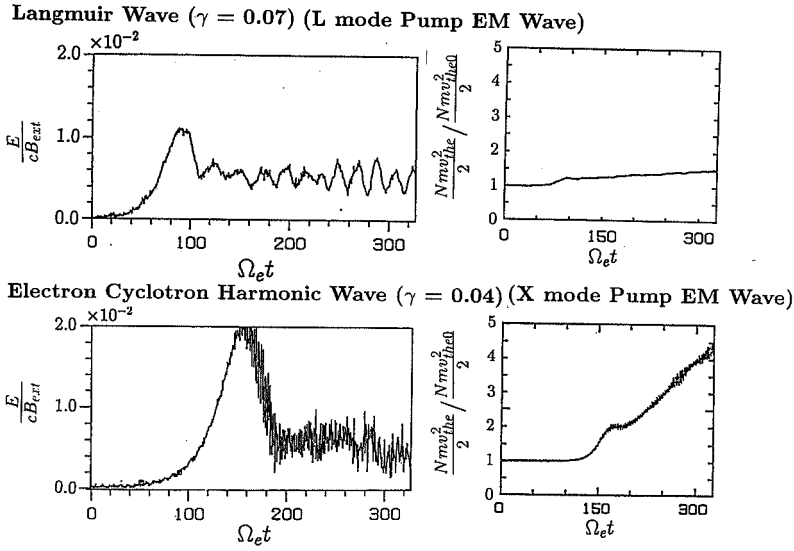


Fig. 14 - The left column shows time evolution of the electric field of electrostatic (ES) plasma waves excited nonlinearly by the pump electromagnetic (EM) waves. The right column shows the corresponding time history of particle thermal energy. The upper and lower panels are for the parallel and perpendicular cases, Langmuir Wave is excited by an L-mode EM pump wave while ECH Wave is excited by an X-mode EM pump wave in the perpendicular case.

extract detailed information on particle dynamics which is normally very difficult to measure with a sufficient time resolution. On the right panels in Figure 14, plotted are the time history of the particle kinetic energy in the simulations. In both the Langmuir wave and ECHW cases, the particle thermal energy increases exponentially according as the electric field E grows exponentially, until the electric field reaches the saturation. This initial exponential increase of the kinetic energy, however, does not actually cause the thermalization of electrons, but reflects the sloshing motion synchronizing with the electric oscillation of the ES wave. However, at the time of the wave saturation, those particles begin to be thermalized through phase-mixing. The rise in the particle thermal energy and the fall in the electric field intensity after the saturation show that the particles start to extract energy from the ES wave and are heated.

Examination of the thermal energy increase helps us to understand why the saturation levels of the electric field of the ES waves are much lower than those estimated by theory, and why the measured Langmuir wave intensity is lower than that of the ECHW. To confirm the result of the first single simulation on the growth rate and saturation levels, we ran a series of simulations with different frequencies of the pump EM wave. The simulation parameters for the simulation series are listed in Table 2. The result is shown in Figure 15. The solid and dashed lines show theoretical values of the LW and ECHW, respectively. Those of the LW and ECHW observed in the simulations are plotted by square and circular symbols, respectively. The agreement of the growth rates (Figure 15(a)) between theory and simulations were confirmed by this series of simulations.

<Plasma Parameters>

Speed of Light	c	50
Electron Plasma Angular Frequency	Π_e	2.0
Electron Cyclotron Angular Frequency	Ω_e	1.0
External Magnetic Field Strength	B_{ext}	1.0
Charge to Mass Ratio of Electrons	q_e/m_e	-1.0
Total Number of Electrons	N_p	32768
Parallel Thermal Speed of Electrons	$v_{th\parallel}$	$1.0\ 10^{-4}$
Perpendicular Thermal Speed of Electrons	$v_{th\perp}$	$1.0\ 10^{-4}$
Dielectric Constant	ε_0	1.0

<System Parameters>

Time Step	Δt	0.01
Grid Spacing	Δx	1.0
Number of Grids	N_x	2048

<Parameters of Injected EM Wave>

Angular Frequency	ω_0	5.0 ... 22.5
Wave Number (Mode Number)	k_0	33 ... 109
Wave Magnetic Field Strength	B_ω	0.5

Table 2 - Numerical parameters for a series of simulations presented in Figure 15

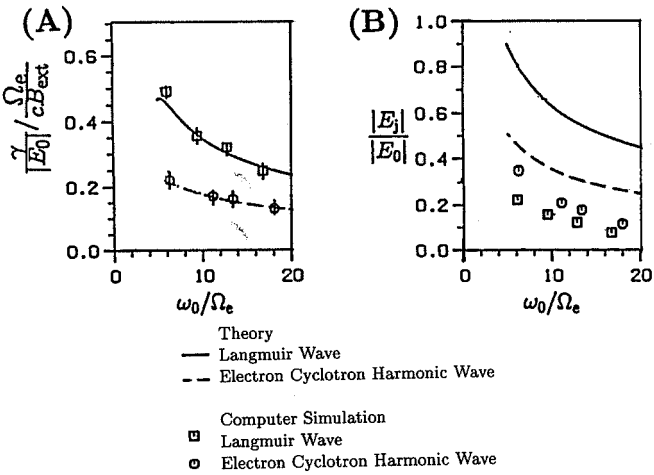


Fig. 15 - Comparison of the growth rates of the ES waves and their saturation levels between theory (indicated by lines) and computer simulations (indicated by symbols). The solid line and square symbols are for Langmuir Wave. The dashed line and circular symbols are for the ECHW.

On the other hand, the saturation levels observed in the series of simulations are much lower than the theoretical prediction shown in Figure 15(b). The larger gap between the theory and simulation for the Langmuir wave compared to the ECH wave results in the reversal of the saturation levels thereby contradicting the theoretical prediction. The saturation at lower intensity or at earlier time of interaction for the Langmuir waves turns out to be the result of the breakdown of the resonant

condition (Eq.(2)). As seen in Figure 14, the plasma is heated by the nonlinearly growing ES wave leading to heating of the plasma. Therefore, the fact that the earlier saturation than predicted by the three-wave coupling theory should be related to the plasma heating. We then examined the change of the dispersion characteristics as a function of the electron temperature. The result shows that the dispersion relation of the Langmuir waves has higher susceptibility to the change of the electron temperature. Figure 16 shows the change of the dispersion relation in the ω-k diagrams when the electron temperature is doubled. As indicated in the figure, the increment of the frequency $\Delta\omega$, which represents the frequency mismatching, is larger for the LW than for the ECHW. The frequency mismatching given by $\Delta\omega = \omega_0 - \omega_1 - \omega_2$ influences the growth of the excited ES wave through [44],

$$\gamma = \sqrt{|\beta_1\beta_2|E_0^2 - \Delta\omega^2} \tag{6}$$

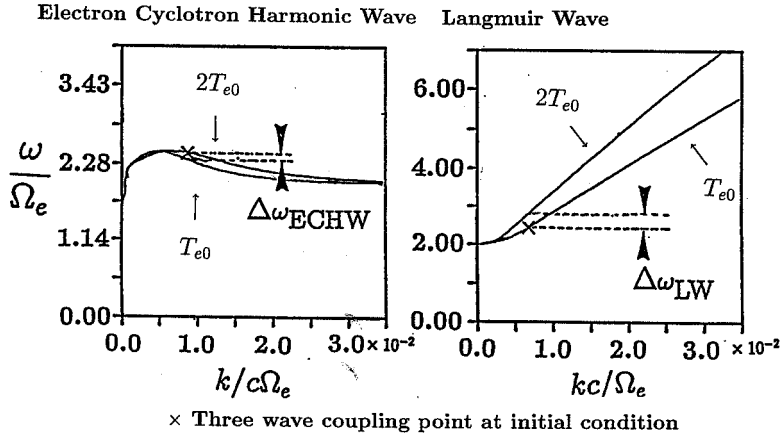


Fig. 16 - Comparison of the change of the dispersion characteristics of LW and ECHW due to doubling of the electron temperature. The quantity Dw gives a measure of the frequency mismatching from the resonant condition of the nonlinear three-wave coupling.

Therefore, the change of the dispersion relation, or the breakdown of the resonance condition for three-wave coupling due to plasma heating explains why the Langmuir waves reached saturation earlier and thereby stayed at a lower intensity than ECHW.

Another discrepancy concerning the nonlinear excitation of plasma waves by the microwave power beam revealed by the MINIX rocket experiment was the broad band nature of the excited Langmuir waves. The excited Langmuir waves did not show the line spectrum as predicted by the theory (Eq. (2)), but showed a broad spectrum. This can be understood in the light of the results of the computer simulation. The previous example of the simulation shows an effective plasma heating by the excited ES waves which self-quenches the wave-amplitude of the ES waves. This

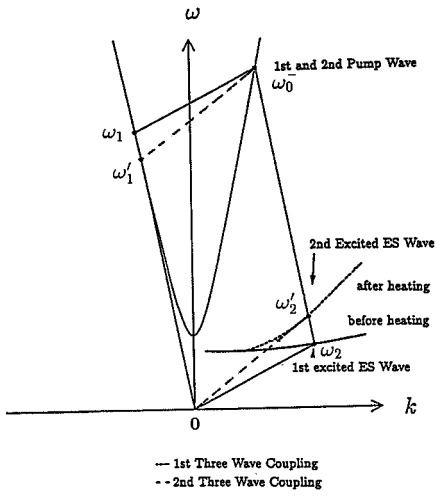


Fig. 17 - A schematic illustration of the change of resonance condition of the three-wave coupling involving Langmuir Waves. Due to the change of electron temperature, a new triplet (ω_0, ω_1', ω_2') is found automatically, thus leading to the frequency broadening.

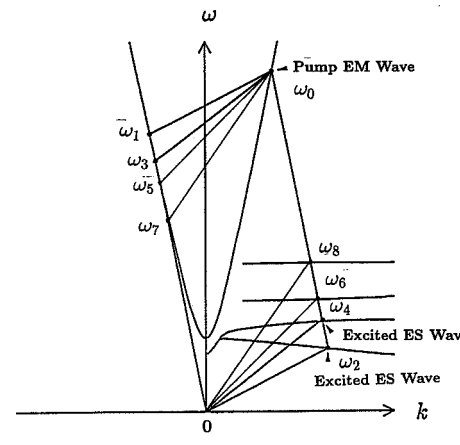
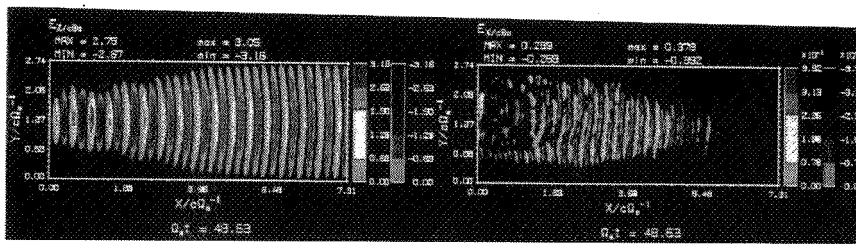


Fig. 18 - A schematic illustration of simultaneous nonlinear three-wave coupling by one EM pump wave feeding energy and momentum into multiple ECH waves.

means that the frequency and wave number of the triplet (ω_0, ω_1, ω_2) and (k_0, k_1, k_2) changes in time so that a new triplet automatically satisfies Eq. (2) for the heated plasma. A schematic illustration of this interpretation is given in Figure 17. Such sequential shift of the resonance frequency can explain the broad band nature of the observed Langmuir waves in the MINIX. Concerning the ECHW, another type of nonlinear three-wave coupling is possible for the electron cyclotron harmonic waves, as illustrated in Figure 18. In this case, multiple triplets (ω_0, ω_1, ω_2), (ω_0, ω_3, ω_4), (ω_0, ω_5, ω_6) can satisfy the energy and momentum conservation simultaneously as a result of multiple branches of the dispersion relations in the ω-k diagram. This feature can explain the MINIX multiple spectral peaks of the ECHW with a spacing of the order of the local electron cyclotron frequency.

As discussed above, even a simple one-dimensional computer simulation based on the particle model of the plasma turns out to be sufficiently effective to fill a gap between theory and experiment. Moreover the simulation solves an apparent discrepancy between the simple nonlinear theory and rocket experiment. Extending the simulation model from 1-D to 2-D, we are able to study the nonlinear interaction more realistically. Figure 19 is an example of such a 2-D simulation. Snap shots are shown of the spatial intensities of the pump wave (left upper panel), of the excited ES plasma waves (right upper panel) and of the thermal velocity of the plasma. The pump EM wave is radiated from the left boundary by an array of current sources placed on the left boundary, thus reproducing the spatially inhomogeneous intensity distribution of the MINIX microwave power beam. The ES waves shown in the upper right panel have already been damped in the vicinity of the antenna where the plasma is heated effectively.



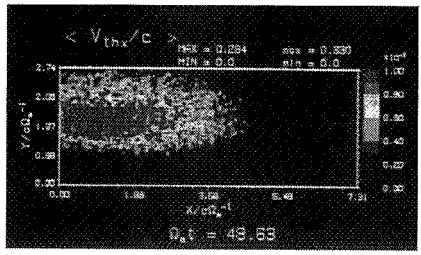


Fig. 19 - A set of the result of two-dimensional computer experiment by KEMPO on the nonlinear effects caused by the injected intense EM wave. The upper left panel shows the intensity contour of the electric field of the pump EM wave injected from the left boundary. The upper right panel shows the intensity of the excited ES waves. The lower left panel shows the spatial distribution of thermal speed of the plasma. It is interesting to note that the plasma electrons are heated in a region where the ES waves are intense.

The microwave power beam used for the future SPS has to pass through the magnetosphere and the ionosphere. In addition to the Ohmic heating and large-scale thermal instabilities discussed in Section 3, nonlinear excitation of electrostatic plasma waves is highly possible taking into account the theory, computer simulation and the rocket experiment MINIX. Though the power absorbed by these ES plasma waves and resultant plasma heating is very small, the impact of the microwave power beam onto the ionospheric plasma is not negligible. Nevertheless knowing the plasma wave characteristics excited by the microwave power beam as well as the physical plasma process involved in the excitation, we should be able to avoid possible interference to the HF communication network.

6. Application of Microwave Power Transmission to Microwave-Driven Airplane

In the late 1980's, a program to develop a long endurance high altitude platform called SHARP (Stationary High Altitude Relay Platform) was proposed in Canada [45]. The idea is to float an unmanned light-weight airplane for a long period, circling at an altitude of about 21 km for the purpose of relaying radio communications signals over a wide area. To maintain the platform floating for weeks or months, a fuel-less airplane powered by microwave energy transmitted from the ground was proposed and experimented on [46]. On September 17, 1987, a 1/8-scale prototype SHARP flew on beamed microwave power for 20 minutes at an altitude of about 150 m. Figure 20 shows a photo of the prototype SHARP with a 4.5 m wingspan. The microwave beam was transmitted by a 4.5 m diameter parabolic antenna transmitting 10 kW microwave with a

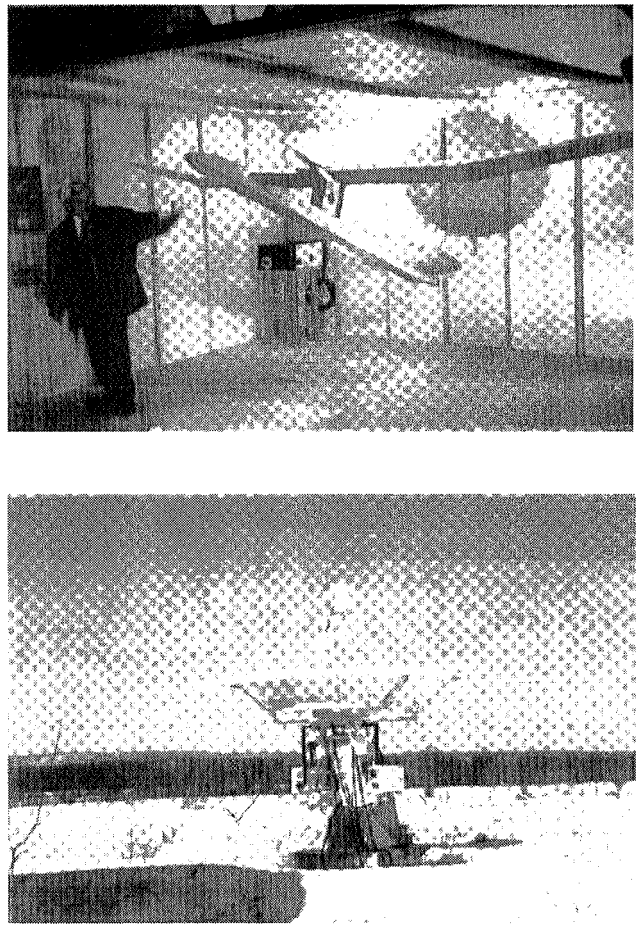
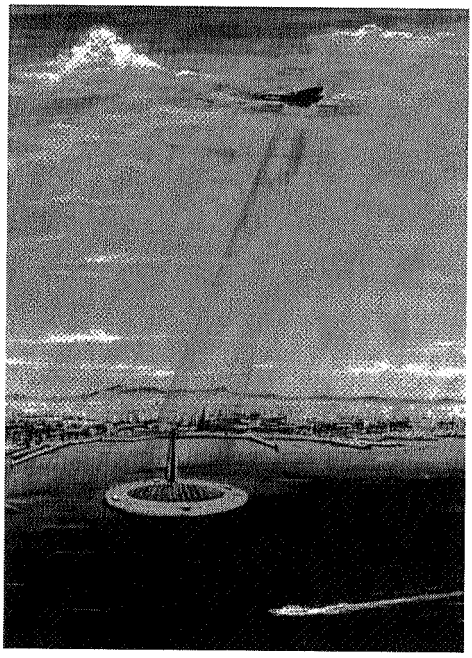


Fig. 20 - A 1/8-scale SHARP Airplane and a parabolic antenna which will be used for 1/4-scale SHARP experiment.

frequency of 2.45 GHz. Two water-cooled magnetrons each with 5 kW output power were used. The parabolic antenna mechanically tracked the airplane which flew inside a 50 degree cone. The power density at the airplane altitude was 400 W/m^2. A dual polarization rectenna with two orthogonal linearly-polarized dipole arrays was developed. The rectenna diodes used in the first flight were Silicon Schottky diodes (HP2835). Its power handling capability was 1 W/element, and its microwave-to-DC conversion efficiency was about 70% [46]. The rectenna received sufficient power to feed 150W to the electric motor of the 4.1 kg weight SHARP airplane [46].



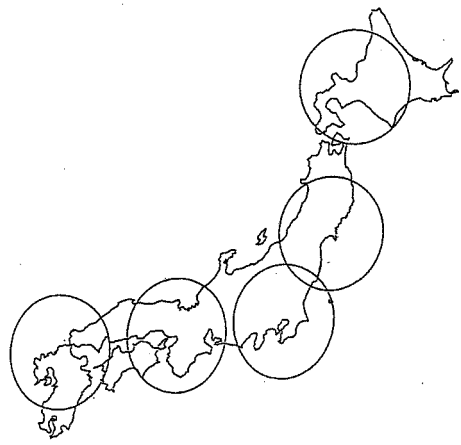


Fig. 21 - A schematic illustration of Stratospheric Radio Relay System (SRRS). Five SRRS's will cover most of the heavy communication demands in Japan.

A similar project was carried out in Japan in the early 1990's. The project was called Stratospheric Radio Relay Systems (SRRS), and was studied by a working group under the Ministry of Posts and Telecommunications of Japanese government [47]. The objectives of the SRRS are similar to those of Canadian SHARP. In the SRRS, it is planned to launch five such unmanned airplanes over Japan as depicted in Figure 21, so that these five platforms can cover most of areas where communication demands are heavy. In parallel with the SRRS working group, a microwave-driven airplane experiment was planned and conducted successfully on August 29, 1992 by a joint team organized by the present author [48]. The team members were from Kyoto University, Kobe University, Communications Research Laboratory, Nissan Motor Co. Ltd., Fuji Heavy Industries Ltd. and Toshiba Co. The experimental project was called MILAX meaning MIcrowave Lifted Airplane eXperiment, and was partly sponsored by ISAS of Japan. The MILAX airplane is a balsa-based light-weight (~ 4 kg) airplane with a 2.5 m wingspan and has a shape as shown in Figure 22. The MILAX flew successfully for 40 seconds (or 400 m distance over a straight course for car driving test) at an altitude of about 15 m. The testing scenery is shown in Figure 23. Because of the limits of the maximum microwave power (~ 1 kW) and of the aperture of the transmitting antenna (~ 1.2 m) , the flight altitude had to be as low as 15 m in order to guarantee the power density of 200 W/m$_2$ at that altitude. The microwave power beam was radiated toward the fuel-free MILAX airplane by an active phased array antenna. The MILAX active phase array transmitter was composed of five-stage Gallium-Arsenic (GaAs) semi-conductor amplifiers (see Figure 24), 4-bit digital phase shifters and circular microstrip antennas (see Figure 25). The transmitter is divided into 96 sub-arrays, each consisting of 3 antennas, one phase-shifter and one

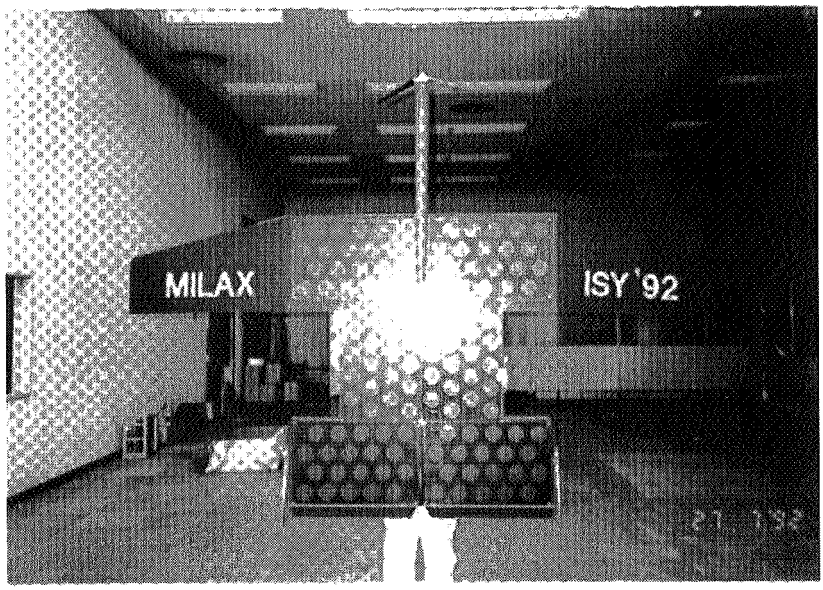


Fig. 22 - An outlook of MILAX airplane (Bottom side). Circular patches are microstrip antenna used for the antennas.

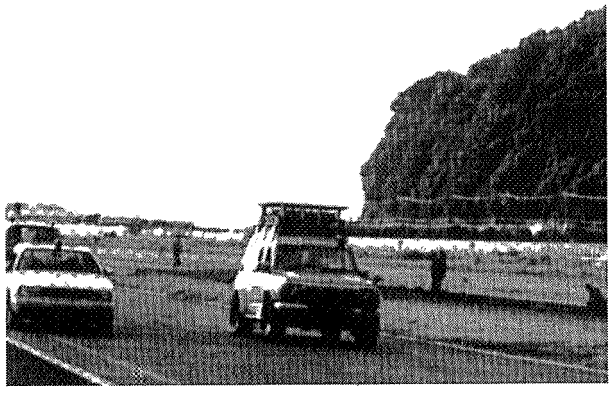


Fig. 23 - MILAX demonstration flight. The MILAX airplane flew only by the microwave power transmitted from the transmitter car running panelled to the airplane. The MILAX was conducted on Aug. 29, 1992 at Oppama driving test course of Nissan motor Co, Japan.

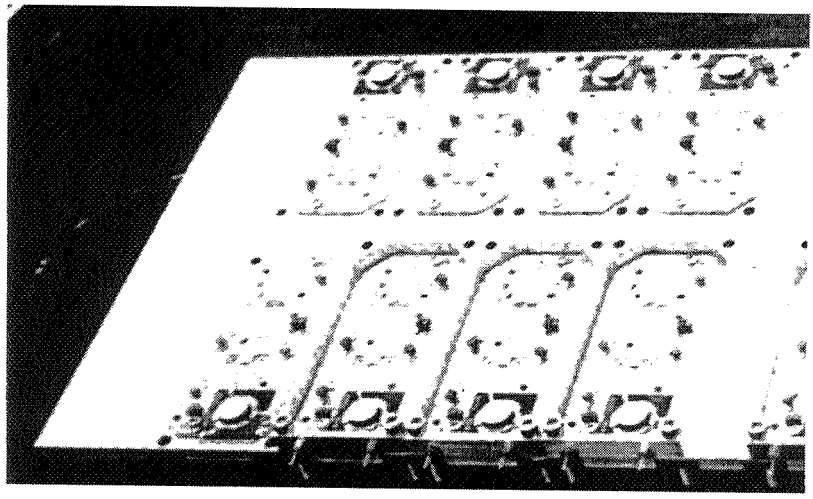


Fig. 24 - GaAs-based semiconductor amplifiers used in the MILAX. Each amplifier supplies 13 W microwave output.

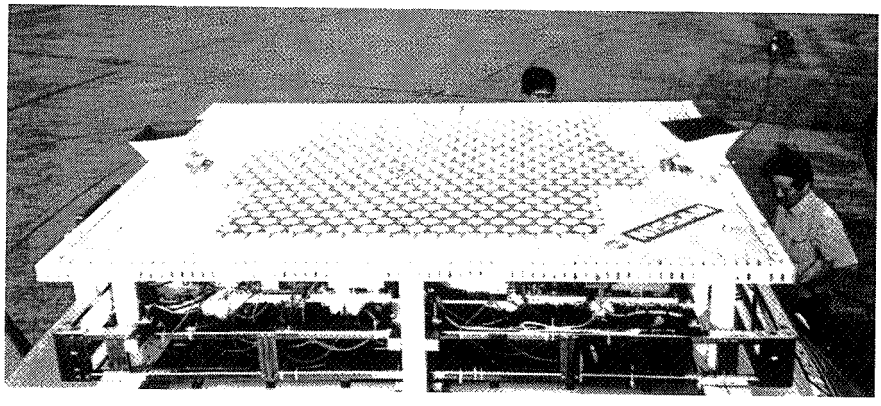
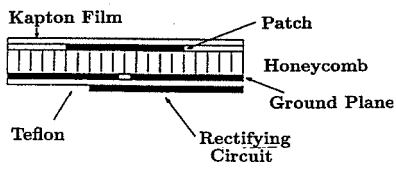


Fig. 25 - A view of the transmitting antenna array installed on the roof of a transmitter car. The antennas were of circular microstrip type.

GaAs amplifier. Each sub-array can supply 13 W microwave output resulting in the total radiation capability of 1.25 kW. The frequency used in the MILAX was 2.411 GHz in the ISM frequency band. The transmitter system was installed on the roof of a transmitter car (see Figure 25).

Six rectenna subarrays, each consisting of 20 rectennas are installed on the flat-bottom of the MILAX airplane. Prior to the development of the MILAX rectenna, several rectenna researches had been done in Japan [49, 50, 51]. Based on these studies, the receiving antennas used for the MILAX rectenna were not of the dipole-type, like these used in the JPL/Goldstone Ground-to-Ground Power Transmission Experiment and in the MINIX and SHARP, but were of a new type



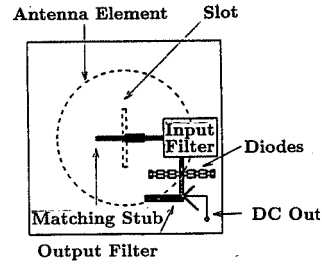


Fig. 26 - A microstrip circular antenna-based rectenna used for the MILAX project [50].

of microstrip circular patch antennas. The circular patch antennas have the advantage of a non-resonant nature at integer multiple harmonic frequencies, thereby having the capability of suppressing spurious radiation from the rectennas. The disadvantage of heavier weight as compared to dipole antennas was overcome by introducing a paper honeycomb structure [52], as shown in Figure 26. The diodes used for the MILAX rectenna are eight HP5082-2350 Schottky diodes in 2-series / 4-parallel combination. The power handling capability was 1 W per element, and the microwave-to-DC conversion efficiency was about 52 % [52].

The main reason of the adoption of the active phased array in place of a conventional parabolic antenna is its higher steerability of the microwave power beam. The power beam can be controlled and steered electronically in contrast to the mechanical control of a parabolic antenna. In the MILAX, we monitored the location of the MILAX airplane by two CCD cameras which were installed on the edge of the roof-transmitter antenna looking upward. In Figure 27, a system of identifying the location of the airplane is shown. A micro-computer, after recognizing the pattern

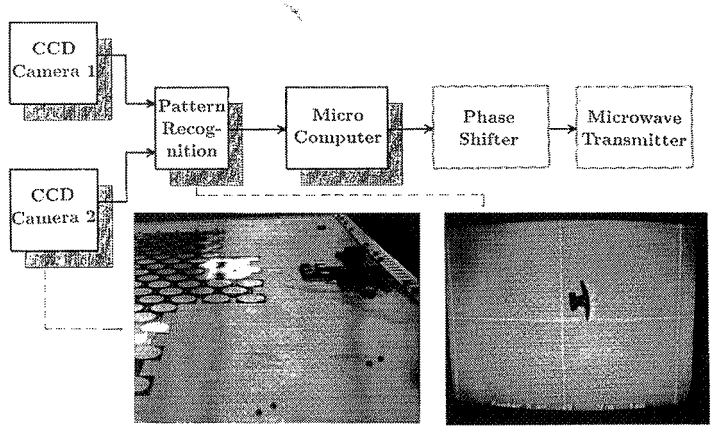


Fig. 27 - A computer-controlled beam steering system for microwave power transmission toward the MILAX airplane. The image of the airplane is captured by two CCD cameras installed on the roof of the transmitter car. Then the computer recognizes the airplane location and height by a pattern recognition software. According to the information on height and location, the computer controls the phase shifters of the microwave amplifiers of the active phased array.

of the airplane image and calculating the x-y coordinates and the altitude of the airplane, sends the control signals to the phase shifters of the microwave amplifiers so that the microwave beam is accurately directed toward the airplane. This system worked perfectly in the MILAX.

Though we adopted computer-control for steering the microwave power beam in the MILAX due to time and budget limitations, we have also attempted to develop the retro-directive beam control system. Figure 28 shows a photograph of a 90 W microwave power transmitter with the use of the retro-directive control method [53]. Seven dipole antennas are connected to semi-conductor microwave amplifiers. The 90 W microwave power transmitter has capability of transmitting the microwave power beam automatically in the direction of incoming pilot signals. A unique feature of the developed unit shown in Figure 28 is that it uses a new phase conjugate circuit (PCC) using asymmetric two pilot frequencies. The PCC used for the unit is shown in Figure 29. As we use two asymmetric pilot signals $\omega_0 + \Delta\omega$ and $\omega_0 + 2\Delta\omega$ instead of $\omega_0 \pm \Delta\omega$, we need no division of the phase ϕ_t of the pilot signal after the first mixer. Therefore we can determine the phase $\phi_1 t$ uniquely without the ambiguity of π radians.

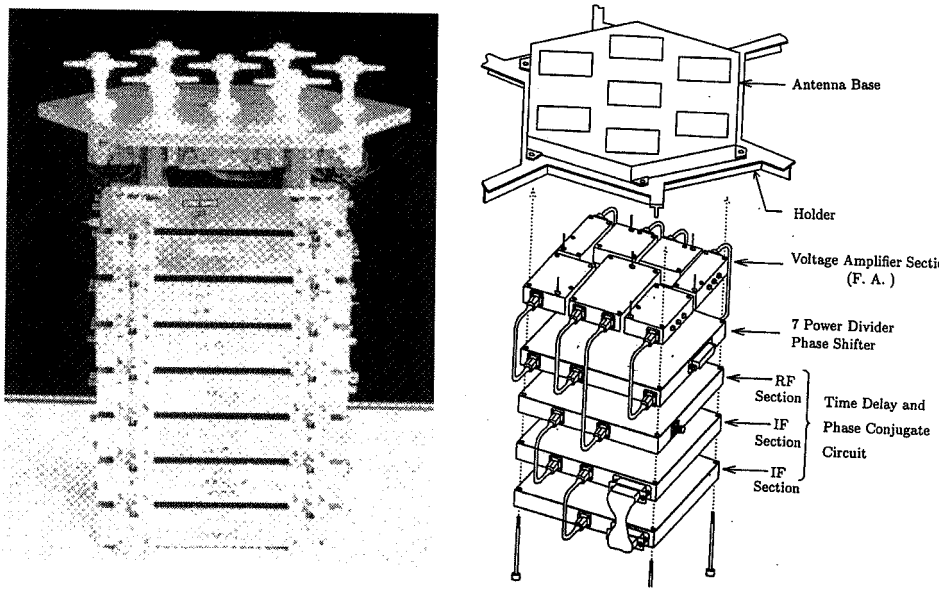


Fig. 28 - A 90 W microwave power transmitter unit composed of seven dipole antennas, seven GaAs semiconductor amplifiers and retrodirective phase conjugate circuits.

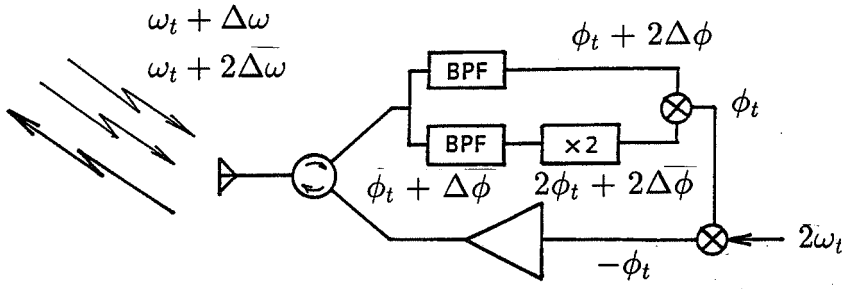
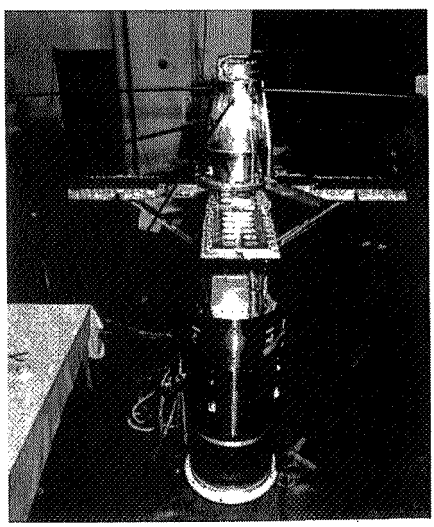


Fig. 29 - A new phase conjugate circuit used for the 90 W unit in Figure 28. Two pilot signals with asymmetric frequencies of $\omega_0 + \Delta\omega$ and $\omega_0 + 2\Delta\omega$ are used.

7. Other Recent Experiments on Microwave Power Transmission in Japan

On Feb. 18, 1993, a second rocket experiment of microwave power transmission between mother and daughter units was carried out by the S-520-16 sounding rocket [54, 55]. The rocket experiment was given the name of ISY-METS meaning Microwave Energy Transmission in Space during the International Space Year. The ISY-METS had objectives of investigating nonlinear effects of the high power microwave onto the ionospheric plasma in more or less a similar way to that of the MINIX. In this sense, the ISY-METS is an advanced version of the MINIX. However, it has another mission to verify a newly developed active phased array microwave transmitter which had been modified from that used in the MILAX in the space plasma environment. The total power of approximately 800 W was transmitted from the microstrip array



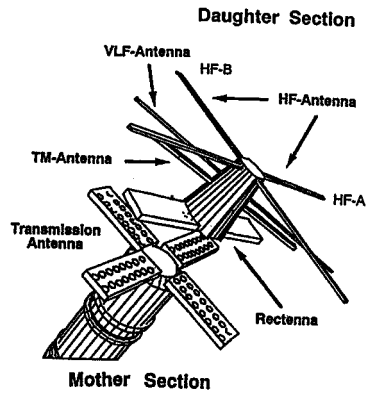


Fig. 30 - Picture and illustration of the payload section of the ISY-METS rocket experiment. Four deployed paddles mount 16 transmitting microstrip antennas each. The daughter unit carried diagnostic packages and rectennas with various sensors.

antennas mounted on four deployed paddles. The configuration of the transmitting antenna paddles and the sensors extended outward from the daughter unit are shown in Figure 30. The phase of the transmitted microwave from each antenna on the deployed paddle was controlled by the same 4-bit digital phase shifter as used in the MILAX. The phase shifters were controlled by an onboard computer providing a variable transmitted power density and direction of the microwave beam. Figure 31 shows an example of power concentration at a point of approximately 4.5 m away from the center of the transmitting antennas. Such power-concentrated points were determined by the computer using pre-set parameters and the onboard real-time data of the relative direction the daughter unit. Figure 32 shows a data set during a pre-flight test in the radio anechoic chamber. The left panel shows the antenna pattern of the microwave transmitter. The right panel shows the measured power at concentrated points which were varied from 1 m to 10 m from the transmitting antenna. The measured data are shown by the solid line in the figure. The dots in the same figure are the power density measured onboard the ISY-METS rocket during the flight; these measurements and agreed well with the pre-flight test values [55].

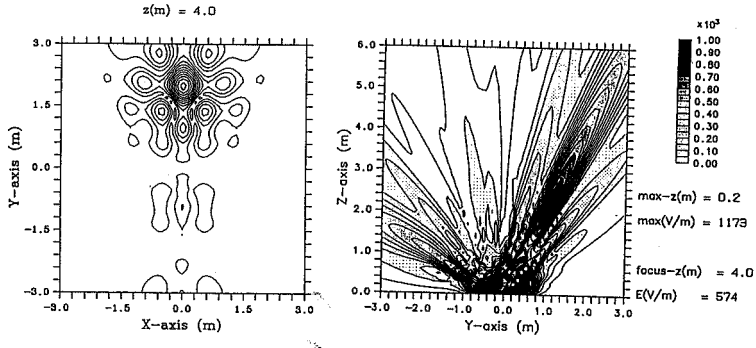


Fig. 31 - Computed power concentration map in the x-z and x-y plane, where the contour of the ISY-METS transmitting antenna is placed at x = 0, y = 0 and z = 0 in the x-y plane. In the specific example, the power-concentration point is set at (x, y, z) = (0 m, 2 m , 4 m)

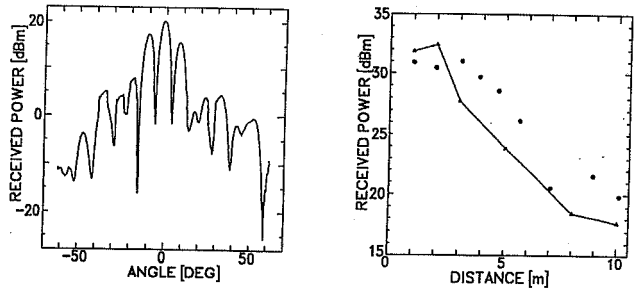


Fig. 32 - Antenna pattern of the ISY-METS microwave power transmitter (left), and the measured power (right) at power-concentrated points at distances from 0 m to 10 m. The solid line and dots represent the values measured in the pre-flight radio anechoic chamber, and in the flight in space, respectively.

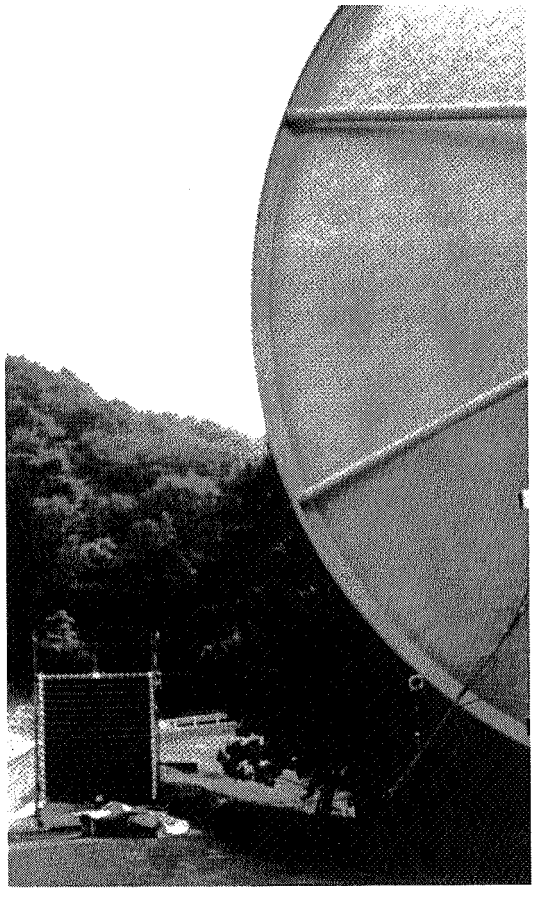


Fig. 33 - A photo showing a field experiment of MPT at Yamazaki Experimental Site of Kansai Electric Power Company Inc (KEPCI). The experiment is a joint collaborative research between Kyoto University, Kobe University and KEPCI (K^3 project).

Two types of rectennas were installed on the daughter unit. One was developed by a research group at Texas A&M University, while the other was by the CRL group. The former used an orthogonally placed pair of three dipole antennas, as used in the MILAX. The structure and configuration are found in Figure 9 and Figure 10 in [55]. The initial results of the ISY-METS rocket experiment are now under analyses and will be published in detail elsewhere.

Other applications of microwave power transmission have recently gathered interests in Japan in the practical world in addition to the academia. One of them is a small-scale ground-to-ground power transmission without wires toward a distant place where wired power distribution networks are either unavailable not or very poorly available. In order to collect fundamental data on microwave power transmission under varying weather conditions, the Kansai Electric Power Company Inc. began a collaborative field experiment with Kyoto University and Kobe University. Figure 33 shows a photo of the experimental site showing a parabolic antenna with a 3 m diameter driven by a 5 kW magnetron and a rectenna array of a size of 3.5 m x 3.2 m placed 42 m away from the parabolic transmitting antenna. The rectenna array was built by Kyoto University and is composed of 2304 rectenna elements. A preliminary test started in October, 1994 and is still being conducted to collect fundamental data on the characteristics of power transmission and reception by the system.

8. Discussion and Conclusion

The Tesla idea of wireless power transmission was revived by the NASA/DOE SPS studies program in the 1970's. Following the pioneering work on microwave power transmission by W. C. Brown, many engineers and scientists have conducted the related research and developed applications stimulated by the SPS studies program in

the US, former Soviet Union, France, Germany, Japan and other countries. However, even after the SPS boom subsided following the suspension of the SPS research in the US in 1981, both fundamental academic research and application-oriented developments and experiments in microwave power transmission have continued. Radio scientists have been a core of such research and development because the power transmission by radio inevitably involves problems in radio science such as those on antennas, rectennas, propagation characteristics and nonlinear interaction of microwave

power beam with plasmas and neutral environments. In this paper, we have emphasized the development of such studies after 1980 as studies before that time was well documented (e.g.[1]). Since 1980, many feasibility studies have been made and a variety of new ideas of utilization of microwave power transmission have been proposed. However, as far as the present author knows, very few experiments have actually been carried out except for those discussed in the present paper except for a laboratory-based development and research of rectennas.

There are many plans and proposals for MPT applications, e.g., a French plan of a medium-scale wireless power transmission (~ 100 kW, 3 km) on Reunion island, a proposal named WISPER (Wireless Space Power Experiment) in the US which aims at ground-to-space power transmission of the order of 100 kW, and a test project called ALASKA 21 which plans to transmit electric power via microwave to scattered villages in Alaska. As well as these, private industries have their own MPT projects, while national research institutes will continue fundamental researches.

In Japan, MITI (Ministry of Trade and Industry) has re-examined the SPS feasibility based on recently developed technologies after 1980. Japanese research groups at ISAS and national universities are accordingly pursuing a conceptual study on an orbiting 10MW-scale Power satellite called SPS2000 which beams down the electricity to equatorial countries [56] (Figure 34) and on a power satellite called PSS (Power Supplying Satellite) which feeds electricity of the order of 100 kW to other orbiting customers [57] (Figure 35).

In summary, microwave power transmission has been one of the interesting topics in radio science both technologically and scientifically However, the use of radio waves as a means of transmitting electric power and energy is still in a very immature phase. More research and development will be needed before the dreams of Hertz and Tesla became reality. Radio scientists and engineers have great challenges to face in this new field.

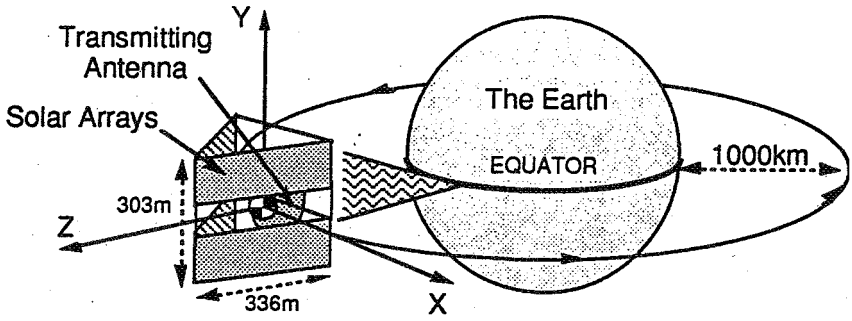


Fig. 34 - A concept of SPS2000 which will beam down electrical energy to equatorial countries from an orbiting satellite with a shape of triangular structure.

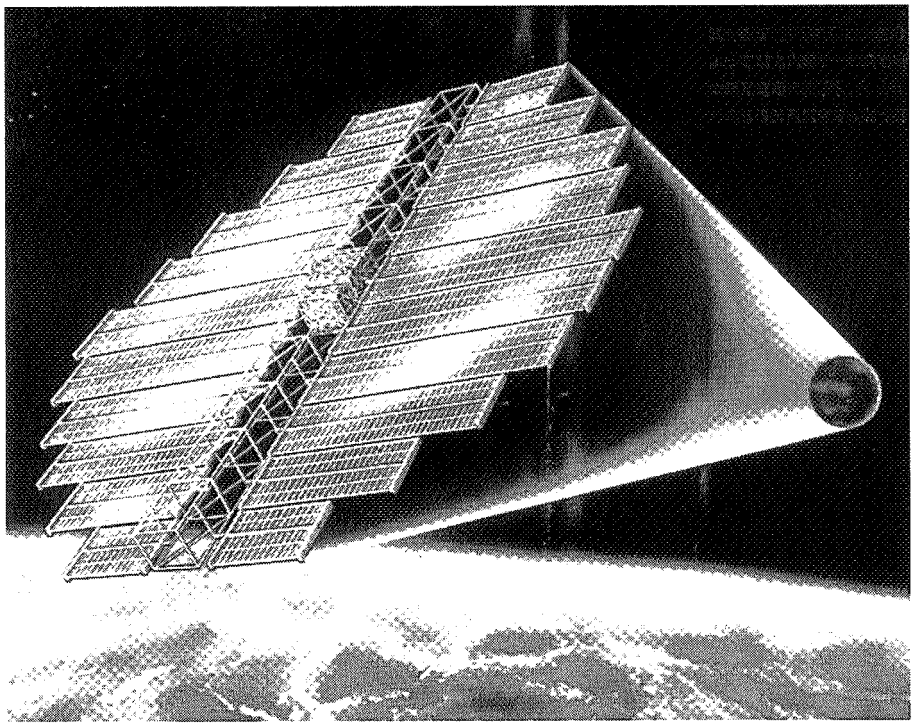


Fig. 35 - A concept of Power Supplying Satellite. A 100 kW power will be transmitted to customer satellites.

Acknowledgments

I would like to express my thanks to N. Kaya of Kobe University for his collaboration on this topic for many years. Without his contribution most of the experimental projects described in this paper would not have been successful. I also thank N. Shinohara, T. Miura and K. Miwa for their help in preparing this manuscript, and D. Summers for his careful reading of the manuscript.

References

[1] W. C. Brown, The history of Power transmission by radio waves, IEEE transaction on microwave theory and techniques, MTT-32, 1230-1242, 1984.

[2] I. C. Maxwell, Treatise on Electricity and Magnetism, Dover Publications (reprinted), New York, 1954.

[3] H. Hertz, Electric Waves, MacMillan and Co., New York, 1893.

[4] G. Marconi, Nozioni Elementari de Radio Telegrafia, Ufficio Marconi, Roma, 1916.

[5] R. A. Fessenden, Wireless Telephony, 25th annual convention of the American Institute of Electrical Engineers, Atlantic City, NJ, June, 1908.

[6] J. S. Belrose, Fesssenden and the Early History of Radio Science, The Radio Scientist and Bulletin, URSI, 5, 94-110, 1994.

[7] N. Tesla, The transmission of electric energy without wires, The thirteenth Anniversary Number of the Electrical World and Engineer, March 5, 1904.

[8] N. Tesla, Experiments with Alternate Current of High Potential and High Frequency, McGraw Pub. Co., N.Y., 1904.

[9] A. W. Hull, Effect of uniform magnetic field on the motion of electrons between co-axial cylinder, Phys. Rev., 18, 31-57, 1921.

[10] K. Okabe, Generation of Ultra short wavelength electromagnetic waves by magnetic with divided anodes, J. Institute of Electrical Engineering of Japan, 49, 284-290, 1928.

[11] H. Yagi and S. Uda, Feasibility of Electric Power Transmission by Radio Waves, Third Pan-Pacific academic conference, Tokyo, 1926.

[12] R. H. Varian and S. F. Varian, A high-frequency oscillator and amplifier, J. Appl. Phys., 10, 321, 1939.

[13] S. Nakajima and S. Yamazaki, Application of high power magnetron at Naval Research Laboratory during World War II time (in Japanese), History of Electronic Tubes, Ohm Pub. Co., Japan, 157-165, 1987.

[14] W. C. Brown, Experiments involving a microwave beam to power and position a helicopter, IEEE Trans. Aerosp. Electron. Systems, AES-5, No.5, 692-702,1969.

[15] W. C. Brown, Thermionic diode rectifier, in Microwave Power Engineering, vol. 1, E. C. Okress, Ed., N.Y. : Academic, 295-298 , 1968.

[16] W. C. Brown, The combination receiving antenna and rectifier, in Microwave Power Engineering, vol. II, E. C. Okress, Ed. N. Y.: Academic, 273-275, 1968.

[17] W. C. Brown, Progress in the design of rectennas, J. Microwave Power, vol.4, 168-175, 1969.

[18] R. M. Dickinson and W. C. Brown, Radiated microwave power transmission system efficiency measurements, Tech-Memo 33-727, Jet Propulsion Lab., Cal. Inst. Technol., March 15, 1975.

[19] Final Proc. Solar Power Satellite Program Rev. DOE/NASA Satellite Power System Concept Develop. Evaluation Program, Conf.-800491, July, 1980.

[20] P. E. Glaser, Power from the Sun, Science, 162, 857-886, 1968.

[21] F. W. Perkins and R. G. Roble, Ionospheric heating by radio waves : Predictions for Arecibo and the Satellite Power Station, J. Geophys. Res., 83, 1611-1624, 1978.

[22] L. M. Duncan and J. Zinn, Ionosphere-Microwave interactions for Solar Power Satellite, Los Alamos Scientific Lab. Report, LA-UR-78-758, March 1978.

[23] L. H. Holway and G. Meltz, Heating of the Laser ionosphere by powerful radio waves, J. Geophys. Res., 78, 8402-8408, 1978.

[24] L. M. Duncan and W. E. Gordon, Final Report, Ionosphere / Microwave beam interaction study, Rice University, Houston, TX, Final Report NAS9-15212, Sept., 1977.

[25] C. M. Rush, SPS simulated effects of ionospheric heating on the performance of telecommunication systems

: a review of experimental results, Space Solar Power Review, 2, 355-366, 1980.

[26] S. Basu, S. Bau, A. L. Johnson, J. A. Klobuchar, and C. M. Rus, Preliminary results of scintillation measurements associated with ionospheric heating in overdense and underdense modes, Geophys. Res. Lett., 7, 609-612, 1980.

[27] F. W. Perkins and E. J. Valeo, Thermal self-focusing electromagnetic waves in plasmas, Phys. Review Lett., 32, 1234-1237, 1974.

[28] J. E. Drummond; Thermal stability of Earth's ionosphere under power transmitting satellites, IEEE Transactions of Plasma Science, PS-4, 228, 1976.

[29] G. D. Thome and F. W. Perkins, Production of ionospheric striations by self-focusing of intense radio waves, Phys. Rev. Lett., 32, 1238-1240, 1974.

[30] L. M. Duncan and R. A. Behnbe, Observation of self-focusing electromagnetic waves in the ionosphere, Phys. Rev. Lett., 41, 998, 1978.

[31] B. L. Cragin, J. A. Fejer and R. L. Showen, Theory of coherent parametric instabilities excited by two or more pump waves, Geophys. Res. Lett., 5, 183-186, 1978.

[32] R. L. Showen, L. M. Duncan, and B. L. Gragin, Observations of plasma instabilities in a multiple pump ionospheric heating, Geophys. Res. Lett., 5, 187-190, 1978.

[33] H. Matsumoto, Theory of nonlinear interaction of SPS microwave with the ionosphere (in Japanese), Report of Inst. Space Astro. Science, Univ. of Tokyo, 15, 407-430, 1979.

[34] H. Matsumoto, Numerical estimation of SPS microwave impact on ionospheric environment, Acta Astronautica, 9, 493-497, 1982.

[35] H. Matsumoto and T. Kimura, Nonlinear excitation of electron cyclotron waves by a monochromatic strong microwave, Space Power, 6, 187-191, 1986.

[36] M. Nagatomo, N. Kaya and H. Matsumoto, Engineering aspect of the microwave ionosphere nonlinear interaction experiment (MINIX) with a sounding rocket, Acta Astronautica, 13, 23-29, 1986.

[37] N. Kaya, H. Matsumoto, S. Miyatake, I. Kimura, M. Nagatomo and T. Obayashi, Nonlinear interaction of strong microwave beam with the ionosphere, Space Power, 6, 181-186, 1986.

[38] I. Kimura, H. Matsumoto, N. Kaya and S. Miyatake, Plasma wave excitation by intense microwave transmission from a space vehicle, Advances in Space Research, 8, 291-294, 1988.

[39] N. Kaya and H. Matsumoto, Space chamber experiments of ohmic heating by high power microwave from the Solar Power Satellite, Geophys. Res. Lett., 8, 1289-1292, 1981.

[40] H. Hirata, Nonlinear wave-wave-particle interaction induced by intense EM waves in a magnetized plasma, Master Thesis (supervised by H. Matsumoto), Kyoto Univ., 1989.

[41] H. Matsumoto, H. Hirata, Y. Hashimoto and N. Shinohara, Theoretical analysis of nonlinear interaction of intense electromagnetic wave and plasma waves in the ionosphere (in Japanese), Trans. IEICE on Communications, vol. 78-B-II, March, 1995.

[42] H. Matsumoto, Computer experiment of space plasmas, Computer Space Plasma Physics, Simulation Techniques and Software, ed. by H. Matsumoto and Y. Omura, Terra. Sci. Pub. Co., Tokyo, 3-18, 1993.

[43] H. Matsumoto and Y. Omura, Particle simulation of electromagnetic waves and its application to space plasma, Computer Simulation of Space Plasma, ed. by H. Matsumoto and T. Sato, Terra Sci. Pub. Co. Tokyo, 43-102, 1985.

[44] for example, T.Taniuchi and K.Nishihara, Non-linear Waves (in Japanese), Iwanami Pub.Co., pp.116-134, 1977.

[45] G. W. Jull, A. Lillemark and R. M. Tunner, SHARP: Telecommunications Missions and Systems, Proceedings IEEE Globecom 85, 955-959, 1985.

[46] J. J. Schlesak, A. Alden and T. Ohno, A microwave powered high altitude platfome, IEEE MTT International Microwave Symposium, N.Y. Vol.I 283-286, 1988.

[47] Report of Studies on Stratospheric Radio Relay Systems, SRRS Working Group, Ministry of Postage and Communications, March 1991.

[48] H.Matsumoto, N. Kaya, M. Fujita, T. Fujiwara and T. Sato, Microwave Lifted Airplane Experiment with active phased array antennas, to be submitted to IEEE Trans. MTT, 1995.

[49] S. Adachi and Y. Shimanuki, Theoretical and experimental study on rectenna array for microwave power transmission, Space Solar Power Review, 5, 127-129, 1985.

[50] K. Itoh, Y. Akiba and Y. Ogura, Fundamental study on SPS rectenna printed ona sheet of Copper Glad Laminate, Space Solar Power Review, 5, 149-162, 1985.

[51] K. Itoh, T. Ohgane adn Y. Ogawa, Rectenna composed of circular microstrip antenna, Space Power, 6, 193-198, 1986.

[52] Y. Fujino, M. Fujita, H. Sawada and K. Kawabata, A rectenna for MILAX, 12th Space Energy Symposium, ISAS, Tokyo, March, 1993.

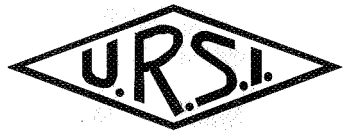
[53] H. Matsumoto, N. Kaya, N. Nagatomo, T. Hashizume,and K. Nakatsuka, Feasibility Study on Retrodirective System for METS, The Sixth ISAS Space Energy Symposium, 24, March 1987.

[54] H. Matsumoto and N. Kaya, Initial results of ISY-METS rocket experiment (in Japanese), RASC report Kyoto University, May, 1993.

[55] R. Akiba, K. Miura, M. Hinada, H. Matsumoto and N. Kaya, ISY-METS rocket experiment, The ISAS Report No.652, 1-14, September, 1993.

[56] M. Nagatomo, 10 MW Satellite Power System: A space station mission beyond 2000, Space Power 6, 299-305, 1986.

[57] H. Matsumoto, N. Kaya, S. Kinai, T. Fujiwara and J. Kochiyama, A feasibility study of Power Supplying Satellite (PSS), Space Power, 12, 1-6, 1993.



Symposium Chair: Dr. Pierre Bauer, URSI President
Technical Programme Committee: Prof. Paul Delogne
Local Organizing Committee: Prof. Jean Van Bladel
Secretary : Prof. Peter Van Daele

URSI Sercretariat:
Universiteit Gent - INTEC, St. Pietersnieuwstraat 41, B-9000 Gent, BELGIUM,
Tel: + 32 (9) 264 33 20, Fax: + 32 (9) 264 42 88, e-mail: inge.heleu@intec.rug.ac.be