AIS message extraction from overlapped AIS signals for SAT-AIS applications

Andis Dembovskis

DISSERTATION

zur Erlangung des Grades

eines Doktors der Naturwissenschaften

– Dr. rer. nat. –

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen

im März 2015

Gutachter: Prof. Dr. rer. nat. Marc Kesseböhmer (Universität Bremen)

Prof. Dr. rer. nat. Hansjörg Dittus (Universität Bremen)

Abstract

The AIS (Automatic Identification System) is a communication standard for ships traveling the seas and oceans. It serves as a collision avoidance system by identifying nearby ships, thus assisting in safe navigation.

The SAT-AIS (Satellite based Automatic Identification System) is a communication technology for ship traffic surveillance from space and is under active research and development worldwide. The basic principle of the SAT-AIS system is to monitor AIS channels.

The motivation for using terrestrial AIS technologies with space applications is of great interest to safety organizations that monitor ship traffic in high seas and oceans. These regions far away from coastal zones are unreachable from the terrestrial antennas, which have a usual range of 40 kilometres. Successful application of the SAT-AIS could provide AIS data to coast guards and other agencies, with an hourly ship location update from every place on the planet.

The first trials of SAT-AIS in 2006 suffered from some serious difficulties. As AIS was initially designed to be a terrestrial traffic avoidance application for ships, with the traffic participants communicating among their neighbours and the nearby coast guard, it was developed without resistivity against effects which arise when applied for space applications. Apart from signal strength and Doppler shift effects, which could be constructively handled, the demodulation of overlapped AIS messages proved to be a great challenge.

This work analyses the problem of overlapping AIS signals and proposes innovative approaches for reconstructing these based on L^2 norm orthogonalization and projections. Moreover, the work showcases results of demodulation efficiency analysis for simulated real world application of satellite passes over a dedicated shipping region based on AIS channel simulation in noisy environment

For more reliable AIS data reception in space, new dedicated frequencies are allocated for channels AIS3 and AIS4, which are being affirmed for all AIS transceiver installations from 2013. These new frequency channels carry dedicated messages with a ship position report, encapsulated into smaller data packets at lower report rates, which promises to partly eliminate the packet overlapping problem. Since the new Space-AIS format does not completely solve the packet collision problem and as the steady growth of interest on terrestrial-AIS message content received from space continues to persist, the topic of solving overlapped AIS signals remains vital for SAT-AIS applications.

Zusammenfassung

AIS (Automatisches Identifikationssystem) ist ein Kommunikationsstandard für Schiffe, die auf den Meeren und Ozeanen fahren. Indem es einander nahe Schiffe identifiziert, dient es als Kollisionsvermeidungssystem und trägt zur sicheren Navigation bei.

SAT-AIS (Satellitenbasiertes AIS) ist eine Kommunikationstechnologie zur Überwachung des Schiffsverkehrs aus dem Weltraum, an der weltweit in Forschung und Entwicklung gearbeitet wird. Das Grundprinzip von SAT-AIS besteht darin, AIS-Kanäle aus dem Weltraum zu überwachen.

Der Ansatz, terrestrische AIS-Technologie mit Weltraumanwendungen zu kombinieren, ist von großem Interesse für Sicherheitsorganisationen, die den Schiffsverkehr auf hoher See überwachen. Diese Gebiete fernab der Küstenzonen sind für terrestrische Antennen unerreichbar, die üblicherweise eine Reichweite von 40 km haben. Erfolgreiche Anwendung des SAT-AIS könnte Küstenwachen und andere Agenturen mit AIS-Daten versorgen, mit stündlicher Aktualisierung der Schiffsstandorte für jeden Ort des Planeten.

Die ersten Versuche mit SAT-AIS in 2006 zeigten einige schwerwiegende Probleme auf. Da AIS ursprünglich als terrestrisches System zur Überwachung des Schiffsverkehrs ausgelegt wurde, wobei die Verkehrsteilnehmer mit ihren Nachbarn und der nahen Küstenwache kommunizieren, wurde es entwickelt ohne Schutz gegen typische Effekte, die bei der Benutzung im Weltraum auftreten. Neben der Signalstärke und Doppler-Verschiebungs-Effekten, die konstruktiv gelöst werden konnten, erwies sich die Demodulation von überlappenden AIS-Signalen als große Herausforderung.

Diese Arbeit analysiert das Problem überlappender AIS-Signale und schlägt innovative Ansätze zu ihrer Rekonstruktion vor, basierend auf L^2-Norm Orthogonalisierung und Projektionen. Darüber hinaus stellt die Arbeit die Ergebnisse einer Wirksamkeitsanalyse der Demodulation vor, die auf der Simulation von realen Satellitenüberflügen über eine Schifffahrtsregion basiert mittels der Simulation von AIS-Kanälen in einer Umgebung mit starkem Rauschen.

Für einen zuverlässigeren Datenempfang im Weltraum wurden neue Frequenzen für die Kanäle AIS3 und AIS4 zugewiesen, die seit 2013 in allen neu installierten AIS-Transceivern Anwendung finden. Diese neuen Frequenzkanäle übertragen die Nachricht mit der Schiffsposition verpackt in kleinere Datenpakete bei geringeren Berichtsraten, womit sie eine teilweise Lösung des Problems der Datenpaket-Überlappung versprechen. Da aber das neue Weltraum-AIS-Format das Problem der Datenpaket-Kollision nicht vollständig löst, und da ein ständig wachsendes Interesse am Weltraumempfang von terrestrischen AIS-Daten weiterhin besteht, bleibt das Thema der Auflösung von überlappenden AIS-Signalen von zentraler Bedeutung für alle SAT-AIS-Anwendungen.

Contents

The pathway through this work

This PhD thesis work is written for extensive reading to those who are working within topic of space based AIS (Automatic Identification System), called SAT-AIS. This work addresses processing of overlapped message for space application – how to separate and successfully demodulate the overlapping AIS signals for a single sensor satellite antenna.

The reader is assumed to have general background in statistics, radio frequency signal processing and programming.

The first two chapters of this work present preliminaries to the AIS topic and are composed using external information sources. The further chapters hold development work which is constructed by the author of this document.

Chapter 1 of this thesis gives an introduction to AIS and space-AIS related technical details, answering questions such as what is AIS, how it works, what are AIS message contents as well as introduces to problems arising for space reception of the messages.

Chapter 2 guides the reader through the construction of the AIS transmissions signal, showing how the radio frequency signal of AIS looks like and revealing its physical and mathematical properties.

Chapter 3 describes message overlapping scenarios, which helps in a very intuitive way to see how different collision cases can be grouped for purposeful further processing. This also helps to understand which message overlap scenarios are the most challenging ones.

Chapter 4 leads the reader into some theoretical detail, showing how a new joint demodulation scheme using shape recognition can be set up as well as demonstrates its performance. The Viterbi decoder with bit decision probability analysis is also introduced here.

Similarly, the Chapter 5 provides the reader with some more technical details, revealing how a new joint demodulation scheme with frequency extraction can be constructed, analyses when it is suitable for application and shows how it performs in different signal environments.

Chapter 6 summarizes results of the two previous more theoretical chapters into a new AIS signal processing algorithm, which implements all the developed approaches into highly advanced overlapped message extraction solution. Finally, this chapter reveals the gains in number of demodulated messages for the discussed AIS demodulation methods. Based on a simulated SAT-AIS reception scenario, it shows what improvements can be awaited from implementation of the new developed demodulation schemes from Chapter 4 and Chapter 5.

Chapter 7 gives an outlook to further work needed to proceed with the implementation and further investigation of the overlapped AIS message demodulation methods.

The references are given at the end of this work, along with a glossary of acronyms, and acknowledgments for the supporters of this work.

1 Introduction

This chapter will briefly introduce general as well as technical readers to the preliminaries of this work. This chapter includes the concepts of AIS, the working principles of AIS, the space application issues of AIS and the solution approaches for improving data reception.

1.1 The concept of AIS

The AIS (Automated Identification System) was developed as a traffic collision avoidance system for ships. Its technical properties are defined in ITUR (International Telecommunication Union Regulation) document [1]. In contrast to earlier means of communication, AIS autonomously provides the basic identification information of ship traffic participants within a range of 40km. Regulation 19 in Chapter V of [2], the International Convention for SOLAS (Safety of Life at Sea) describes onboard carriage requirements for shipborne navigational systems and equipment, including AIS. From 2000 the IMO (International Maritime Organisation) has adopted the requirement for all SOLAS ships to be equipped with AIS [3].

AIS mobile stations are divided in the following two classes.

- Class A – Fully complies with IMO AIS carriage requirements and works SOTDMA (Self Organized Time Division Multiple Access) based as specified in Annex 2 of [1].
- Class B – Not necessarily in full accordance with IMO AIS carriage requirement and works SOTDMA or CSTDMA (Carrier Sensitive Time Division Multiple Access) based, as specified in Annex 2 or Annex 7 of [1] respectively.

The Class A mobile station AIS equipment should be carried by all ships to which the IMO applies its regulations based on SOLAS, and according to SOLAS Chapter V, paragraph 2.4, it is related to:

> "All ships of 300 gross tonnage and upwards engaged on international voyages and cargo ships of 500 gross tonnage and upwards not engaged on international voyages and passenger ships irrespective of size shall be fitted with an automatic identification system (AIS)" [2].

The Class B mobile station AIS equipment is intended for non-SOLAS vessels. It is not mandated by the IMO. It is usually used by leisure yachts and fishing boats.

The following three paragraphs are excerpts from the IMO website [5] and summarize the basic principles of the AIS regulation:

> "The requirement became effective for all ships by 31 December 2004.
>
> Ships fitted with AIS shall maintain AIS in operation at all times except where international agreements, rules or standards provide for the protection of navigational information.
>
> The regulation requires that AIS shall:
>
> - provide information - including the ship's identity, type, position, course, speed, navigational status and other safety-related information – automatically to appropriately equipped shore stations, other ships and aircraft;

- receive automatically such information from similarly fitted ships; monitor and track ships;
- exchange data with shore-based facilities. “[5]

Among other data, this IMO regulation requires ships to constantly broadcast their position in intervals ranging from 2 seconds to 3 minutes, depending on the ships’ cruising speed. For coast guards and safety agencies it is of very high interest to gather this information for post processing, including applications of ships’ driving trajectory analysis.

1.2 AIS sentence format

The naval AIS equipment has a specific format for how it outputs received messages to the navigation system. The following is an example of an AIS position message:

```
!AIVDM,1,1,,B,19NS7Sp02wo?HETKA2K6mUM20<L=,0*27
```

This message contains AIS data encapsulated into an AIVDM sentence according to the NMEA (National Marine Electronics Association) standard for serial communication of naval onboard electronic devices [4]. The AIS message contains AIS data, which can take one or multiple sentences to be transmitted.

To briefly introduce the NMEA format for AIVDM sentences, details of the message above can be considered. It is a single sentence message and is constituted from multiple fields, separated by the symbols “,” or “*” allowing to interpret its contents as follows.

1. “`!AIVDM`” – sentence identifier for AIS messages,
2. “`1`” – total of sentences needed to transfer the message,
3. “`1`” – current sentence number,
4. (empty) – sequential message identifier,
5. “`B`” - AIS Channel identifier,
6. “`19NS7Sp02wo?HETKA2K6mUM20<L=`” – AIS raw data in 6-bit character representation, in this case containing the received AIS position report of type 1,
7. “`0`” – number of fill bits for last character,
8. “`27`” – the CRC8 checksum, calculated over the whole AIDM message between symbols “!” and “*”.

This NMEA standard is used on the majority of AIS receivers, even for those onboard of satellites. One could directly notice promising chances for AIVDM message storage optimization, especially on reconverting 6-bit characters back to 8-bit bytes.

Throughout this work, the words “AIS messages” and “AIS reports” will be used as synonyms for two meanings: AIS raw data and AIS raw data encapsulated in NMEA sentences, depending on context. The “AIS sentence” will be used to point to NMEA sentences. “AIS data” will be referenced for any

form of AIS related information, both in raw data format as well as encapsulated or decoded and gathered in databases.

1.3 Properties of the AIS communication channel

This section summarizes AIS channel specifications, which are relevant for AIS channel simulation. For a more detailed description please refer to the source document [1].

The terrestrial AIS communication takes place in two channels:

- 161.975MHz, called Channel A, also AIS1 or Channel 87B and
- 162.025MHz, called Channel B, also AIS2 or Channel 88B.

The regular and maximum transmission power is 12.5 Watts. The modulation used is GMSK (Gaussian Minimum Shift Keying) with 9600 bps (bits per second) bit rate and 0.4 BT (Bandwidth Time). The channel access scheme is TDMA (Time Division Multiple Access) based. It means that time is divided into slots of 256bits, thus one minute, called an AIS-Frame, contains 2250 transmission time slots, called AIS-Slots. There are 27 types of messages which can be used to transmit information. Most of them are one slot long and are transmitted by ships. Other than ships, AIS messages can be transmitted by base stations, search and rescue vehicles, beacons or buoyage systems from AtoN (Aids to Navigation) systems.

Please refer within document [1] to the Annex 8, Table 43 “Message Summary” for further details on message types, their lengths, contents and corresponding channel access schemes, as well as transmitting instances.

1.4 AIS message types

The position report messages, corresponding to message types 1, 2 and 3, further denoted as Message 1, Message 2 and Message 3 respectively, are of highest interest to be received, as they contain ship position data. A setup of 5 AIS receiving stations, installed by DLR in northern Germany in 2012, have obtained evidence that the three position report message types together make up about 90% of the total data communicated within the AIS channels near a shore. With Message 4, Base Station Report, which is transmitted by AIS shore stations, it sums to 95%. Adding Message 5, Voyage Information Report, transmitted by ships in reporting intervals of 6 minutes, this sum increases to 98% of total channel traffic. Therefore, within this work, only these five message types will be considered for analyses and simulations. The reporting percentages for the message types are depicted in Fig. 1.

The message types 1, 2 and 3 possess almost similar contents. They do differ by the application and access scheme and are used as follows.

- Message 1 transmission is SOTDMA (Self Organized TDMA), RATDMA (Random Access TDMA) or ITDMA (Incremental TDMA) based, scheduled by the ship on its own.
- Message 2 transmission is SOTDMA based, when the slot is assigned from a nearby ground station.

- Message 3 transmission is ITDMA based and is used when a ship is announcing its new reporting slots during initial communication start-up phase or a need for extra slots due to increased reporting rate, usually related to a change in speed or cruising direction.

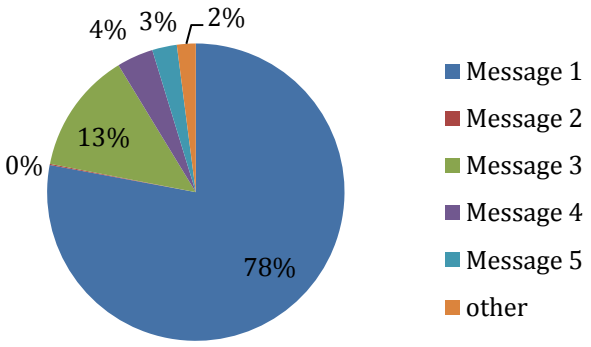


Fig. 1. Observed AIS message type statistics by five AIS receivers in coastal zone of German Bight.

The next table, taken from [1], page 4, summarizes position reporting intervals of the A category ships, depending on their traveling speed.

Ship's dynamic conditions	*Nominal reporting interval*
Ship at anchor or moored and not moving faster than 3 knots	3 min
Ship at anchor or moored and moving faster than 3 knots	10 s
Ship 0-14 knots	10 s
Ship 0-14 knots and changing course	3 1/3 s
Ship 14-23 knots	6 s
Ship 14-23 knots and changing course	2 s
Ship >23 knots	2 s
Ship >23 knots and changing course	2 s

Table 1. Class A shipborne mobile equipment reporting intervals.

The above table shows that the maximum reporting rate of a ship is 30 times per minute, whilst for an ordinary cruising ship the report interval is expected to be 6 seconds during static travel and 2 seconds when changing course.

As noted before, there are two dedicated channels (Channel A and Channel B) for transmitting AIS data. They are both used one after another. The transmission of the position report is a good example – each sequential message is transmitted on the opposite channel to the previous one. Meaning that if a station was only listening to channel A for a Class A ship driving at a speed of 20knots, it would receive a message every 12 seconds.

For more detailed information on AIS message types and their transmission properties, please refer to [1].

The use of SOTDMA protocol implies that the ships self-organize into groups, which communicate within an approximately 40km radius neighbourhood. The communication distance strongly depends on an antenna type and its installed location on the ship, especially the height of the communication

point on the spar. In normal operation mode the transmission power is 12.5 Watts. With Message 22 a ground station can command the levels of surrounding mobile stations to be set to “low level” which is a 1 Watt power setting. This could be used in local base station regions in order to reduce the size of the TDMA communication cell in high density traffic zones, like harbours. The low level mode is also used for other applications, for example in burst transmissions of AIS-SART (AIS Search And Rescue Transponders).

Within this work the term AIS-Cell will be used with meaning the area of the surrounding SOTDMA cell, where a sender ship transmits and receives AIS signals in overlap-free communication. The term Satellite-AIS-Cell, will be used with meaning the area of a satellite antenna footprint, where overlapping of signals can occur due to observing multiple SOTDMA cells. This is an important distinction because different SOTDMA cells may use the same reporting slots.

1.5 The packet structure of an AIS message

This section will introduce the reader to the AIS packet structure, as far as it is necessary to understand the basics of the packet transmission, identification at the receiver side and an error checking. The information source is the ITUR document [1].

As was previously mentioned, the time in AIS communication is organized into frames, which are each 60 seconds long. Since a single slot is 256 bits long and bit transmission speed is 9600bps, this means that one frame contains 2250 slots. A single message can take one to five slots, but all of the AIS transmissions share the same packet structure. The following two images are taken from the ITUR technical specification document to show the magnitude of RF transmission power and the packet structure of a single AIS packet.

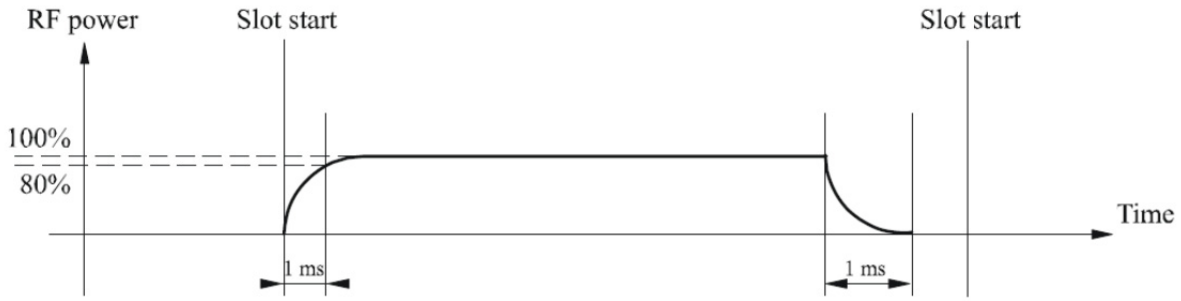


Fig. 2. Transmission power magnitude of an AIS packet [1].

Ramp up	*Training sequence*	*Start flag*	*Data*	*FCS*	*End flag*	*Buffer*

Fig. 3. The structure of an AIS packet [1].

The first image shows that at the beginning of transmission within a timing slot a dedicated time is available to switch on the transmitter – within the first millisecond the transmitter should reach at least 80% of transmission power. The switch off timing is also 1ms long. In case the packet is exactly 256bits long, this switch off timing would overlap with the switch on timing of the next message, which is tolerable for both packets.

For an example of the packet structure, an AIS Message type 1 packet, which is 1 slot long, would possess the lengths and functions of its fields as shown in Table 2 below.

Packet data	Length	Functionality
Ramp up	8 bits	Switch on time for AIS transceiver, 1ms long.
Training sequence	24 bits	Necessary for synchronization
Start flag	8 bits	The start flag in accordance with High-Level Data Link Control (HDLC) protocol, '7E' in Hex or '01111110' in binary
Data	168 bits	Data for the packet
FCS	16 bits	In accordance with the HDLC, CRC-16-CCITT used over all Data bits
End flag	8 bits	The same as start flag
Buffer	24 bits	Bit stuffing, distance delays, repeater delay and jitter
Total	256 bits	

Table 2. Packet structure of an AIS Message 1 [1]. The size of the Data field will vary for different message types.

Fig. 3 together with Table 2 allows the following important points to be summarized:

- Every AIS packet has a start and an end flag, also called the HDLC flag, which serve for a packet data field identification;
- Every AIS packet has a 16 bit checksum.

The HDLC flag for AIS packets is constituted of a unique bit sequence '01111110'. To ensure that nowhere in the data there will never be 6 consecutive '1's present, a bit stuffing of the data is performed prior to its modulation. The bit stuffing for AIS means putting a '0' bit after every 5 consecutive '1' bits within data field. This ensures that the packet will not contain 6 consecutive '1' bits nowhere apart of the HDLC flags, thus allowing the flags to serve for a unique packet start and finish identification.

After encapsulating the AIS data, NRZI (Non Return To Zero Inverter) encoding is used to ensure better clock synchronization for signal receivers. NRZI gives a change of the signal level when a '0' bit is encountered in the bit stream. An example of this procedure will be shown in Section 2.1.1.

The use of a buffer field, depicted in Fig. 3, is made available for various needs when packet length exceeds beyond the allocated data field size. For Message 1 case, theoretically the 168 bits could be enlarged by 33 additional bits due to bit stuffing. However, as stated within the ITUR document, it is not possible for the size of the bit stuffing to exceed 4 bits in real world situations. The following is an excerpt from the ITUR document about buffer functions:

- "bit stuffing: 4 bits (normally, for all messages except safety related messages and binary messages)
- distance delay: 12 bits
- repeater delay: 2 bits
- synchronization jitter: 6 bits" [1]

To justify the choice of size for distance delay bits, the ITUR document states “A buffer value of 12 bits is reserved for distance delay. This is equivalent to 202.16 nm (nautical miles). This distance delay provides protection for a propagation range of over 100 nm” [1]. The distance delay properties for a satellite application will be inspected in Chapter 1.8.

1.6 The concept of AIS reception from space

As already mentioned before, the AIS was developed for terrestrial applications shortly before the year 2000. Although the AIS technology is fulfilling its initial expectations well, the evolving time and space technologies have given a rise to new opportunities, interests and challenges. In 2008 ORBCOMM was the first to launch a satellite constellation, called Orbcomm-QL, with AIS receivers as a secondary payload on board **Error! Reference source not found.**. It was followed by LuxSpace in 2009 [7], and many universities, research institutes and satellite companies since then. It has become a very popular payload for small satellites, due to the fact that even naval commercial AIS equipment with a small antenna can perform data reception in space [8]. Already, a commercial company, exactEarth, is offering AIS data captured from space to safety and maritime organisations [9].

The interest in global maritime surveillance is growing with each new launch of AIS equipped satellites. The satellite based AIS reception does offer an opportunity to meet the demand of global AIS data availability. It’s most important advantages over terrestrial AIS receiver system are global coverage including deep seas and oceans, which are out of reception range of static antennas installed in coastal zones. Receiving data from orbit also grants royalty-free access to AIS data and offers the choice of multiple data providers.

Receiving terrestrial AIS transmissions from space offers new applications for an existing technology that it was not initially designed for. Terrestrial AIS was developed to be used mainly by ships to identify local surrounding transport vehicles at sea and provide safety services based on local data availability, such as Collision Avoidance, Aids to Navigation, Maritime Security, Safety and Rescue etc.; whereas the new space-application, built on the existing AIS technical base, offers global monitoring of ship activity across seas and oceans, lonely islands, fishing zones etc. Many national agencies worldwide have shown significant interest in global AIS data availability for building various safety applications and services based on the global data.

This interest in global AIS data availability has given a rise to the development of a space-application dedicated technical add-on to serve the needs for better AIS reception from satellites. Since the year 2010, the topic of AIS data reception from space is evolving in two pathways:

- The first elaborates on improving the satellite reception capability of the usual AIS messages, which are used in ship-to-ship communication on the ground. It will be further referred as “terrestrial-AIS” or “tAIS” technology.

- The second pathway works on defining an additional standard, to be adopted for space reception of ship position data. It will be further referred to as “space-AIS” or “sAIS” technology. This pathway includes introduction of two new additional AIS transmission channels, dedicated solely space use. This pathway is described in more detail in Section 1.7.

The main working subject of the terrestrial-AIS pathway is advancing signal processing capabilities, as well as technical antenna constellations, to better succeed in receiving separated AIS messages and extracting AIS messages out of overlapped signals.

Contrastingly, research on space-AIS is working to find the most optimal settings for data contents and reporting intervals on the ship side, as well as the most appropriate antenna and processing algorithms on the satellite side to build efficient AIS data gathering solution for space applications.

The sub topic of radio frequency data processing for signal extraction from overlapped AIS messages is a working direction for both approaches, since neither in terrestrial-AIS nor in space-AIS there are means to ensure that the transmissions of ships occur in different time slots for such large fields of view as observable from a satellite. Improvements in overlapped signal extraction would provide a direct benefit for both the terrestrial-AIS and space-AIS signal reception from satellite. This means that the future AIS receivers in orbit will need to be equipped with the ability to receive signals from both AIS standards, the terrestrial-AIS and the space-AIS. The naming conventions for this dual application of AIS receiving on board of a satellite are “satellite-AIS”, “SAT-AIS” and “S-AIS” and involve the passive listening of all of the four AIS channels on a satellite. The mentioned terms are also often used in literature to reference the technical solution involving terrestrial-AIS receiver adopted for use in space.

1.7 Space-AIS standard

The work on a new Space-AIS standard to improve AIS reception capabilities from space is still under way at both technical and political levels. The ITU-R M-1371-4, issued in 2010, together with ITU-R M.1084-5 issued in 2012 define the planned technical characteristics of the space-AIS technology as follows:

- A new message type, Message 27 is introduced and named Long Range AIS Message. It is to be transmitted from ships to satellite at a 3 minute interval. It is described in Annex 4, Section 3 of the document [1].
- Two additional frequencies are allocated for ships to transmit exclusively the Message 27 AIS reports. These new channels are called AIS3 and AIS4 or according to WRC-12 named channels 75 and 76 at frequencies 156.775MHz and 156.825MHz respectively.

The following table shows the field structure of the AIS Message 27:

Slot composition	Bits	Notes
Ramp up	8	Standard
Training sequence	24	Standard
Start flag	8	Standard
Data field	96	Data field is 168 bits for other single-slot AIS messages. This field is shortened by 72 bits to support the long-range receiving
CRC	16	Standard
End flag	8	Standard
Long-range AIS receiving system buffer	96	Bit stuffing = 4 bits, Synch jitter (mobile station) = 3 bits, Synch jitter (mobile/satellite) = 1 bit, Propagation time delay difference = 87 bits, Spare = 1 bit
Total	256	Standard (NOTE – Only 160 bits are used in the 17 ms transmission)

Table 3. Packet bit structure for long-range AIS message reception [1].

And the following table summarizes the contents of the data field to be included in Message 27 transmission.

Parameter	**Number of bits**	**Description**
Message ID	6	Identifier for this message; always 27
Repeat indicator	2	Always 3
User ID	30	MMSI number
Position accuracy	1	As defined for Message 1
RAIM flag	1	As defined for Message 1
Navigational status	4	As defined for Message 1
Longitude	18	Longitude in 1/10 min (±180º, East = positive, West = negative
Latitude	17	Latitude in 1/10 min (±90º, North = positive, South = negative
SOG	6	Knots (0-62); 63 = not available = default
COG	9	Degrees (0-359); 511 = not available = default
Status of current GNSS position	1	0 = Position is the current GNSS position; 1 = Reported position is not the current GNSS position = default
Spare	1	Set to zero, to preserve byte boundaries
Total number of bits	**96**	

Table 4. Data field contents for AIS Message 27 [1].

This paragraph summarizes some of the important properties of Message 27 content. The structure of Message 27 resembles to the structure of position report Message 1, but with lower geographical position resolution as well as excluded SOTDMA channel status data. This leads to having less data bits per message – just 96 bits in Message 27 compared to 168 bits in Message 1, allowing to have as large as a 96bit buffer for each time slot. Channel access scheme is RATDMA, the modulation and bit rates are preserved as for terrestrial-AIS application. The 87 bit propagation delay time ensures avoidance of consecutive message overlaps for range differences up to 1467nm, which equals to

2717km. In the case of a satellite operating at 560km altitude, it would obtain a complete earth horizon view with a 134 degree conical antenna opening. In the more common 700km altitude, it would achieve this with a 128 degree antenna opening. Assuming a basic satellite antenna possessing a field of view of 90 degrees, the available propagation delay would serve satellites up to approximately 1600km altitudes. Thus the embedded propagation delay suits very well the satellite application needs.

Another important property of Message 27 is that its reporting can be faded out in the proximity of coastal zones. Since the Revision 4 of the document [1], released in 2010, the Ground Station Report Message 4 contains a switching bit, which can command the AIS transmitter of a ship to switch off space-AIS transmission. This allows for a greater cooperative opportunity between terrestrial and space AIS networks – if ships switch off space-AIS transmission near densely cruised coastal zones, it would enable satellite's to receive less filled space-AIS channel. This would help to diminish unnecessary interferences from the ships which are already monitored by shore stations, including the densely occupied areas around harbours. The terrestrial network would then gather data from ships surrounding coastal zones while the satellites from space would gather AIS reports from more deep sea cruising vessels, building a mutually beneficial ship monitoring network.

It should be noted that the space-AIS is not the only space dedicated technology for ships to report their position globally. With the resolution MSC.210(81) [10], IMO "Performance Standards and Functional Requirements for the Long-Range Identification and Tracking of Ships", the IMO has required that all SOLAS vessels be LRIT (Long-Range Identification and Tracking) equipped by 2009 or on the first radio installation survey after July 2009 [11], [12]. The LRIT is a technology which reports a ship's position and some related data once every 30 minutes through a continuous Inmarsat satellite link. This ensures safe and robust ship tracking around the entire globe. However, the data of the LRIT system is under a strongly regulated distribution control, prohibiting some data analysis applications which royalty-free AIS can easily offer. Additionally, for some data analysis applications the LRIT reporting rate is too low. Lastly, it is a commercial service – there is a certain cost per message transmitted through the Inmarsat data network, paid for by ship owners, while the gathering of AIS data in space is free of any charges.

As LRIT technology has a very different technical realization and purpose than the AIS technology, the LRIT will not be discussed here, but and an interested reader for further investigations is advised to consult the documents [10], [11], [12].

1.8 The difficulties for receiving AIS from space

An overview of the difficulties related to the Satellite-AIS is well described in many publications [14], [15], [16] and [17], especially in [13]. Here we will briefly review the issues in order to obtain a common understanding for the needs of this thesis work.

Doppler shift – This is the first obvious challenge to be noted, which comes into play due to the high relative speed of a satellite relative to the ship. A usual satellite speed at 700km altitude is approximately 7.5km/s. Whereas a high speed cargo ship on the open sea can cruise at 20kn (37km/h or equally 0.01km/s), which is a fraction of the satellite's speed, and therefore can be neglected. The effect of Doppler shift could be a disturbance in applications of wide angle antenna and no adaptive frequency filter presence, as there are expected to be carrying frequency shifts for the received

signal. On the other hand, the Doppler shift property comes in handily, when there are advanced frequency filters available for receiver's RF-frontend or signal processing, to help in separating different signals. Since the Doppler shift for a 700km LEO satellite varies within the range of ±3.6 kHz, it can be used selectively to separate signals coming from different relative speed areas. In the case of a 60° opening angle antenna, the Doppler shift frequency range is ±2 kHz and for a 38° antenna the range is ±1.3 kHz. Therefore, a satellite-AIS receiver for high performance should be equipped with algorithms, and or a RF-frontend to take advantage of AIS signals of different frequencies to distinguish their sources and minimize message collisions. Yet another option is to use a narrow angle antenna, to minimize the probability of message collisions.

In Fig. 4 is shown a visual sketch of the maximum Doppler shift case, occurring when the satellite arises just above the horizon with respect to the ship, and Fig. 5 reveals a mesh plot of the Doppler shift magnitude for a wide angle satellite footprint.

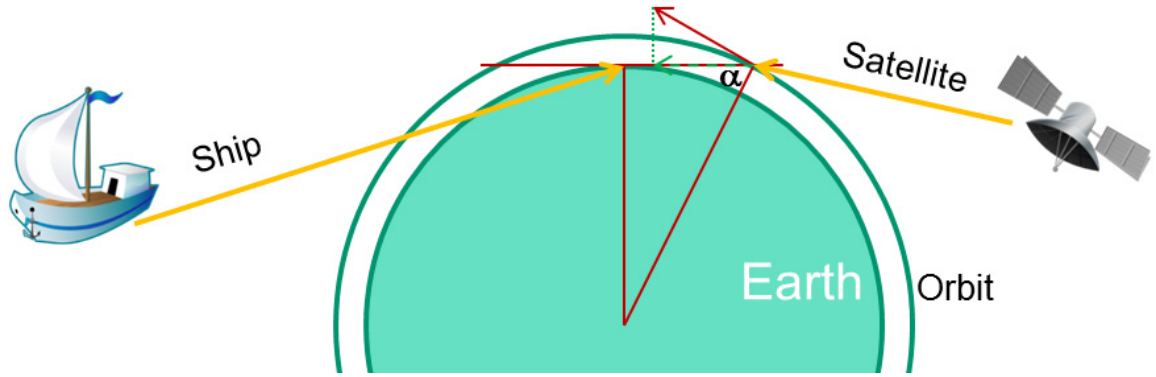


Fig. 4. A ship reception scenario from a satellite in the case of maximum Doppler shift. Alpha is the half-cone opening angle of the satellite's antenna.

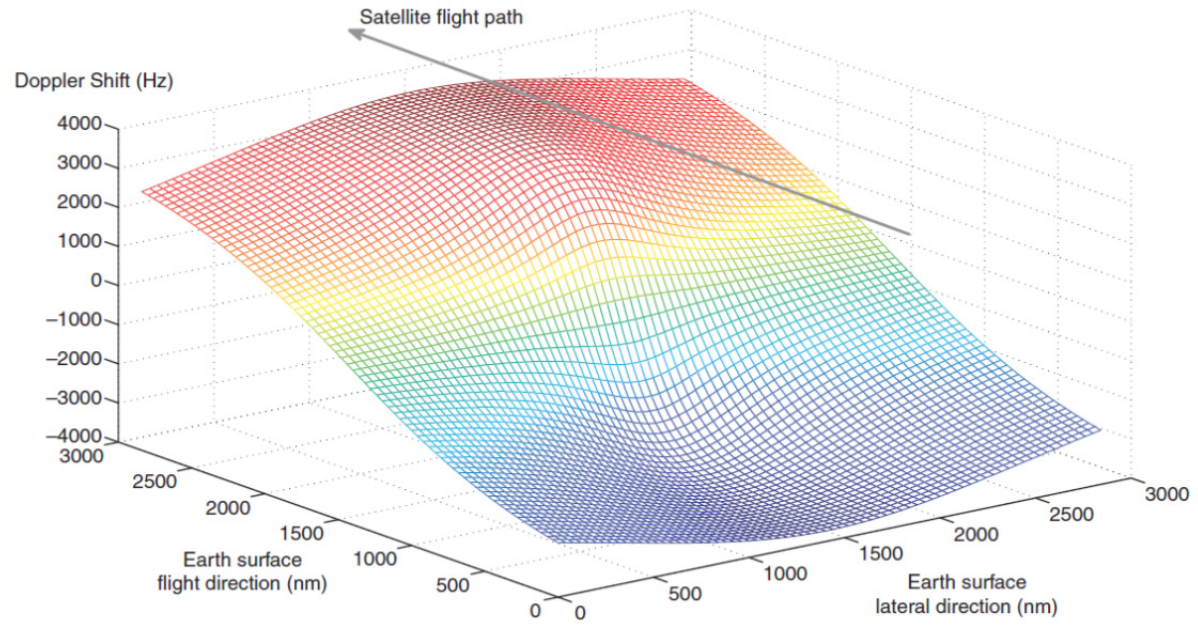


Fig. 5. A typical distribution of the absolute Doppler shift in the presence of a 600km altitude LEO satellite and uniform ship distribution within the coverage [13].

Overlapping messages – This is truly the central issue for building reliable space-AIS technology. Fig. 9 in the next section schematically shows how packet collisions occur. The origin of this problem is caused by the fact that the SOTDMA cells of ships are about 80km in diameter, depending on spar and antenna characteristics. At the same time the FOV (Field of View) of a satellite orbiting at an altitude of 700km and possessing an antenna with a 60° opening angle has a footprint of 824km surface diameter. This is effectively an increase of 10 times the diameter of the signal reception area. For yet another example, a 38° opening angle antenna would encompass an area with 485km surface diameter, which is approximately 6 times wider than that of the AIS SOTDMA cell of ships. These aspects point to the problem that a satellite antenna will see multiple SOTDMA cells, which inevitably leads to the presence of overlapped signals. Although signals do not overlap for terrestrial receivers, they do within the much broader satellite antenna view. The next section will dive more into this topic.

Faraday rotation – This is introduced due to signals passing through the Earth's ionosphere. For a general case of randomly polarized signals entering an antenna, a 3dB power loss should be expected [13], [17].

Path attenuation – There are relatively long signal travel paths from ships to a satellite in orbit. These are much longer than initially planned for ship to ship or ship to base station applications in terrestrial communication. The following formula can be used to calculate signal's attenuation for its travel distance.

$$L = 20 \cdot \log_{10}\left(\frac{\lambda}{4\pi d}\right), \tag{1}$$

where L is path loss in dB, λ is wavelength in meters and d is distance in meters [18]. Since ships are transmitting AIS messages with 12.5 Watt (11dB) power, usually satellites at altitudes below 800km are used for AIS signal reception. For example, a ship's signal received by a satellite in 700km orbit directly above the ship (the shortest possible path), would suffer 133.4 dB of signal attenuation. This leads to a maximum -123dB arriving signal power for an AIS message transmitted with 12.5 Watt power. For a 38° FOV antenna at the outer most angle this would lead to 134dB path attenuation, and for a 45° FOV antenna 137dB attenuation. However, the path attenuation is not the only factor influencing the strength of the arriving signal on satellite. A precise calculation of the AIS signal reception power would require knowing additional information about a ship's antenna type and its mounting point on the specific ship, the ship's geometry and deck coverage material, as well as the signal receiving angle, antenna and RF-frontend characteristics of the signal reception hardware on the satellite. For the low-power setting, when AIS messages are transmitted at a 1 Watt power level, the received signal at orbit would be correspondingly lower. This leads to concerns about the strength of the received AIS signals on a satellite, which can be addressed by a dedicated high gain antenna design, for example like it was built for AISat1 mission, or sensitive, low noise electronics equipment.

According to [14], assuming the transmission antenna on a ship is a dipole with a toroidal radiation pattern and using the formula from (1) for the estimation of the signal path attenuation, the received signal power curve as shown in Fig. 6 can be calculated.

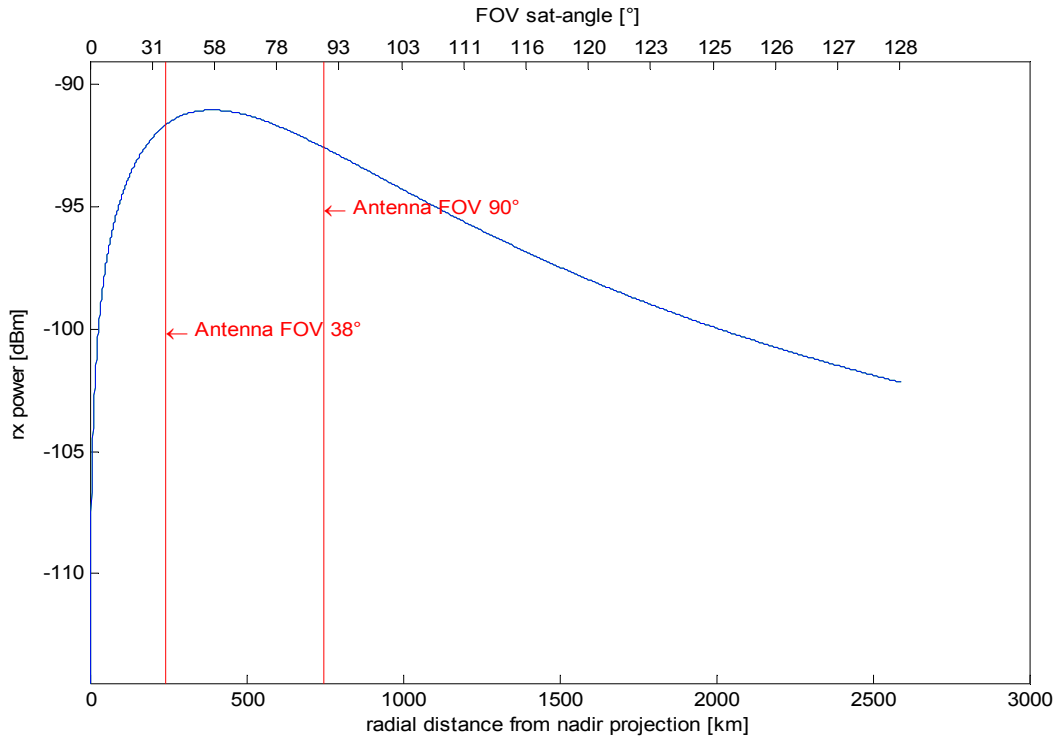


Fig. 6. Signal power versus ground range at 700km for 3dB toroid antenna gain at transmitter and receiver.

As can be seen in Fig. 6, the signal coming from exactly the nadir direction, which is directly below the satellite with the shortest signal path from ground, has a minimal signal strength. This is to be explained by the torroidal shape radiation pattern of the ship's dipole antenna, which has its strenght in horizontal radiation direction but not in upwards vertical direction.

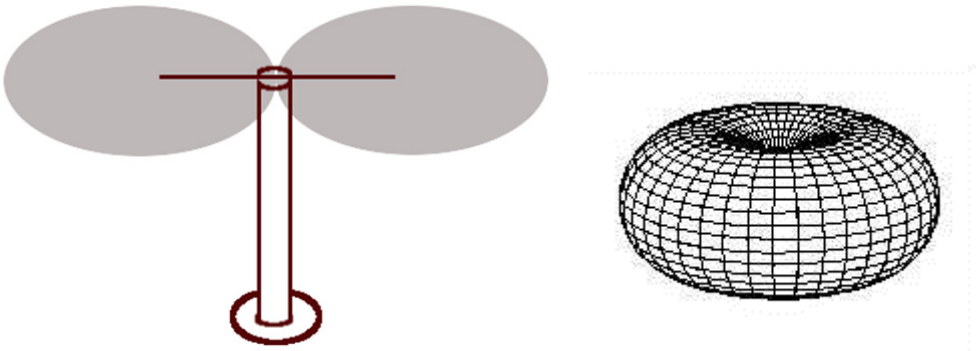


Fig. 7. a) Dipole antenna and b) its radiation pattern of toroidal shape [45].

However, this signal reception approach is too simplified to use it as a solid reference. For the model with the assumption of ships possessing a dipole antenna, the toroid radiation model does not include reflections from the water's surface or the ship's deck. An explicitly modelled transmission pattern around a ship with the "CST Antenna Magus" software is shown below in Fig. 8. If the satellite antenna possesses a narrow angle, then by design it usually has a better gain for the central objects, which to a certain extent will compensate the disadvantageous power distribution of dipole

antennas used on ships. Since a comprehensive modelling for an antenna radiation is outside the scope of this dissertation, this topic will not be covered within this work.

As multiple AIS cube sat missions have shown, including Norwegian AISSat-1 and AAUSAT3, a basic monopole or dipole antenna can be suffiecient to receive AIS signals at satellite to some extent.

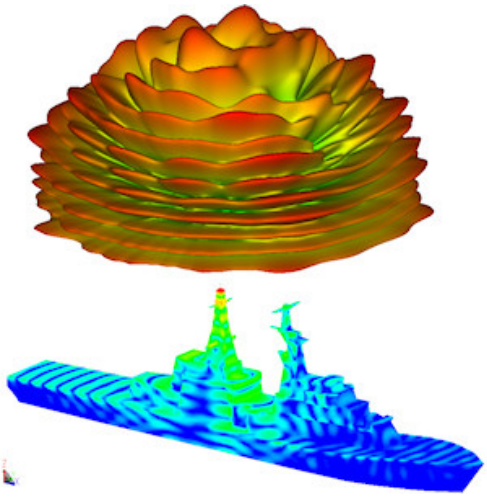


Fig. 8. A modeled transmission power of a ship antenna. Credit: Martin Drobzyk, DLR Bremen.

Atmospheric attenuation – Due to the fact that AIS systems use the VHF frequency range below 1GHz, the atmospheric attenuation is very small – approximately 0.05dB [18], [41], [42]. Therefore, this issue can be neglected.

Signal disturbances – Signal disturbances are to be expected in different world regions, where the AIS ship frequencies are not strictly restricted for AIS applications, e.g. inland regions. Although the AIS frequency use for specific purposes inlands does not disturb terrestrial AIS receivers of ships in seas, some radar signal within continental zones using the AIS frequencies may prove to be an obstacle to receiving ship signals from a satellite with a relatively large FOV. An example of such a case is the Russian Antiballistic Missiles Radars around Moscow, which are known to have an impact on satellite-AIS reception over the Baltic Sea [19].

1.9 The issue of overlapped messages

The next figure, Fig. 9, visually shows the different cell sizes of SOTDMA and satellite-FOV and the resulting signal overlaps as the consequences of that.

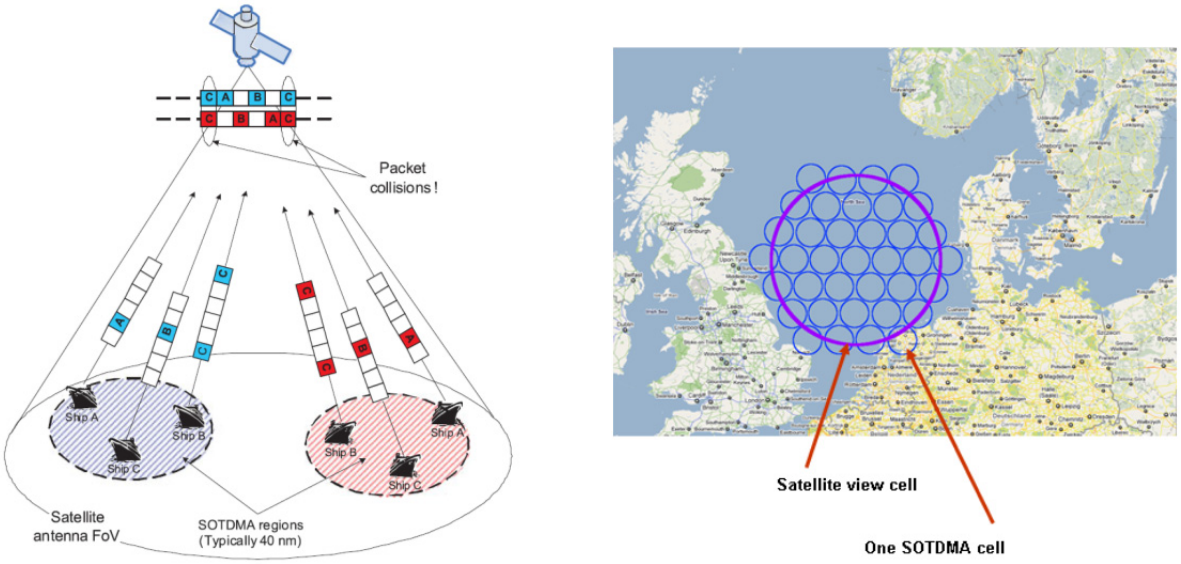


Fig. 9. a) Scenario of colliding messages as two organized SOTDMA cells viewed from satellite with bigger reception area [13]. b) Satellite FOV and SOTDMA comparative sizes. [Source of the map is Google Maps.]

As shown in Fig. 9 a), the ship C from the blue SOTDMA cell doesn't hear the ship B from red SOTDMA cell, thus considers the selected time slots free and thus transmits. On the other hand, a satellite in orbit sees both of the SOTDMA cells and thus receives overlapped signals within the same time slots.

Assuming that the differentiation frequency due to Doppler shift between red and blue SOTDMA cells is too low to separate the signals using frequency filters and assuming the signals' amplitudes are of approximately equal levels, a problematic situation arises – how does one extract the overlapped signals?

There can be distinguished two sources for message collisions:

1. Collision occurrences due to multiple transmissions within the same time slot, as discussed in the previous paragraph;

2. Collision occurrences due to path delay – as described in Chapter 1.5, the terrestrial-AIS message buffer incorporates an extra 12 bits for supporting path delays for up to 375km relative distances. Larger distance differences, like the ones present in the satellite applications, can lead to a rise in message overlaps due to different propagation times. An example of a messages overlap scenario due different propagation times is shown in Fig. 10 below. The packet from cell 2 is assumed to come from a far distant ship, thus arriving very late at the receiver and colliding with packets from cell 1 and cell 2.

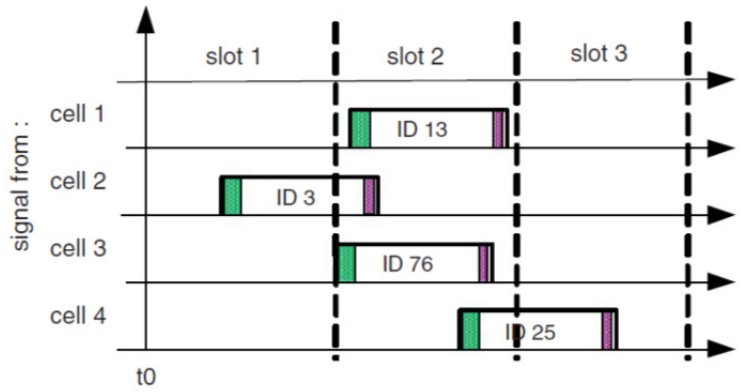


Fig. 10. The burst overlap issue as a consequence of the large satellite path delays [13].

Whilst the first type of collision source is a fundamental issue of a communication channel, multiple workarounds can be applied for second collision source. It should be noted, that a 375km difference of path delays arise only in the case the antenna opening angle is greater than 92^0 for a satellite at 700km altitude or higher. Essentially for Message 27, the message type foreseen for use in space-AIS, the message collision due to relatively large path delays is truly eliminated by introducing additional path delay time within the message buffer, as described in Section 1.7.

An important question is, how often and where such situations occur. According to [15], [20], the AIS marine traffic can be divided into 3 categories:

1. LTZ (Low Traffic Zones):
    1.1. South hemisphere
    1.2. North Pacific
    1.3. North-West Atlantic
2. MTZ (Medium Traffic Zones) – intermediate zones
3. HTZ (High Traffic Zones):
    3.1. Europe, i.e. North Sea, Mediterranean, Channel Coast
    3.2. Caribbean, Gulf of Mexico
    3.3. East Asia, i.e. Chinese Sea, Yellow Sea

The next two pictures, Fig. 11 and Fig. 12, show global world maps with ship densities.

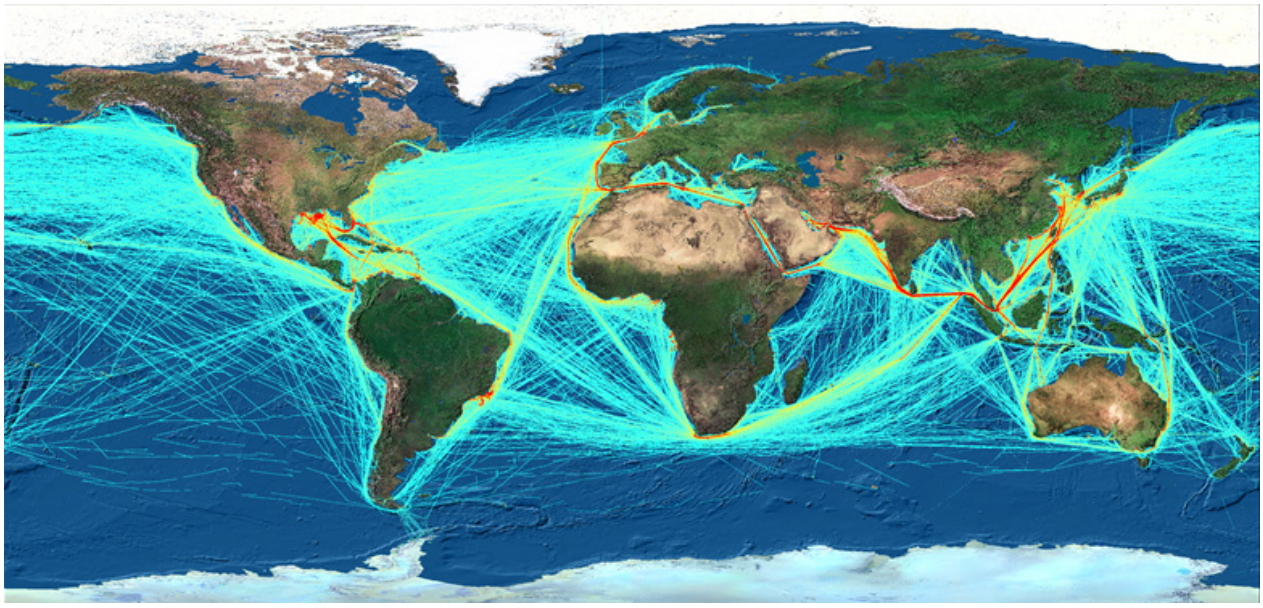


Fig. 11. Global Ship density map, January 2013. Source: United States Coastal Guard [21].

The definitions of the colours in the map of Fig. 11 are as follows [21].

- Red Cells: The monthly plot totaled over 50 vessels.
- Orange Cells: The monthly plot totaled between 15 and 49 vessels.
- Green Cells: The monthly plot totaled between 5 and 14 vessels.
- Blue Cells: The monthly plot totaled 4 or fewer vessels.
- Empty Cells: No vessels.

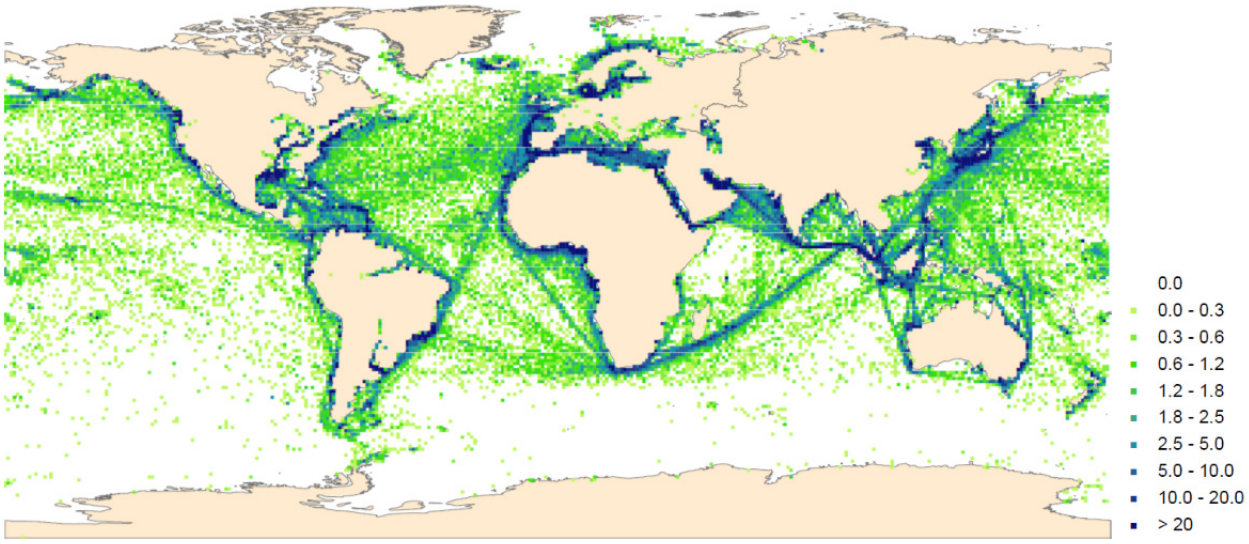


Fig. 12. Average ship density per grid cell [22].

Another study by OHB-Systems for simulating satellite reception of AIS data from multiple sources yielded the following ship density map [15].

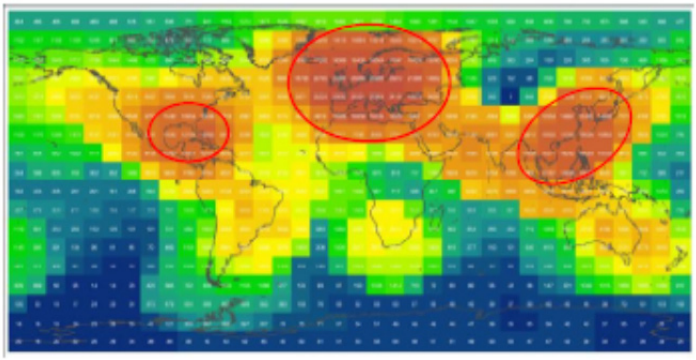


Fig. 13. Plot of maritime traffic distribution. Cells are of satellite narrow antenna footprint size.

On the world map of Fig. 13 one can observe the three density categories as previously denoted. The bluish areas represent low ship traffic, green and yellow medium traffic, and the red regions, especially the ones identified with the three red circles, represent high traffic areas.

Due to the extensive number of ships present in HTZ areas, satellites flying over those regions experience AIS channel overload, causing the receiver to fail demodulation, due to a high AIS message interference [13], [14], [15], [23], [24]. This gives rise to the question of what can be done to improve demodulation capabilities over these high density ship areas. The following approaches are usually taken to cope with the overlapping message issue:

- Interference Cancellation – demodulating a received signal, subtracting the successfully demodulated result from the initial signal and then applying the demodulation algorithm again. An advanced implementation of this approach by ESA is called Remod-Demod method [16], which is also referenced in literature as CSC (Consecutive Signal Cancellation) [17].

- Bandwidth separation in sub-zones – utilizing Doppler shift presence, the signal bandwidth can be separated into three demodulation frequency zones: negative Doppler shift extreme, positive Doppler shift extreme and the central frequency. This process is called a Zonal-Demodulator, and is implemented by the ESA (European Space Agency) [16].

- Phased array antennas – using an antenna grid for signal reception. This gives an ability to isolate signals coming from different angles [25], [26].

- Utilize varied Faraday rotation angles – suppress certain signals coming from different angles with respect to the antenna [13].

- Reduce AIS channel load – as the Space-AIS standard is working toward to, introducing additional channels with lower reporting rate, optimizing the AIS transmission intervals for best performance according to ships' positions and improving coastal zone infrastructure for Space-AIS collaborative support [27].

While the first two methods for solving overlapped terrestrial-AIS packets are software based, the second two are hardware based and need additional technical equipment. The last approach is political and can be approached only with an international collaboration.

Since this dissertation is dedicated to development of algorithms for signal processing software of single sensor antennas, the hardware solution approaches will not be considered in this work.

Chapter 4 and 5 will review algorithms developed by the author of this paper to enhance demodulation capabilities of overlapping signals, as received by a single beam antenna, focusing on applications over high density traffic zones.

2 Modelling AIS waveforms

This chapter will show how an overlapped AIS signal can be constructed. It will start with showing how the bits contained in an AIS message are formed, then how GMSK (Gaussian Minimum Shift Keying) modulation is applied and demonstrating how a single AIS message signal look like. After that, it will be shown how overlapping AIS messages can be composed from two single messages.

2.1 Building a basic AIS waveform

2.1.1 Modelling AIS message contents

This section addresses more closely the technical details of how a single AIS message is composed with step by step carrying out an example for building AIS Message 1. While demonstrating the process, there will be data extracted from a NMEA sentence with AIVDM message, encapsulated in AIS packet and prepared for modulation. In each step there will be shown the MATLAB code, to assist understanding and to allow for reprogramming by interested readers. These MATLAB procedures were developed in accordance with the ITU-R standard [1].

To start modelling, let us take a ready AIVDM message with the MATLAB code given as:

```
%% Extract data from AIVDM message
aismsg = '!AIVDM,1,1,,B,19NS7Sp02wo?HETKA2K6mUM20<L=,0*27'
di = find(aismsg==',',2,'last');
msgdata = aismsg(di(1)+1:di(2)-1);
```

The variable `aismsg` shows how a usual AIVDM message of NMEA format looks, with the AIS data part underlined. The operation above will crop the AIS data from the provided NMEA message to the data part. For interpreting the encapsulated data, please refer to the introduction chapter, Section 1.2. To interpret the AIS data, please refer to Section 1.5.

To obtain AIS data bits out of the AIS data characters of NMEA format, the data symbols should be converted from 6-bit characters, also called character-bytes, to 8-bit bytes as described in [4]. While the 6-bit format serves as a convenient way for presenting data using 64 symbols, for data interpretation the incorporated bit sequence should be extracted. This is done by subtracting 48 from the character's Unicode ASCII integer code, if it is smaller than 88 [4]. Otherwise 56 should be subtracted. The following example demonstrates the procedure for the first 4 characters, giving the first 24bits or 3 bytes.

`unicode2nativ('1')` - gives 49, subtracting 48 gives 1, converted to binary gives '000001'
`unicode2nativ('9')` - gives 57, subtracting 48 gives 9, converted to binary gives '001001'
`unicode2nativ('N')` - gives 78, subtracting 48 gives 30, converted to binary gives '011110'
`unicode2nativ('S')` - gives 83, subtracting 48 gives 35, converted to binary gives '100011'

A code which accomplishes the task can be written as follows.

```
%% Convert msg data to six-byte stream
msgd_b = unicode2native(msgdata); %contains values [48..119]
msgd_bs = [];
for i=1:length(msgd_b)
    if (msgd_b(i)<40+48)
        msgd_b(i) = msgd_b(i)-48;
```

```
    else
        msgd_b(i) = msgd_b(i)-56;
    end
    msgd_bs = [msgd_bs,dec2bin(msgd_b(i),6)];
end
msgb = str2num(msgd_bs')';
```

Applying the NMEA character decoding to the message above, the following bit array is obtained.

```
000001001001011110100011000111100011111000000000000010111111110111001
111011000010101100100011011010001000010011011000110110101100101011101
000010000000001100011100001101
```

In the bit sequence above it can be observed that the first 6 bits, coloured in red, are composing "1" i.e. telling the type of the message is 1. The 8th to 38th bit, coloured in blue, compose number "636012431" – the MMSI number of the ship which transmitted the message.

A bit sequence constructed from character-bits can describe only bit sequences of length which is multitude of 6. In the NMEA sentence, after the data part follows integer number, which denotes how many bits from the last character-byte are relevant. These are called fill-bits and the last character-byte has to be cut up to number of the used bits. Since number of fill-bits within last character-byte for Message 1 is always 0, in this particular case the last six bits should be left unchanged [1].

In case of AIS stations composing data themselves, as it is the case for ship transmitters, the bit sequence derived above is the starting point for the AIS message encapsulation. The reverse engineering as demonstrated above with extracting bits from NMEA formatted AIS data was shown here just for illustration of processing a real AIS message.

The following steps are to be done to encapsulate AIS data bits prior to GMSK modulation.

- Flip bits.

```
%% Flip bits (least significant bits are output first)
for i=1:8:length(msgb)
    msgb(i:i+7) = fliplr(msgb(i:i+7));
end
```

It builds the following bit sequence:

```
001000001110100111000101011110000111110000000001101000010111111111100
111000011010011010110110000010001001100100011000110110101111101010010 0
0010000000001110001110110000.
```

- Add checksum. The checksum is a 16-bit field, which is calculated as per CRC-16-CCITT standard and added at the end of the message. The checksum is taken over all the binary data.

```
%% Add FCS: CRC-16-CCITT standard
% Set parameters
crcw = 16;
poly = str2num(dec2bin(2^16+2^12+2^5+2^0)')'; % it is 0x1021 in HEX
initVal = ones(1,crcw);
finalXOR = ones(1,crcw);

% Registry initialization
am = [msgb,zeros(1,crcw)];
```

```
am(1:crcw) = xor(am(1:crcw),initVal);

% CRC calculation
reg = [0,am(1:crcw)];
for i=crcw+1:length(am)
    reg = [reg(2:end),am(i)];
    if reg(1)==1
        reg = xor(reg, poly);
    end
end
mcrc = reg(2:end);
mcrc = xor(mcrc,finalXOR);

msgb = [msgb, mcrc]; %add checksum at the end of message
crch = dec2hex(bin2dec(num2str(mcrc)),4) % for the sample it is 0x9920
```

It builds the following bit sequence, with the newly added part underlined:

```
00100000111010011100010101111000011111000000000110100001011111111110011
11100001101001101011011000001000100110010001100011011010111110101001000
00100000000011100011101100001001100100100000.
```

- Perform bit-stuffing. This means, adding a “0” after every 5 consecutive “1”. This is used to distinguish data bits from the HDLC flag, which is “01111110” bit sequence with 6 consecutive ones to denote the start and the end of the transmitted packet data.

```
%% do bit-stuffing - put zero after every 5 consecutive 1's
onesl = 5;
i = onesl;
while i<=length(msgb)
    if (msgb(i-(onesl-1):i) == ones(1,onesl))
        msgb = [msgb(1:i),0, msgb(i+1:end)];
        i = i+onesl+1;
    else
        i = i+1;
    end
end
```

It builds the following bit sequence:

```
00100000111010011100010101111000011111000000000110100001011111011111
00011100001101001101011011000001000100110010001100011011010111110010100
01000010000000001110001110110000100110010010000.
```

- Build the AIS packet. This puts a series of Ramp-Up bits, Preamble bits and the Start-Flag at the beginning of the message and adds the End-Flag and buffer at the end of the data packet.

```
%% build packet: Ramp-Up, Preamble, Start-flag, Message&FCS, End-flag,
Buffer
preamble = [0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1]; %24-bit
flag     = [0,1,1,1,1,1,1,0];  %%8-bit HDLC flag
msgb = [zeros(1,8), preamble, flag, msgb, flag, zeros(1,24)];
```

It builds the following bit sequence:

```
00000000010101010101010101010101111111000100000111010011100010101111000
01111110000000001101000010111110111110001110000110100110101101100000100
01001001100100011000110110101111100101001000010000000001110001110110000
00100110010010000001111111000000000000000000000000000.
```

- Perform the non-return to zero inversion. The algorithm goes through all bits and gives a state transition if a 0 is observed in the data.

```
%% do NRZI
% init params
binmsg = msgb;
icase  = 0;      %means, invert when 0 observed

% do the work
vstate = 0; % voltage state, assume initial=0
for i=1:length(binmsg)
    if binmsg(i)~=icase
        binmsg(i)=vstate;
    else %% invert voltage
        if (vstate==0)
            vstate=1;
        else vstate=0;
        end
        binmsg(i)=vstate;
    end
end
msgb = binmsg;
```

producing following bit sequence.

```
10101010110011001100110011001100111111110100101011110010000101100111110
10111111010101010001101011000000111111010000101000110111001110001010
11010010001001011101000111001111111011001001010010101010111101000011101
01101110110110101011111110101010101010101010101010.
```

By this procedure the AIS data bit encapsulation in an AIS packet is accomplished and the packet can be further passed to a GMSK modulator.

In general, a common naval AIS receiver will demodulate any data sequence from the AIS channels, which lies between two HDLC flags and possesses a correct checksum. Some receivers have an option to also display AIS messages with wrong checksum as PTHAR sentences in NMEA format, with the same structure as AIVDM sentences. The AIS packet building approach shown in this section above may serve as a practical tool for beginners of AIS programming, especially for understanding how the 16-bit checksum should be composed.

2.1.2 GMSK signal modulation

This section shortly reviews how GMSK modulation works [28], [29]. In Fig. 14 below is shown an overview of the GMSK modulation steps.

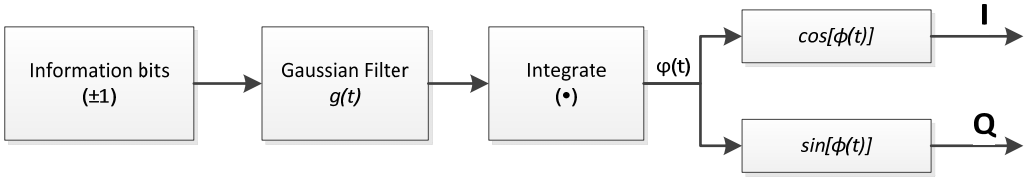


Fig. 14. The GMSK modulation graph [39].

The GMSK modulation process shown in Fig. 14 assumes that the input bits are provided in format of +1 and -1 values corresponding to 1 and 0 bits respectively. It applies Gaussian filter to the bit sequence, then integrates it and uses the signal as a phase in sine and cosine functions. The product of the modulation is an IQ signal, which is a complex number sequence. The real component of the signal is called I for in-phase signal and the complex component as Q for quadrature signal. A multiplications of the modulated IQ data with a carrier frequency, as it will be later shown in (6),

yields transmission ready IQ signal, ready to be sent to antenna. The real component of the IQ signal corresponds to the electric field applied to the antenna and the Q for the magnetic.

Let us start the description of GMSK modulation process with having given four data bits:

$$b \coloneqq \{1, 0, 1, 1\}.$$

To prepare the bit sequence b for filter application, the bits should be converted to values +1 for bit=1 and -1 for bit=0. This gives the following 4 bit NRZ (Non-Return-to-Zero) converted bit sequence [43]:

$$\hat{b} \coloneqq \{+1, -1, +1, +1\}.$$

To prepare the bits further for modulation, a time scale parameter needs to be introduced, since in data transmission bits can be transmitted only one by one. Thus the ordered sequence of bits is de facto transformed to a time series of bits.

Since the AIS standard uses 9600bps (bits per second), it means that time per bit T_b in the AIS case is $T_b = \frac{1}{9600} = 1.041(6) \cdot 10^{-4}$ seconds. Additionally, there should be chosen an osr (oversampling rate) for bits, which denotes how many discrete time points per bit should be used to approximate the signal under modulation. The higher the oversampling rate, the smoother will be the modulated signal. The Nyquist-Shannon sampling theorem tells that the sampling rate should be at least twice the bit rate [28]. Let us choose here 50 samples per bit, which is 25 times more than the minimum, resulting to $9600 \cdot 50 = 480000$sps (samples per second). This sets time per sample for IQ data to $T_s = 2.08(3) \cdot 10^{-6}$ seconds. It ensures good conditions for smooth signal modulation and detailed analysis for the purpose of this thesis. To convert the originally provided four bits in $\hat{b}$ to signal samples with the chosen oversampling rate as required for subsequent filter application, there are inserted 49 zeros after each bit, denoted further as $z_{49} = \{0,0, \dots ,0\}$.

$$\begin{aligned}\tilde{b}(t_s) &\coloneqq \{+1, z_{49}, -1, z_{49}, +1, z_{49}, +1, z_{49}\} = \\ &= \{+1,0,0, \dots ,0, \ -1,0,0, \dots ,0, \ +1,0,0, \dots ,0, \ +1,0,0, \dots ,0\}.\end{aligned}$$

The tilde above a symbol within this section will be used to denote sampled data with a discrete time step T_s. The t_s within this section will be used to denote the discrete time index for a samples sequence.

The next preparatory step is to generate the GMSK filter. This can be accomplished by defining the filter $h(t)$ as follows [39].

$$h(t) \coloneqq \frac{1}{\sqrt{2} \cdot \pi \cdot \delta \cdot T} e^{\frac{-t^2}{2 \cdot \delta^2 \cdot T^2}}, \tag{2}$$

where

$$\delta \coloneqq \frac{\sqrt{\ln(2)}}{2 \cdot \pi \cdot BT} \quad \text{and}$$
$$BT \coloneqq 0.4 \, .$$

The function $h(t)$ in (2) is called a GMSK filter and its parameter BT is called bit time and was introduced in Section 1.3. The BT influences the smoothness of Gaussian filtered data bits. In general,

the filter length is infinite as the time component T, but since for real digital signal processing there must be used filters of finite length, a modified filter $\tilde{h}(t_s)$ is introduced, shown in Fig. 15. It is discretized with time step T_s and is a finite version of the $h(t)$, cropped to length of four modulation symbols [28]. The chosen length is very reasonable, since for the particular filter of GMSK modulation for the AIS standard, the relative magnitude of the lowest filter value to the maximum filter value (corresponding to the sample 1 and the sample 100 in Fig. 15) is $\frac{\tilde{h}(t_1)}{\max\left(\tilde{h}(t_s)\right)} = 1.7 \cdot 10^{-8}$. This is small enough to be neglected, thus cropping the filter width to four symbols is chosen and will be used throughout the rest of this work. A filter length of only two modulation symbols would have been a too rough approximation, since the relative magnitude of the first sample to the maximum one would be 0.0105 and this does introduce some undesirable side effects for signal processing due non-smooth curves in the modulation waveforms. The obtained Gaussian filter $\tilde{h}(t_s)$ is shown in Fig. 15 below.

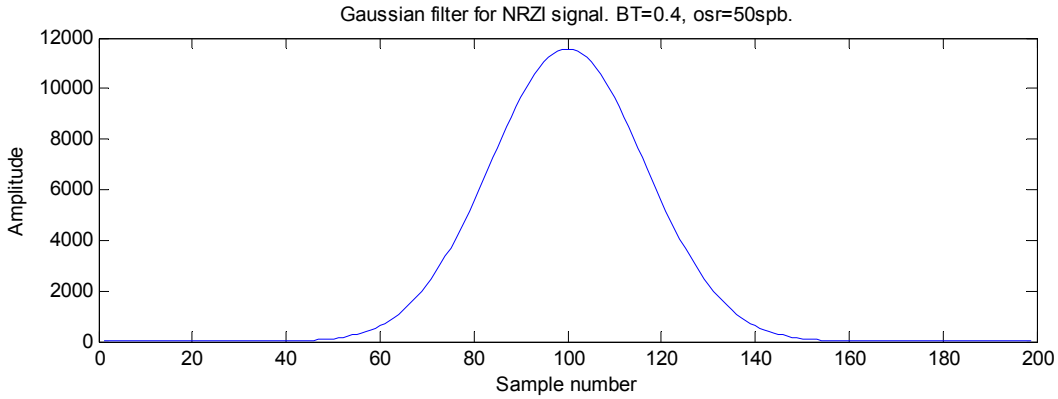


Fig. 15. Gaussian filter shape, with a length of four symbol durations and an oversampling rate of 50 samples per bit.

The next step is to apply the derived Gaussian filter to the bit samples sequence $\tilde{b}(t_s)$. This can be mathematically written as convolution of two functions, the samples sequence $\tilde{b}(t_s)$ with the derived Gaussian filter $\tilde{h}(t_s)$, as shown in (3). A visual example of the filtering procedure is illustrated in Fig. 16.

$$\tilde{b}_h(t_s) \coloneqq \left(\tilde{b} * \tilde{h}\right)(\mathrm{t_s}) = \sum_{\tau=1}^{M} \tilde{b}(t_s) \cdot \tilde{h}(t_s - t_\tau)\,, \tag{3}$$

where M is the length of the filter $\tilde{h}(t_s)$ and t_i is the discrete time index using time step T_s.

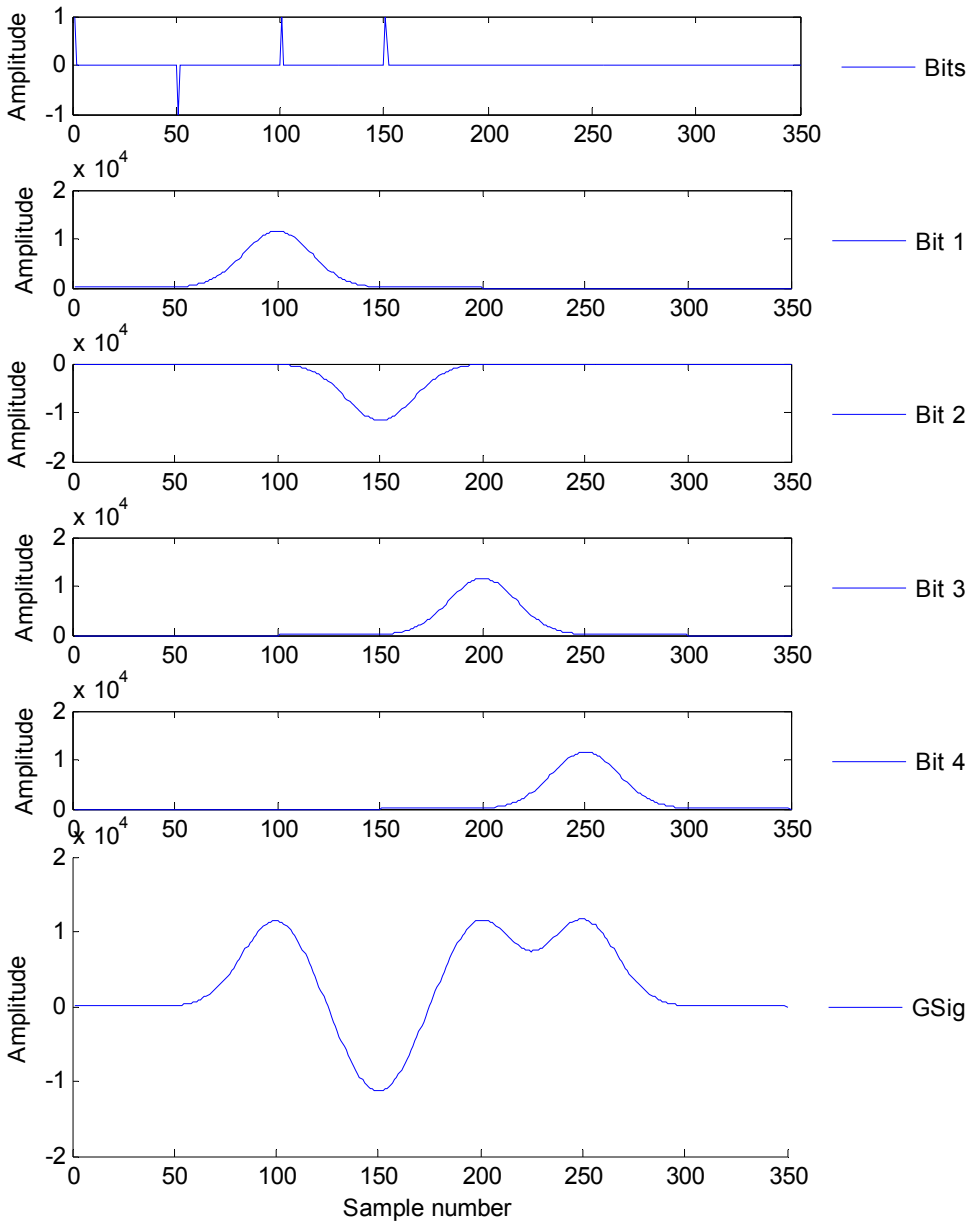


Fig. 16. Gaussian filter application for the four bit sequence {+1,-1,+1,+1}. a) provided bits $\tilde{b}$, b) bit 1 c) bit 2 d) bit 3, e) bit 4, d) resulting gaussian filtered siganal for the four bit sequence.

The next step is to integrate the Gaussian filtered signal. This step is written with a Riemann integral in (4) and the result is shown in Fig. 17.

$$\tilde{b}_{hI}(t_s) = \sum_{\tau=1}^{s} \tilde{b}_h(t_\tau)\Delta t \ , \tag{4}$$

where Δt is the time interval between samples, also introduced as sampling time T_s previously.

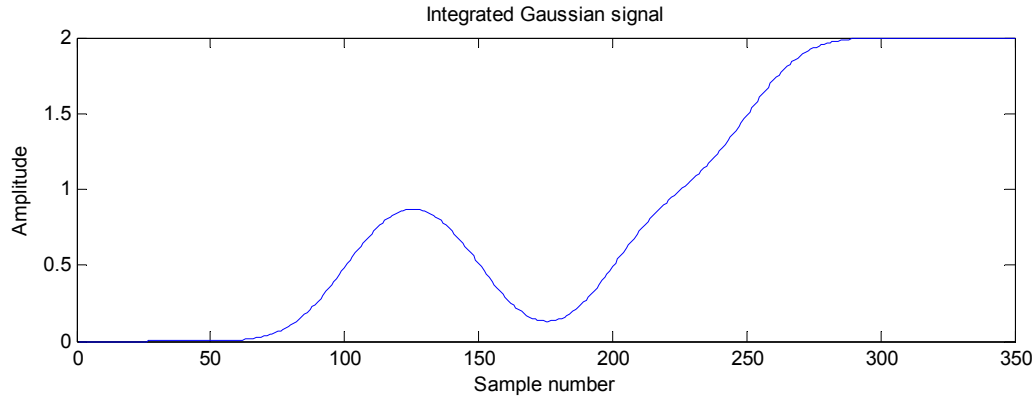


Fig. 17. Integrated Gaussian-filtered signal.

Since all the practical signal processing work done within this theses is performed with discrete time signal, in further text the discrete time denotation t_s will be shortened with t for easier writing.

The final step in GMSK modulation is to apply the integrated signal as the phase of a sinusoidal signal. It can be accomplished as follows.

$$\begin{aligned} \varphi(t) &\coloneqq \tilde{b}_{hI}(t) \cdot \frac{\pi}{2}\,, \\ I(t) &\coloneqq \cos\big(\varphi(t)\big)\,, \\ Q(t) &\coloneqq \sin\big(\varphi(t)\big)\,, \\ IQ(t) &\coloneqq I(t) + iQ(t) = \cos\big(\varphi(t)\big) + i\sin\big(\varphi(t)\big) = e^{i\varphi(t)}\,. \end{aligned} \tag{5}$$

The (5) serves as the definition for the IQ data. It is valuable to note that the data are complex valued and can be written as exponential function. This property will be used in the next section.

This step generates complex valued IQ data samples as shown in Fig. 18.

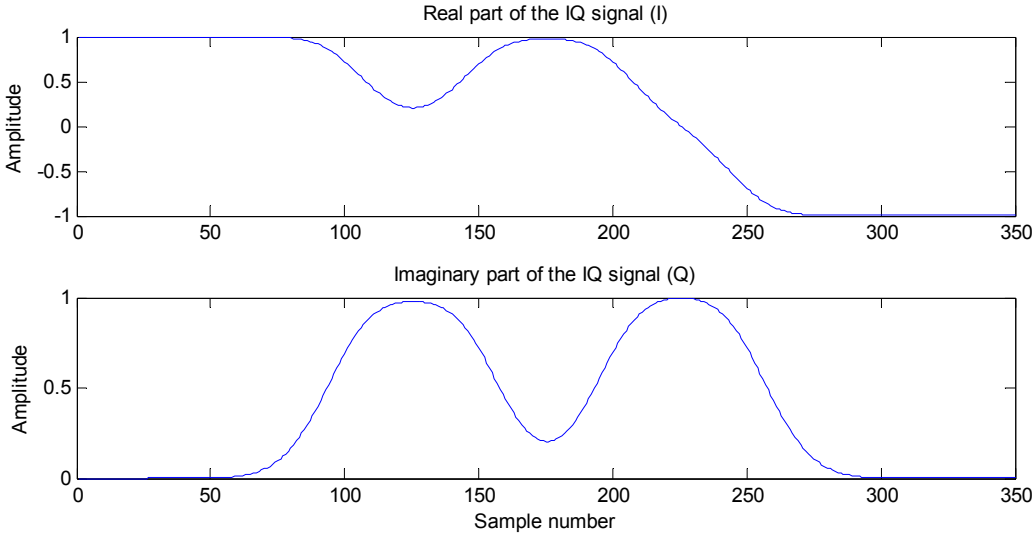


Fig. 18. Real and Imaginary parts of the IQ signal, containing GMSK modulated bits {+1,-1,+1,+1}.

The MATLAB code for the whole GMSK modulation work described in this section can be written as follows.

```
%% do GMSK
% init params
osr     = 64;     % Oversampling Rate for bits
br      = 9600;   % Bound rate [bits per second]
BT      = 0.4;    % Bit-Time
bindata = msgb;   % binary voltage data

% do NRZ - translates all '0' to '-1'
for i=1:length(bindata)
    if bindata(i)==0
        bindata(i)=-1;
    end
end

% resample the signal
rbindata = [];
for i=1:length(bindata)
    rbindata = [rbindata, bindata(i), zeros(1,osr-1)];
end
rbindata = [rbindata, zeros(1,osr)];

% calculate Gaussian FIR filter
T  = 1/bps;
Ts = T/osr;
k  = [-osr+1:osr-1];
alpha = sqrt(log(2))/(2*pi*BT);
h = exp(-(k*Ts).^2/(2*alpha^2*T^2))/(sqrt(2*pi)*alpha*T); % Gaussian
FIR shape

% do Gaussian filtering
gfsignal = filter(h,1,rbindata);

% do integration
isignal = zeros(size(gfsignal));
for i=1:length(gfsignal)-1
  isignal(i+1)=isignal(i)+gfsignal(i)*Ts;
end

% do IQ modulation
signalI  = cos(isignal*(pi/2));
signalQ  = sin(isignal*(pi/2));
signalIQ = complex(signalI, signalQ);
```

Obtaining the IQ data accomplishes the GMSK modulation process. Another example of a step by step IQ data generation with a different bit sequence can be found in Section 4.2.1.

2.1.3 Modelling and transmitting the AIS signal

The GMSK modulation process as shown in Section 2.1.2 applied to the derived AIS message bit sequence in Section 2.1.1, leads to generating IQ data of a complete AIS message, shown in Fig. 19 below.

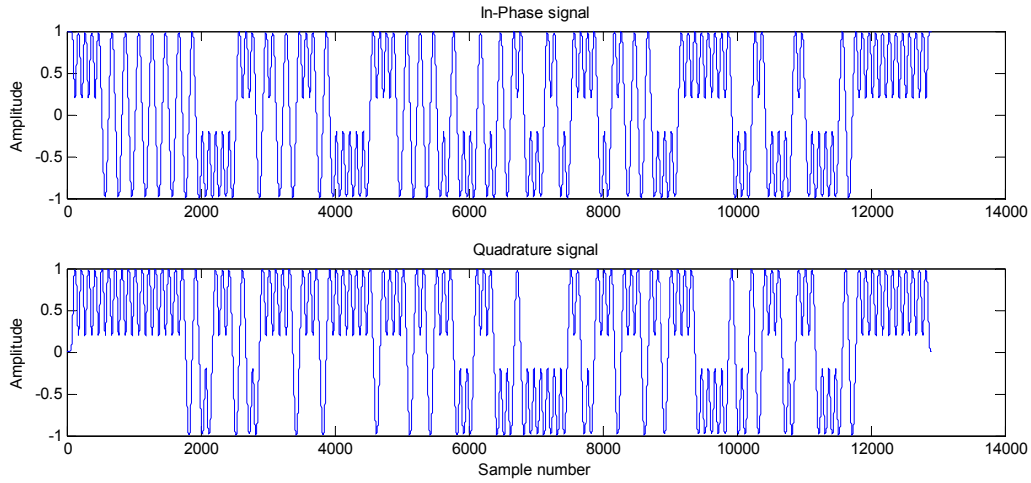


*Fig. 19. IQ data waveform of the AIS message !AIVDM,1,1,,B,19NS7Sp02wo?HETKA2K6mUM20<L=,0*27.*

To transmit the generated IQ waveform, it should be turned form a digital signal into analogue one by DAC (Digital to Analogue Converter), then passed to frequency multiplication and power amplification, and finally sent out to antenna. These steps are shown below in Fig. 20.

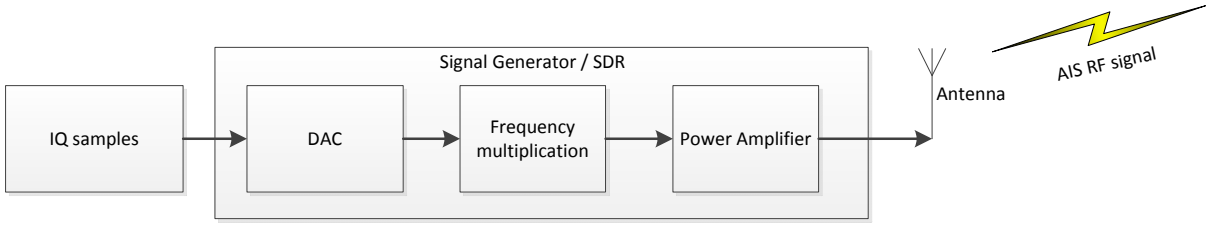


Fig. 20. Signal path from IQ data form to transmission.

The modulation of the IQ data to carrier frequency is performed by hardware which converts digital IQ data in RF signal. Equipment, used by author of this dissertation, is the signal generator Rohde&Schwarz SMBV100A and the SDR (Software Defined Radio) Ettus Research USRP N210. These devices can be fed with the generated IQ data waveforms and accomplish the signal modulation onto the carrier frequencies 161.975 or 162.025 MHz, for channel A or B respectively. Both of the named devices are capable to convert the digital signal into the analogue one and transmit over the air, which is then ready to be received by any commercial AIS receiver. The SDR solution is very convenient for local laboratory tests, while the SMBV solution can be applied also for powerful signal transmission as real AIS transmitter does.

The resulting transmitted RF signal can be written as in (6), by applying a frequency multiplication to the GMSK signal modulation from (5).

$$RF(t) = IQ(t) \cdot e^{i\omega_c(t+\nu_c)} = e^{i\varphi(t)} \cdot e^{i\omega_c(t+\nu_c)} = e^{i(\varphi(t)+\omega_c\tau_c)} \tag{6}$$
$$\omega_c \coloneqq 2\pi f_c \ ,$$

where f_c is carrier frequency, which in the AIS case is 161.975 or 162.025 MHz, and $\tau_c \coloneqq t + \nu_c$ is phase time-shift of the carrier frequency. Here is an ideal transmitter assumed, which does not create noise.

2.1.4 Usage of the IQ Waveform

In case of transmitting the derived IQ waveform, a small waveform modification is required, due to real world hardware being unable to switch “on” instantly, requiring some transition time. In case of automatic conversion of an unmodified IQ waveform by DAC (Digital-to-Analogue Converters), this would create some unexpected spikes in signal magnitude and its frequency spectra. As a counter measure, the AIS standard defines 8 bits Ramp-Up time and finishing time as shown in Fig. 3, taken from [1]. This was implemented in the IQ data by adjusting its magnitude, leading to the following waveform representation, shown in Fig. 21.

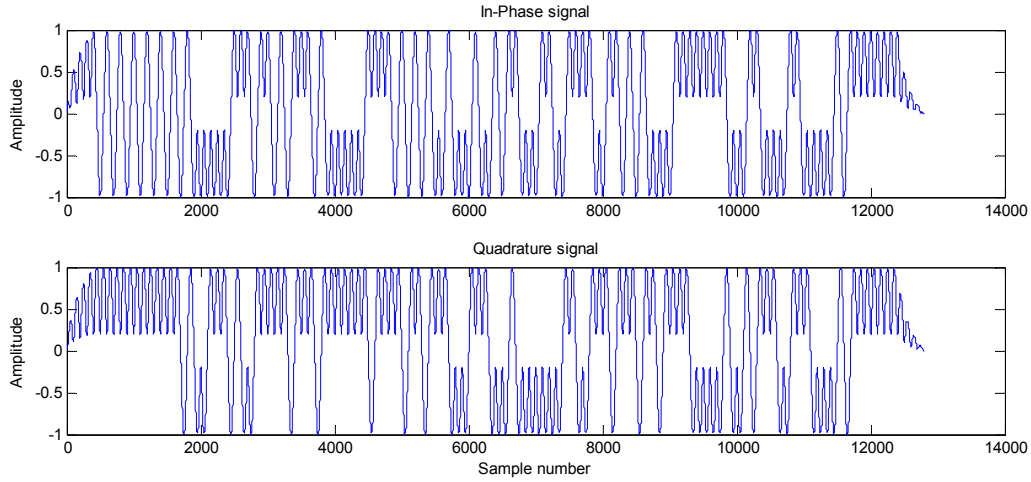


Fig. 21. IQ data samples of a real AIS message, with smooth amplitude change at start and finish of the packet.

To transmit simultaneously channel A and B signals by single SDR equipment, one need to combine both waveforms into single one, taking into account the carriage frequency difference. To include the IQ waveform of B channel into IQ waveform which will be modulated onto A channel carrying frequency, one multiplies the initial B channel IQ data with a 50kHz frequency and adds them onto the A channel IQ data. Thus the IQ waveform data of the B channel will be sent out at the A channel frequency plus 50kHz. The 50kHz value corresponds to the difference between the AIS channel A and AIS channel B carrying frequencies. A care should be taken to have a high enough bit oversampling rate, which is capable of holding the targeted frequencies as well as ensure that RF equipment supports required frequency bandwidth. Fig. 22 shows how the B channel waveform looks after multiplication with the 50kHz frequency. Due to the high oscillation rate, only the amplitude border is recognizable in the IQ data signal.

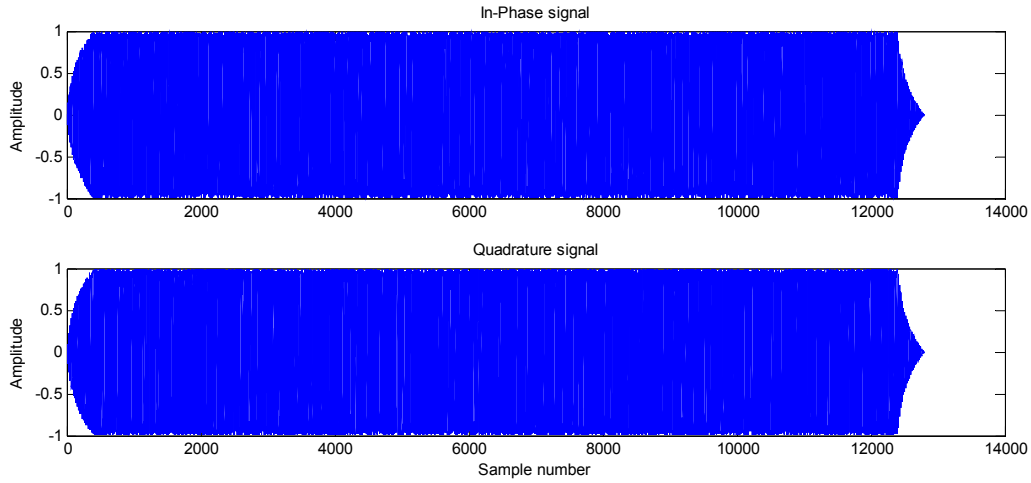


Fig. 22. IQ data sample for B-channel AIS message, to be modulated on channel A carrier frequency. With smoothed Ramp-Up and finishing amplitudes, as required in [1].

It is very practical to know how to visually recognize the start and finish of the AIS data by observing the waveform. For this purpose, let us consider the beginning section of the IQ waveform from Fig. 21 in more detail.

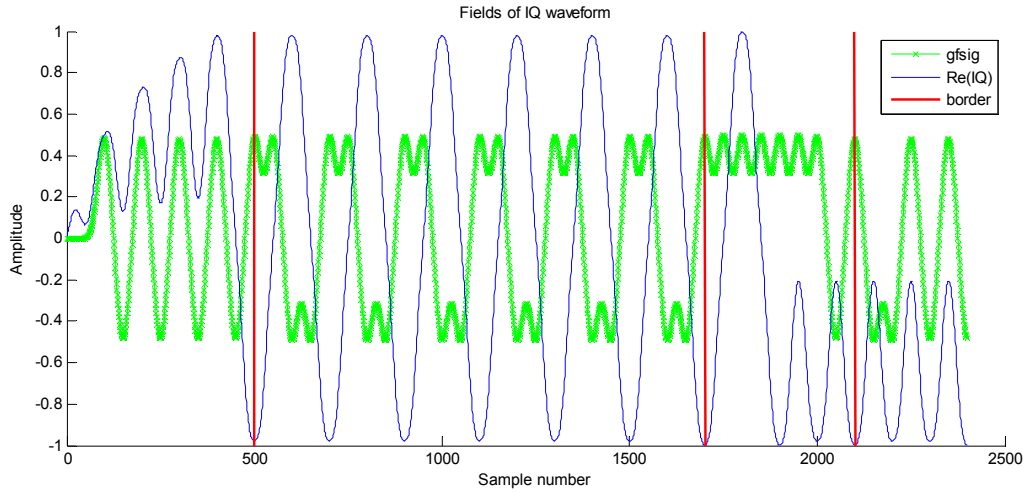


Fig. 23. First 48 bits of the In-Phase data of the IQ waveform, overlapped with scaled Gaussian filtered signal bits. The three red lines separate Ramp-Up time, Preamble, HDLC-Flag and data bits.

The respective bits, representing +1 and -1 voltage states respectively, are as follows:

```
10101010110011001100110011001100111111110010101
```

containing Ramp-Up, Preamble, HDLC-Flag and Data-start respectively. One should note that the HDLC flag presence in the IQ waveform in Fig. 23 may not to be recognizable. This is due the NRZI used:

```
01111110-01010101010101010101010101-01111110-00100000 - before NRZI
10101010-11001100110011001100110011-11111110-10010101 - after NRZI
```

One can see the HDLC-Flag bits can be easily recognized before NRZI application, but not afterwards. However, the very nicely oscillating wave of the preamble, just before the HDLC-Flag, can be easily noticed within IQ data without Doppler shift, by visually inspecting the signal. This preamble is also an important part for signal processing, as it is used for symbol clock synchronization and channel properties' estimation.

2.2 Receiving signal

This section addresses processing steps of the signal reception. As shown in the short overview of the AIS signal modulation in (5) in Section 2.1.2 GMSK signal, one can interpret an AIS signal from the signal processing point of view as a sinusoidal function with a variable phase, steered by data bits under modulation.

For a physical receiver, which can receive only the real part of the transmitted signal, (7) describes the received signal with its subparts.

$$Re\big(RF(t)\big) = \cos(c_1 + c_2 + c_3 + c_4 + c_5)\,, \tag{7}$$

where

$c_1 \coloneqq \varphi(t)$ – data modulated phase
$c_2 \coloneqq \omega_c t + \varphi_c(t)$ – carrier frequency and its phase offset
$c_3 \coloneqq \omega_c t + \varphi_c(t)$ – doppler shift and the phase offset
$c_4 \coloneqq \omega_d t + \varphi_d(t)$ – error of transmitter/receiver hardware
$c_5 \coloneqq \omega_e t + \varphi_e(t)$ – noise and signal path distorsions

The ω_e and φ_e can be assumed to be very small and thus neglected. The φ_d and φ_c are constants and can be summed and substituted with ξ. The c_5 can be substituted with N, for an explicit notation as noise. This allows (7) to be rewritten as (8).

$$Re(RF(t)) = \cos\left(\varphi(t) + \omega_c t + \omega_d t + \xi\right) + N \tag{8}$$

Let us initially consider the received signal containing no noise and no hardware induced errors.

As soon as a signal is received on the antenna, the signal passes through a band-pass filter, which lets through only a certain frequency range around targeted central frequency, for example, +5MHz around the AIS central frequency of 161.975MHz. This is usually called band-pass signal and noted with X_{BP}, with the contained received IQ data noted as X_{IQ}. Rewriting the expression of IQ as defined in (5) with the carrier frequency multiplication $\omega_c = \omega_0$ as defined in (6) and then taking only the real part of the received RF signal as done in (8) gives the following definition for the band-passed signal.

$$
\begin{aligned}
X_{BP}(t) &:= Re\big(RF(t)\big) = Re\big(X_{IQ}(t)\cdot e^{i\omega_0 t}\big) = \\
&= Re\left(\big(X_I(t)+iX_Q(t)\big)\cdot(\cos(\omega_0 t)+i\sin(\omega_0 t))\right) = \\
&= Re\big(X_I(t)\cdot\cos(\omega_0 t)+iX_I(t)\cdot\sin(\omega_0 t)+iX_Q(t)\cdot\cos(\omega_0 t)-X_Q(t)\cdot\sin(\omega_0 t)\big) = \\
&= X_I(t)\cdot\cos(\omega_0 t)-X_Q(t)\cdot\sin(\omega_0 t) \qquad (9)
\end{aligned}
$$

The substitution of carrier frequency multiplication $IQ(t)\cdot e^{i\omega_c(t+\nu)}$ from (6) with the $IQ(t)\cdot e^{i\omega_0 t}$ as used in deriving (9) can be done without loss of generality for a purpose of easier writing. The ω_0 is used to note a phase-shift less Doppler shift frequency. Since the phase ν of the Doppler shift builds a constant, it can be multiplied and hidden within the IQ data, as shown in the following equations, which introduce $\overline{IQ}(t)$.

$$
\begin{aligned}
IQ(t)\cdot e^{i\omega_c(t+\nu)} &= \big(IQ(t)\cdot e^{i\omega_c\nu_c}\big)\cdot e^{i\omega_c t} = \\
&= \big(I(t)\cdot e^{i\omega_c\nu_c}+iQ(t)\cdot e^{i\omega_c\nu_c}\big)\cdot e^{i\omega_c t} = \overline{IQ}(t)\cdot e^{i\omega_0 t} \\
\overline{IQ}(t) &:= \bar{I}(t)+i\bar{Q}(t) \\
\bar{I}(t) &:= I(t)\cdot e^{i\omega_c\nu_c} \qquad (10) \\
\bar{Q}(t) &:= Q(t)\cdot e^{i\omega_c\nu_c} \qquad (11)
\end{aligned}
$$

An extraction of the in-phase and quadrature signals from the X_{BP} signal is usually accomplished by multiplying the received signal with a dedicated cosine or sine function and then applying a LPF (Low Pass Filter).

A LPF is a device which processes an input signal and allows to pass through only low frequency signal parts of it. A characteristic property of the filter is its cut-off frequency, above which the signal components will be removed. Mathematically for a LTI (Linear, Time-Invariant) system the filter can be written as a transfer function $H(s)$, which is a linear mapping of the Fourier transform of the input $\hat{X}_{in}(s) = \mathcal{F}\{X_{in}(t)\}$ to the Fourier transform of the output $\hat{X}_{out}(s) = \mathcal{F}\{X_{out}(t)\}$:

$$
H(s) = \frac{\hat{X}_{out}(s)}{\hat{X}_{in}(s)}\,. \qquad (12)
$$

In symbolic notation the low pass filtering process for a discrete signal can be written as

$$
X_{in}(t) \xrightarrow{LPF_\omega} X_{out}(t),\ \text{with}\ \mathcal{Z}\{X_{out}(t)\} < \omega\,. \qquad (13)
$$

In (13) the $X_{in}(t)$ and $X_{out}(t)$ are discrete time input and output signals correspondingly and $\mathcal{Z}\{X_{out}(t)\}$ is the z-transform of the discrete time signal from time domain into frequency domain, similar to Fourier transform for a continuous time signal in (12). The ω in (13) is the cut-off frequency for the filter. For an ideal LPF, the frequency spectrum of X_{out} would contain only frequencies which are lower than the cutting frequency ω. As the properties of LPF is an extensive topic by itself, for further details and notations an interested reader is advised to consult signal processing textbooks [28], [29].

The low pass filtering is often used to obtain the I and Q components of the enclosed data in a signal containing a carrying frequency. To extract the I component, one needs to multiply the sampled signal X_{BP} as defined in (9) with the $\cos(\omega_0 t)$ function, as shown in (14). Passing the signal through a

LPF will do the single component extraction by filtering out high frequency parts signal parts, including the $2\omega_0$ frequency components. For getting the Q component, one proceeds in the same way, except taking $\sin(\omega_0 t)$ for the multiplication, as shown in (15) below.

$$X_{BP}(t)\cdot\cos(\omega t) = X_I(t)\cdot\underbrace{\cos(\omega_0 t)\cdot\cos(\omega_0 t)}_{\frac{1}{2}+\frac{1}{2}\cos(2\omega_0 t)} - X_Q(t)\cdot\underbrace{\cos(\omega_0 t)\cdot\sin(\omega_0 t)}_{\frac{1}{2}\sin(2\omega_0 t)} =$$

$$=\frac{1}{2}X_I(t) + \frac{1}{2}X_I(t)\cdot\cos(2\omega_0 t) - \frac{1}{2}X_Q(t)\cdot\sin(2\omega_0 t) \tag{14}$$

$$X_{BP}(t)\cdot\sin(\omega t) = X_I(t)\cdot\underbrace{\cos(\omega_0 t)\cdot\sin(\omega_0 t)}_{\frac{1}{2}\sin(2\omega_0 t)} - X_Q(t)\cdot\underbrace{\sin(\omega_0 t)\cdot\sin(\omega_0 t)}_{\frac{1}{2}-\frac{1}{2}\cos(2\omega_0 t)}$$

$$= -\frac{1}{2}X_Q(t) + \frac{1}{2}X_Q(t)\cdot\cos(2\omega_0 t) + \frac{1}{2}X_I(t)\cdot\sin(2\omega_0 t) \tag{15}$$

Applying a LPF_ω to the equalities (14) and (15) yield the I and Q components for the input signal X_{BP} as follows:

$$X_{BP}(t)\cdot\cos(\omega t) \xrightarrow{LPF_\omega} \frac{1}{2}X_I(t)\,, \tag{16}$$

$$X_{BP}(t)\cdot\sin(\omega t) \xrightarrow{LPF_\omega} -\frac{1}{2}X_Q(t)\,. \tag{17}$$

The X_I and X_Q components in (16) and (17) represent the I and Q components respectively, obtained from the X_{BP} signal. For a noiseless RF signal with a small Doppler shift applying an ideal LPF the IQ extraction procedure shown above would provide the low frequency part of the input signal, incorporating the IQ data transmitted along with Doppler shift frequency and phase constant as defined in (10) and (11). Using the definitions from (10) and (11), it would mean $\bar{I}(t) = X_I(t)$ and $\bar{Q}(t) = X_Q(t)$.

A real world low pass filters will introduce some noise as well as let some higher frequencies with dumped amplitudes to pass through. Therefore at the receiver side one can generally assume to have a scaled form of IQ signal similar to the one derived in (5) in Section 2.1.2 with some Doppler shift and noise applied as shown in (8), (10) and (11).

To generate a simulated signal containing a Doppler shift frequency, there will be applied a multiplication of the IQ signal as derived in (5) with the corresponding exponential function $e^{i\omega_d(t+\nu_d)}$. During further work there will be unity amplitude for the IQ signal usually assumed. A scaling of the IQ signal amplitude for some specific overlapping scenarios will be shown and used while describing demodulation methods in Chapter 4 and Chapter 5.

2.3 Overlapping signals

2.3.1 Modelling overlapped signals with Doppler shifts

To find how overlapped signals can be modelled, let us start by summing two single RF signals on a single sensor antenna, using single signal model as derived in (6) with a Doppler frequency addition.

$$RF^1(t) \coloneqq IQ^1(t) \cdot e^{i\omega_{d1}(t+\nu_{d1})} \cdot e^{i\omega_c(t+\nu_{c1})} = \mathrm{e}^{i\varphi_1(t)} \cdot e^{i\omega_{d1}(t+\nu_{d1})} \cdot e^{i\omega_c(t+\nu_{c1})} \quad (18)$$

$$RF^2(t) \coloneqq IQ^2(t) \cdot e^{i\omega_{d2}(t+\nu_{d2})} \cdot e^{i\omega_c(t+\nu_{c2})} = \mathrm{e}^{i\varphi_2(t)} \cdot e^{i\omega_{d2}(t+\nu_{d2})} \cdot e^{i\omega_c(t+\nu_{c2})} \quad (19)$$

$$\widehat{RF}(t) \coloneqq RF^1(t) + RF^2(t) = \left(e^{i\gamma_1(t)} + e^{i\gamma_2(t)}\right) \cdot e^{i\omega_0 t} = \widehat{IQ}(t) \cdot e^{i\omega_0 t} \quad (20)$$

$$\widehat{IQ}(t) \coloneqq e^{i\gamma_1(t)} + e^{i\gamma_2(t)} \quad (21)$$

The hat symbol above the RF and IQ notations in (20) and (21) is introduced to denote signal samples containing overlapped transmissions.

The γ_1 and γ_2 introduced in (20) define the modulated signal phase as follows

$$\gamma_1(t) \coloneqq \varphi_1(t) + \omega_{d1}(t + \nu_{d1}) + \omega_c \nu_{c1} \,,$$
$$\gamma_2(t) \coloneqq \varphi_2(t) + \omega_{d2}(t + \nu_{d1}) + \omega_c \nu_{c2} \,.$$

To make the writing of the $\gamma_i(t)$ even shorter, following additional notions can be introduced:

$$\nu_1 \coloneqq \nu_{d1} + \frac{\omega_c \nu_{c1}}{\omega_{d1}} \,,$$
$$\nu_2 \coloneqq \nu_{d2} + \frac{\omega_c \nu_{c2}}{\omega_{d2}} \,,$$
$$t_{\tau 1} \coloneqq t + \nu_1 \,,$$
$$t_{\tau 2} \coloneqq t + \nu_2 \,.$$

Then it can be written

$$\gamma_1(t) = \varphi_1(t) + \omega_{d1}(t + \nu_1) = \varphi_1(t) + \omega_{d1}(t_{\tau 1}) \,,$$
$$\gamma_2(t) = \varphi_2(t) + \omega_{d2}(t + \nu_2) = \varphi_2(t) + \omega_{d2}(t_{\tau 2}) \,.$$

As the $\widehat{RF}(t)$ in (20) hold the same structure as $RF(t)$ in (6), there can be applied the same procedure of low pass filtering to remove carrier frequency components as performed for single signal containing RF data in (16) and (17). It is shown below in (22) and (23) respectively.

$$\hat{X}_{BP}(t) \coloneqq Re\left(\widehat{RF}(t)\right)$$

$$\hat{X}_{BP}(t) \cdot \cos(t) \xrightarrow{LPF_\omega} \frac{1}{2}\hat{X}_I(t) \quad (22)$$

$$\hat{X}_{BP}(t) \cdot \sin(t) \xrightarrow{LPF_\omega} -\frac{1}{2}\hat{X}_Q(t) \quad (23)$$

$$\hat{X}_{IQ}(t) \coloneqq \hat{X}_I(t) + i\hat{X}_Q(t)$$

Rewriting the $\hat{X}_{IQ}(t)$ more explicitly with the substitutions done previously, we conclude that the received IQ signal of two overlapped messages, when arriving at DSP unit, can be represented as follows.

$$\hat{X}_{IQ}(t) = e^{i\varphi_1(t)} \cdot e^{i\omega_{d1}(t+\nu_1)} + e^{i\varphi_2(t)} \cdot e^{i\omega_{d2}(t+\nu_2)} = \\ = IQ^1(t) \cdot e^{i\omega_{d1}(t+\nu_1)} + IQ^2(t) \cdot e^{i\omega_{d2}(t+\nu_2)} \quad (24)$$

From (24) it is easy to identify that the received overlapped signal can be viewed as sum of IQ components of the signals it contains, each multiplied with an exponential function corresponding to a frequency and a phase of the Doppler shift and a phase of the carrier frequency. This property will be used throughout the rest of the work to model and construct IQ data of simulations.

2.3.2 An example of two overlapped AIS signals

Without loss of generality, the following two AIS messages can be assumed to be provided:

```
!AIVDM,1,1,,A,17PBtb0P?ENT8;2PaNUa@TD0P0Go,0*76
!AIVDM,1,1,,A,17P0Qh0P?INIVW`OugAa4;:0P3mp,0*0F.
```

And let us look how an overlap of the two messages can be made and what adjusting parameters are available. We start with generating IQ waveforms for both of the AIS packets. Then, as it comes to adding the two IQ waveforms together, the following parameters can be adjusted.

- **Relative time shift** sets by how many samples the second waveform should be shifted with respect to the first waveform. This would mean the delay time, measured in samples, bits or seconds. For most of the analysis within this work, the time shift is assumed to be one bit. This is the most favourable situation, giving the most unique phase of each signal, since in that position the minimum phase magnitude of the first signal matches the maximum magnitude of the second signal's phase, since the phase of the AIS IQ signal has periodicity of two bits length. The worst case would be a zero phase shift, as then the signal phases would overlap, making in certain cases it impossible to relate a bit membership to its source signal. Since the bit phases are periodic, investigation of time shift properties within a single bit range is sufficient for bit shift analysis of an overlaps containing waveform.

- **Amplitude ratio** sets the Signal1 amplitude relative to Signal2. Within this document, the amplitude sum is usually normalized to 1. A highly interesting case is a Signal1 to Signal2 amplitude ratio of 0.5:0.5, which means exactly the same amplitudes, as well as with a very small difference ratio, like 0.6:0.4.

- **Doppler frequencies** set the Doppler frequency shift values for each of the signals. This can be realized by multiplying both IQ waveforms with sinusoids of desirable Doppler shift. Expected values for space application range from 0 to ±4kHz.

- **Phases of the Doppler frequencies** set the Doppler frequency phase values for each of the signals. This can be realized by multiplying both IQ waveforms with a complex number, corresponding to the Doppler shift phase offsets of the signals.

- **Noise** sets the noise amplitude with respect to signal amplitudes. An AWGN (Additive White Gaussian Noise) is assumed. To simulate noise amplitude, the value of its standard deviation is used. In Matlab it can be generated as follows.

```
% add noise to sigIQ
noiseweight = 0.1;
noise = randn(2,length(sigIQ));
```

```
noiseiq = complex(noise(1,:), noise(2,:));
sigIQn = sigIQ+noiseiq*noiseweight;
```

Let us look on an overlapped signal example, containing the two AIVDM messages introduced at the beginning of this section.

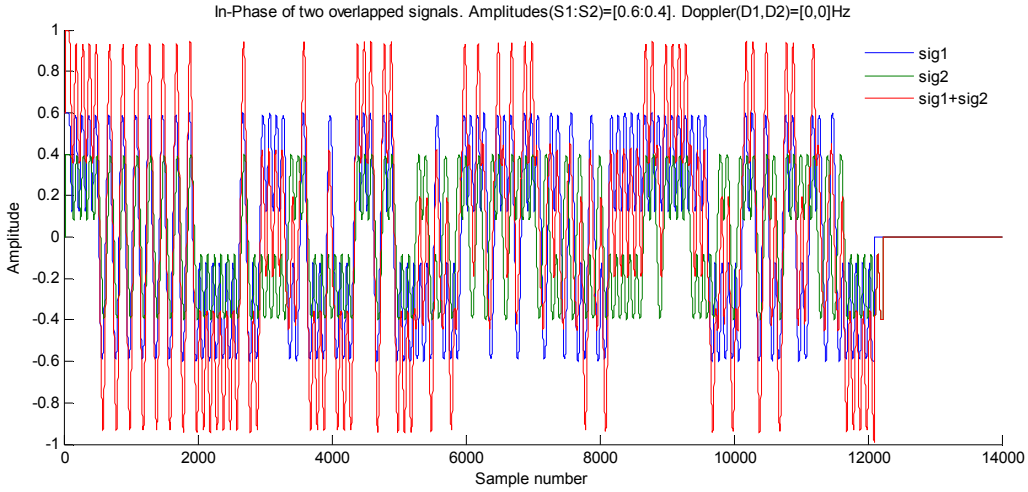


Fig. 24. An example of an In-Phase signal of overlapped IQ data. Relative amplitudes are 0.6:0.4 and relative time shift is 1 bit. Noise and Doppler shifts are not included in the signals.

In Fig. 24 there can be observed a simplistic example of overlapped signals. The Chapter 3 will deal with various overlapped signal scenarios and the chapters 4 and 5 will reveal its overlapping parameter analysis in more detail.

2.3.3 Signal power calculation

In the further work there will be used various notations of signal amplitudes and its relation to noise level. Let the notations be shortly introduced here.

The SIR (Signal to Interference Ratio) is used to compare the powers of two signals. The power of a single signal is represented by a standard deviation of the signal's amplitude squared and denoted as σ^2. For an AIS signal with unity amplitude its standard deviation was calculated to be approximately 0.7, depending on data contained within a transmitted message. The σ^2 can be used to calculate two following parameters for comparing powers of two overlapping signals.

$$SIR := \frac{\sigma^2_{stronger\_signal}}{\sigma^2_{weaker\_signal}} = \frac{(A_{ss} \cdot 0.7)^2}{(A_{ws} \cdot 0.7)^2} \tag{25}$$

$$SIR_{dB} := 10 \cdot \log_{10}(SIR) \tag{26}$$

The SIR_{dB} characterizes the strongest signal power over weaker one in dB (decibel) units. This is a comfortable and a widely used notation. For example, $SIR_{dB} = 3$ dB tells that the strongest signal is two times stronger in magnitude over the weaker one.

A similar approach can be taken to calculate SNR (Signal to Noise Ratio), which characterizes by how much a signal power is higher than a noise level.

$$SNR \coloneqq \frac{\sigma^2_{stronger\_signal}}{\sigma^2_{noise}} = \frac{(A_{ss} \cdot 0.7)^2}{\sigma(N)^2} \tag{27}$$

$$SNR_{dB} \coloneqq 10 \cdot \log_{10}(SNR) \tag{28}$$

For simulations within this work the AWGN, denoted as N_{AWGN}, was always assumed and was generated by Matlab with a unity standard deviation $\sigma(N_{AWGN}) = 1$. Multiplying the N_{AWGN} with a wished noise amplitude A_N, the amplitude of the additive noise for IQ data can be adjusted:

$$\sigma(N) = \sigma(N_{AWGN}) \cdot A_N = A_N \, .$$

Often in work there will be used a following notation of signal and noise amplitudes

$$[S1\colon S2\colon \sigma(N)] = [0.6\colon 0.4\colon 0.1].$$

The above line means that the amplitudes of Signal1 and Signal2 are 0.6 and 0.4 respectively. And the standard deviation of the Noise is 0.1. Thus leading to following calculations of SIR_{dB} between both of the signals and SNR_{dB} for the Signal1.

$$SIR_{dB} = 10 \cdot \log_{10}\left(\frac{(A_{s1} \cdot 0.7)^2}{(A_{s2} \cdot 0.7)^2}\right) = 10 \cdot \log_{10}\left(\frac{(0.6 \cdot 0.7)^2}{(0.4 \cdot 0.7)^2}\right) = 10 \cdot \log_{10} 2.2 = 3.5[dB]$$

$$SNR_{dB} = 10 \cdot \log_{10}\left(\frac{(A_{s1} \cdot 0.7)^2}{\sigma(N)^2}\right) = 10 \cdot \log_{10}\left(\frac{(0.6 \cdot 0.7)^2}{(0.1)^2}\right) = 10 \cdot \log_{10} 17.6 = 12.5[dB]$$

3 Modelling message collisions in space environment

3.1 Statistics based message reception analysis

3.1.1 Single message detection model

An extensive single AIS message detection probability analysis for a satellite application is covered in [30]. Our aim in this section is to shortly review the modelling approach.

The FOV of a satellite covers multiple SOTDMA cells of ships. Although the area of a communication cell is defined by ship itself, there can also be used a model which divides the area of FOV of satellite into smaller cells, similar to the size of a common SOTDMA cell of a ship. For easier modelling the smaller cells are assumed to be of rectangle shape, continuously covering the FOV of satellite. Also the FOV can be assumed to be quadratic. This approach is depicted below in Fig. 25.

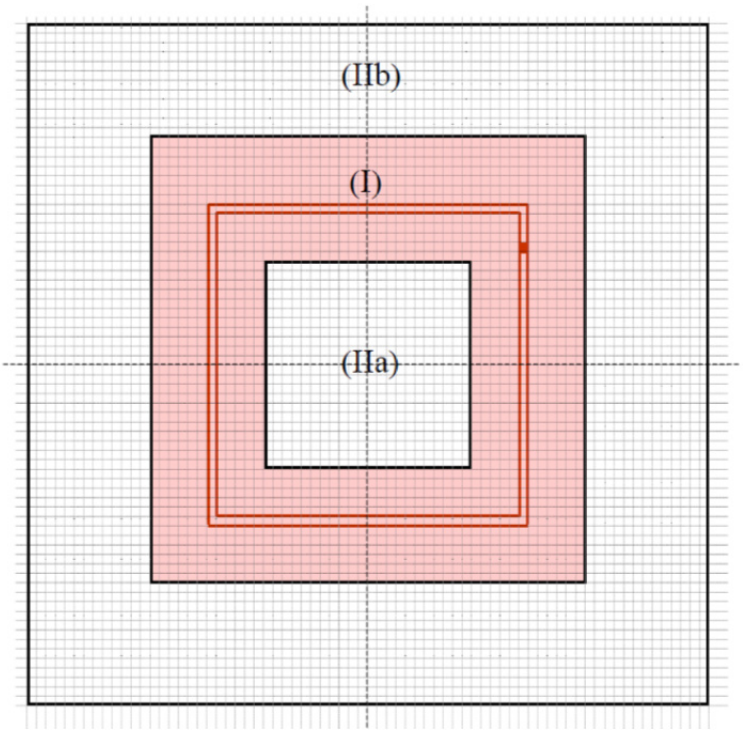


Fig. 25. The quadratic observation area for an AIS sensor at 600 km altitude with field of view to the horizon. The small filled red rectangular dot on the right upper part of the image reflects an area of a single SOTDMA cell. The image source is Figure 3.5 in [30].

The areas shown in Fig. 25 have following interpretation, cited from page 17 of [30]:

- "Main area (I) (marked by pink in the figure) contains the organized area being observed and all other organized areas that the AIS sensor receives transmissions from within the distance delay buffer relative to that organized area.
- Main area (IIa) contains all the organized areas that the AIS sensor receives transmissions from outside the distance delay buffer relative to the organized area being observed, and that arrives at the AIS sensor too early.

- Main area (IIb) contains all the organized areas that the AIS sensor receives transmissions from outside the distance delay buffer relative to the organized area being observed, and that arrives at the AIS sensor too late."

Further an additional assumption can be made, that within all cells there are the same number of ships, and that all of the contained ships are transmitting with the same reporting interval. For the case of swath width, which is the width of FOV of satellite's antenna, being below 800nm, translating to about 90^0 opening angle antenna for satellite of 700km altitude, following simplified formula was derived to calculate probability of receiving non-overlapped packet [30]:

$$P_s \coloneqq 1 - \left[1 - \exp\left(-\frac{N_{tot}}{37.5 \cdot n_{ch} \cdot \Delta T}\right)\right]^{\frac{T_{obs}}{\Delta T}}, \tag{29}$$

where

- P_s denotes detection probability of overlap free packet,
- N_{tot} denotes total number of ships within FOV of satellite,
- n_{ch} denotes number of channels used for reporting,
- ΔT denotes reporting interval of ships,
- T_{obs} denotes observation time of ship from satellite and
- 37.5 denotes constant for single channel capacity in messages per second.

Depending on ship speed and cruising characteristics, its reporting interval of its position broadcast message varies. However, it is assumed to be constant for all of the ships. It is also assumed that all the rest of AIS message types apart of position reports are minor and can be neglected. This also supports the assumption of packets being of 1 slot length.

The number of channels used for reporting relates to the reporting intervals of the ships. For example, a Class A ship travelling at 20 knots would report its position every 6 seconds in sequentially changing channels A and B. In other words, within each of the two AIS channels the ship will be reporting once in 12 seconds. The constant 37.5 in (29) denotes the number of available message time slots per second in a single channel. It is derived from taking 1 second interval and dividing it into 256 bit slots for data speed of 9600bps. Thus, the expression $\frac{N_{tot}}{\Delta T}$ means number of messages transmitted per second by the ships within the satellite's FOV and the expression $37.5 \cdot n_{ch}$ means number of total message slots available per second.

Let us take a very optimistic case and assume a 13 minute observation time per satellite pass for ships. Fig. 26 unveils detection probability analysis of AIS messages in satellite's FOV, showing (29) plotted for different N_{tot} and ΔT.

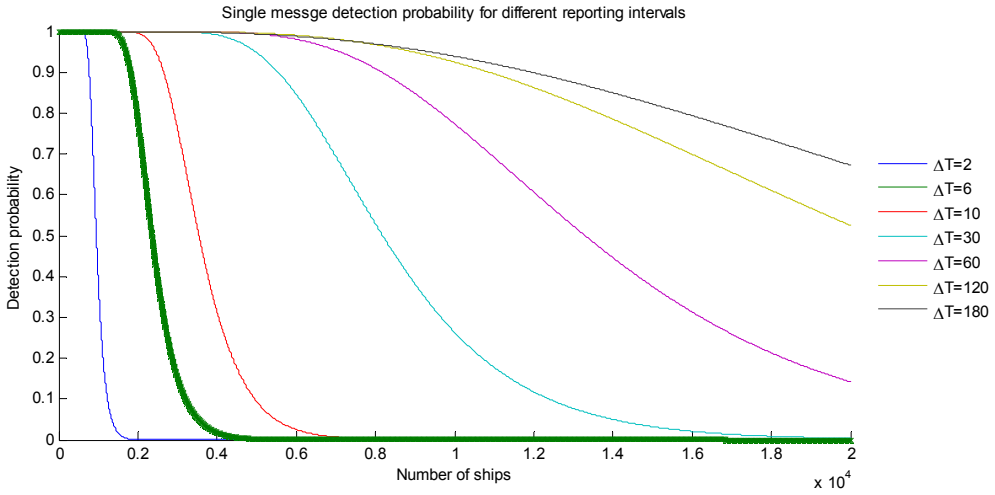


Fig. 26. Detection probabilities of single message arrival at satellite, covering 1 to 20'000 ships at certain reporting intervals ΔT. The observation time T_{obs} is 13 minutes.

Let us consider the case when 6 second's reporting intervals are used for the ships, as the plotted line coloured in thick dark green depicts in Fig. 26. One can observe that when ship density in the FOV of satellite has reached 4000 ships, the single message detection probability has almost reached zero. This signalizes that a satellite passing over high density zones, which hold more than 10'000 of vessels within antenna's FOV, will barely receive a single slot AIS message.

For a single SOTDMA cell where the report rate of all participating vessels is $\Delta T = 6$ seconds, the maximum capacity $2250/(60/\Delta T/2) = 450$ ships per cell. This would mean that a condition of just 9 fully used SOTDMA cells within FOV of satellite is enough to render the reception from orbit to nearly zero.

The $\Delta T = 180$ line represents detection probability for Space-AIS, which has a 3 minute reporting interval. At the first glance it appears to perform very robustly. However, one should take into account that the assumption of 13 minutes observation time is very optimistic. If the passing LEO satellite has wide angle antenna, this situation could be encountered once per day. The other satellite passes would be very near to horizon limiting communication time to few minutes. Thus, in the case the Space-AIS packets would collide, it would mean a missed single chance of message reception per its orbit. The question of an optimal transmission interval for the Space-AIS messages is still open. The 3 minute interval was proposed and got ITU acceptance from analysis of ship traffic simulation by the FFI (Norwegian Defence Research Establishment) working group [27], [31].

3.1.2 Overlapped message detection model

The simplicity of the ship detection probability model presented in the previous section relied on assumption that if a single message is received in a single slot, it can be demodulated.

However, more advanced demodulation techniques may take advantage of different signal levels of the arriving signals, frequency shifts or other means to split overlapping messages by advancements in signal processing algorithms or antenna technologies. Thus, a question arises about signal reception probability estimate for signals containing overlaps for receivers capable to process interfering signals.

There are three research articles available from James K.E. Tunaley regarding probability analysis of interfering AIS detection in orbit [32], [33], [34].

In [32] there is derived a probability estimation formula with additional parameters, which enable more detailed estimate of interference impact to signal reception. There are the following constants introduced:

- k – toleration factor, the number of message collisions that can be tolerated in each trial;
- q – corruption factor, the probability that an additional signal does not corrupt the message;
- s – overlap factor, the probability that nearby slot messages overlap, due big path delays.

In [32] it was assumed that the probability of receiving exactly *n* messages is given by Poisson distribution P_n, which is given by

$$P_n \coloneqq e^{-\lambda\tau_0}\left(\frac{\lambda\tau_0}{n!}\right)^n, \text{ where } \lambda \coloneqq \frac{N_{tot}}{n_{ch}\Delta T}.$$

This yielded that within a sequence of independent trials over observation time T_{obs}, the probability of receiving at least one uncorrupted message is equal to

$$p_k \coloneqq 1 - \left(\sum_{n=k}^{\infty} P_n\right)^{\frac{T_{obs}}{\Delta T}}.$$

To implement the idea that an additional overlapping message does not always disturb some already correctly demodulated ones, the q^n was introduced to denote probability that original message stays uncorrupted by presence of n additional signals. This yielded to the following equation

$$\sum_{n=0}^{\infty} q^n P_n = e^{-\lambda\tau_0}\sum_{n=0}^{\infty}\frac{(q\lambda\tau_0)^n}{n!} = e^{\lambda\tau_0(1-q)},$$

which together with the overlap factor s gives that

$$p \coloneqq 1 - \left(1 - e^{-(1+s)\lambda\tau_0(1-q)}\right)^{\frac{T_{obs}}{\Delta T}}. \quad (30)$$

The overlap factor s was introduced to account the case when messages from nearby slots overlap due to big signal path distances, occurring for a satellite's swat range being more than 800km [30]. The following Fig. 27 unveils single message detection probability analysis for various collision factors using (30).

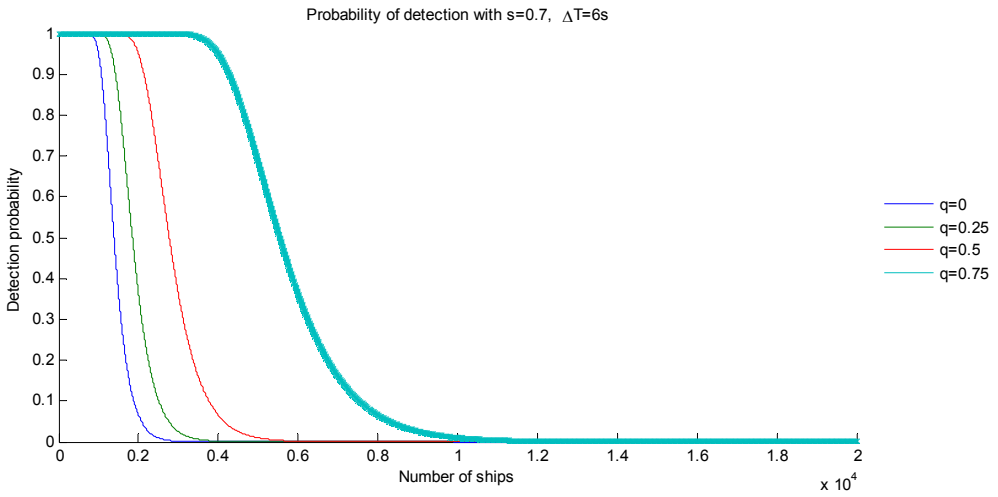


Fig. 27. Message detection probabilitiy with incorporated enhancemt of overlapped message processing capability. Plotted are various corruption factors q with overlap factor $s = 0.7$, observation time $T_{obs} = 13$ minutes, reporting interval $\Delta T = 6$ seconds.

Comparing the bold dark green coloured curve of Fig. 26 with the bold turquoise coloured curve of Fig. 27, one can observe that in the first case at 4000 ships per FOV there could be almost no ship detection predicted. Whereas in the second figure, accounting for improved demodulation algorithms, the ship density of 4000 gives nearly 100% message detection probability and tends to zero only as the number of ships in satellite's FOV approaches toward 10'000. This points to a much greater message detection probability for receivers with an overlapped message processing capability. However, the k, q and s parameters should be determined individually for a dedicated antenna, receiver and satellite constellation, leaving the question of the implemented AIS message reception capability open. Meanwhile, the named parameters enable to analyse reception rate statistics for some specific AIS detection scenarios in a very comfortable way.

3.2 AIS data simulation based model

While statistics is a very convenient tool for modelling and predicting events, its assumptions are very much dependent on the assumed constants incorporated in source equations. However, the prediction or estimation of the constants for space application is certainly not trivial.

Motivated by needs to test different AIS demodulation algorithms and to evaluate and test its performances at certain message collision scenarios as well as to estimate and predict the characteristics of message overlaps more precisely than the statistical methods described in the previous two sections can offer, an AIS data simulation software was developed by author if this dissertation. Its application as an AIS-Testbed was published in ASMS/SPSC2012 conference [49] and also presented C-SIGMA workshop [50].

The DLR AIS Simulator software is composed the following 3 parts.

- **Ship Generator** – generates AIS messages of ships, taking into account SOTDMA protocol and satellite orbit. Data source can be either randomly generated or an AIS message data file in the NMEA format could be provided. To generate the random data, the wished longitude-latitude rectangle coordinates and number of ships within the selected area should be selected, along with the uniform reporting rate of choice. Whereas for a file data source the

persistent ship positions, cruising speeds and directions will be read and simulated accordingly. The satellite positon in orbit is calculated from an input of the TLE (Two Line Elements) data file for a selected time range. Having the ships, its packets and satellite in their places, signal receiving angles, attenuations, Doppler shifts and time shifts are calculated and written to a file. The Ship Generator software also allows to model different SOTDMA cell sizes as well as to apply specific types of ship antennas and transmission powers.

- **Signal Modulator** – it reads the formatted data generated by Ship Generator, translates each of the AIS messages into IQ signal and incorporates the simulated distances, angles, signal delays and attenuations for each of the AIS packet for its reception on the satellite. The simulated IQ signal can then be modulated onto a carrier frequency using a Software Defined Radio or other means of Digital to Analog Converter for a real world signal processing test. Alternatively, the generated IQ waveform can be also directly used for demodulation processing and statistical analysis.
- **Scenario Visualizer** – it loads the model provided by Signal Modulator in AGI/STK satellite modelling software environment and visualizes the satellite flight path in orbit, ship cruising paths along waters, as well as shows available communication contact times.

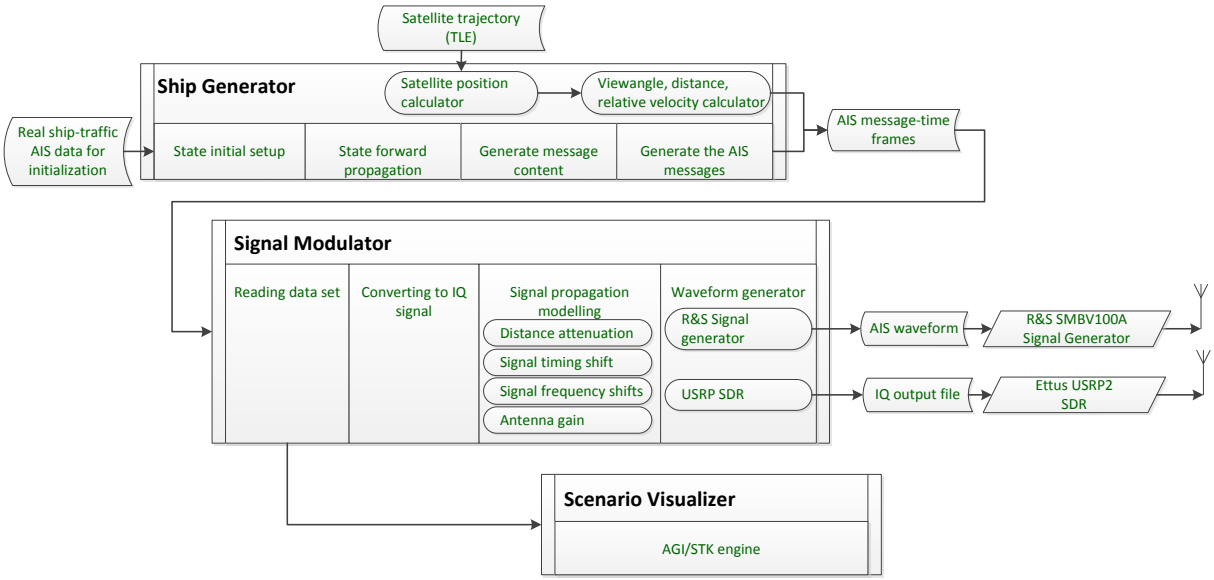


Fig. 28. A diagramm of the architecture of the DLR AIS Simulator software, developed by the author of this work.

The Ship Generator has following input parameters for case with Random input Data.

- Longitude range,
- Latitude range,
- Number of ships within selected area,
- Maximum number of SOTDMA cell overlaps available for the random generator,
- SOTDMA cell radius, usually set to 40km,
- Should the position reports be generated (AIS Message 1),
- Should voyage data repots be generated (AIS Message 5),
- Reporting interval, summed over A and B channels

- Channel for which data should be generated,
- Satellite TLE input data file,
- Time offset from the epoch of the TLE data,
- Duration the simulation.

For the case when a persistent AIS data set is available, for example, the received AIS data over a dedicated area from a flight campaign with airplane, the AIS messages can be input by a source file instead of providing first three parameters for the random generator. The report rate in this case can be either left to be default or set to uniform value. In the default case the ship reporting rate will be calculated according to its travel speeds while in the uniform case the selected reporting rate will be applied to all of the simulated ships.

To account for different antenna installations on ships, the ship antenna transmission power is assumed to be randomly set between 11 and 12.5 Watts. For the simulations described in this work, the installation of a common dipole antenna with a unity gain is assumed.

The number of SOTDMA cells tells the maximum permitted number of message overlaps. To reach realistic simulation conditions, it should be set higher than the SOTDMA cells required to fit all data in. The size of SOTDMA cell around each ship is assumed to be uniform. Within a single SOTDMA cell radius, for each of the ships the simulator calculates Nominal Start Slot, Nominal Slot, Report Rate, Nominal Transmission Slot according ITUR regulation for the AIS communication [1]. Below is a simplified version of Ship Generator GUI (Graphical User Interface) for data input. Along with some generated results.

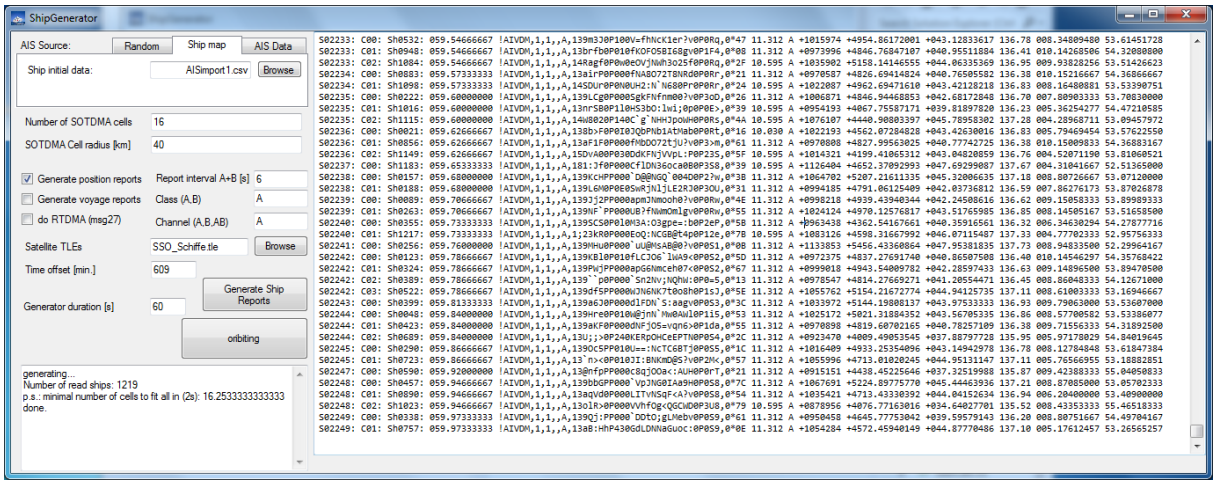


Fig. 29. The GUI view of the Ship Generator taking as input detected AIS messages and generating transmissions with uniform report rate 6 seconds.

Signal Modulator does signal processing part for the data generated by Ship Generator. This software performs GMSK modulation of the AIS messages, adjusts its timing, signal strength and Doppler shift according to satellite position and antenna characteristics. Finally it sums the IQ signals in single IQ data array, enabling the simulated data to be exported and transmitted. This software was developed for a Space-AIS testbed purpose and it can serve to derive following receiver characteristics:

- Sensitivity to noise levels,
- Permissible Doppler shift range and processing capability within the range,

- Tolerance against disturbances from overlapping signals,
- Behaviour in selected scenarios of packet collisions with specific Doppler, timing and amplitude characteristics.

Signal Modulator is a comprehensive and useful tool for testing AIS receivers' capabilities to perform in a space environment and to evaluate the receivers' tolerances against definite packet collision scenarios expected in a space application. This software was used also for testing the DLR satellite AISat1.

The generated and modulated data can be passed to Scenario Visualizer, which using AGI/STK environment shows the satellite pass over ship region, its antenna footprint and timeline of contact times.

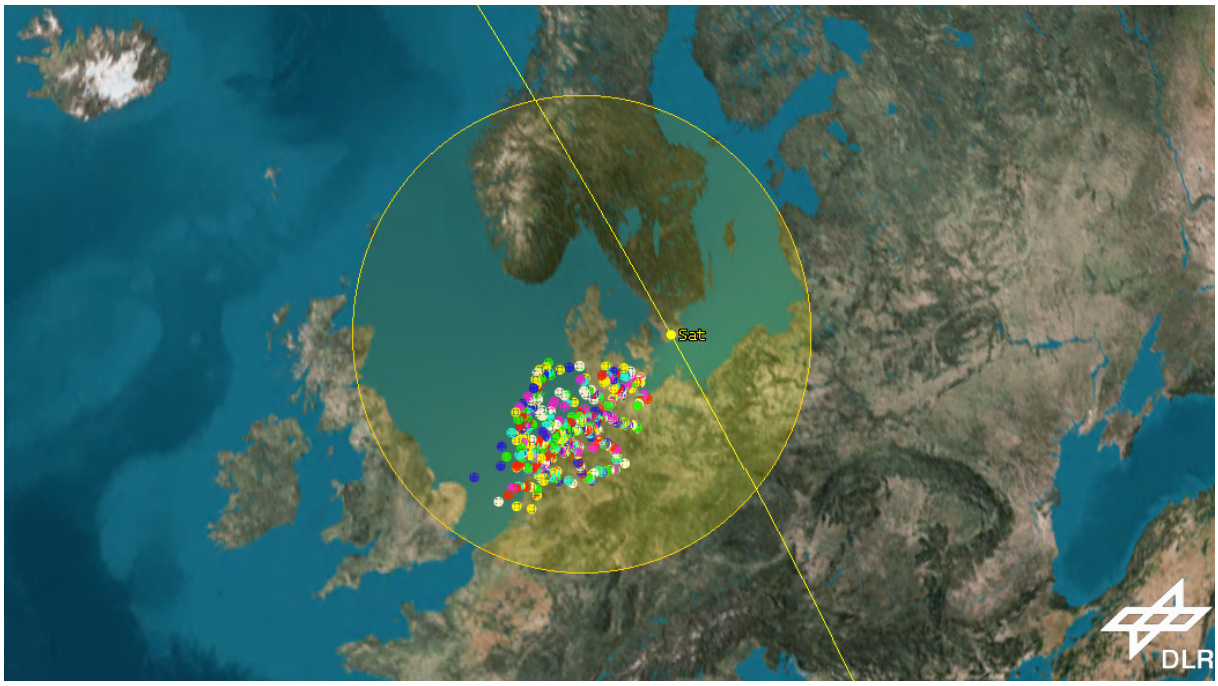


Fig. 30. Satellite pass over the German Bight, with 90^0 opening angle cone antenna. Ship data source is AIS data gathered by a dedicated flight cmapaign with airplane over the region, accomplished by DLR in 2010. Overall 1219 ships were detected.

The real ship simulation approach enables Ship Modulator software tool to calculate very accurate and deeply detailed statistics for packet collisions and channel load. Below in Fig. 29 are shown analysis of packet overlaps per slot for a simulation case, when the observed ships from German Bight were made to report at 2 seconds and at 15 seconds constant intervals for time duration of 60 seconds.

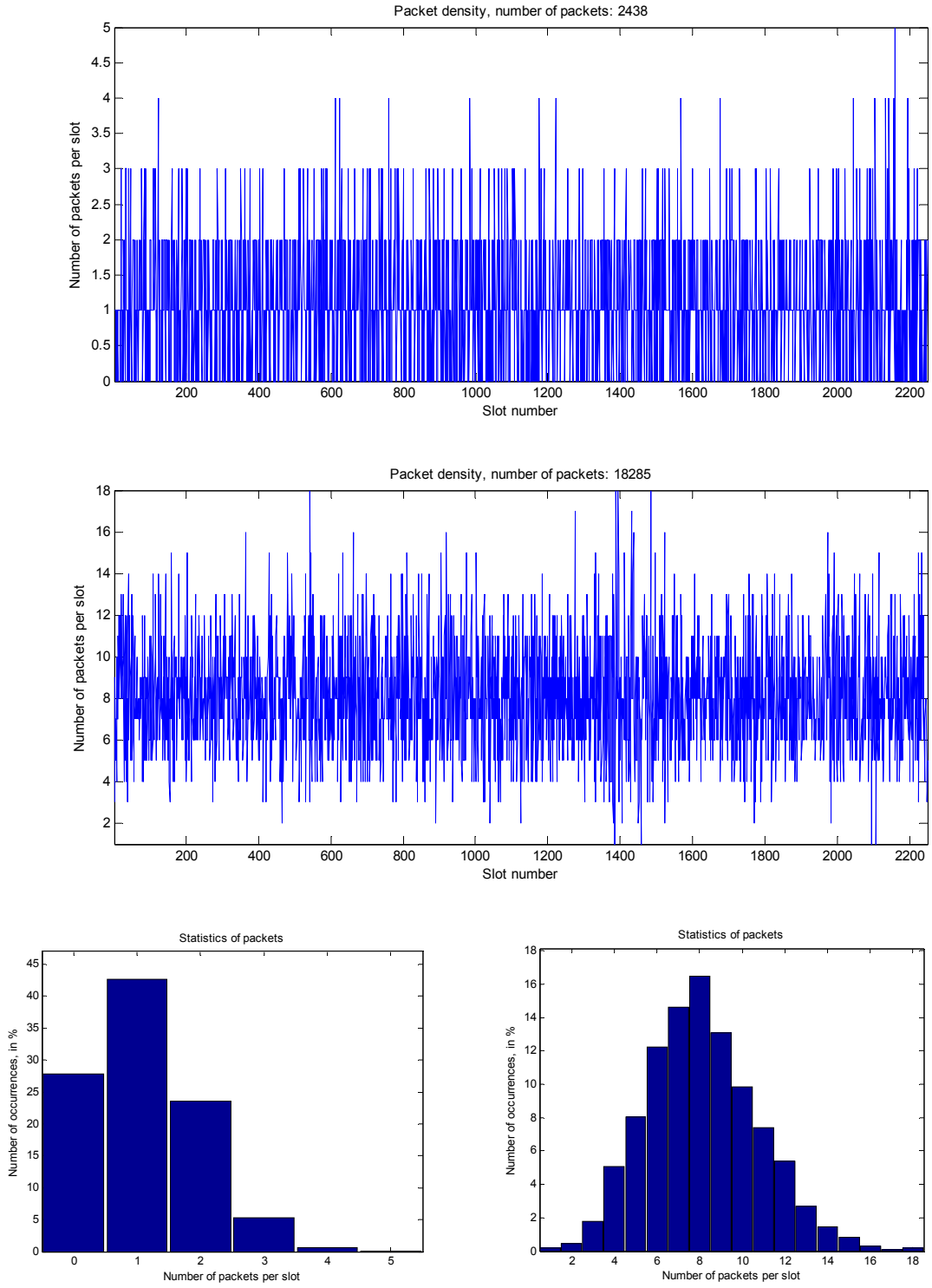


Fig. 31. Statistiscs of packet overlaps, as generated by simulated ships in German Bight with static reporting intervals 15 seconds (Figures a, c) and 2 seconds (Figures b, d).

Finally, by simulating 5 minute pass of a satellite over the ship region, the received power for each of the received AIS packets can be calculates and is shown below in Fig. 32.

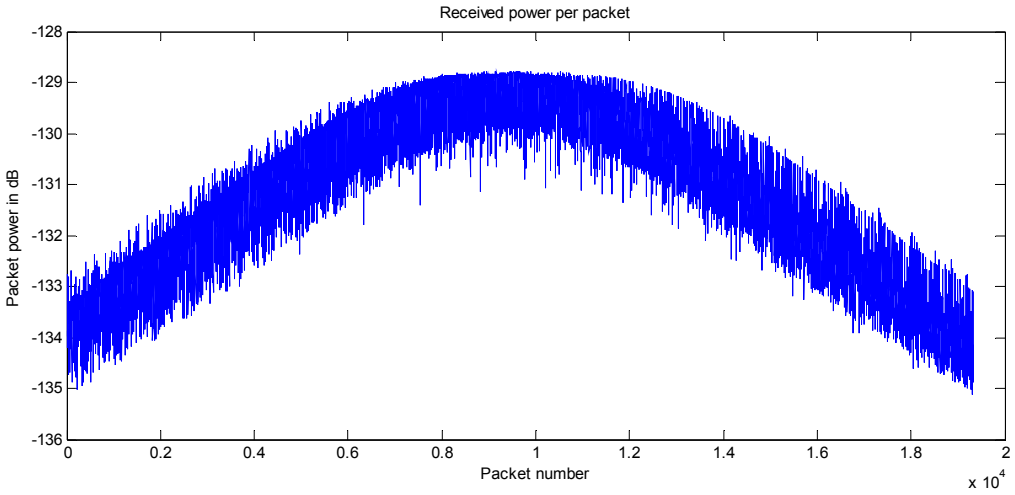


Fig. 32. Packet power per satellite pass. A 120° opening angle antenna assumed.

The figures within this section show that simulations based on a real ship AIS data can be used for a very detailed analysis of collision statistics in AIS channel. Further use of this simulation tool will be addressed in Chapter 6.

The comprehensive and detailed information of packet overlap properties made available with the help of the DLR AIS Simulator tool helped to further pinpoint elaboration issues in packet collision processing. Collisions cases were decided to be divided into groups, to effectively approach message extraction algorithm development. This topic is addressed in the next section.

3.3 Message collision cases

For an analysis on overlapping properties of two messages and to categorize these to help estimating an appropriate processing algorithm, the author of this work proposes to divide message collision scenarios in the four following groups.

I. **CC1 (Collision Case 1): Two AIS messages completely overlap having very similar amplitudes.**

We consider two ships are having a similar signal path length to a satellite in orbit, transmitting with the same power and having the same antenna patterns, with the consequence that both transmissions will be received by a satellite at the same time and with equal signal powers. Additionally, these two ships may be separated in a distance greater than that of the radius of SOTDMA cells and AIS position report messages transmission in the same time slot. The collision case as shown in Fig. 33 is expected.

A similar situation to that given directly above is when the messages overlap with a slight time shift of few bits, as shown in Fig. 34, will also be considered under this collision case. As a reference example, this could reflect the situation when two ships have minor differences in distances to a satellite, yet the one being more near having a slightly lower transmission power.

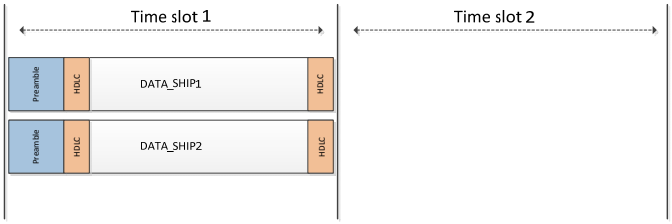


Fig. 33. Schematical drawing of two AIS packets colliding in the same time slot with no time shift (CC1). The parts of preamble and HDLC flag of the messages are shown in blue and orange respectively.

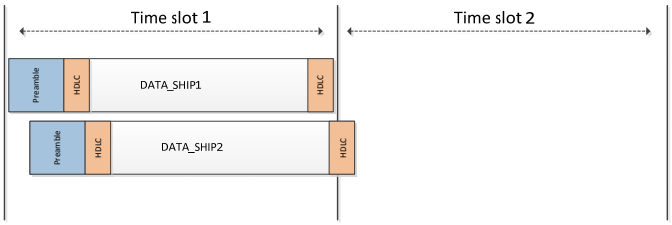


Fig. 34. Schematical drawing of two AIS packets colliding in the same time slot with very slight time shift (CC1).

II. CC2 (Collision Case 2): two AIS messages partly overlap having very similar amplitudes.

In the second scenario the two ships shall have different signal path lengths to a satellite, varied transmitting powers or antenna structures, so that signals will arrive timely shifted but with equal or almost equal amplitudes. The AIS data reception would appear as that given Fig. 35.

A similar situation is when the messages overlap just at the end of one and beginning of second one, as shown in Fig. 36, will also be considered under this case. As a reference example can be considered the situation when the start and end bits of two packets in two consecutive slots are overlapping due to different signal traveling time intervals. This can occur, when a distant ship sends an AIS packet in the direction of a satellite with a high power and a nearby ship with a lower power setting.

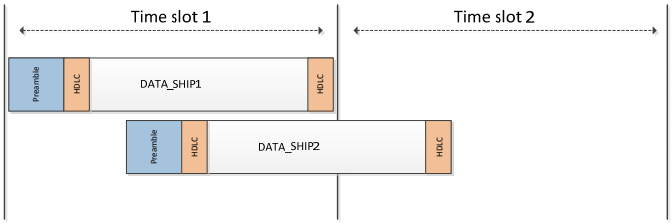


Fig. 35. Schematical drawing of two AIS packets transmitted in the same time slot and colliding over two time slots. Packets have significant time shifts (CC2).

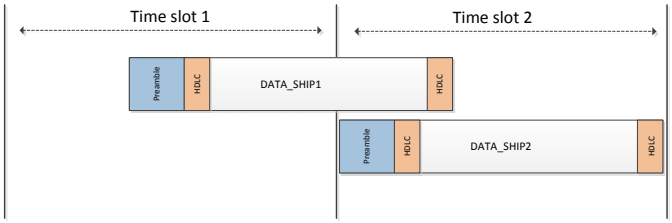


Fig. 36. Schematical drawing of two AIS packets transmitted in two disinct time slots and colliding over two time slots. Packets have significant time shift (CC2).

III. CC3 (Collision Case 3): two AIS messages partly overlap, but data bits are free of overlaps

The third scenario represents a packet collision case, when the preamble or HDLC flag fields of packets do overlap, creating a disturbance in bit recognition and packet identification. As this can be solved by using sliding checksum checking algorithm, avoiding the need to have HDLC flag bits, it will not be in more detail considered in this work.

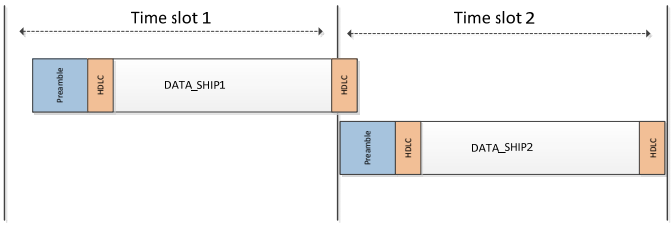


Fig. 37. Schematical drawing of two AIS packets colliding over two time slots. Data parts safe of overlaps (CC3).

IV. CC4 (Collision Case 4): two AIS messages partly or fully overlap, having significant frequency or amplitude differences

The fourth scenario reflects the cases when significant Doppler shift or amplitude differences are present, independent of overlap timing properties.

An example reflecting big Doppler shift difference is a scenario when a satellite with wide angle antenna is receiving one ship in far front and one in far back relative to its flight vector direction. Here the first could have a +2.4kHz Doppler shift and the second a -2.4kHz Doppler shift, together building a 4.8kHz frequency gap between the signals, which could be used to separate them using sensitive frequency filtering algorithms.

A significant signal amplitude difference could account for a two ship constellation when the first ship is sending in a lower transmission power, like a class B ship, a ship within a port or a ship with an unfavourable antenna pattern, while the second ship is transmitting with a high power setting and a beneficial signal transmission environment. For this case a dedicated CIC (Consecutive Interference Cancellation) methods could be used.

Solution methods for cases CC3 and CC4 will not be analysed in further detail here, leaving the focus of this work to the more sophisticated and more often occurring packet collisions of cases CC1 and CC2. For the analysis of CC1 and CC2 scenarios, the collision scenarios with Doppler shifts below 2.4kHz for wide angle antenna and below 700Hz for narrow angle antenna will be considered. The upcoming Chapter 4 addresses solving the packet collisions of the CC2 scenarios and the Chapter 5 does that for the CC1 based collision scenarios.

4 Approaching joint demodulation with shape recognition

This chapter will introduce a new overlapped AIS signal demodulation approach based on a ML (Maximum Likelihood) estimation of IQ signal waveforms. The developed algorithm is named Joint Shape Based Demodulator, shortened as JShBD, by author of this document.

The theoretical base of JShBD method has evolved from learning MLSE (Maximum Likelihood Sequence Estimation) application in VDA (Viterbi Decoding Algorithm), which is widely used in DSP (digital signal processing) applications to find the bit sequence with the highest probability in a signal containing ISI (Inter-Symbol Interference) [17]. In the first section, Section 4.1.1, the idea of VDA is introduced from literature references. The further sections present how the author of this work extends and applies the idea of VDA to overlapped AIS signal joint estimation.

We begin this chapter with a short introduction to VDA and a demonstration of the ML estimation as means to shape recognition. We will outline the JShBD principles and apply optimizations for signal estimation like noise reduction and basis transformation. Finally, we will derive a demodulation method and show its performance analysis for two overlapping Doppler shifted signals.

4.1 Introduction to Viterbi Decoding Algorithm

4.1.1 Single sample based approach

In this section we provide an introduction to the framework about trellis. The material of this section is very much influenced by the work of [39].

Let us consider the case of AIS signal, with the GMSK modulation scheme in use. The FSM (Final State Machine) in this case uses binary signalling by outputting two states +1 and -1, each of them being one bit time long T_b. In [39] to explain symbol equalization the output filter of $3T_b$ length was chosen, showing how “smearing” of the output filter can be accounted to perform valid symbol equalization. It involved searching for a path through decoder trellis diagram to find the transition path with the highest probability. There was also introduced a formulation that for random scalar values at input APP (a posteriori probability) of a decision on certain value can be written as

$$P(d = i|z) := \frac{p(z|d = i)P(d = i)}{p(z)} \quad \text{for} \quad i = 1 \dots M, \tag{31}$$

where the a posteriori probability $P(d = i|z)$ is the conditional probability on the data d belong to the i-th signal class. In the case of GMSK, which uses binary signalling, there are two classes corresponding to 0 and 1 bit signals. The $p(z|d = i)$ is the PDF (probability density function) of the observation z conditioned on the signal class $d = i$. The $P(d = i)$, called a priori probability, is the probability of occurrence of the i-th signal class. The function $p(z)$ is unconditional PDF of the received signal z, taken over the entire space of signal classes.

When making a decision, the a posteriori probabilities can be compared and the signal class with highest probability can be selected so that

$$P(d_k = +1|z_k) > P(d_k = -1|z_k)\,. \tag{32}$$

If the inequality holds, then the sample z_k equals to bit +1 and otherwise is equal to -1. Substituting the APP evaluation from (31) within its comparison in (32) one obtains that

$$\frac{p(z_k|d_k = +1)}{p(z_k|d_k = -1)} > \frac{P(d_k = -1)}{P(d_k = +1)}\,, \tag{33}$$

where the denominator of the APP equation cancels out. In case there is no information of probabilities for signals taking the values +1 or -1, both the a priori probabilities *P(d=i)* can be assumed to be equal, thus from (33) giving that

$$\frac{p(z_k|d_k = +1)}{p(z_k|d_k = -1)} > 1\,. \tag{34}$$

This means that signal class with the highest probability will be selected for z_k. This approach is known as the ML criterion. The following graph gives a visualization of the inequality given in (34).

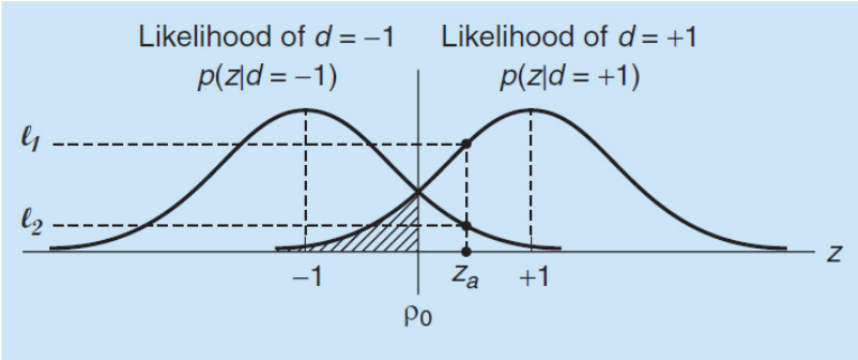


Fig. 38. Likelihood functions for bits +1 and –1 [39].

The graph above can be interpreted as follows. Plotted are two probability functions. The rightmost function $p(z|d = +1)$, called the likelihood of $d = +1$, shows the PDF of the random variable z conditioned on $d = +1$. The leftmost function illustrates the PDF for $d = -1$. The random variable z at presence of Gaussian noise can take values between $-\infty$ to $+\infty$. Therefore, the probability functions are infinite. For some input value z_a there can be two points found on each of probability functions, corresponding to the likelihoods of z_a to belong to the $d = +1$ or $d = -1$ signal classe. The class with the highest likelihood should be selected. In the case of Fig. 38 above, the $p(z|d = +1) > p(z|d = -1)$, leading to decision that z_a belongs to the symbol $+1$.

The above explanation holds for the case when input samples correspond to random variables. However, in the case of signals being "smeared" due ISI (Intersymbol Interference), the sample under consideration is influenced by previous and next bit values, making the decision of the probability functions more involved. To summarize the idea presented in [39]: instead of considering single sample and relating it to single symbol class, a sequence of symbols is considered, corresponding to a minimum PDF from all possible bit variation cases. For a received sample sequence $\boldsymbol{Z} = \{z_k,\ z_{k+1}, \dots\}$ to find the corresponding most likely random sequence $\boldsymbol{U} = \{u_k,\ u_{k+1}, \dots\}$, where the u_k is a voltage level of the FSM during the time interval $(t_k,\ t_{k+1})$, the ML rule (34) has to be modified to consider sample sequences:

$$p\left(\boldsymbol{Z}\middle|\boldsymbol{U}^{(j')}\right) = \max_{j} p(\boldsymbol{Z}|\boldsymbol{U}^{(j)}) \ . \tag{35}$$

To avoid considering probabilities of long sequences, which have number of variations growing exponentially relatively to contained values, there is introduced encoder and decoder trellis diagrams. Branch words of encoder trellis diagram show voltage states for transition from one bit state to another. The decoder trellis diagram at its branch words show a distance between received sample voltage value at time t_k and that of possible transitions. Thus, creating a graph with weights, corresponding to probabilities of state transitions. Selecting the path with the smallest weight gives also the sequence with the highest probability. The trellis graph of bit transitions should represent just a limited number of transitions, in binary case having just four: 0 to 0, 0 to 1, 1 to 0 and 1 to 1 transitions. The graph can be made as a sliding window of the smallest weight path, by selecting the minimum weight path to each of the states for each of the oldest transitions within the window. Thus, using a small amount of memory, while finding the most probable symbol sequence.

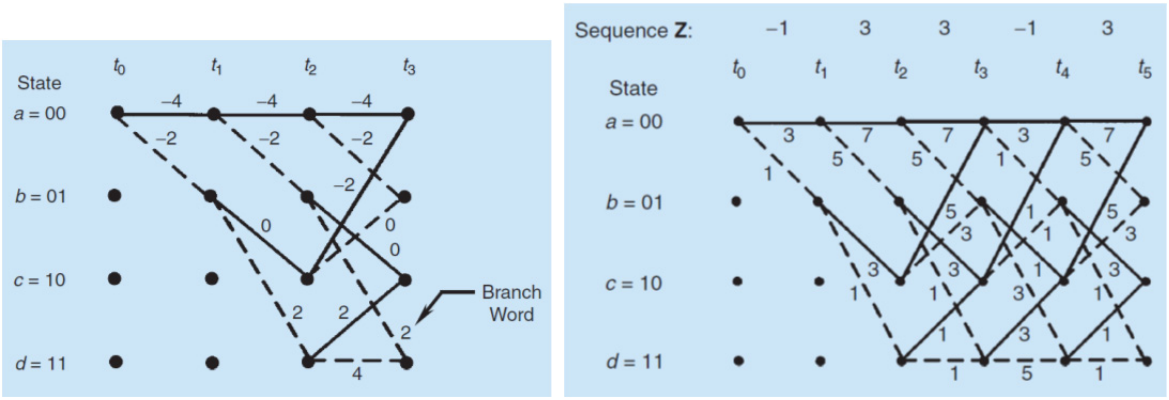


Fig. 39. (a) An encoder trellis diagram. (b) Decoder trellis diagram [39].

4.1.2 Shape based approach

Instead of using ML for single IQ signal samples as outlined in previous section, the author of this work takes his own approach to implement VDA using ML estimation for a range of IQ samples, where each of the selected ranges correspond to voltage state transition curves. The Section 4.2 will dive into more detail on an approach of selecting curves for signal approximation purposes.

Let us investigate the concept of the shape-based approach to VDA and let us start by considering an IQ waveform of an AIS signal. In Fig. 40 below the blue line highlights the voltage state transition from +1 to -1 within some part of the IQ waveform.

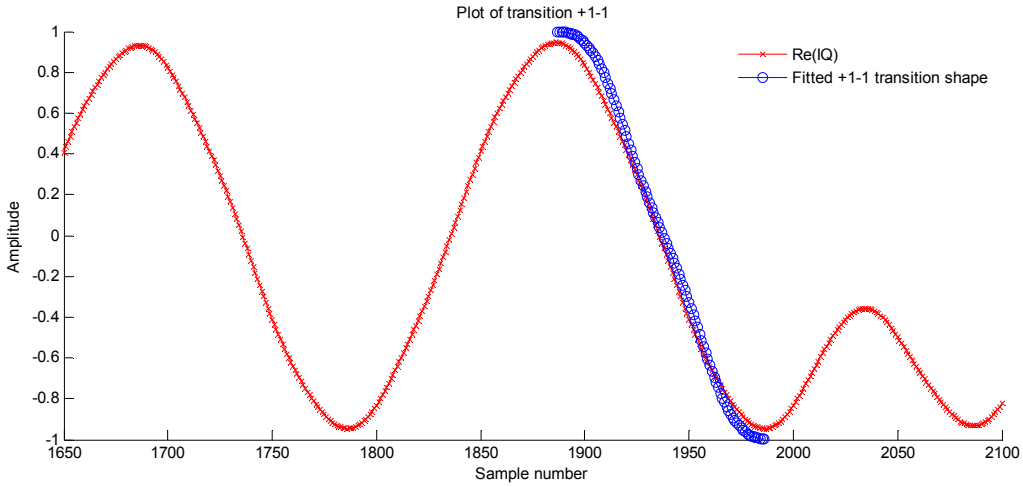


Fig. 40. IQ data of AIS signal in red with a fitted transition shape from +1 to -1 state in blue.

As one can observe in the IQ data of an AIS signal in Fig. 40, there are 4 voltage state transitions to be expected in the IQ signal under analysis, namely the +1+1, -1-1, +1-1 and -1+1. These can be similarly interpreted as the bit transitions in the trellis diagrams shown in Fig. 39. From this observation it seems reasonable that the whole IQ signal could be divided into small parts, here called chunks, corresponding to the length of a single transition. For each chunk of the provided IQ data the most appropriate transition from the four ones could be assigned, leading to approximation of the IQ data with the transition shapes. Finding the best approximating curve can then be accomplished by calculating the L^2 distance of the fitting transition shapes to the samples in the IQ signal chunks. The distance between the approximated and input signal samples can be written in L^2 norm as follows.

$$\Delta^{k,j} := \left\| C^k - S^j \right\|_2 = \sqrt{\left(c_1^k - s_1^j\right)^2 + \left(c_2^k - s_2^j\right)^2 + \cdots + \left(c_n^k - s_n^j\right)^2} \tag{36}$$

The $\Delta^{k,j}$ in (36) is the L^2 norm distance between chunk k and shape j, $C^k = \{c_1^k, c_2^k, \ldots, c_n^k\}$ are the n samples from the input IQ signal starting at sample k, $S^j = \{s_1^j, s_2^j, \ldots, s_n^j\}$ are the n samples from the j-th shape and n is the number of samples for single transition required. For the case shown in Fig. 40, where the bit oversampling rate is 50, the length of the transitions is $n = 100$ samples.

With the equality given in (36), there can be calculated the distances between the input IQ signal and all the four possible transition shapes and the one with the minimum distance selected,

$$\Delta^{k_j} := \min_j \Delta^{k,j} \,. \tag{37}$$

Meaning, that the Δ^{k_j} denotes the choice of transition shape j as the best approximation for the chunk k. The finding of the transition sequence Δ^{k_j} directly leads to a recognition of the bits contained within the input IQ data.

While the approach described above will select the transition shape with a minimum distance to one specific chunk, the resulting approximated signal may contain non valid discontinuities. For example, if the transition -1+1 has taken place before, then as the next transitions can only be +1-1 or +1+1, in order to preserve signal continuity. The usage of the continuity property of the signal can help to improve the signal approximation. Thus, we need to introduce a continuity rule to force the transition sequences of Δ^{k_j} to be a continuous voltage curve along a time axis. For this purpose the

beauty of trellis decoder with branch words, distance metrics and minimum path selection, as was introduced in the VDA framework in the previous section, can be very conveniently applied.

The previously calculated distances $\Delta^{k,j}$ of all four shapes j to all chunks k can be assigned as branch words on the trellis diagram, building a graph. The path with the least weight is expected to represent the best approximating curve to the input IQ signal. To force the continuity rule into action, a transition matrix should be introduced and used to select the valid paths and calculate its weights. Table 5 represents a valid shape transition matrix for the shapes derived in 4.2.2. It can also be seen as probability matrix of sequential transitions.

To / From	Sh1	Sh2	Sh3	Sh4
Sh1	0	1	0	1
Sh2	1	0	1	0
Sh3	1	0	1	0
Sh4	0	1	0	1

Table 5. Transition matrix for transition shapes. 1 means possible, 0 impossible.

The matrix from Table 5 above can be read as follows. The transition from Sh1=[+1-1] to Sh1=[+1-1] is not possible, since the finishing state of first shape does not match the beginning state of second shape. While the transition from Sh1=[+1-1] to Sh2=[-1+1] is possible. Using this matrix to select valid paths and assessing its distances, the continuity rule of the IQ data can be imposed and the best approximating path of minimum distance transition shapes be found, revealing the initially incorporated bit sequence of the analysed IQ data.

A question arises for how many bit times does a single bit influences its surrounding transition shape values.

Initially VE was thought to be means of evaluating “smeared” bits, when bit value within one interval influences voltage value in surrounding bit intervals. The width of influence region would then depend on the wideness of the bit filter, which in [39] was chosen to be $3T_b$.

For the AIS application the bit filter is the Gaussian filter used in the GMSK modulation of the AIS messages. As the description of the filter in Section 2.1.2 noted, its chosen width is $4T_b$ for a convenient programming and good signal approximation purpose. Therefore, although within the Gaussian filtered signal, as shown previously in Fig. 16 d) and in the first row of Fig. 41 below, one can observe unity amplitude hill or valley on every bit clock interval, the actual pulse response for every bit stretches to $2T_b$ rise time and $2T_b$ fall time before and after the peak. In Fig. 41 there are shown influence zones of the nearby bits for a single transition, corresponding to a transition shape +1-1.

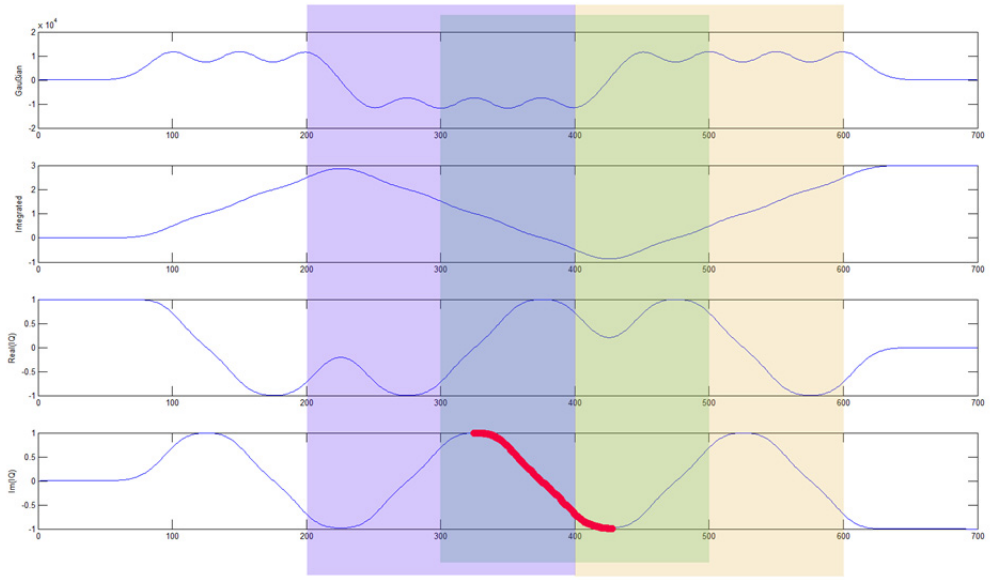


Fig. 41. Bit influence regions for single transition marked in red. Modulated bit sequence: [+1,+1,+1, –1, –1, –1, –1,+1,+1,+1,+1]. Meaning of vertical axis is amplitude, horisontal axis is sample number. The content of this figure is simmilar to Fig. 43, where it is shown in higher zoom and detailed axis description.

As Fig. 41 above reveals, the bit number 7, experiencing its peak -1 value in Gaussian filtered signal at sample 400, gives the main influence for transition chunk C^{325}, shown with a red thick line in the imaginary part of IQ data with samples 325 to 425. One can observe that the value for the red line is influenced by bit zones of bit 5, in violet colour, and very slightly of bit 9, in orange colour.

In this works it will be assumed that the transition shapes are influenced only by bit values of curve starting and ending positions, meaning violet and green zones in Fig. 41. The small interference from the following bits will be discarded. This approximation is reasonable, since the neglected part of the yellow zone corresponds to 0.00034% of the total area covered by the $4T_b$ Gaussian filter, used in GMSK modulation for AIS application. This approximation allows considering transition shapes as interference-free from passed or following transitions.

This answers also the above posed question about the range on influence of a single bit to transition shapes. Therefore the processing memory for finding the minimum path selection from trellis diagram with the continuity rule applied, suffices to be of two transition shape length.

The VDA approach outlined in this section for finding a valid voltage state transition path with highest probability will be used in the sequel work and the results calculated with it will be referenced as calculations with the VDA in use. The VDA is expected to improve demodulation process in presence of a high noise level, with the neighbouring bits helping to recognize the best fitting bit transition shape.

4.2 Joint-Shape-Based demodulator (JShBD)

This section will deal with a derivation of basic shapes which serve as building blocks for IQ data curves. These will be extensively applied for shape recognition throughout this and the next chapter. The derivation will be done for Signal1 first and a time shifted version for Signal2 will be introduced afterwards.

4.2.1 The principle of shapes within IQ data containing overlaps

Let us take a look at IQ data of two overlapped AIS signals with some small noise and without a Doppler shift. A short excerpt of such data is shown below in Fig. 42.

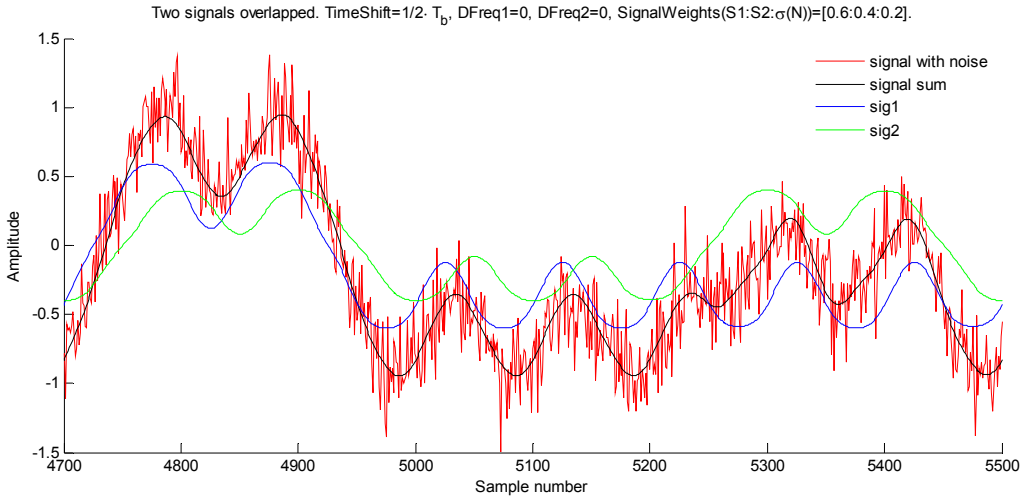


Fig. 42. Overlap of IQ data of two signals with very similar amplitudes.

As Fig. 42 discloses, the two IQ signals summed construct a signal line, in the figure the green and the blue lines summed construct the black line, which contain characteristics of both underlying signals. It is recognizable even in presence of noise, as the red line presents.

From the observation above there was the following idea developed by the author of this dissertation. If the IQ signal under analysis contains overlapped transmissions, a correct knowledge of its parts would give in its sum a similar IQ signal to the initial one. And each of the individual sub-signals can be separated into fractions, corresponding to the state transitions between +1 and -1 with its corresponding amplitude. Thus it means that a Doppler-less IQ signal containing overlaps could be projected on a basis composed of such transitions shapes, called basic shapes. A wise construction of such a basis could help to reconstruct the overlapping signals. This idea is being followed within upcoming sections.

4.2.2 Constructing basis shapes for Signal1

The first step to start with is deriving basis shapes which constitute IQ waveform of a single GMSK modulated AIS message signal. In further text the term “basis shapes” of Siganal1 will be also referred as “basic shapes”, as they are the elemental building blocks for signal waveforms. A stepwise representation of waveform construction from bits to IQ signal is shown below in Fig. 43. For details on GMSK modulation, please refer to Section 2.1.2.

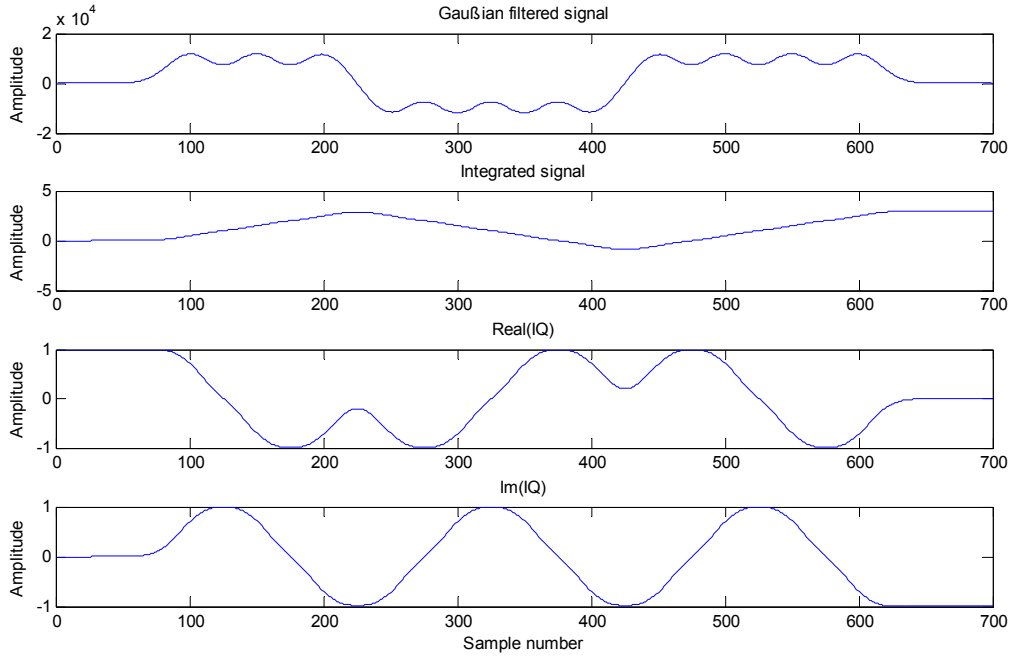


Fig. 43. The GMSK modulation steps of the bit sequence [+1,+1,+1,-1,-1,-1,-1,+1,+1,+1,+1]. a) Gaussian filtered bits b) Integration of the Gaussian filtered signal c) cosine of the integrated signal d) sine of the intergrated signal.

Assumed the Doppler frequency to be zero, the receiver is getting IQ data in input with central frequency $f_c = 0$ and we should search for basis shapes within the IQ data. Since the GMSK signal is binary, only +1 and -1 symbols are present, thus giving four possible signal level transitions for the contained waveform transitions: +1-1, -1+1, +1+1 and -1-1.

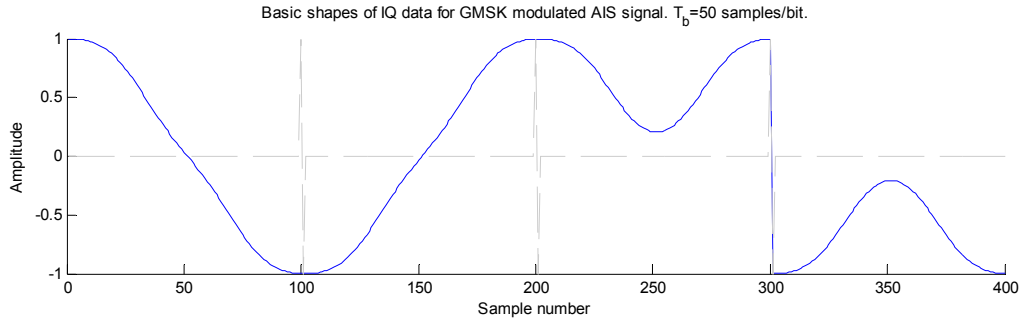


Fig. 44. Basic shapes of IQ data from GMSK modulated signal, each 100 samples long. The shapes are separated by the grey vertical and shown in order [sh1, sh2, sh3, sh4].

The four shapes from Fig. 44 above will be further referenced as basic shapes, basis shapes of the first signal or Shapes1. From the figure it can be noted that each shape has a length of twice the samples per bit. That is due to the fact that the every next bit is encoded in sequentially changing imaginary and real parts of the IQ signal. However, as long as we analyse a single part of the signal, for example the real part, it is comfortable to work with these four shapes of the $2T_b$ length.

With the approach of shapes being $2T_b$ samples long, whereas the filter in use being $4T_b$ long, the ISI effect between the shapes is partly cut, as noted in Section 4.1.2. It is still a reasonable approximation for GMSK signals. To better approximate the ISI in the waveforms of choice, the shapes were taken from the middle of the generated IQ signal shown in Fig. 43, not from the

beginning or end, where the ISI would be absent on one of the sides of the IQ signal. The following parts were taken as Shapes1 from the IQ waveform from the GMSK modulation steps in Fig. 43:

```
sh1 = shI(istart+8*osr:istart+10*osr-1); %transition +1-1
sh4 = shI(istart+2*osr:istart+4 *osr-1); %transition -1-1
sh2 = shI(istart+4*osr:istart+6 *osr-1); %transition -1+1
sh3 = shI(istart+6*osr:istart+8 *osr-1); %transition +1+1
```

where

- `shI` denotes the real part of the IQ signal,
- `istart` denotes the first peak position,
- `osr` denotes the oversampling rate ($osr = 50$ samples per bit used here).

The derivation of the basic shapes as shown above allows the following approach to be considered: if one would take the real part of an IQ waveform of a single signal with centre frequency $f_c = 0$, then synchronize the shapes with the IQ data at a bit phase so that the peak of the signal matches beginning of the shapes, the four shapes could reconstruct the IQ signal under analysis. This could be similarly repeated also for the imaginary part of the IQ signal, to conclusively derive all signal bits.

4.2.3 Constructing time-shifted basic shapes for Signal2

After deriving the basic shapes for Signal1, the basic shapes for the Signal2 should be found. Since Signal2 is also GMSK signal, the direct answer is that the basic shapes of the Signal2 should look the same. However, since signal samples are going to be processed in the blocks corresponding to sample subsets of Shapes1 from Signal1, there is a time shift present between the Signal1 and Signal2 basic shapes. To account such setting, a second collection of shapes is constructed as transitional combinations of the two valid sequential transitions. For example, the transition -1 to 1 (sh2) can occur only if previously 1 to -1 (sh1) or -1 to -1 (sh4) has taken place. Thus, the collection of the basic shapes for Signal2 can be constructed as all feasible transition pairs from Shapes1.

```
shps2 = [[sh3,sh1];[sh4,sh2];[sh2,sh3];[sh1,sh4];...
         [sh2,sh1];[sh1,sh2];[sh3,sh3];[sh4,sh4]];
```

The constructed Shapes2 are of the double length of the Shapes1 and are shown below in Fig. 45.

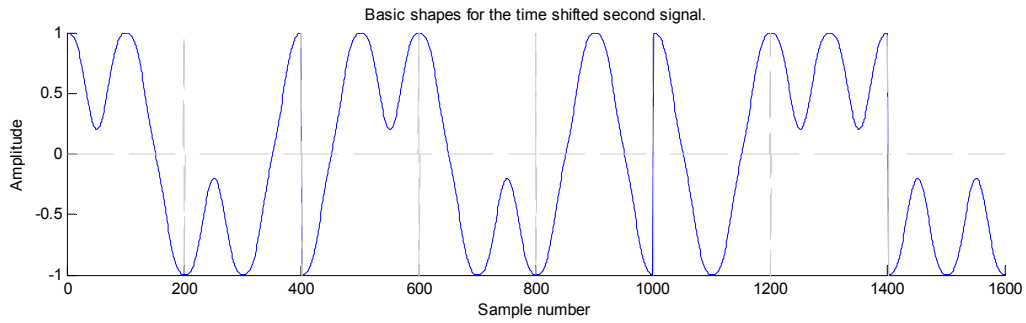


Fig. 45. The eight basic shapes for the second signal, constructed using pairs of Shapes1.

To give an example of how these transition waveforms are to be used, the following example can be considered. At time sample $t_s = 275$ the first peak of the Signal1 occurs. Thus, the beginning of Shapes1 is synchronized to that sample. Let us further assume that Signal2 has arrived slightly earlier by half-bit time. That means that the peak of the second shape should be time shifted by half bit time

$\frac{1}{2}T_b$, making the first peak of Signal2 to occur in sample $t_s = 300$. The transition shift of the 25 samples is due to the oversampling rate 50spb being used. Consequentially, for each of the Shapes2 a sub-range cut should be applied, corresponding to taking samples 75 to 175 from its length of 200 samples. For the first of the Shapes2 shapes, as shown as the first in Fig. 45, it would mean taking the last 25 samples from the first +1+1 transition (sh3) and then concatenating them with the first 75 samples from the second transition +1-1 (sh1). This shape along with the other seven similarly cut shapes, prepared for the overlapping signal according to its bit time shift, are shown in Fig. 46 below.

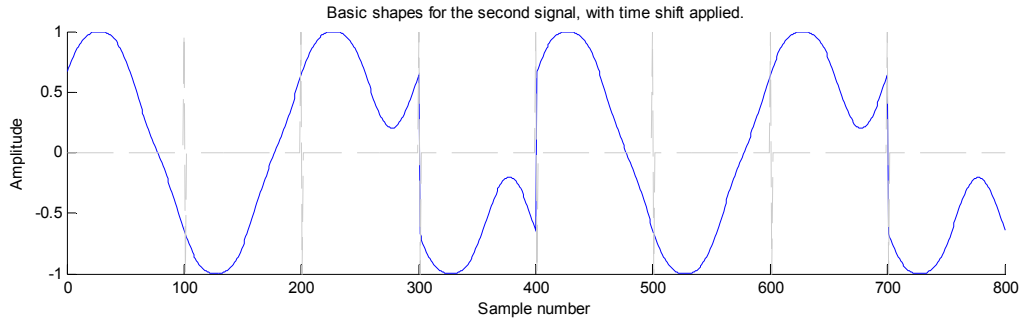


Fig. 46. The eight basic shapes for the overlapping singnal, cut to respective bit clock shift of Signal1 to Signal2.

Let us further reference the Shapes2 with the sub-range cuts corresponding to the bit time shift of the second signal as Shapes2t. As it can be seen comparing Fig. 46 with Fig. 44, the length of the Shapes2t waveforms is equal to the ones of Shapes1, thus the basic shapes of the both signals can be summed and analysed for the best joint fit.

With the preparatory work of deriving the Shapes1 and Shapes2 completed, the algorithm of identifying two bit sequences of two overlapped AIS signals would be as follows. Assume the information of the bit clock and amplitude of the first and the second signals being available. Assume the two overlapping signals both having centre frequency $f_c = 0$. Finding the best fitting shape pairs from Shapes1 and Shapes2t in the sense of L^2 norm will build the ML approximation of the real and imaginary parts of the provided IQ signal and will reconstruct the provided IQ waveform. Due to the knowledge of the bit compositions contained within each of the shapes used for fitting, the bit sequences of the both signals can be directly derived. The following diagram summarizes the JShBD algorithm.

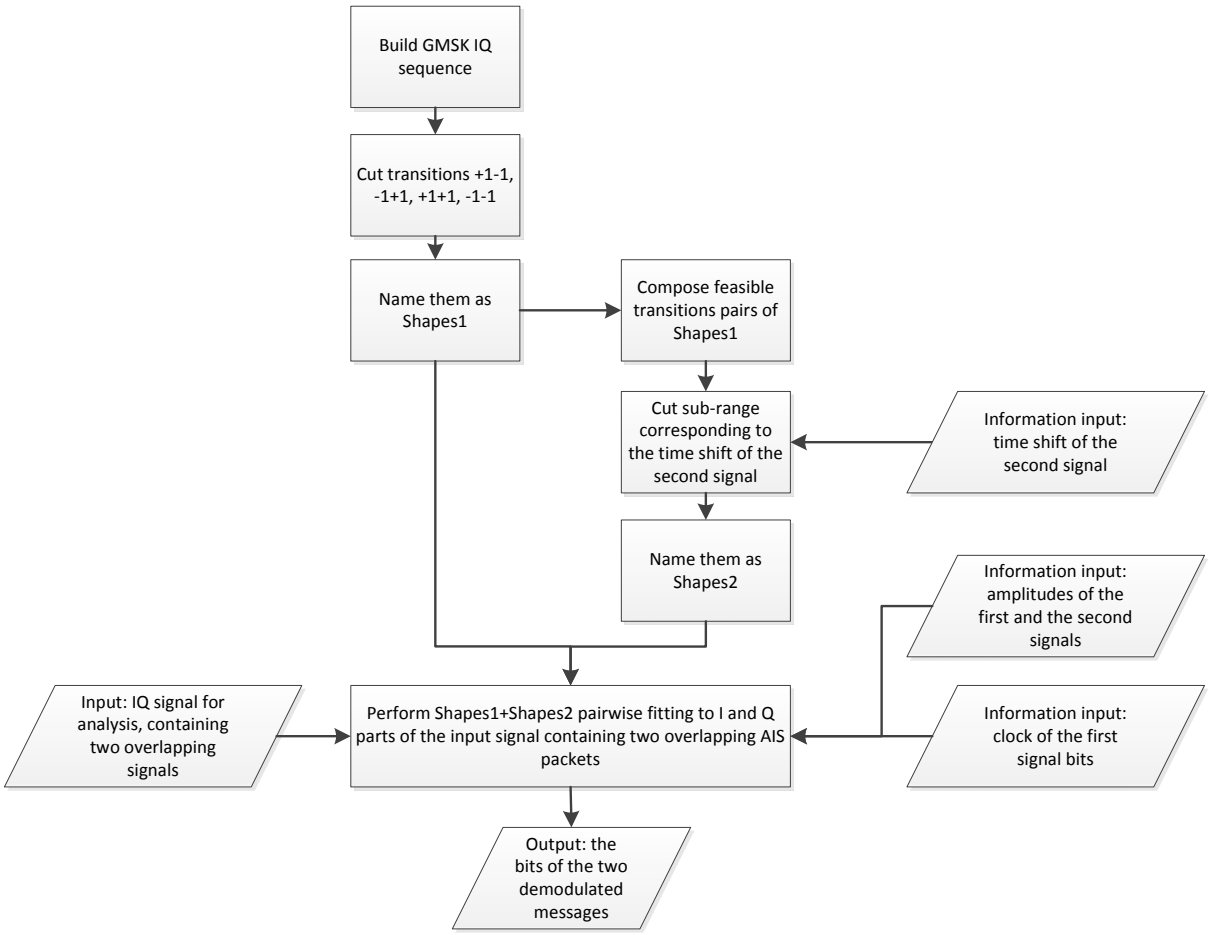


Fig. 47. Diagram of derriving basic shapes for two overlapping singals, used for JShBD algorithm.

A practical example of this algorithm will follow in further sections.

4.3 Eye-Diagram averaging method for noise suppression

As it can be seen within the diagram of Fig. 47 above, for each of the signals there should be two information inputs provided to perform shape estimation: symbol clock and amplitude. The estimation of the Signal2 shift relative to Signal1 could be effortful, especially in the presence of noise. However, since the symbol timing is constant throughout whole bit overlap time, multiple bit frames constituting eye diagram could be used to find combined basic shapes by grouping and averaging them. The approach would also serve for noise suppression. This algorithm is shown as a diagram in Fig. 48 below.

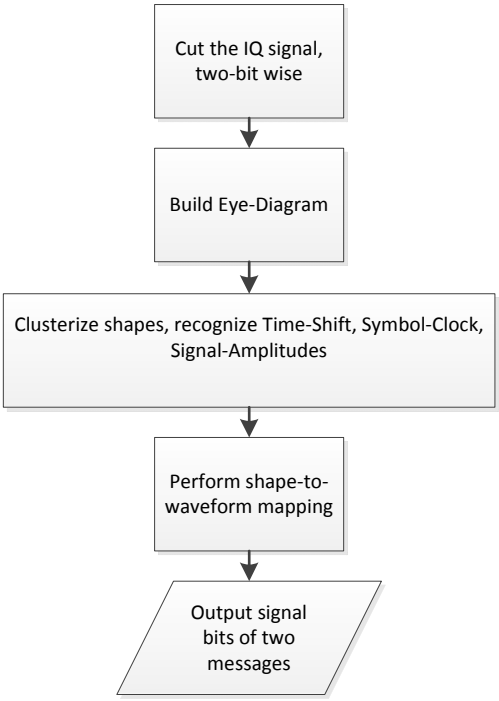


Fig. 48. Conceptual diagram of the shape recognition algorithm.

In contrast to the approach outlined in Section 4.2, which used cutting shapes out of a dedicated IQ signal, deriving and building the basic shapes of Signal1 and Signal2 for joint use in pairwise sums, here the joint shapes are obtained directly from the input signal. The eye diagram of the IQ signal of a single AIS message with $\sigma(N) = 0$ and $\sigma(N) = 0.1$ noise is shown below in Fig. 49.

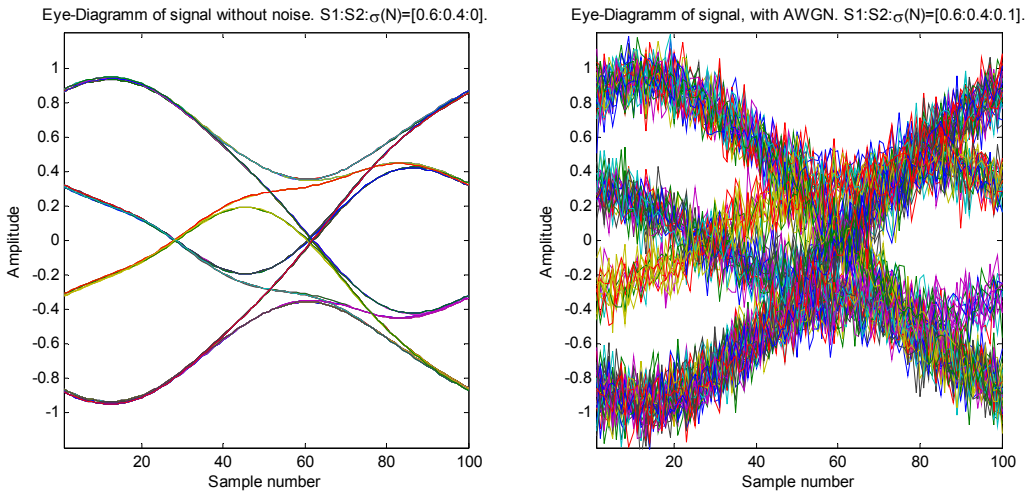


Fig. 49. Eye-Diagram of IQ signal containing two overlapping AIS signals. a) Noiseless, b) with noise.

In Fig. 49 a) there can be observed how the input signal of $256T_b$length looks, when it is cut in chunks of $2T_b$ samples and overlaid over each other. As expected, there can be observed that the diagram contains distinct signal paths, which are characteristic to the modulation scheme. While these can be clearly identified in Fig. 49 a), it is more difficult to be determined in Fig. 49 b). Observing the both figures and having the knowledge that in the both cases the modulated signal is the same, thus the underlying signal paths should be the same, the author of this work hold it reasonable to develop the following algorithm. Introduce a repository of shapes. Then consider each of the chunks from the input IQ signal. If there exists a similar shape in the repository, add it to the

existing one and take the mean weight average. If there is no similar shape present, then put the input chunk as a new shape in the repository.

To estimate if two shapes are alike to each other, the L^2 norm could be used. Let us also introduce a constant ε, which denotes if the two signals are similar enough to be considered having the same underlying signal path. This can be written as follows.

$$\left\| C^k - R^j \right\|_2 < \varepsilon, \quad \text{for } j = 1 \text{ to } N \,,$$

where N is the number of shapes within a repository, C^k is the input chunk and R^j is the j-th shape from repository. As shown in previous section, each signal has 4 basic shapes. Taking the transition cases for each of the two signals, would mean having 8 shapes per signal, meaning 64 variations. If the phase could be synchronised to the peak of the strongest signal, this would reduce to 32 shapes. Due modulation characteristics the preceding transition has very less impact near the signal peak under consideration, which is in detail shown later in Fig. 53, the number of expected shapes within eye diagram reduces to 16.

If the input signal chunk is similar to some of the repository shapes, it is added to the most similar repository shape with weight one divided by the number of similar shapes already found. In this way the shapes in repository update with every incoming alike signal shape by averaging themselves, thus suppressing also noise. The waveforms finally achieved in the repository are further referenced as "characteristic shapes". They can be interpreted as being the pairwise sum of basic shape waveforms of Shapes1 and Shapes2, thus representing the characteristics of the IQ signal under analysis and also the individual signals contained within the IQ data.

A schematic of basic implementation of the shape clustering is shown in the Fig. 50 below and the result of the algorithm application is shown in Fig. 51 and Fig. 52.

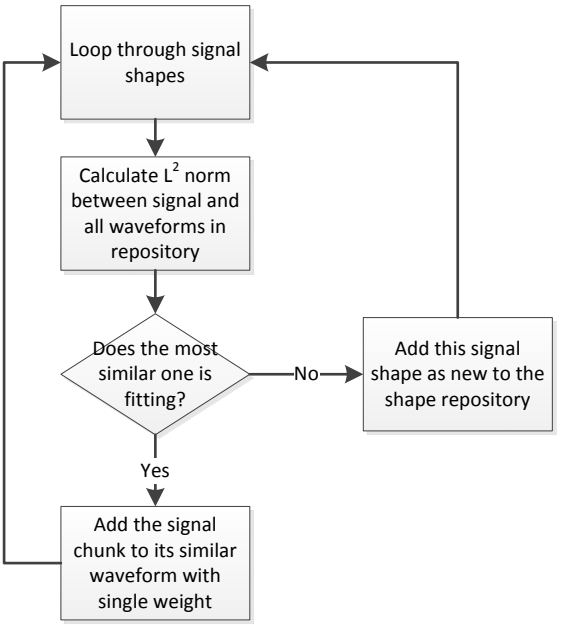


Fig. 50. Waveform clustering algorithm.

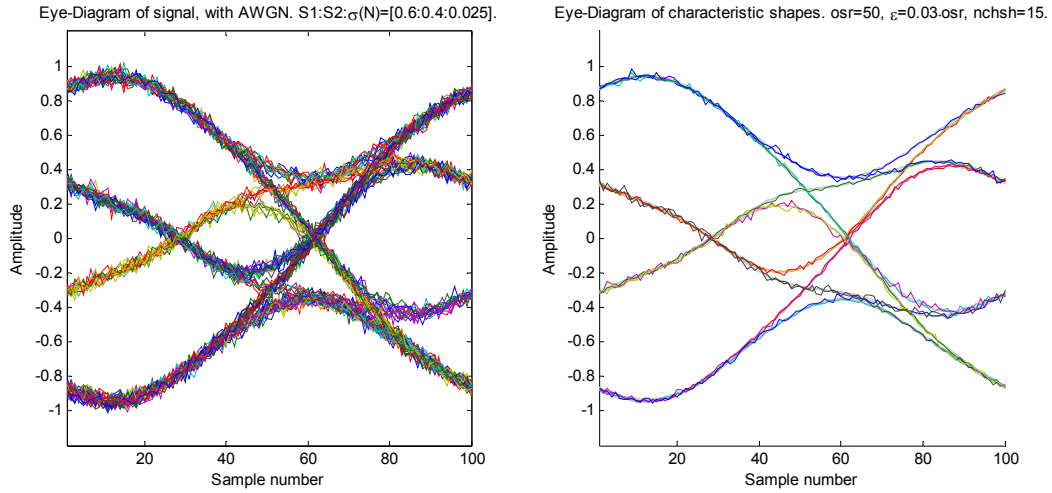


Fig. 51. Comparison of eye diagrams: a) Eye-diagram of input signal b) the characteristic shapes found (in colors) overlyed with noiseless eye-diagram(in grey).

As it can be observed in Fig. 51, for the applied noise ratio, the number of shapes found in the input signal is not exactly 16, as it was expected. In the case under consideration, it is not error and has emerged from the fact of absence of one characteristic signal path variation in the Signal1 and Signal2 relation. Considering number of shapes within each of characteristic shape group found, it can be noticed that some shapes are more than 20 pieces found, while other are just 2 or 1 or can be completely absent.

In case the input signal contains high noise, by increasing ε, one could minimize the number of introduced shapes, and by lowering ε, do vice versa. Below is another example with more noise and another adjusted ε.

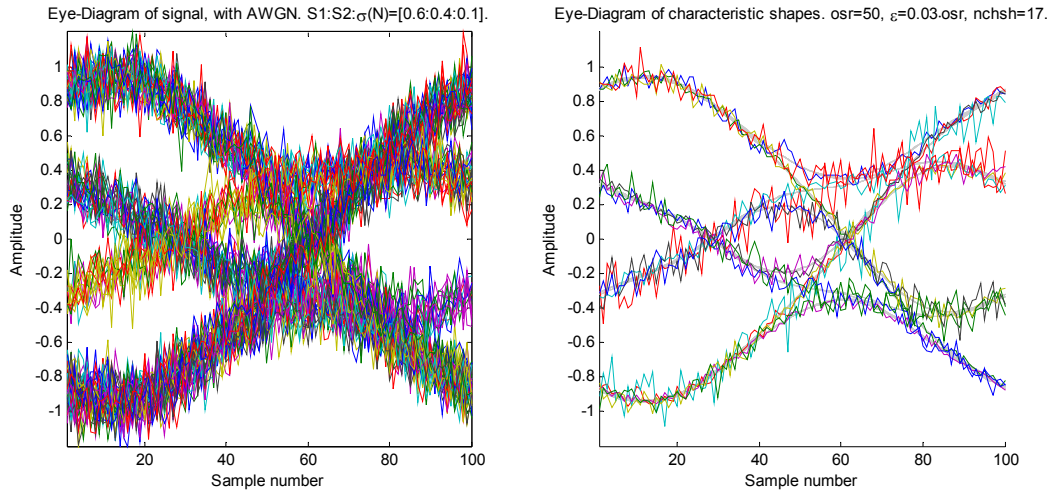


Fig. 52. Comparison of eye diagrams, with noise weight 0.1. a) Eye-diagram of input signal b) the eight characteristic shapes found (in colours) overlyed with noiseless eye-diagram(in grey).

The characteristic shapes in Fig. 52 b) could be considered to approximate the initial eye-diagram of Fig. 52 a) sufficiently. Nevertheless, some curve smoothing improvements could be applied with dedicated filters to improve results.

The derived characteristic shapes can be related to the variations of pairwise sums of the Signal1 and Signal2 basic shapes, providing the following parameters:

- Symbol clock of Signal1;
- Time shift of symbol clock of Signal2 relative to Signal1;
- Amplitude of Signal1 ;
- Amplitude of Signal2 relative to Signal1.

Having the four parameters above derived, a mapping of each signal chunk to the Signal1 and Signal2 basic shapes can be done, leading to obtaining contained bits within both of the signals.

An overview of efficient algorithms for relating eye diagram signal paths to the basic shapes and thus finding the clock and amplitude information of the underlying signals is skipped here to keep the focus of this written work on the demodulation process.

4.4 Joint Shape-Based demodulator with waveform orthogonalization

We previously observed derivation of basic shapes and characteristic shapes which can be used to recognize contents of IQ signal. An improvement in shape recognition might be achieved by setting up a new orthonormal basis from characteristic shapes. Then, by projecting signal parts onto the new basis, one could directly interpret which of the provided characteristic shapes is best approximated by signal chunk under consideration. This approach would lead to identifying the underlying basic shapes more efficiently, giving improvement in bit recognition. The algorithm is named JShBD with waveform orthogonalization and shortened as JShBDO.

The next three sections will review process of introduction signal shape based basis, signal analysis with it and arising problems for its application.

4.4.1 The Gram-Schmidt orthogonalization

The Gram-Schmidt orthogonalization allows building up a new Hilbert space composed of a function basis and will be used for signal basis derivation here. The mapping of a signal samples' vector to the new function basis space is defined by a projection operator as shown in (38).

$$\mathrm{proj}_u(v) \coloneqq \frac{\langle u, v\rangle}{\langle u, u\rangle} u \,, \tag{38}$$

where the inner product $\langle \cdot\,, \cdot \rangle$ of two vectors is defined as

$$\langle u, v\rangle \coloneqq \sum_i (u_i \cdot v_i) \,. \tag{39}$$

The definitions (38) and (39) is to be interpreted as follows: projection of vector v onto base vector u equals the inner product of the two vectors, divided by the square of base vector u and directed in the direction of vector u.

A derivation of further basis vectors with Gram-Schmidt orthogonalization is shown in (40).

$$
\begin{aligned}
u_1 &= v_1 & e_1 &= \frac{u_1}{\|u_1\|} \\
u_2 &= v_2 - \mathrm{proj}_{u_1}(v_2) & e_2 &= \frac{u_2}{\|u_2\|} \\
u_3 &= v_3 - \mathrm{proj}_{u_1}(v_3) - \mathrm{proj}_{u_2}(v_3) & e_3 &= \frac{u_3}{\|u_3\|} \\
u_4 &= v_4 - \mathrm{proj}_{u_1}(v_4) - \mathrm{proj}_{u_2}(v_4) - \mathrm{proj}_{u_3}(v_4) & e_4 &= \frac{u_4}{\|u_4\|} \\
&\vdots & &\vdots \\
u_k &= v_k - \sum_{j=1}^{k-1} \mathrm{proj}_{u_j}(v_k) & e_k &= \frac{u_k}{\|u_k\|}
\end{aligned}
\tag{40}
$$

From (40) the Gram-Schmidt orthogonalization process can be described as follows: for provided data vectors v_i the first data vector v_1 is taken as the direction, normalized and defined as the first basis vector e_1. The second data vector is projected on the e_1 and the remainder of the projection is used to define the second basis vector. This process is continued until all basis vectors are defined.

This work results in deriving orthonormal vectors, which can be used as basis vectors in signal recognition. If basic shapes are used for v_i vectors and the signal under analysis is projected onto the corresponding basis, the result directly points to the shapes contained in the signal, as shown in the next section.

4.4.2 Applying Gram-Schmidt orthogonalization to characteristic curves

Let us apply Gram-Schmidt orthogonalization to an IQ signal space, by choosing v_i vectors to be constituted of the basic shapes of the GMSK modulated IQ signal as derived in Section 4.2, which jointly build the characteristic shapes as shown in 4.3.

Let us look at the basic shapes for Signal1, which are shown in Fig. 44. Since by projecting the signal on the basis vectors, one can get positive and negative values, the number of elements in Shapes1 is about to be reduce by two. By looking at Fig. 44: the transitions +1-1 (sh1) and -1+1 (sh2) are the same just with opposite sign and the same holds for the pair +1+1 (sh3) and -1-1 (sh4). Thus, it is enough for Shapes1 to contain the transitions -1+1 (sh1) and +1+1 (sh3).

The same simplification also holds for the basic shapes of Signal2. However, there is a further simplification possible, which will reduce the needed shapes by half due similarity. Let us consider transition triples +1+1-1 (Shapes2-1) and -1+1-1 (Shapes2-5). This is plotted below in Fig. 53.

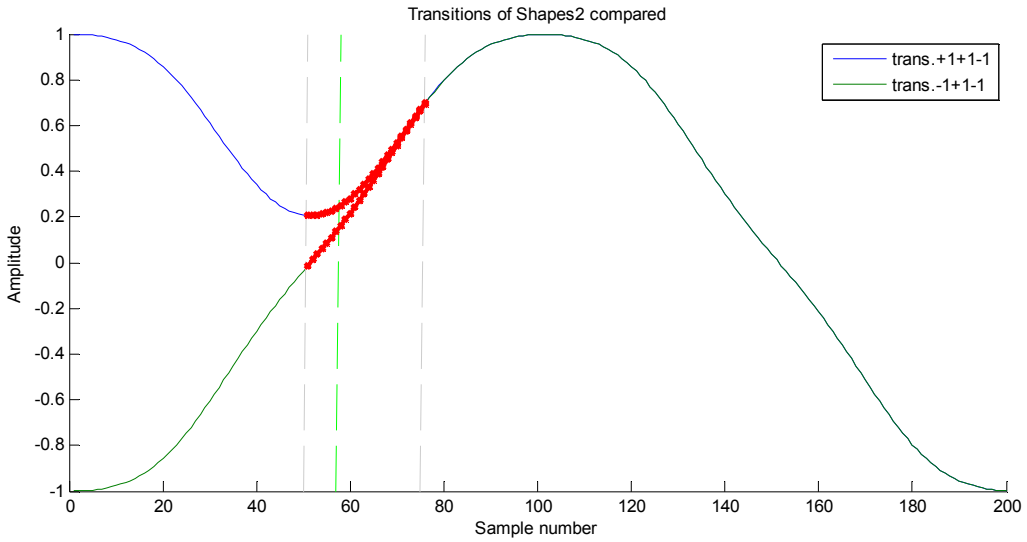


Fig. 53. Comparision of Shapes2-1 and Shapes2-5. Absolute difference between the curves: 0.243 at T_s=50, 0.101 at T_s=57, 0.004 at T_s=75.

Assuming the Signal2 being time shifted by ½-bit, this would mean that cutting the Shapes2 should be made by taking 25 samples from preceding curve and 75 samples from the following one, corresponding the samples $T_s = 75$ to $T_s = 175$ in Fig. 53. For the two plotted transitions at the maximum deviation point $T_s = 75$, the difference between the amplitudes of the two shapes would be only 0.4% of unity amplitude.

Since the signal under analysis in a real receiving application will contain some noise and it can take values of 10% amplitude or even greater, the small difference in the two shape amplitudes can be neglected. The 10% barrier would translate to 0.86 bit delay, meaning a shift of 43 samples. Depending on permitted noise levels for signal processors and the targeted further signal analysis methods, one can assume the curve differences marked in red to be small enough for approximation as a single curve. In general, the bit time shift is not expected to be greater than a single bit, which is here 50 samples, since then the curve identifications can be interchanged, thus Signal2 is not 80 samples in front, but Signal1 (then named Signal2) is 20 samples in front.

For Gram-Schmidt orthogonalization, it would be disturbing to take the two separate curves for the signal delay below ½-bit, since the curves are very similar and their orthogonalization would require big normalization coefficients, thus leading also to unstable projections, especially in the presence of noise. This case will be shown by an example later in this section.

Therefore, it is enough and recommended to take just one of the two very alike curves in case of permissible small time shift between the signals. Let us perform analysis on ½-bit delay with the proposed approximation due curve similarity.

Finally we arrive at the observation that for Signal1 only 2 curves out of 4 are needed, due to the opposite sign reduction. For Siganl2 instead of 8 transition variations, we can have half of them due to the same argument of opposite signs, and even half of those ones, due to the similarity approximation for the given bit delay case. Thus, deriving to have just four basis shapes for building orthonormal basis, where the provided IQ signal can be projected and searched. The four basis shapes along with its orthogonalizations are shown below in Fig. 54 a) and Fig. 54 b) respectively.

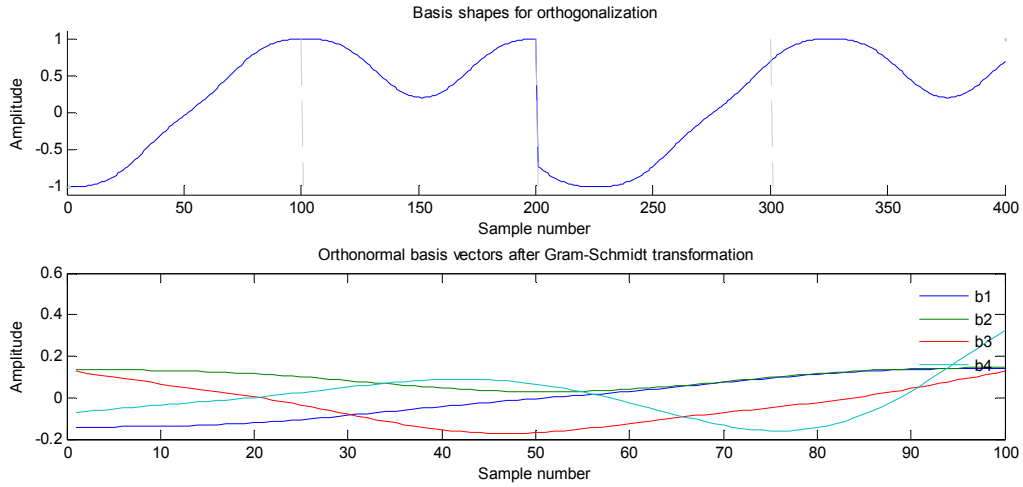


Fig. 54. a) Basis shapes Shapes1-2, Shapes1-3, Shapes2c-6, Shapes2c-7 , b) basis shapes orthogonalized with Gram-Schmidt process.

It can be noted that the b1 waveform in Fig. 54 b), corresponding to the $\boldsymbol{e_1}$ vector in the equation (40), is the same as the first basis shape in Fig. 54 a), just a scaled version. The normalization constants of the new base unity vectors in the particular case are following

$$1/[\ \|u_1\|, \|u_2\|, \|u_3\|, \|u_4\|\ ] = [\ 0.1414 \quad 0.1398 \quad 0.2816 \quad 0.6216\ ].$$

The next step is to calculate matrix indices for transformation, which will allow samples of single transition chunks to be projected on the new basis. This means sequentially taking provided input shapes and calculating how much of previous basis it contains. For a case of two basis shapes this would mean the following. The basis shape 1 is taken as basis1 as with normalization and the basis shape 2 will be projected first on the basis1, its projection will be subtracted. This leads to the interpretation of "how much basis2 is in basis shape 2", which will be then normalized to unity.

The transformation matrix can be calculated as follows:

```
for i=1:k
    c(i,i) = 1/norm(uv(i,:),2);
    uv(i,:) = uv(i,:)*c(i,i);
    for j = i+1:k
        c(j,i) = uv(i,:)*uv(j,:)';
        uv(j,:) = uv(j,:)-c(j,i)*uv(i,:);
    end
end
```

where k is the number of basis shapes, uv is initialized from the input basis shapes v as in (38) and afterwards transformed to orthonormal basis, as vectors e_i in (40).

This calculation results in following matrix.

```
c =    0.1414         0         0         0
      -0.2828    0.1398         0         0
       5.1377   -3.3150    0.2816         0
      -2.5830    5.9500   -2.5623    0.6216
```

This constitutes obviously well-defined projections, since the matrix elements of C are deviating from each other below the second order of 10.

To carry out a misleading projection construction example, let us choose for basis shapes Shapes2-6, Shapes2-7, Shapes2-2 and Shapes2-3, where Shapes2-6 differ from Shapes2-2 by fact that the first one is composed of transitions +1-1+1 (sh1-sh2) and second one of transitions -1-1+1 (sh4-sh2). As noted before and showcased in Fig. 53, there is very small absolute value difference for ½-bit shift. In this case, the basis shapes look as follows:

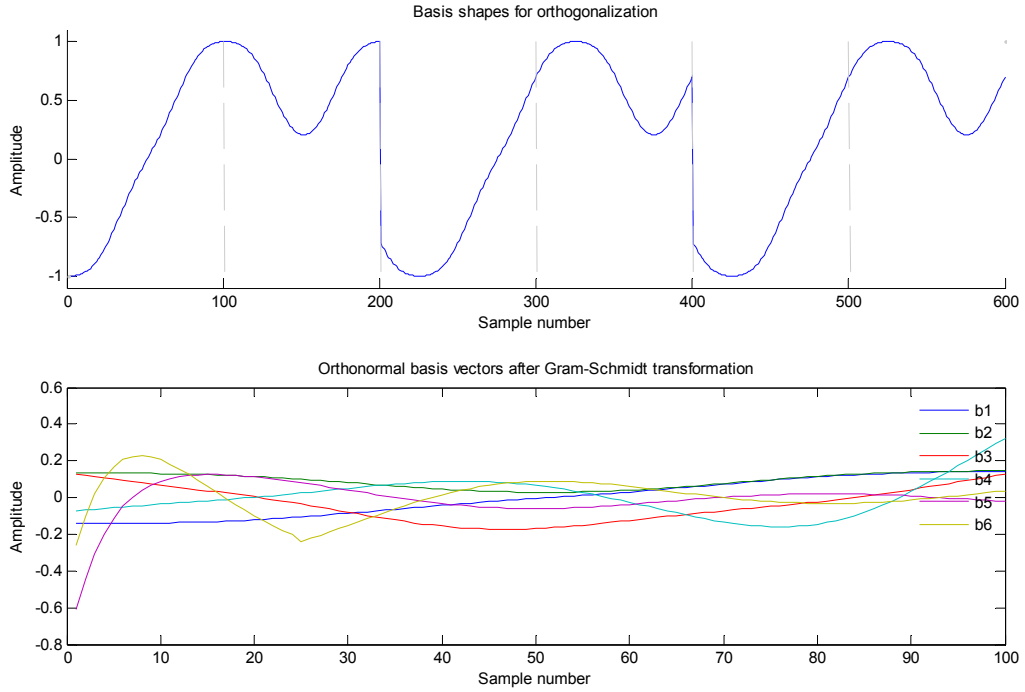


Fig. 55. a) Basis shapes Shapes2-6, Shapes2-7, Shapes2-2 and Shapes2-3. b) Orthogonal basis vectors developed with Gram-Schmidt process.

and constitute the following transformation matrix:

```
C =
   1.0e+04 *
    0.0000         0         0         0         0         0
   -0.0000    0.0000         0         0         0         0
    0.0005   -0.0003    0.0000         0         0         0
   -0.0003    0.0006   -0.0003    0.0001         0         0
    0.0005   -0.0003    0.0004    0.0000    0.0322         0
   -0.0003    0.0006   -0.0003    0.0002    0.0000    7.7341
```

As one can note, the last two elements on the diagonal are very big compared to the other entries in matrix, which indirectly points to expected instabilities in using the matrix for projections.

This example explicitly shows why the idea of basis shape approximation to as less items as possible in the basis vector development case is indeed a very necessary and important.

Yet another approach for basis construction could have been to use characteristic shapes, as derived in Section 4.3. So far the constructed basis was built using two signal vectors from each of the two

underlying signals, a subset of basic shapes derived in Section 4.2. That meant that each projected signal chunk containing two overlapping signals is expected to give one value on the Signal1 basis and one on the Signal2 basis. An alternative with use of the characteristic shapes would require from the 16 initial shapes to identify 8 shape pairs with opposite signs and take just one shape per pair. Then, for each projected signal chunk, the projection operation would yield a corresponding single vector with the largest magnitude in projection space. Thus, avoiding the need to search signal projections for waveform pairs corresponding to those two orthogonal vectors, which prescribe the signal at best. Due preferred compactness of the basis shapes, the author of this work chose the approach of using basic shapes for the basis construction, with two basis vectors for each the two overlapping signals, as described at the beginning of this section.

As the preparatory work of basis construction is completed, the IQ samples can be projected onto the new basis using basic matrix multiplications:

```
for i=1:(sp-1)*osr2
    projI(i,:) = (uv*ssigI(i:i+osr2-1)')';
    bsI(i,:) = (N*(Ah*(N*projI(i,:)')))';
    projQ(i,:) = (uv*ssigQ(i:i+osr2-1)')';
    bsQ(i,:) = (N*(Ah*(N*projQ(i,:)')))';
end
```

where `projI` and `bsI` are respectively projections and basis magnitude of the real part of the input IQ signal. The same holds for `projQ` and `bsQ` for the imaginary part of the signal.

The following two pictures in Fig. 56 show basis magnitudes `bsI` and `bsQ` for the I and Q signals.

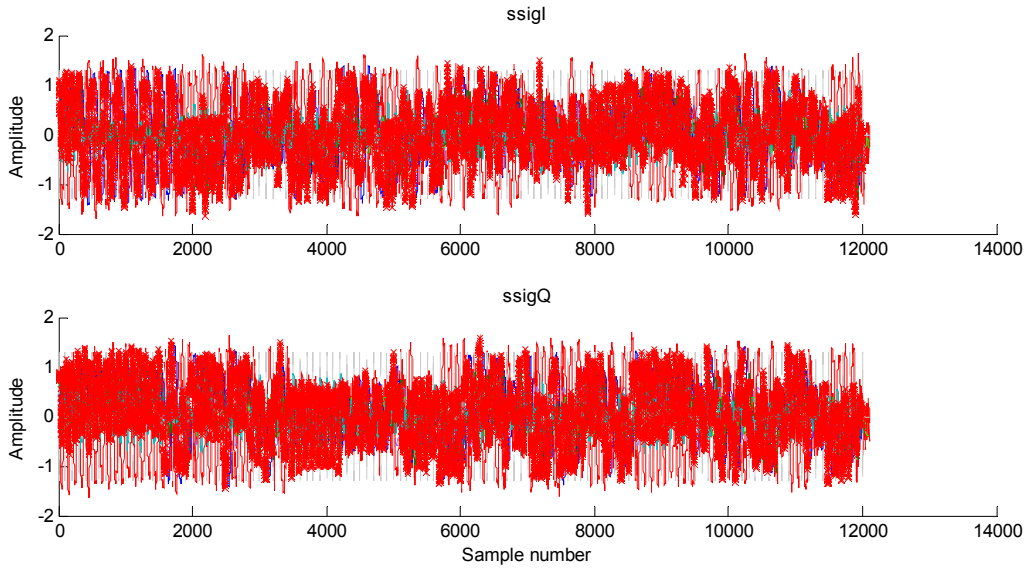


Fig. 56. Magnitudes of signal projection on the developed basis, for I and Q samples.

Taking a zoomed in view of I signal of Fig. 56 reveals in Fig. 57 and Fig. 58 in detail how the basis vectors projection is performed.

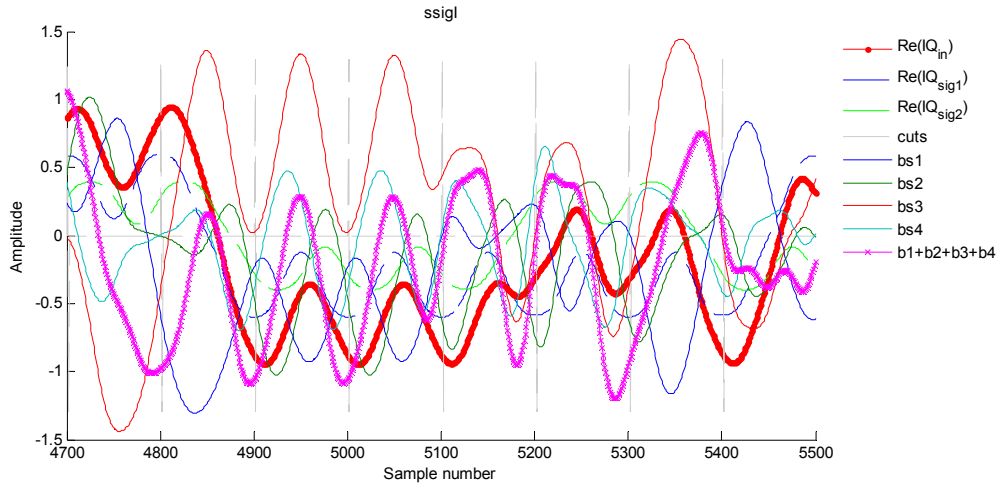


Fig. 57. Magnitudes of signal projection on developed basis, for Re(IQ) samples. Zoomed.

A further zoom of the signal is shown below.

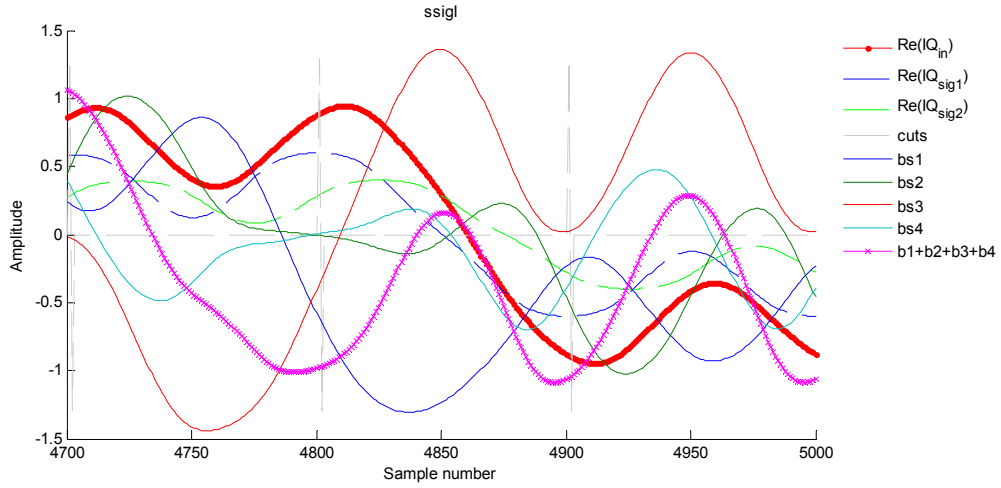


Fig. 58. Magnitudes of signal projection on developed basis, for Re(IQ) samples. More zoomed. The bs1 to bs4 lines correspond to signal projection onto the corresponding basis vector.

In the figures Fig. 56, Fig. 57 and Fig. 58 the grey vertical lines identify borders of chunks, corresponding to the clock of Signal1, thus also beginning of Shapes1. The clock of Signal1 shapes are input based on preknowledge from construction of a waveform and in real application would have to be determined using dedicated algorithms. The grey lines help to observe shape projections and identify its' correctness.

In Fig. 58, at sample 4800 one can notice that bs2 and bs4 lines are close to zero, while bs1 plus bs3 sum to 1, as the magneta curve identifies. However, as it can be observed in Fig. 57, clear identification of unity sum can be seen only when other bases are clearly zeros. Thus, what we would like to analyse is summing possible Signal1+Signal2 pairs: bs1+bs3, bs1+bs4, bs2+bs3 and bs2+bs4. There is expected to be exaclty one sum-curve equalling to 1 at each shape border point, identified with the vertical grey lines. These pairwise sums are shown in the next three following pictures.

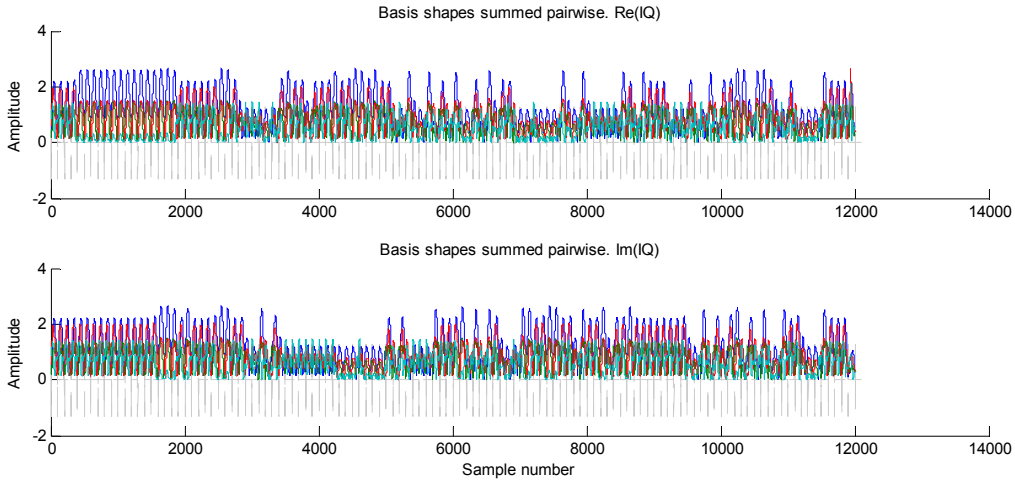


Fig. 59. Absolute sums of permissible basis-pairs.

A zoomed view of the I part of the signal from Fig. 59 can be observed in more detail in Fig. 60 and Fig. 61 below.

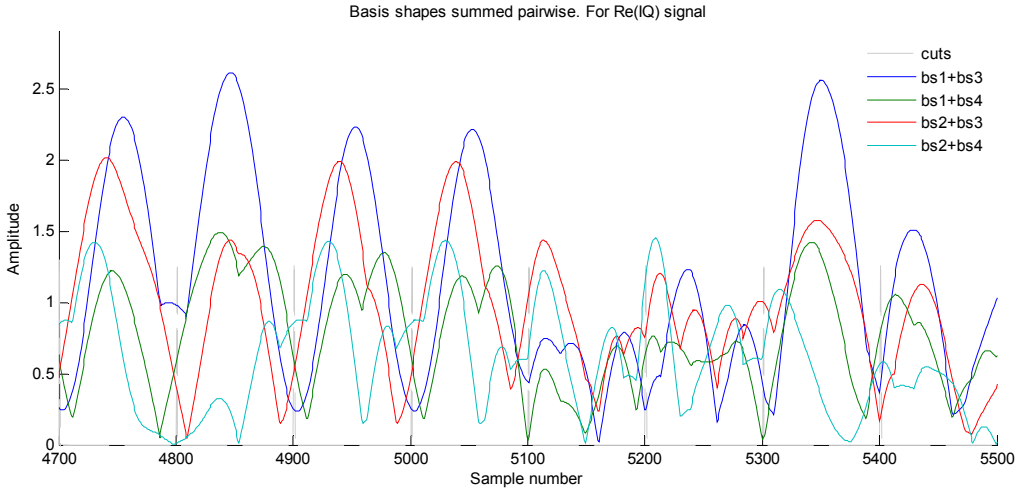


Fig. 60. Absolute sums of permissible basis-pairs. Zoomed.

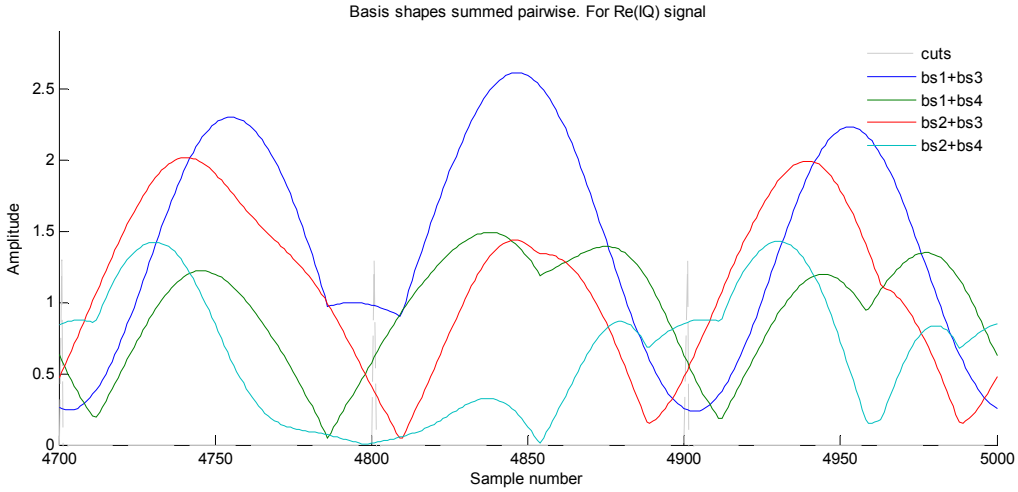


Fig. 61. Absolute sums of permissible basis-pairs. More zoomed.

Observing the figures Fig. 59, Fig. 60 and Fig. 61 one can clearly observe, that the blue curve, corresponding to the sum of the projection on bs1 and bs3, is approximately 1 at samples around $T_s = 4800$. This serves as a visual feedback to verify the correctness of projections. Also this can serve for signal clock determination, since it is expected that repetitively in time range of two bits there will be one pair from the four, which summed will represent the signal magnitude, which is a unity in the case under analysis.

Let us name these permissible basis projection pairs as characteristic orthonormalized curves. Since, similarly to characteristic curves before, they are composed from sub-shapes. But this time the underlying shapes are orthogonalized basis projections instead of basic shapes.

As the next step in the JShBDO algorithm processing is to search and select in each transition range the projection pair which produces the maximum magnitude onto the projected orthogonal basis. As shown in Fig. 61 above, at the $T_s = 4800$ the choice corresponds to the blue curve (bs1+bs3) and for $T_s = 4900$ – the light blue one (bs2+bs4). After the pair selections, the characteristic orthonormalized curves can be mapped to their bit source, revealing the incorporated bit transitions for each of the containing signals. This mapping from characteristic orthonormalized curves to bits has to be done for the I and the Q parts of the signals, with the timing of the Q shifted by a half-phase than that of the I. The mapping endeavours finally complete the input IQ signal approximation with shape transitions of determined bits, shown in Fig. 62 and as a zoomed-in view in Fig. 63.

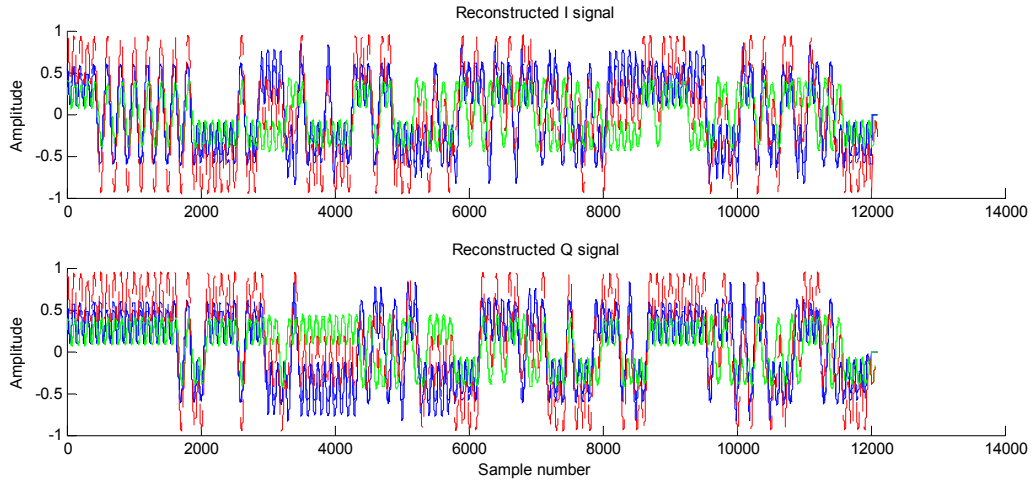


Fig. 62. IQ signal reconstruction with basis shapes.

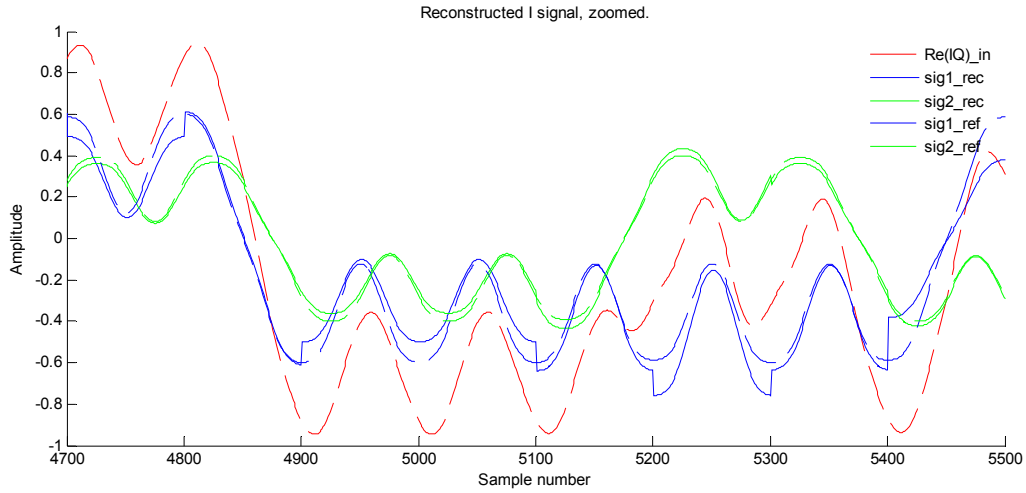


Fig. 63. IQ signal reconstruction with basis shapes, zoomed. Meanings of the subscripts for legends are as follows: "in" means input signal to processing, "rec" means recovered signal by processing, "ref" means the reference signal not known to processor, used for visual feedback to estimate correctness of the recovered signal.

As one can note in Fig. 63, the continues blue and green lines of Signal1 and Signal2 respectively approximate the initially modulated overlapping signal, shown for reference with blue and green dashed lines, correctly.

Fig. 63 also illustrate that further optimization could be implemented to restrict each signal's amplitude to a single constant. For example, the blue curve between samples 5000 to 5100 has an amplitude of about 0.5, whereas the shape in samples 5200 to 5300 has amplitude of about 0.75. The assumption that the signal amplitude is constant along one single AIS message is safe. The switch on and switch off signal power transitions are skipped here for a simplicity.

Finally, after cutting the reconstructed signal shapes, the bit estimation can be performed. For a performance evaluation of the JShBDO algorithm the derived bits can be compared to the modelled ones by subtracting one from another, showing if errors are encountered during the demodulation process. By visible means it is shown in Fig. 64 below.

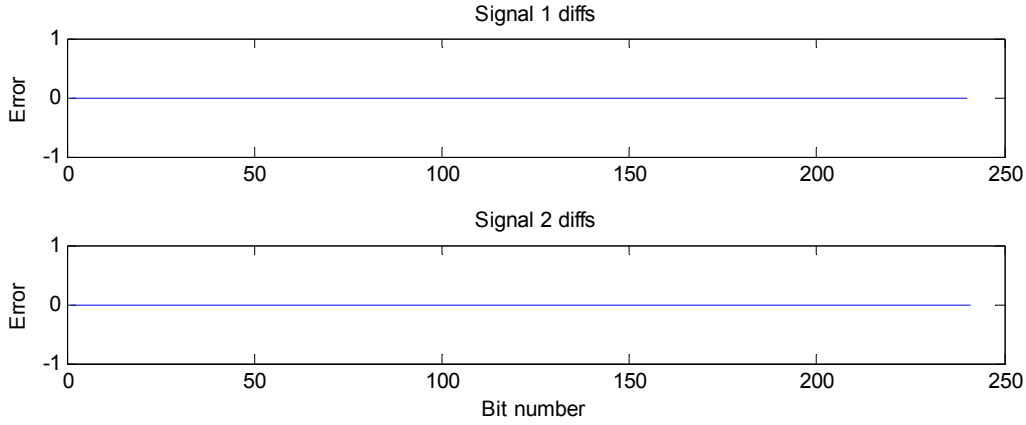


Fig. 64. Comparison of the demodulated bits from Signal1 and Signal2, showing if bit errors are encountered. Overall zero means all are accurate.

The following figures Fig. 65, Fig. 66 and Fig. 67 show ShBDO algorithm performance in application on signals containing white Gaussian noise.

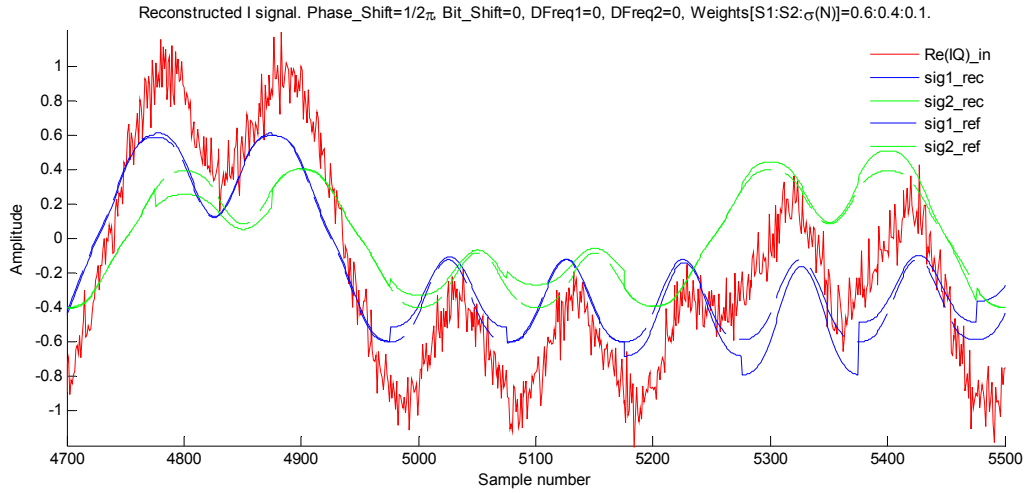


Fig. 65. Reconstruction of signal in presence of noise σ(N)=0.1. All shapes are correctly identified.

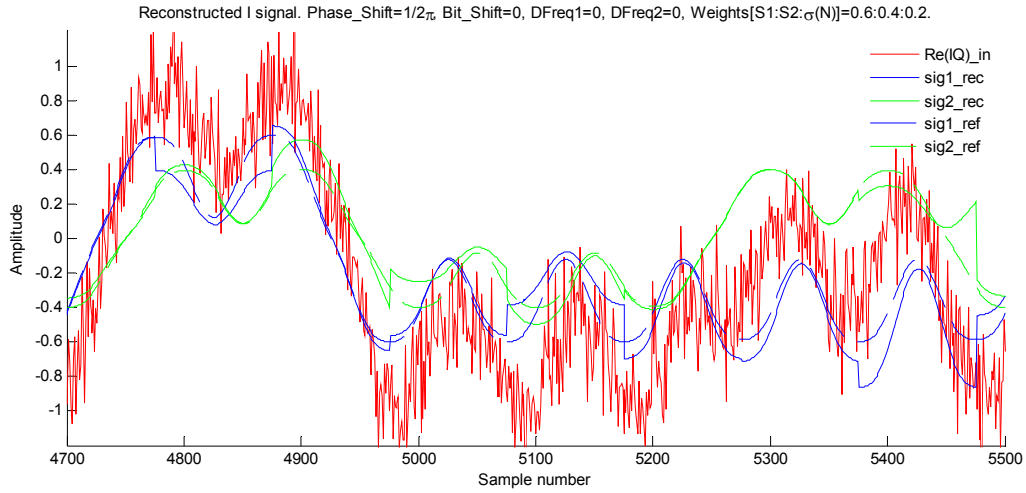


Fig. 66. Reconstruction of signal in presence of noise σ(N)=0.2. Some of the shapes are incorrectly identified.

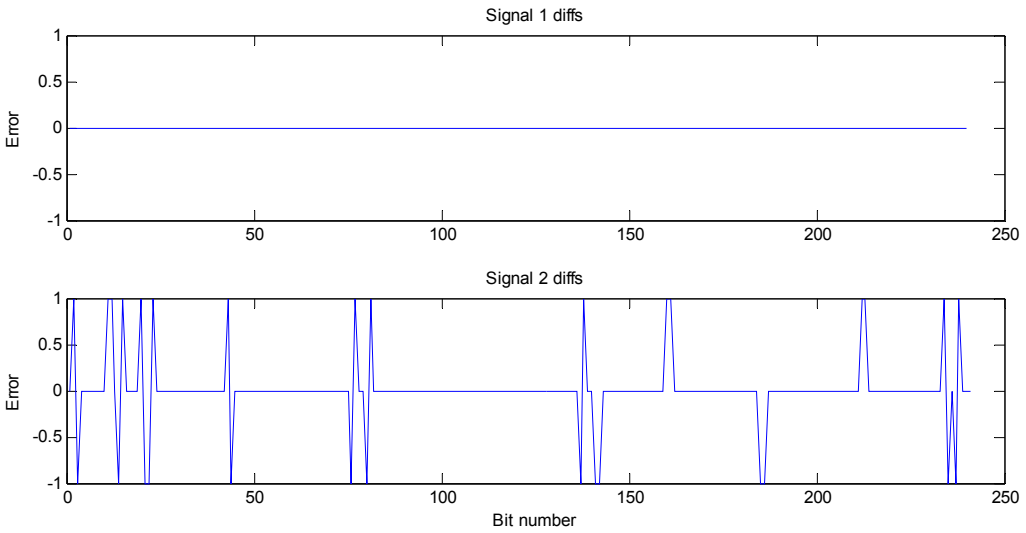


Fig. 67. Error bit lines for shape recognition shown in Fig. 66, for signal with noise σ(N)=0.2. The signal with smaller amplitude has some faulty bits.

To improve this demodulation method, the following work directions could give reasonable improvements:

- Restrict shape amplitudes to a constant – after initial analysis, restrict expected Signal1 and Signal2 amplitudes to two certain values.
- Applying continuity rule – allow only continuous transitions in shape forms to take place. For example, after the transition +1-1 cannot follow the transition +1+1. This could be well achieved with the Viterbi decoder as introduced in Section 4.1. It would require a construction of a matrix of possible shape transitions as shown in Table 5 and apply it to the amplitudes on signal chunk projections on the orthogonal basis, deriving the path weights for the trellis diagram application.

4.4.3 Arising problems for signal shape recognition in presence of Doppler shift

Up to now two overlapped signals were analysed, with some time shift relative to each other, along with Gaussian noise addition. There was not considered a situation when Doppler frequency shifts in IQ data are present. As previously noted in Chapter 1, in practical applications the Doppler shift can have values up to ±3.6kHz and in most cases will be below 1.3kHz in absolute value for a narrow angle helix type antenna. Let us now look at an IQ waveform of a single signal with a Doppler shift $f_{D1} = 200Hz$.

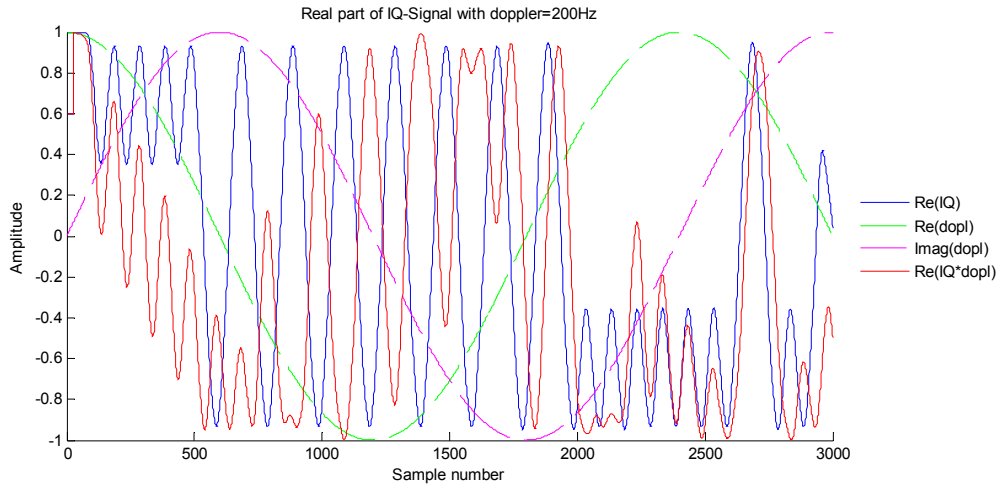


Fig. 68. Single signal with Doppler frequency shift of 200Hz.

In Fig. 68 it can be observed that the initial waveforms of the IQ signal, shown in blue, after multiplication with the Doppler shift, shown in green and magenta for real and imaginary parts respectively, get changed a lot and not directly recognizable, as the red curve identifies. For example, in samples 440…540 the curve is more similar to the transition -1-1 than the correct transition +1-1, as identified by the blue curve.

This example demonstrates that the shape recognition approach cannot be used in a straightforward way for IQ signals containing Doppler shifts. Yet a solution can be constructed in case a precise

Doppler frequency and its phase are provided. Then at each step orthonormal projection basis could be built taking the 4 basis shapes used before and multiplying them with respective Doppler shifts along with their time shifting phases of each signal. Then the same algorithm as outlined before could be used to recognize contained signal transitions and to recognize bits. The orthogonalization framework would be reasonable to be skipped in the Doppler shift case. On the one hand due wish to keep algorithm simple and on the other hand, due to the fact that in each sequential bit step a new basis construction would be necessary, which would then require extra computation resources compared to a single basic shape usage. The approach of JShBD without orthogonalization but applicable to the Doppler shifted overlapping signals is demonstrated in the next section.

4.5 Joint Shape-Based Demodulator in presence of Doppler (JShBDD)

As noted at the very end of the previous section, the Joint Shape-Based demodulator with or without basis transformation in a form developed up to here can only be applied to signals with the centre frequency being precisely zero. For applications within a satellite communication environment, the received signals are expected to have Doppler shifts with various frequencies. Therefore, additional enhancements for the JShBD method should be searched.

An extended version of the JShBD was developed by the author of this paper to work with signals containing Doppler shifts. The method is further referenced as JShBDD. The following sections review the input parameters required for JShBDD use and show the results of its application.

4.5.1 Input parameters

To model waveforms of two overlapping signals with Doppler shifts, the following information should be provided for each of the signals:

- Bit clock
- Amplitude
- Doppler frequency shift
- Phase of the Doppler frequency shift

In contrast to previous JShBD approaches, where intermediate results could be used to find bit clock and amplitudes, here those parameters are needed as input. The shapes are also much different. Due to the presence of a Doppler shift, the basic shapes used during signal reconstruction should be modified to include corresponding Doppler frequency and phase. Following the notation used in (18), the modifications applied to the basic shapes can be written as multiplication of the IQ waveform with a complex number as shown in (41).

$$S_d^i \coloneqq S_b^i \cdot e^{\omega_d(t+\nu_d)}, \tag{41}$$

where S_d^i stands for the i-th shape with Doppler, S_b^i for the i-th basis shape composed of IQ data, ω_d denotes Doppler frequency and ν_d the phase shift of the Doppler.

To derive and demonstrate JShBDD in action the basis transformation technique will be left out, as otherwise it would be necessary to build new basis on every step, negating the computation efficiency of the transformation procedure. This also will ease the application development and allow

interpreting waveforms directly by visual inspection. Nevertheless, in case of interest, the algorithm outlined here in JShBDD could be also applied to the orthogonalized projections.

An example of two overlapping IQ signals containing Doppler shifts, along with their sum and white Gaussian noise added is shown below in Fig. 69.

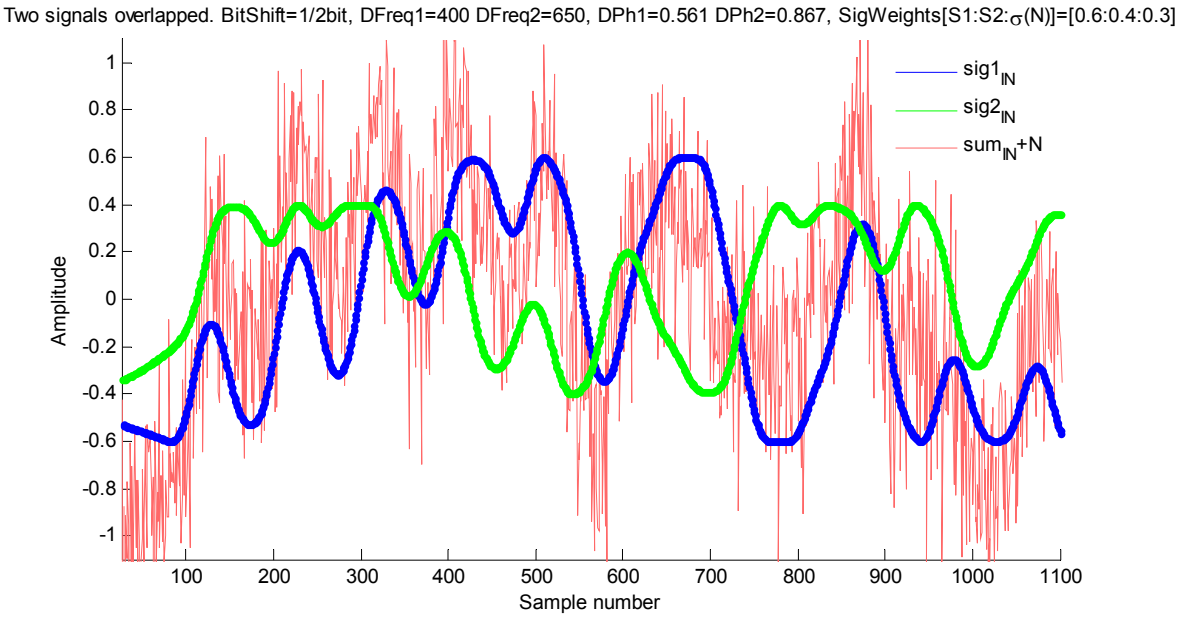


Fig. 69. Two source IQ waveforms and the constructed IQ signal from them with noise.

As noted before, one cannot recognize the basic IQ shapes in a waveform containing the Doppler shift. Also there is no direct approach possible for using basis waveforms with just multiplying them in each step with Doppler frequency – the Doppler sinusoid incorporates real and imaginary parts, and an imaginary part of Doppler multiplied by an imaginary part of signal will also contribute to a real part of the signal. Thus, the real and imaginary parts, which are by 90 degrees in phase shifted, had to be jointly constructed and jointly demodulated. For a single signal case, this would mean taking 4 basic shapes of real parts, multiplying them with Doppler, checking all phase shifted imaginary parts, and building joint blocks.

A much more straightforward approach is to perform complete GMSK modulation bit per bit with a sliding window. Since in such case one can work stepwise, demodulating each new bit, add it to the already recognized part to reconstruct I and Q signals in the same time. This means, on every new bit, one needs to check only 2 possibilities: +1 and -1.

Since the bits of single signal have ISI (Inter Symbol Interference) among themselves and the imaginary part is 90 degrees phase shifted, better signal evaluation can be made if few upcoming bits are also included in the evaluation. This leads to the following algorithm developed for bit extraction:

- Initiate two bit sequences with some starting bits, for example bits1=[0,1], bits2=[0,1].
- Construct all possible bit state variations for 3 sequential bits, which are the following eight cases.

```
bcases = [0,0,0;
          0,0,1;
          0,1,0;
          0,1,1;
          1,0,0;
          1,0,1;
          1,1,0;
          1,1,1];
```

- Start sliding window iteration:
  - Add modulation of all the three bit variations to each of the two signals.
  - Each of the two signals has eight cases; build all variation possibilities (64 cases).
  - For each of the 64 cases calculate its distance to input signal, by L^2 norm.
  - Choose the bit variation case with the minimum norm, thus approximating the input signal with minimal possible deviation.
  - Take the first bit of each of the signal variations and add them to the corresponding recognized bit sequence for each of the signals.
- Repeat the sliding window iteration until expected signal length is reached or end flag is found.

This approach, of modulating waveforms from guessed bits as a replacement for using basis shapes and Doppler frequencies to construct searched waveforms, eases very much the reconstruction of multiple signals in parallel. Since each extra signal has only the two additional bit chances to be modulated. Finally, the approach directly leads to demodulated bit sequence and error weights in each step. The derived L^2 distances can also be used to construct trellis graph and to perform Viterbi decoding.

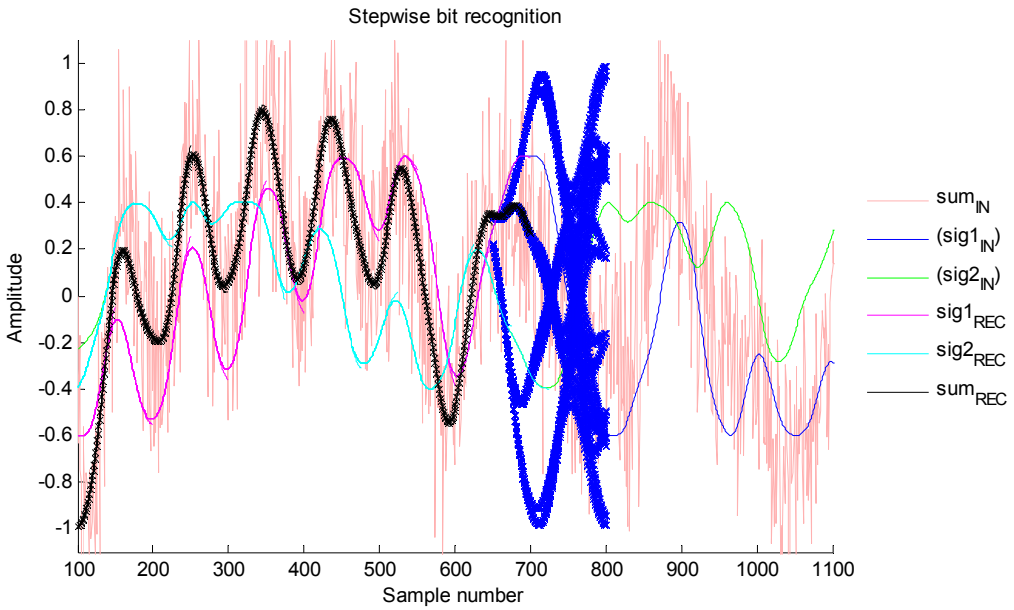


Fig. 70. Joint bit recognition in two overlapped IQ waveforms with noise. The real part of the signal is plotted.

In Fig. 70 the process of bit recognition for JShBDD is shown, using the basis shapes modelled as in (41). On the left side from the blue three bit sliding window with all 64 bit variations plotted, there is

a thick black coloured line representing recognized signal within the contained waveform. It approximates the red noisy one, which is input signal. The magenta and cyan colours show the recovered IQ waveforms of Signal1 and Signal2, with the Doppler signal contained. A visual feedback that the derived curves are correct is the observation that the magenta and cyan lines cover the green and blue tiny curves very good, so that the two initial signal curves are fully covered on the left and visible only on the right from the processing frontier with the blue lines.

Within the blue coloured 3 bit processing window, one can recognize that it begins with two paths and with each further bit these paths diverge to more ones. The black coloured path within the blue section shows the recognized minimum weight distance to input signal. In the case shown, it would also work with a 2 bit propagator, having 4 (instead of 8) bit cases to be modulated for each of the signals and thus just 16 (instead of 64) waveform variations to be checked for the minimum distance search. Even with 1 bit propagator this case would work but the recognition would not flourish for higher noise levels. A general optimum, with respect to processing time and demodulation stability, is expected to be 2 bit propagator length. All further simulations referenced in this dissertation were performed with 3 bit propagator length to more thoroughly assess the performance borders of the algorithm.

4.5.2 Analysis of JShBDD demodulation capability

This section will present the results of demodulating two overlapped AIS signals with the JShBDD algorithm. Performance analyses are presented within the tables in the following pages of this section. All the tables contain two variable parameter analyses. We will start with an Amplitude-Noise analysis, and then continue with an Amplitude-Phase and an Amplitude-Doppler analysis.

For generating IQ data with packet collision conditions for the analysis, two messages of AIS message type 1 with a length of 256 bits and an oversampling rate of 50 samples per bit were used. The amplitudes, Doppler shifts and time shifts were varied. For each scenario, the varied parameters were set, both messages modulated and overlapped. The four input parameters from each of the two signals were passed to the demodulator, which took over the generated IQ data and used its parameters to run the demodulation algorithm, reporting results in number of correctly demodulated bits.

In the following tables the amplitude variation as weights of the both overlapping signals is shown on the left side in multiple units: unity amplitude, amplitude difference in %, SIR (signal to interference ratio) and SIR in dB units. The SIR and SIR_{dB} values are calculated using (25) and (26) respectively.

The bottom grey part of the tables shows the second variable parameter. It is either noise or phase, or Doppler shifts. The e1 and e2 columns within the main part of the analysis tables show the number of errors encountered during the demodulation process of the two messages. The vertical slash separates the demodulation results for cases with Viterbi decoder and those without it. For the shape weights to be input to a Trellis graph, the 3 bit distances from the basic shape transitions to the input data were used. The SNR1 and SNR2 columns represent SNR (Signal to Noise Ratio) for the strongest the interfering signal respectively. The noise levels were calculated using (27) and (28).

For an example, Table 6 can be read as follows: for the noise amplitude of $\sigma(N) = 0.3$ and the two signal amplitude weights 0.9:0.1, which translates to noise ratios of 6.5 and -12.6dB respectively, only the strongest signal could be demodulated. This example is highlighted with an orange outer

line in the table. The highlighted example also shows that the weaker signal was buried too much in noise and could not be recognized and that the application of a Viterbi decoder did not give a correct improved demodulated solution.

4.5.2.1 Resistance to noise

Let us start with Amplitude-noise analysis with simulation results presented in Table 6 below. The amplitudes of AIS signals were varied from 0.5:0.5 to 0.9:01 ratios. The standard deviation of AWGN (Additive White Gaussian Noise) was varied from $\sigma(N) = 0.1$ to $\sigma(N) = 0.8$. The Doppler frequency shift of Signal1 and Signal2 were fixed to 400 and 650Hz frequencies and its respective phase shifts were selected randomly as 0.561 and 0.867 seconds.

AMPL1	AMPL2	e1	e2	SNR1	SNR2	e1	e2	SNR1	SNR2	e1	e2	SNR1	SNR2	e1	e2	SNR1	SNR2	e1	e2	SNR1	SNR2	e1	e2	SNR1	SNR2	e1	e2	SNR1	SNR2
0.900	0.100	0|0	0|0	16.01	-3.08	0|0	0|2	9.99	-9.10	0|0	8|6	6.47	-12.62	0|0	18|20	3.97	-15.12	0|0	42|36	2.03	-17.06	0|0	50|44	0.44	-18.64	0|0	57|56	-0.89	-19.98	0|0	62|58	-2.05	-21.14
0.800	0.200	0|0	0|0	14.98	2.94	0|0	0|0	8.96	-3.08	0|0	0|0	5.44	-6.60	0|0	0|0	2.94	-9.10	0|0	0|2	1.01	-11.04	0|0	4|0	-0.58	-12.62	0|0	8|6	-1.92	-13.96	0|0	7|24	-3.08	-15.12
0.667	0.333	0|0	0|0	13.40	7.38	0|0	0|0	7.38	1.36	0|0	0|0	3.86	-2.16	0|0	0|0	1.36	-4.66	0|0	0|0	-0.58	-6.60	0|0	0|0	-2.16	-8.18	0|0	0|0	-3.50	-9.52	0|0	0|4	-4.66	-10.68
0.600	0.400	0|0	0|0	12.49	8.96	0|0	0|0	6.47	2.94	0|0	0|0	2.94	-0.58	0|0	0|0	0.44	-3.08	0|0	0|0	-1.49	-5.02	0|0	0|0	-3.08	-6.60	0|0	0|0	-4.42	-7.94	0|0	0|0	-5.58	-9.10
0.556	0.444	0|0	0|0	11.82	9.88	0|0	0|0	5.80	3.86	0|0	0|0	2.28	0.34	0|0	0|0	-0.22	-2.16	0|0	0|0	-2.16	-4.10	0|0	0|0	-3.75	-5.68	2|0	2|0	-5.08	-7.02	2|0	2|0	-6.24	-8.18
0.524	0.476	0|0	0|0	11.31	10.48	0|0	0|0	5.29	4.46	0|0	0|0	1.76	0.94	0|0	0|0	-0.73	-1.56	0|0	0|0	-2.67	-3.50	0|0	0|0	-4.26	-5.08	0|0	0|0	-5.60	-6.42	4|0	4|0	-6.76	-7.58
0.502	0.498	0|0	0|0	10.95	10.86	0|0	0|0	4.92	4.84	0|0	0|0	1.40	1.32	0|0	0|0	-1.10	-1.18	0|0	0|0	-3.03	-3.12	0|0	0|0	-4.62	-4.70	0|0	0|0	-5.96	-6.04	2|0	2|0	-7.12	-7.20
0.500	0.500	0|0	0|0	10.90	10.90	0|0	0|0	4.88	4.88	0|0	0|0	1.36	1.36	0|0	0|0	-1.14	-1.14	0|0	0|0	-3.08	-3.08	0|0	0|0	-4.66	-4.66	0|0	0|0	-6.00	-6.00	0|0	0|0	-7.16	-7.16
	σ(N)	0.100				0.200				0.300				0.400				0.500				0.600				0.700				0.800			

Table 6. JShBDD Amplitude-Noise demodulation analysis for two overlapping signals with 400 and 650Hz fixed Doppler shifts. The orange box highlights the explanation example mentioned in the last paragraph of the previous section.

It can be learned from Table 6 that the JShBDD is very resistant to noise. The table also shows that the method performs well for cases where the weakest signal is above -10dB of SNR. The strongest signal can be demodulated even in a -7dB environment, even with the presence of an interfering signal of equal or lower amplitude.

4.5.2.2 Resistance to phase shift

In Section 4.2 during the analysis of JShBD application to signal without Doppler frequency shift it was mentioned that the demodulation is impossible in the case when the phases and amplitudes are similar. This was due having two signals of zeros frequency shift, what for similar bit clocks of the two overlapping signals lead to similar phases in similar timing, without chance to distinguish which basic shape belongs to which signal.

Meanwhile, in the case with Doppler frequency and phase shift present, the waveforms of both signals are different and that allows the signals to be distinguished even if its symbol clocks are at equal timings. This is a strong benefit and can be observed in Table 7 below. The table shows simulation results of the Amplitude-Phase shift variations. To include some error margin, the amplitude of noise was set to $\sigma(N) = 0.5$.

SIR_{dB}	SIR	Diff %	AMPL1	AMPL2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2
19.08	81.00	900%	0.900	0.100	0|0	26|28	0|0	20|36	0|0	30|34	0|0	32|32	0|0	30|24	0|0	40|44
12.04	16.00	400%	0.800	0.200	0|0	2|0	0|0	2|0	0|0	0|0	0|0	0|4	0|0	0|2	0|0	0|2
6.02	4.00	200%	0.667	0.333	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
3.52	2.25	150%	0.600	0.400	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
1.94	1.56	125%	0.556	0.444	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
0.83	1.21	110%	0.524	0.476	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
0.09	1.02	101%	0.502	0.498	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
0.00	1.00	100%	0.500	0.500	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
					50%		40%		30%		20%		10%		0%	Bit phase shift

Table 7. JShBDD Amplitude-Phase analysis for two overlapping signals with 400 and 650Hz fixed Doppler shifts, and fixed noise amplitude to σ(N)=0.5.

It can be learned from Table 7 that the bit phase shift variations do not influence the JShBDD capability to successfully demodulate overlapped signals.

4.5.2.3 Resistance to Doppler shifts

The results of Doppler shift resistance are reported in Table 8 below. As it is expected when considering the design of the JShBDD algorithm, the method can handle a wide range of Doppler shifts very easily and is completely resistant to those within the real world satellite AIS application range of ±2kHz for a 60° opening angle antenna.

SIR_{dB}	SIR	Diff %	AMPL1	AMPL2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2
19.08	81.00	900%	0.900	0.100	0|0	22|28	0|0	38|30	0|0	30|36	0|0	26|38	0|0	28|34	0|0	40|20	0|0	40|36
12.04	16.00	400%	0.800	0.200	0|0	0|2	0|0	0|2	0|0	0|2	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|2
6.02	4.00	200%	0.667	0.333	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
3.52	2.25	150%	0.600	0.400	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
1.94	1.56	125%	0.556	0.444	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
0.83	1.21	110%	0.524	0.476	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
0.09	1.02	101%	0.502	0.498	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
0.00	1.00	100%	0.500	0.500	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0
					400:450		400:650		400:900		400:1150		400:1400		400:1900		400:2400	DF1:DF2
					50		250		500		750		1000		1500		2000	Δ Diff [Hz]

Table 8. JShBDD Amplitude-Doppler analysis for two overlapping signals. The first signal has a fixed Doppler shift of 400Hz; the Doppler shift for the second signal is varying. The noise amplitude is fixed to σ(N)=0.5.

4.5.3 Resistance to input parameter deviations

All the simulations in the previous section were carried out with precise input of the 4 parameters for each signal: amplitude, bit clock, Doppler shift frequency and phase. However, precise derivation of these values from an input signal could be highly complex. In this section we will briefly look at how stable the correct parameter solution is, by looking at deviated waveform "distances" to the initial signal in sense of L^2 norm. The deviated waveforms will be constructed by slightly varying one of the input parameters.

Doppler shift frequency

Let us start with a stability analysis for deviation of the Doppler frequency parameter by slightly varying the Doppler frequency shift of the Signal1. Below in Fig. 71 are shown the real parts of IQ waveforms and resulting L^2 norm distances to the initial signal.

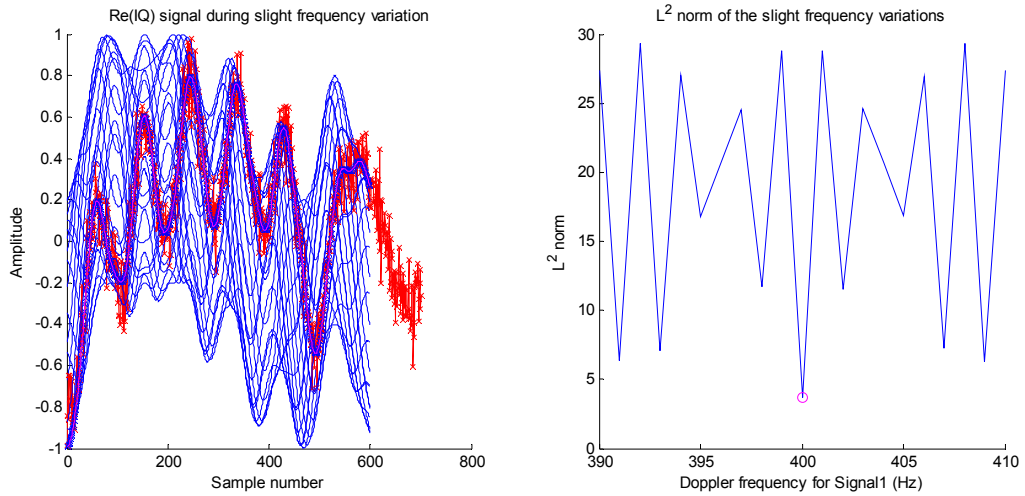


Fig. 71. L^2 norm sensitivity to frequency deviation for a 12 bit long AIS signal. Signal weights are [S1:S2:σ(N)]=[0.6:0.4:0.1]. Doppler frequencies are 400 and 650Hz and Doppler frequency phase shifts are 0.561 and 0.867 seconds for Signal1 and Signal2 respectively. The second signal has a phase delay of ½ bit. On the left side: red is the provided signal, blue are curves for the same signal but with varied Doppler shifts. Magenta line with thick blue curve on the left is the correct input curve and identifies the minimum norm displayed with a magenta circle in the plot on the right.

In Fig. 71 b) it can be observed that L^2 norm has a local minimum of only ±3Hz around the correct central frequency. However, the L^2 norm distance is symmetric to the middle point, which can help to identify it. By observing explicitly the surrounding nearby local distance minimums of ±6Hz, Fig. 72 can be obtained.

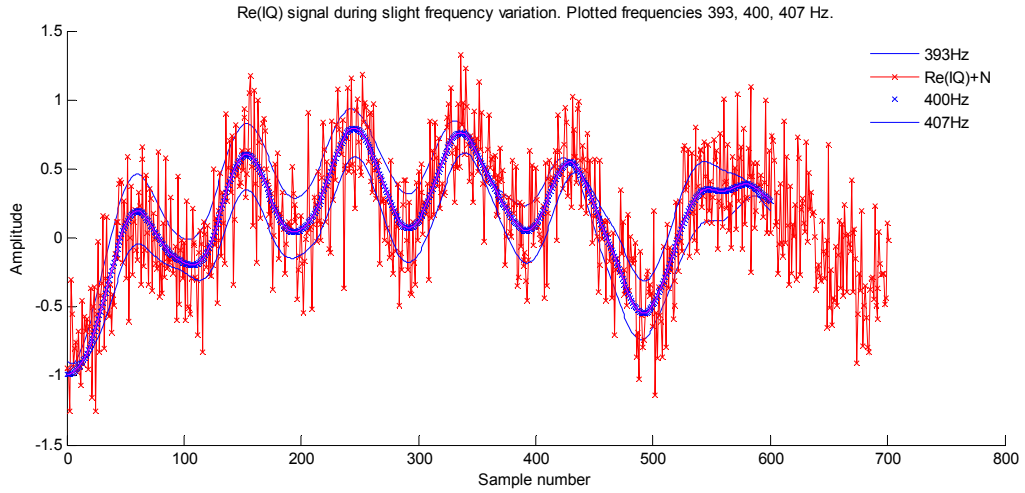


Fig. 72. Plot of the signal multiplied with the other two Doppler frequencies of 393 and 407Hz. These two frequencies can be observed in Fig. 71 b) as the local minimums of the L^2 norm, surrounding the correct one.

As Fig. 72 above reveals, the IQ signals modulated with the two nearby local minimum frequencies having a very similar L^2 norm, mimic the initial signal with either stronger or smaller amplitudes. This

is a positive conclusion, as it shows that the Doppler shifted waveforms with nearby minimum L^2 norms can approximate the initial waveform very well, even if the evaluated Doppler shift parameter has error of few Hz. But this indirectly also suggests that it will be problematic to find a precise signal amplitude and frequency jointly, as they appear to be related. This relation can be observed in Fig. 72 as the signals with 393 and 407Hz Doppler shift frequencies have its amplitudes respectively grater or lower than the initially modulated signal with 400Hz Doppler shift.

Doppler shift phase

Next, let us look at simulation results in which when the phase of the Doppler shift differs from the actual one.

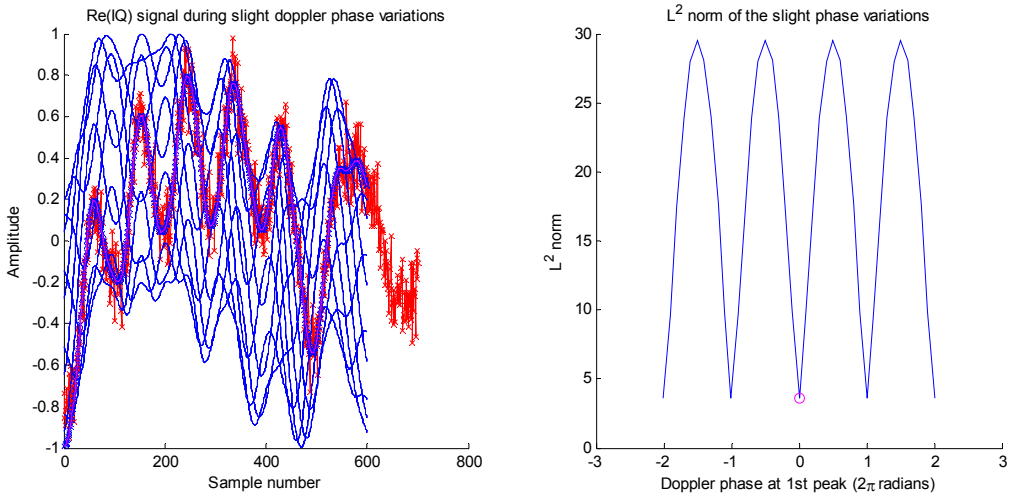


Fig. 73. L^2 norm sensitivity to Doppler shift phase variation for a 12 bit long signal.

As expected, Fig. 73 b) reveals that the minimum norm for Doppler shift phase variation is a cyclic function of time, which corresponds to the period of time taken for one oscillation of the underlying Doppler shift frequency. The positive conclusion is that this parameter is very stable and a convergent within a single oscillation period, thus also convenient for iterative minimum norm finding method applications.

Amplitude

As a further step, let us look at the impact of amplitude deviation. Fig. 74 below reveals the simulation results.

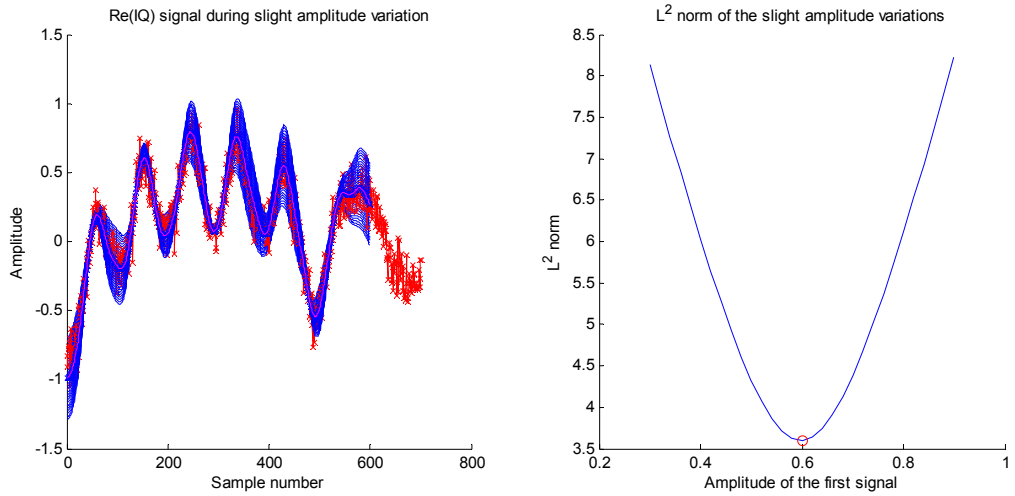


Fig. 74. L^2 norm sensitivity to amplitude deviation for a 12 bit long signal.

As Fig. 74 b) very clearly demonstrates, the amplitude variation is smoothly convergent to the correct setting with the minimum norm exactly matching the input signal.

Symbol clock

Finally, the last analysis parameter is bit clock. Fig. 75 below shows its deviation properties.

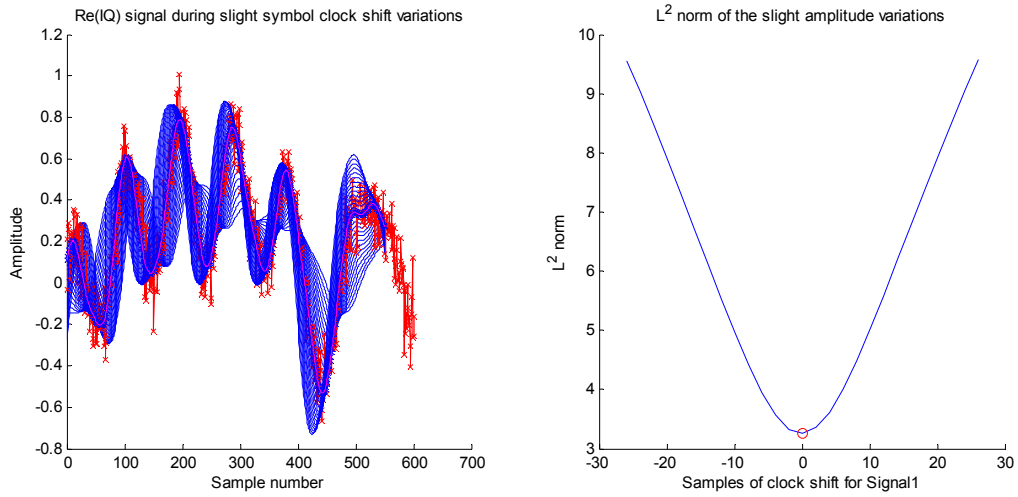


Fig. 75. L^2 norm sensitivity to symbol clock deviation of the first signal for a 11 bit long signal.

Also for this parameter the correct solution is found in the lowest point of a smoothly convergent valley, as observable in Fig. 75 b).

The analysis of the four parameter deviations above identify that the parameter estimation altogether could be a complex task for characterizing input signal. The most difficult task could be to finding the precise frequency, as its nearby values also provide a good approximation. The enlightening property of the frequency parameter is that, if the found minimum lays within few Hz from the correct parameter value, as far as it is local minimum, it still points to a well approximated initial waveform, thus helping to find some further bits. Every additional demodulated bit would increase number of available reference samples, leading to improved estimation of frequency and other parameters.

The deviating parameter analysis concludes this chapter. The application analysis of the JShBDD algorithm is following in Chapter 6.

5 Approaching demodulation with frequency recognition

In this chapter the author of this work will unveil his own developed GMSK signal demodulation approach for overlapping AIS message recognition, named JGSUM algorithm. The method is based on frequency analysis, the framework of which will be introduced in the first section, Section 5.1, using external literature references. The sections 5.2 to 5.3 will present frequency extraction analysis made by author of this work, while the last section, Section 5.5, will introduce the own developed JGSUM algorithm and analyse its characteristics.

5.1 Introduction to Model Based Demodulator (MBD)

The Model Based Demodulator algorithm is an often chosen demodulator for cases having co-channel interferences. It estimates momentaneous frequencies contained in some chunk of signal and tries to decompose them in their counterparts. The main building block behind the MBD algorithm is the consideration that the overlapping signals have different instantaneous signal frequencies due to symbol clock timing shifts and Doppler differences, which could be tried to be separated. Therefore let us start with considering signal frequency estimation.

5.1.1 The Forward Backward Linear Prediction (FBLP) algorithm

To estimate the frequencies contained within a signal, linear prediction methods can be exploited. The following three approaches are widely used [35]:

- Forward Linear Prediction (FLP) – using upcoming samples to estimate frequency components;
- Backward Linear Prediction (BLP) – using past samples to estimate frequency components;
- Forward-Backward Linear Prediction (FBLP) – using past and upcoming samples of signal to estimate frequency components.

The idea behind these methods is taking a certain number of samples, setting up and solving the linear equations to estimate signal frequency using upcoming, past or both tails of the signal.

Let us take a brief look at the algorithm. Without loss of generality, one can always assume that some finite number of samples can be approximated with a finite number of sinusoids:

$$y(n) = \sum_{k=1}^{M} a_k e^{s_k n}, \quad n = 1, 2, \dots N. \tag{42}$$

In (42) the $y(n)$ are samples of the given signal, M is the number of exponential signals contained, a_k is an unknown complex amplitude of the sinusoid with index k and e^{s_k} is the unknown sinusoids. Thus, s_k and a_k should be found.

There can be defined a prediction-error polynomial $G(z)$, which annihilates the $y(n)$, as shown in (43).

$$G(z)\{y(n)\} = 0\ ,\ \text{for all}\ n. \tag{43}$$

To carry out a general example, when $y(n)$ being composed of two sinusoids, meaning $M = 2$, to fulfil the (43) the $G(z)$ can be introduced as the polynomial (44).

$$\begin{aligned} G(z) &= 1 + g_1 z^{-1} + g_2 z^{-2} \\ g_1 &\coloneqq -(e^{s_1} + e^{s_2}) \\ g_2 &\coloneqq e^{s_1 + s_2} \end{aligned} \tag{44}$$

The searched roots (zeros) of the $G(z)$ in (44) are e^{s_1} and e^{s_2}. Therefore, finding g_1 and g_2 for a given sequence of samples $y(n)$ would provide s_1 and s_2. To showcase an example of finding the coefficients, let us write the prediction equation explicitly as

$$\begin{bmatrix} y(2) & y(1) \\ y(3) & y(2) \end{bmatrix} \cdot \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} -y(3) \\ -y(4) \end{bmatrix}. \tag{45}$$

It can equally be rewritten as

$$\begin{bmatrix} y(3) & y(2) & y(1) \\ y(4) & y(3) & y(2) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which solves for g_1 and g_2. Calculating the roots from the $G(z)$ polynomial gives the searched sinusoids e^{s_1} and e^{s_2} for (42). Then (46) lets to obtain the amplitudes a_1 and a_2 searched for (42).

$$\begin{bmatrix} e^{s_1} & e^{s_2} \\ e^{2s_1} & e^{2s_2} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y(1) \\ y(2) \end{bmatrix} \tag{46}$$

The number of samples processed in each step, the M, should be equal or greater than the number of signal components contained within the signal under analysis. Taking M greater than number of frequency components helps to overcome noise influences, but involves bigger matrices for computation.

We will stick to the FBLP method, as it includes benefits from the BLP and the FLP methods. The recommended M is ¾ of the number of samples under processing, denoted further with N [35].

This approach is called the Prony method and it can be summarized as follows: transform finding nonlinearly entering parameters s_k, $k = 1, 2 \ldots M$ into a problem of finding the coefficients $g_1, g_2, \ldots g_M$ of M-th degree polynomial

$$G(z) = 1 + \sum_{k=1}^{M} g_k z^{-k}\ . \tag{47}$$

Now, finding the $g_1, g_2, \ldots g_M$ and $a_1, a_2, \ldots a_M$ would provide an estimation of frequencies and their amplitudes contained within the signal, thus allowing one to decompose the signal.

5.1.2 Extraction of signal roots

One can rewrite (45) as $Ag = b$, concluding that following the Prony method, the g should be searched. The matrix A formed by the linear predication algorithms FLP, BLP or FBLP is a Toeplitz matrix and, in case of noiseless signal, its rank equals to the number of signal components M contained within the provided signal y [35]. However, it is not clear if the matrix A is invertible for all signals. Indeed, for cases when linear prediction in building the A matrix uses more processing samples than the frequencies contained within the signal, which is usually done to improve noise resistance, it is evident that the rank of the matrix will be less than its dimension. In such a case the searched solution is $\tilde{g}$, which is an approximation of g [35].

The following programming approaches for solving the equation $Ag = b$ for an approximation $\tilde{g}$ can be considered:

1) Finding the pseudo inverse A^+ of A [35], which can be written as

$$\tilde{g} = A^+ b.$$

This approach, without an inversion tolerance threshold, is expected to result in a weak performance. Its implementation with `pinv()` function in MATLAB, which uses the Moore-Penrose pseudo inverse, for a noisy AIS signal showed to be unstable, due to low rank matrices in use.

2) Using singular value decomposition (SVD) of matrix $A^H A$ and taking only M significant values of it [47], [35], as shown below:

$$\tilde{g} = (A^H A)^{-1} \cdot A^H b = \left(\sum_{k=1}^{M} \frac{1}{\sigma_k} v_k u_k^H \right) b. \quad (48)$$

Here the number M can be selected based on expected number of signal components, estimations during pre-processing stage or the amount of singular values σ_k, which are greater than an error tolerance ε.

In general, the MATLAB function `pinv()` actually uses SVD to calculate the singular values for inversion, thus both methods are similar. However, doing the work by SVD, allows explicit control over the root composition process.

When the $\tilde{g}$ is found, then using MATLAB the roots can be extracted as follows [47]:

```
G=[1;-g_tilde]
δ=roots(G)
```

The calculated deltas contain the signal roots. They can be plotted and are expected to mimic the underlying Gaussian signal of the initially modulated data, as described before in Section 2.1.2.

3) Using Multistage Wiener Filter (MWF) [48], which leads to the algorithm called JRRMBD (Joint Reduced Rank MBD) [47]. This method is based on finding orthogonal projections of the signal. Similarly to SVD, it uses an error tolerance to stop projection activity and resume

with signal reconstruction. In comparison to previous methods, it doesn't require the full computationally complex SVD calculation, still enabling the extraction of lower power interfered signals.

For further investigations, the SVD approach as shown in (48) is used, unless otherwise stated. This was chosen due the SVD approach being more straightforward to program and verify, while a basic MWF implementation showed only a small computation time improvement.

5.2 MBD Application examples

5.2.1 An example with two sinusoids

Let us look at the application of the MBD method to observe its capability to extract the frequencies of two overlapping sinusoids. For this case we construct two basic signals, each consisting of just a single sinusoid, such as an empty carriage signal with a Doppler component. The signals are modelled with amplitudes 0.4 and 0.6, to account for the case of alike amplitude signals with sum of them normalized to unity amplitude.

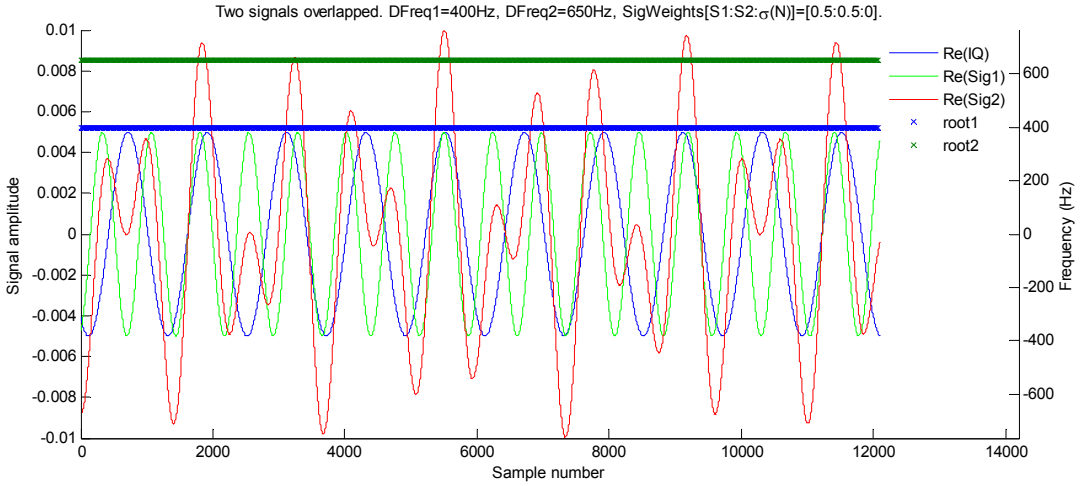


Fig. 76. Two extracted roots of two overlapping sinusoids by MBD method. On the left vertical axis – amplitude of the IQ signal or an anlge of the extraced roots in radians. On the right vertical axis – frequency of the extracted roots, in Hz.

As Fig. 76 above reveals, using the MBD approach with FBLP and SVD, the frequencies of the two sinusoids can be very clearly extracted and identified. An equal extraction capability also holds true for the case of the two sinusoid frequencies being very close, like 400 and 405Hz. The same is also true when both signal amplitudes are very close or equal.

5.2.2 An example with a single AIS signal

Let us consider a practical AIS example for the MBD application with SVD approach. From IQ data of a real AIS message modulated with GMSK and small Doppler shift addition, a sample range of few bits were taken for analysis.

Let us introduce another variable L – the number of roots taken to build the matrix for calculating the $G(z)$ polynomial, as in equation (47). While M was introduced as the number of frequencies contained in a signal, L is used as number of searched frequencies in a signal. In case the signal is composed only of two sinusoids $L = M$ would be the choice. However a single GMSK signal contains multiple frequencies within it, thus for AIS message processing the L is in general greater than the number of GMSK modulated signals contained within an input signal. We will start the analysis by calculating FBLP with 8 samples sliding window with 5 searched roots, performing SVD, calculating 5 roots and picking out just 1 or 2 selected ones.

Let the signal shown in Fig. 77 be an extraction of an IQ signal from a real GMSK modulated AIS message, which will be used as the input IQ signal for analysis within this section.

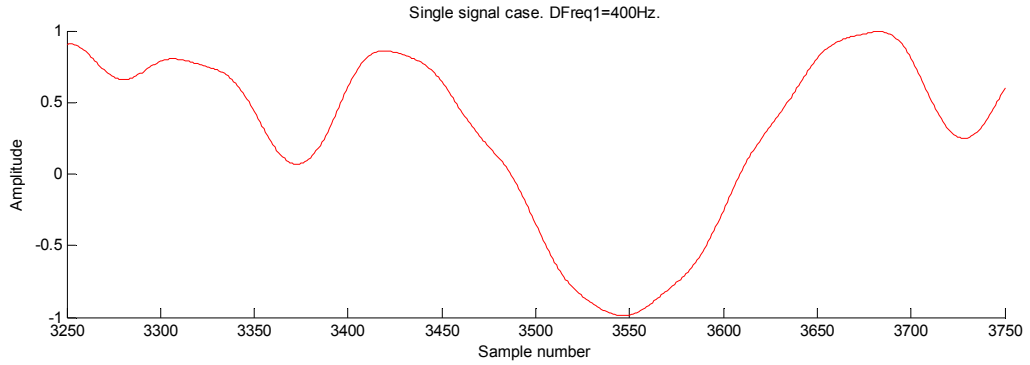


Fig. 77. Real part of IQ signal to be used for analysis, containing doppler frequency 400Hz. OSR is 50spb.

The signal from Fig. 77 processed with FBLP and passed to SVD, generates the roots shown in Fig. 78.

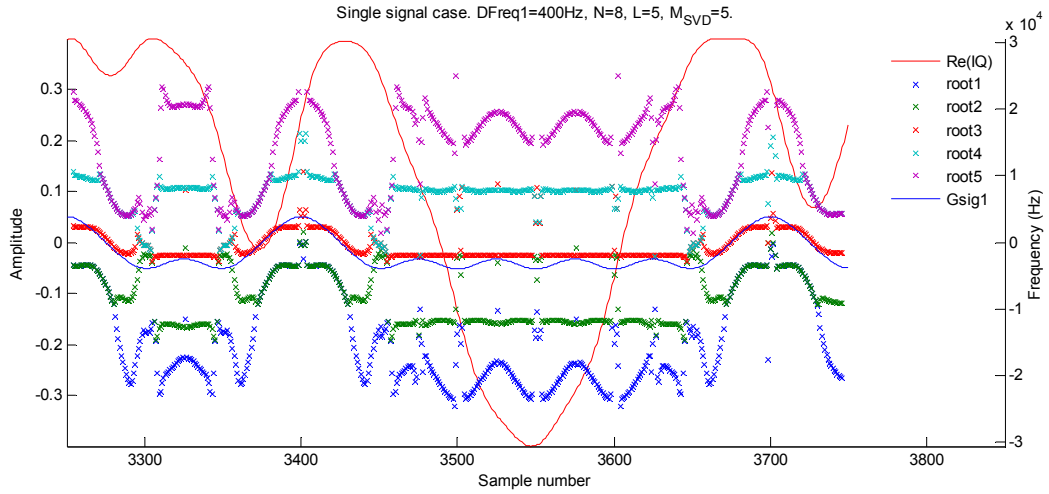


Fig. 78. Five roots of MBD, along with the initial IQ signal, scaled to 0.4, and the initially modulated Gaussian signal, scaled to 0.05.

The five roots shown in Fig. 78 above are ambiguous to associate its lines to the input Gaussian signal. Generally the roots are not wrong, but they approximate different parts of the signal and in some parts are similar.

As the sample under current analysis contains just a single GMSK modulated component, limiting SVD to M=1 signal component had to be consistent for the contained frequency extraction under conditions of high oversampling rate. This can be well ascertained from the following Fig. 79.

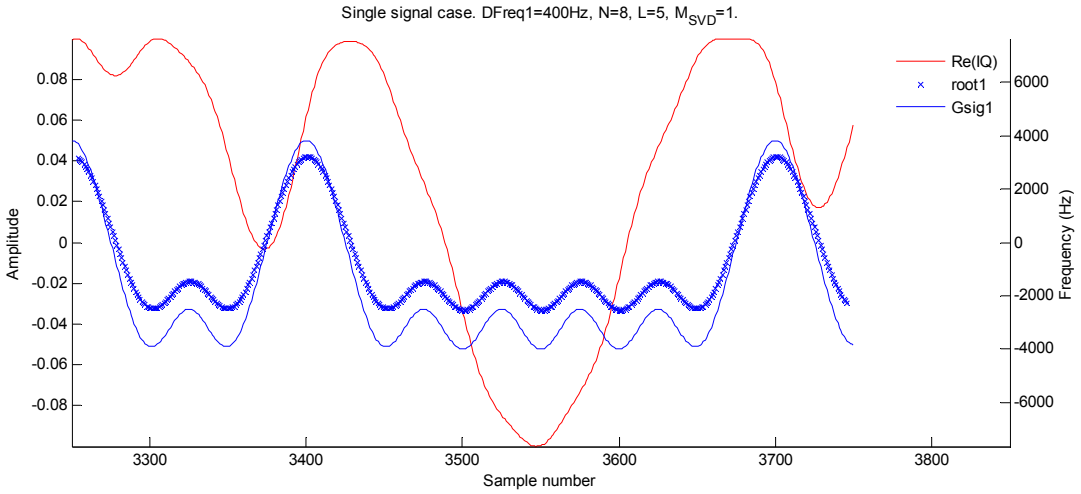


Fig. 79. The first root of frequency extraction, along with the initial IQ signal scaled to 0.1 and initially modulated Gaussian signal, scaled to 0.05.

As expected in an absence of noise, Fig. 79 shows that the first root of SVD reconstructs a complete demodulated signal. One will note that the root-curve in thick blue is shifted a bit up with respect to the Gaussian curve in thin blue. This is due to the Doppler shift of 400Hz contained within the input IQ signal, which makes the signal roots to shift up respectively. The roots from Fig. 79 plotted on a unity circle are shown in Fig. 80 below.

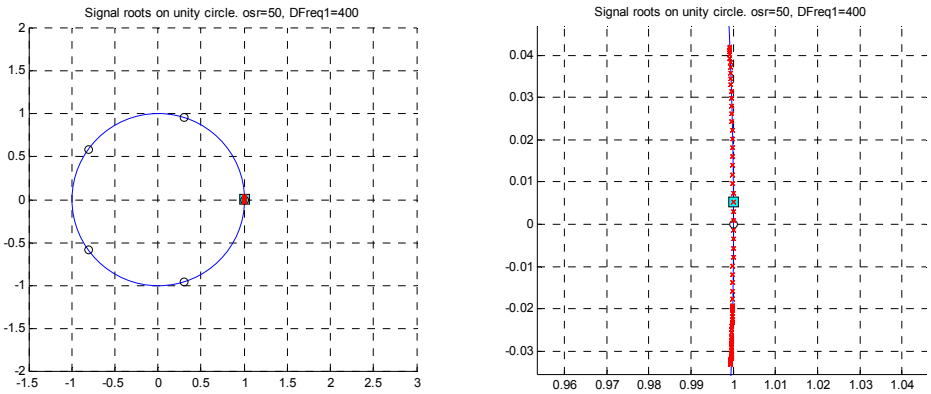


Fig. 80. Roots of M_{SVD}=1 case plotted in red on a unit circle in blue. a) Unity circle plot. b) Unity circle plot zoomed around roots. The bluish square marks the complex Doppler frequency value on the unit circle.

On the unity circle plot in Fig. 80 b), one can see that on the negative angle side the root values only go down to -0.032, whereas on the positive side they go up to 0.042, thus also here pointing to the presence of a Doppler frequency shift. One can also notice that the red roots are equally distant from the Doppler shift point, the blue square slightly above 0. The Doppler shift point plotted in Fig. 80 is based on input parameter value. Alternatively it could be calculated as taking the middle between the amplitudes of the roots.

This demonstrates the comfortable usage of SVD for finding the signal and its included Doppler frequency for interference-free AIS signal cases. Since a GMSK modulated signal contains multiple frequencies due to the Gaussian filter used in signal modulation, it is expected that the single GMSK signal will produce roots in some range rather than at one single point. The width of the GMSK signal

frequency band depends on the bit-time product modulation factor and the bit rate, which in the case of AIS is BT=0.4 and 9600bps respectively [1].

In Fig. 80 a) there could be noted 5 black circles, uniformly distributed from the middle point of the signal. These are additional roots which may mimic some properties of the main signal root but at lower quality. These are neglected by taking roots which are in only $\frac{\pi}{8}$ angle surrounding the zero frequency, as that range is enough to include full GMSK frequency spectrum along with maximal permissible Doppler shift of AIS signals.

This root extraction and selection method is very comfortable for a single signal extraction with a small Doppler – as it allows for slicing bits even without a precise pre-estimate of the Doppler value. However, the comfortable frequency extraction approach has its cost – it is very sensitive to noise. This is investigated in the upcoming section.

5.3 Noise resistivity for single AIS signal extraction

To shortly introduce concerns regarding expected problems using the MBD demodulation with SVD root extraction, let us take a look at the frequency analysis of an IQ signal with noise. The noise used here is additive white Gaussian noise.

Fig. 81 and Fig. 82 below show frequency estimations for signal with noise weight $\sigma(N) = 0.1$ for $M = 5$ and $M = 1$ extraction approaches respectively.

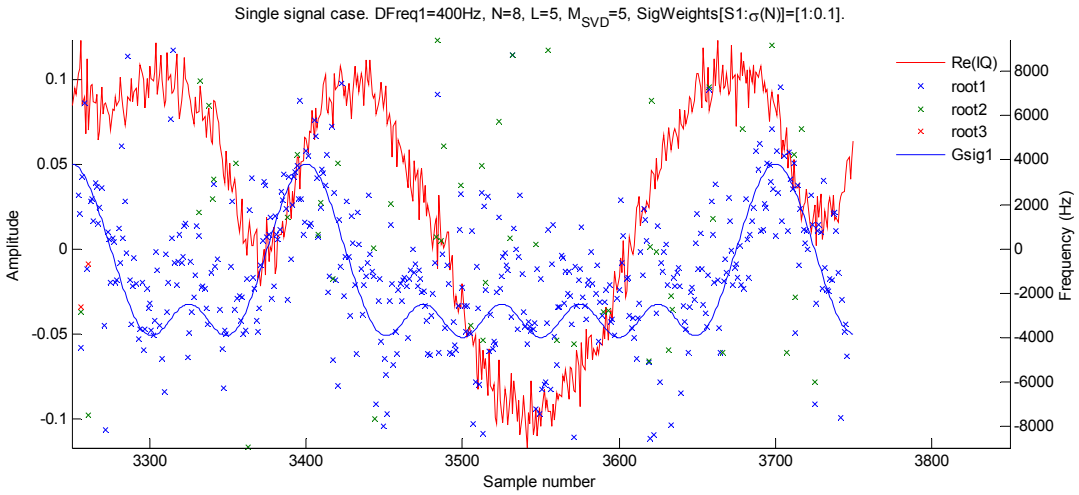


Fig. 81. Three roots from an IQ signal containing noise with weight $\sigma(N) = 0.1$. Some root points are beyond the zoomed window and can be filtered out due laying beyond the expected signal frequency range.

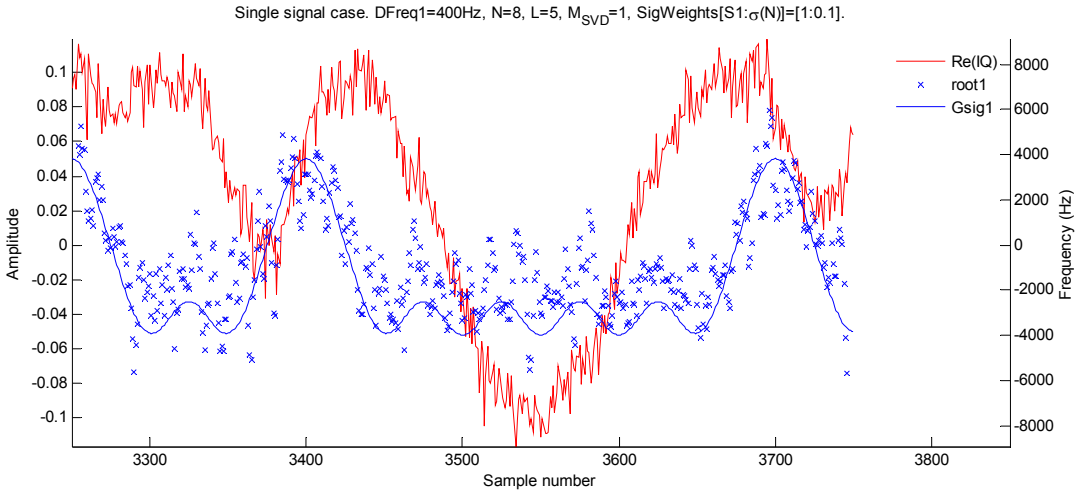


Fig. 82. The first root from the input signal containing noise $\sigma(N) = 0.1$.

The highly scattered roots in Fig. 81 and Fig. 82 clearly point to very pessimistic chances for signal extraction already under small noise conditions. Fig. 82 is cleaner, as there is only the first SVD root taken, lessening the noise impact, but still leads to much undetermined results. To cope with the scattered root issue, there is the option to take bigger N, so that more signal values are included in the frequency estimation, weighting and thus smoothing the estimated frequency value. Let us observe what the usage of a high sample number N can offer. We will preserve $M = 1$, as it demonstrates the best robustness to noise.

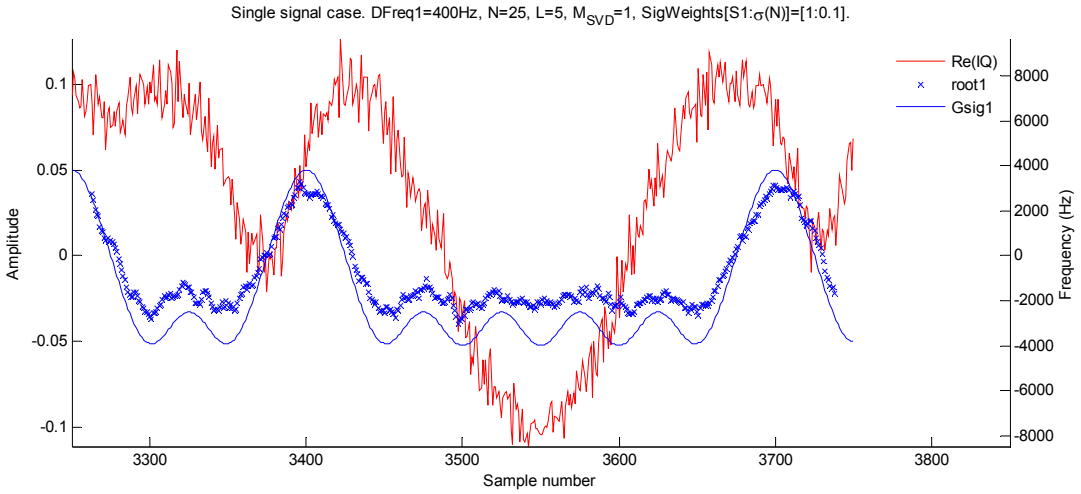


Fig. 83. The first root from signal with noise weight 0.1, with N=25 samples per FBLP processing window.

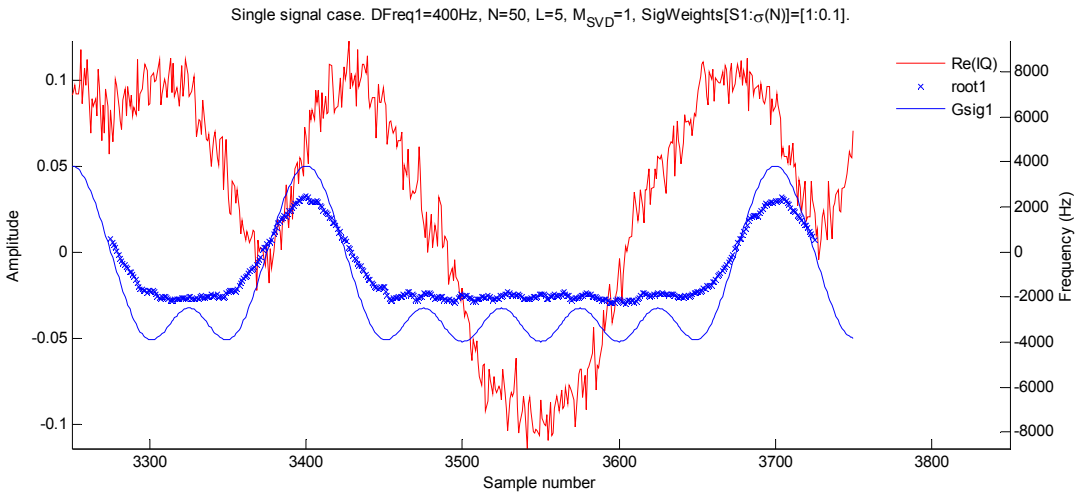


Fig. 84. The first root from signal with noise weight 0.1, with N=50 samples per FBLP processing window.

Observing Fig. 83 and Fig. 84 it can be noted that increasing the number of samples taken per processing window, does help to overcome noise. Increasing N over the number of samples per bit is not advisable, as the taken samples would include impact from the nearby bits, thus mixing the estimates.

From the publications [35], [36], [37] and [38] one can learn that the optimal relation between the number of searched frequency roots and the number of samples is $L = \frac{3}{4}N$. Let us increase the noise weight further to half of the signal amplitude, which corresponds to approximately 3dB SNR, and observe how L, the number of searched frequencies, influences the result under the new noise conditions.

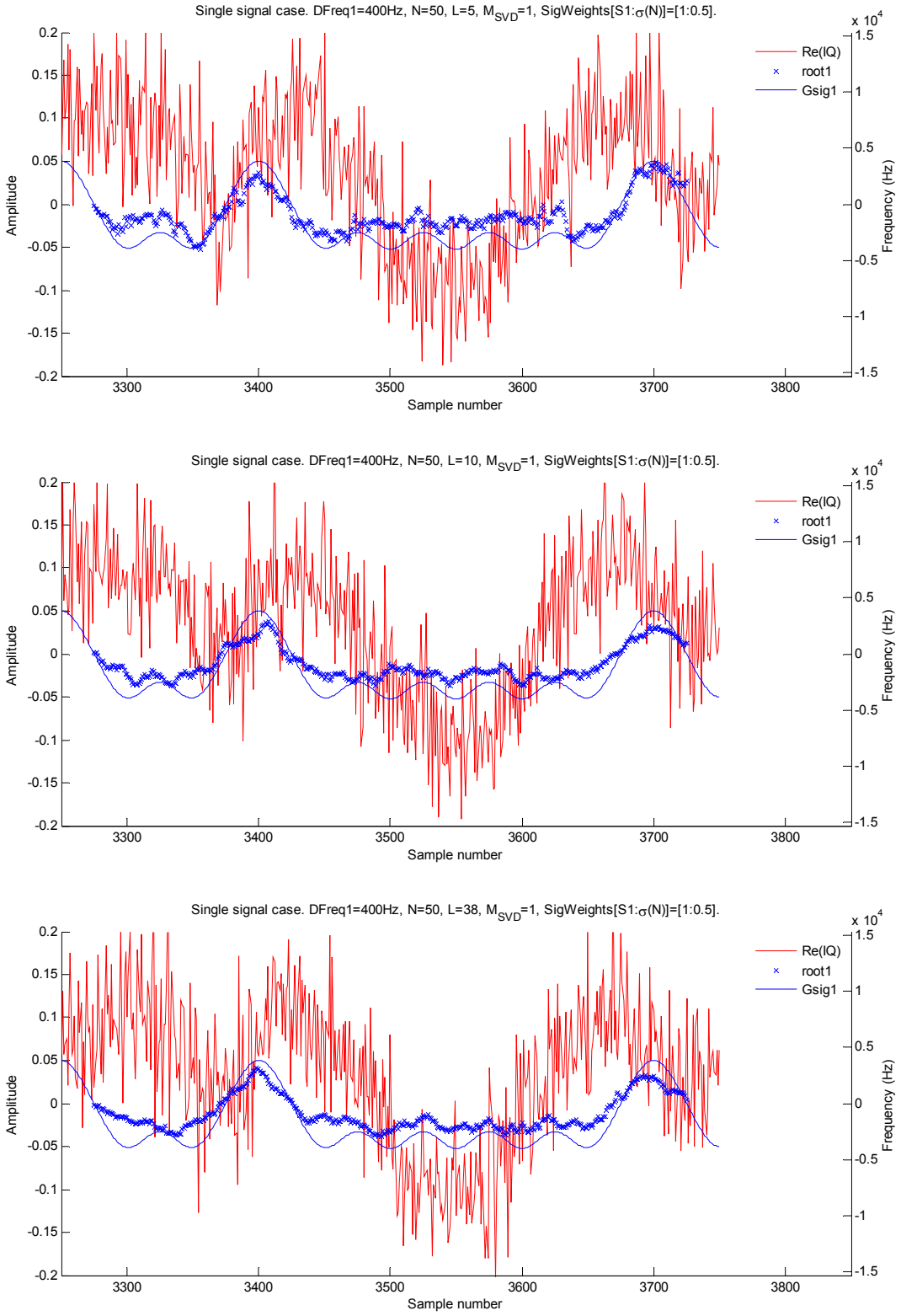
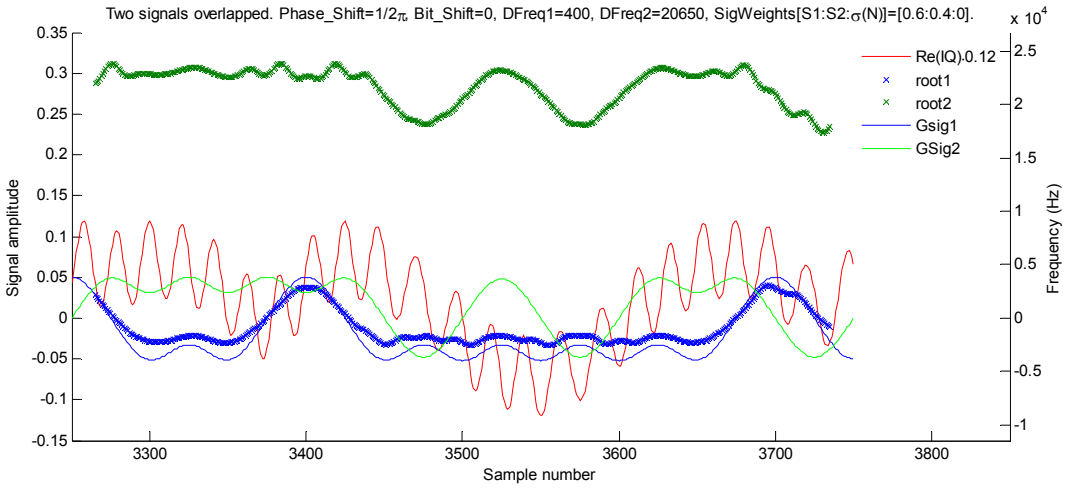


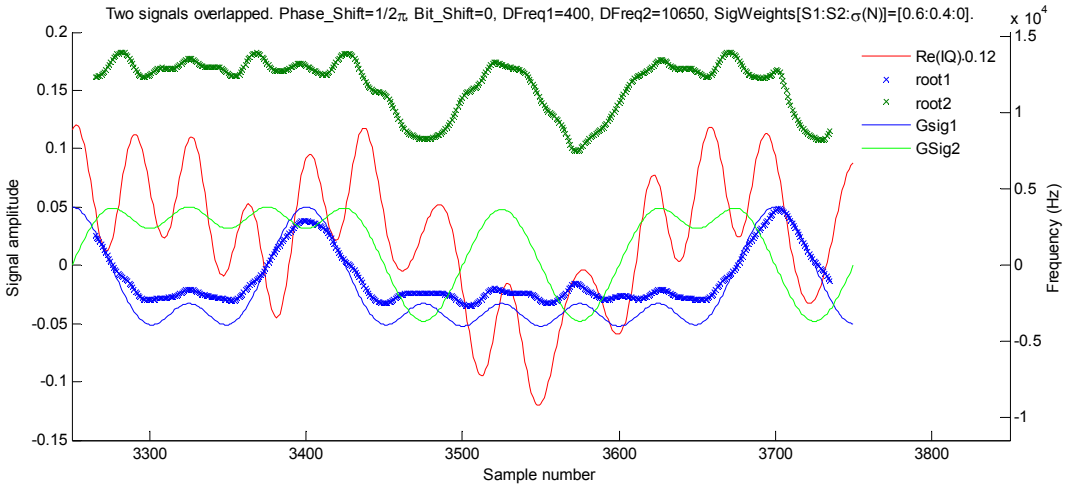
Fig. 85. The first root from signal with noise weight ½ of signal amplitude, SNR=3dB. With N=50 samples per processing window and a) 5, b) 10 and c) 38 frequencies searched.

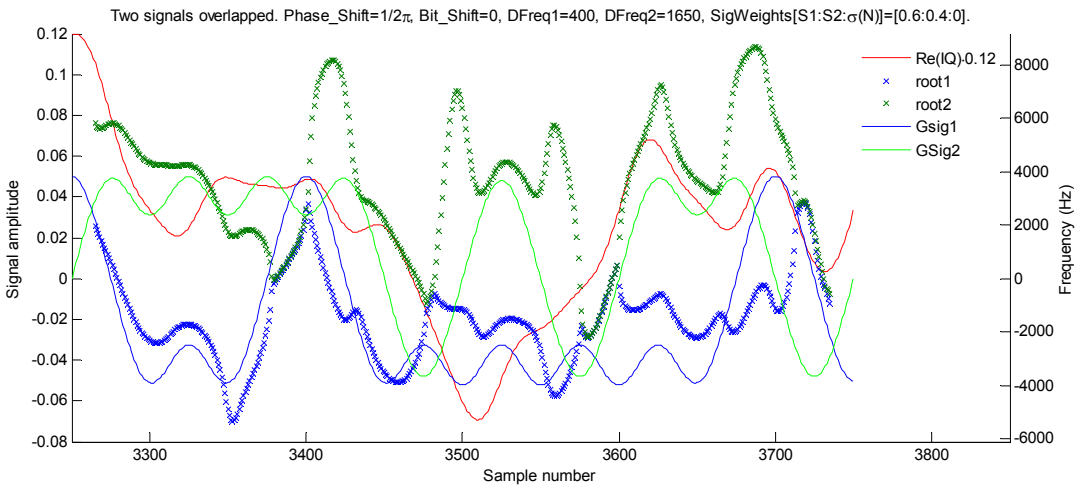
Observing Fig. 83, Fig. 84 and Fig. 85 a), b) and c) it can be concluded that both increasing the number of the samples par sliding window N as well as number of the searched frequencies L helps to approximate the underlying signal more smoothly and accurately.

5.4 Extraction of roots from two overlapping AIS signals with JMBD

For the next MBD application example, let us consider a case with IQ data containing two overlapping AIS signals. Since the bit rate of the GMSK signal is 9600bps and Gaussian shaping bit-time factor is BT=0.4, the frequency band used within 30dB range is nearly equal to bit rate. This means that overlapping signal frequencies should be separable as long as they differ by 10kHz. Below are examples of the first two SVD roots, extracted from the IQ data containing two messages.







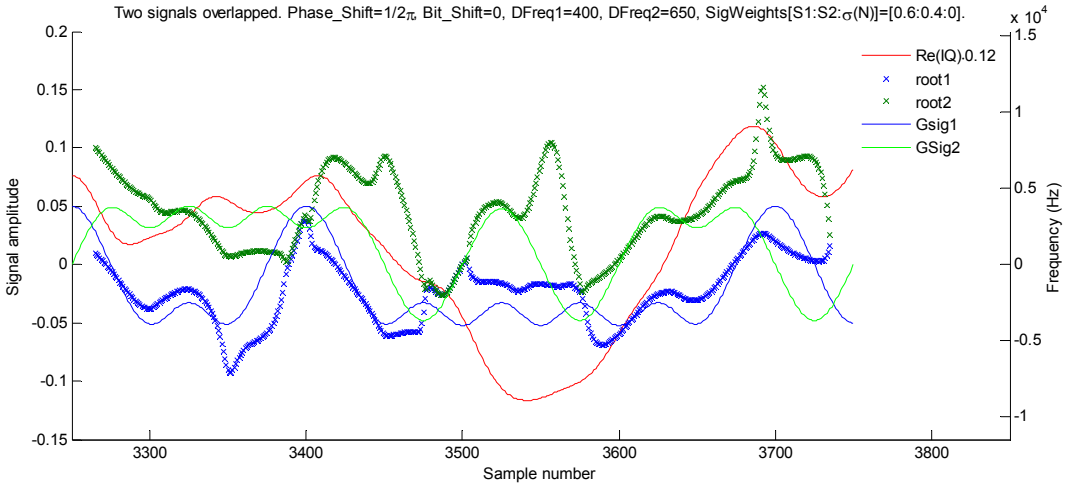


Fig. 86. The first two roots for a chunk of IQ signal containing two overlapping AIS messages, analysed for a) 20.25kHz, b) 10.25kHz and c) 1.25kHz d) 0.25kHz Doppler frequency differences. The thick lines made of crosses are the extracted roots and are expected to mimic the underlying Gaussian filtered signals. The input IQ signal is drawn with a red line, scaled with 0.12. The green and blue curves represent the initially input Gaussian signals, used to build the IQ waveform under analysis. The used MBD parameters: $N = 30$, $L = 5$, $M_{SVD} = 2$.

In Fig. 86 one can observe how MBD frequency separation performs for two overlapping AIS signals, which are altered from a far distant to a very near frequency shifts. Since the input IQ signal contains two AIS signals, the extracted root lines shown as lines from blue and green crosses are expected to mimic the Gaussian filtered signals from Signal1 and Signal2 respectively.

Fig. 86 a), b), c) and d) above demonstrate that for a frequency difference of more than 10kHz the signals can be indeed easily separated. In contrast, when a Doppler frequency difference between the signals nears to 1kHz, the frequency components of both underlying signals become mixed up and are disturbing each other.

Since Space-AIS is targeted to be used from LEO orbits with altitudes 600 to 700km and, as previously noted in Chapter 1, the Doppler range in this case is 3.6kHz maximum for a wide angle antenna and 1.3kHz for a narrow angle antenna, Fig. 86 c) indicates that a direct application of MBD method can't be fruitful, as it mixes up both of the target signals.

The next section will look at the MBD based demodulation from another viewpoint, which promises to be applicable for Space-AIS.

5.5 JGSUM demodulation algorithm

While working with analysis on MBD capability to extract two overlapping signals under conditions of small differences in signal amplitude and Doppler shift frequency, an interesting property of the extracted roots was found. The property served as a spinoff idea for a new demodulation algorithm development. The demodulation algorithm for the special case of equal amplitudes was named JGSUM (Joint Gaussian-Sums MBD). The next sections will explain how this algorithm works and which special overlapping cases it can successfully process.

The root extraction method from IQ signal using the MBD algorithm is an approach which is often implemented for frequency analysis of a signal. The application of the MBD algorithm to the special

case of two overlapping AIS signals which are of almost equal amplitude and using a special configuration to extract a single root from a signal is a new AIS signal demodulation approach, named JGSUM method, which is developed by the author of this document and presented in the following sections of this chapter.

5.5.1 Properties of a single root extraction from two overlapping signals

To seek a solution for cases in which two signals overlap with almost equal Doppler shifts and amplitudes, let us assume an extreme case, when overlapping signals have completely equal amplitudes. If this is the case, single root extraction constructs a signal which mimics a signal equal to the sum of the both contained Gaussian signal curves. This is shown in Fig. 87.

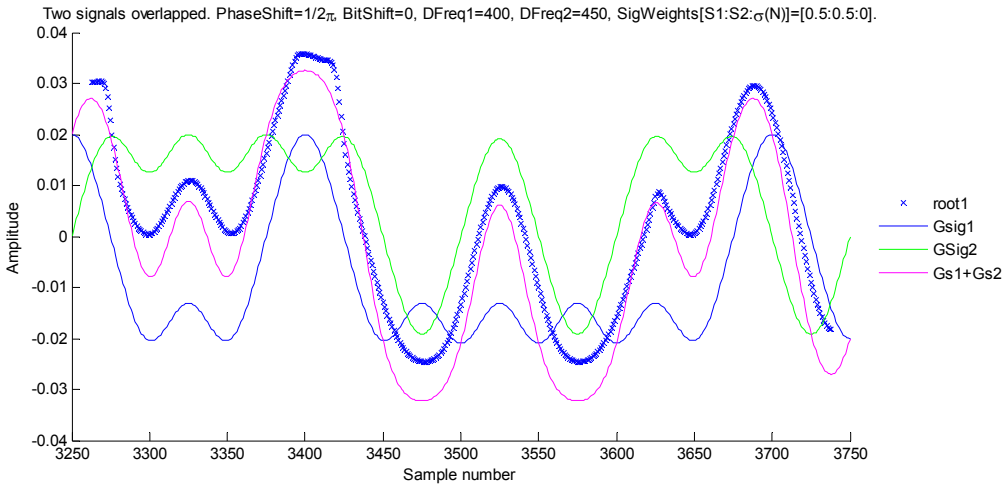


Fig. 87. The first root extracted from signal containing two overlapping AIS messages.

As Fig. 87 above reveals, the thick blue coloured curve of crosses well reflects the magenta coloured line, which in turn reflects the sum of both Gaussian curves. This serves as the motivation for the following algorithm approach: when the processing signal contains two densely overlapping GMSK signal components, apply a single frequency root extraction and search for two summed Gaussian signals within the roots.

However, a direct disadvantage for the JGSUM method is its high sensitivity to a signal amplitude deviation – it works only when signal amplitudes are identical or very close. The following example shows the result of root extraction for amplitudes 0.51 and 0.49, which differ by 4% of the amplitude value or 0.35dB by their powers, with 50 and 250Hz Doppler shift gaps between the two signals.

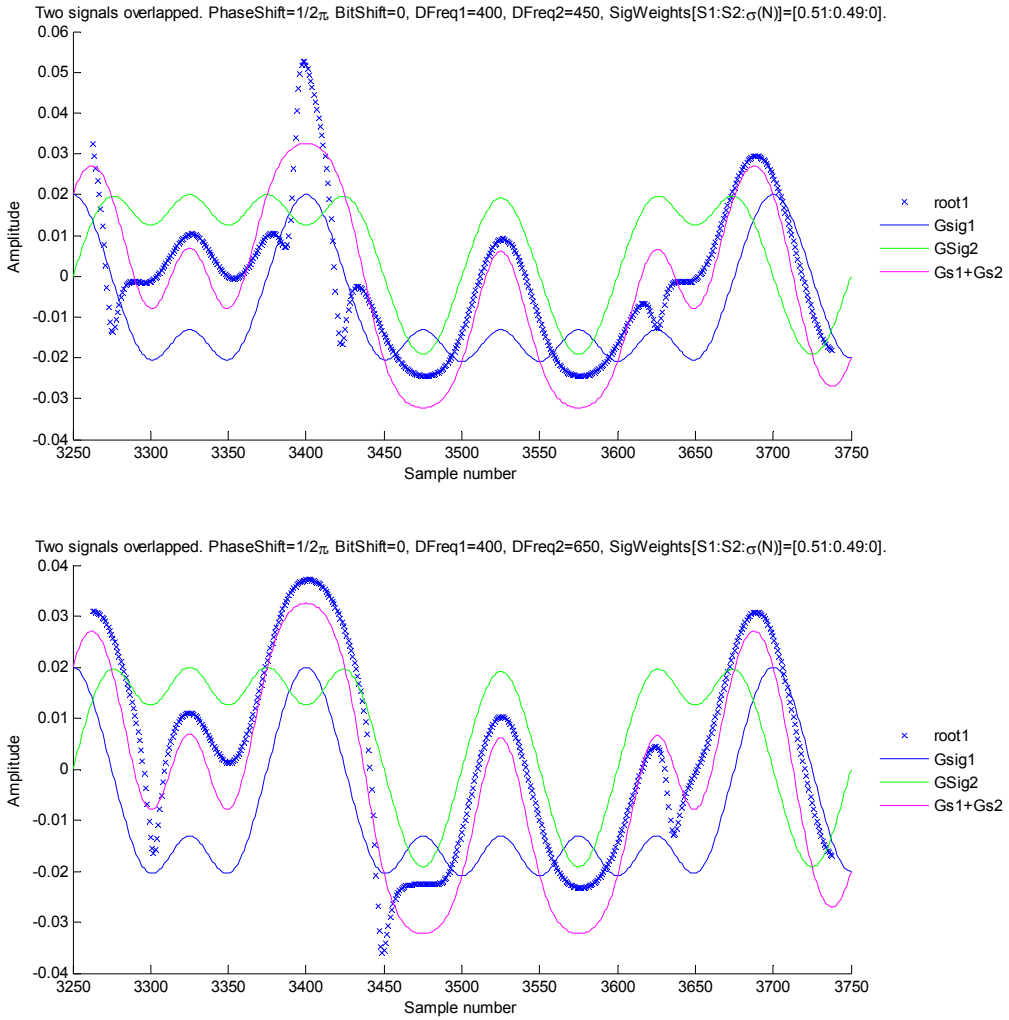


Fig. 88. The first root of two overlapping signals with almost equal amplitudes. The Doppler shifts for the first and the second signals are a) 400 and 450Hz and b) 400 and 650Hz.

As it can be noted in figures Fig. 87 and Fig. 88 above, the bold blue curve composed of crosses generally reflects the sum of underlying Gaussian signals. However it has some ranges with sharp edges upwards or downwards which do not correspond to the underlying magenta curve. The sharp edges differ in position, depending on individual signal Doppler shift frequencies and phases. The edges become more observable with an increased amplitude difference between the signals, as is shown in the following figures.

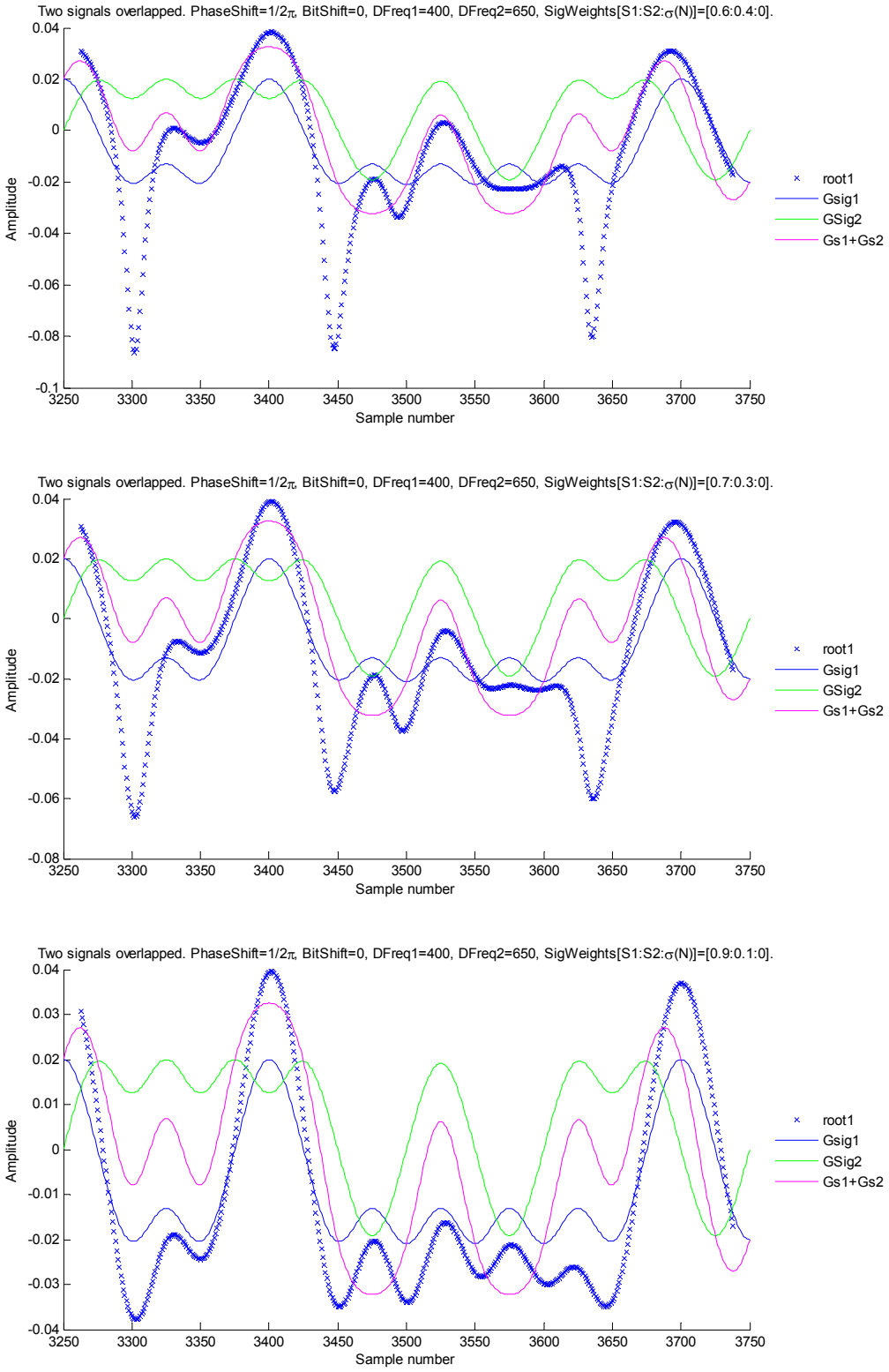


Fig. 89. The single-root of two overlapping signals with Doppler shifts 400 and 650Hz. Cases for amplitudes a) 0.6:0.4, b) 0.7:0.3, c) 0.9:0.1.

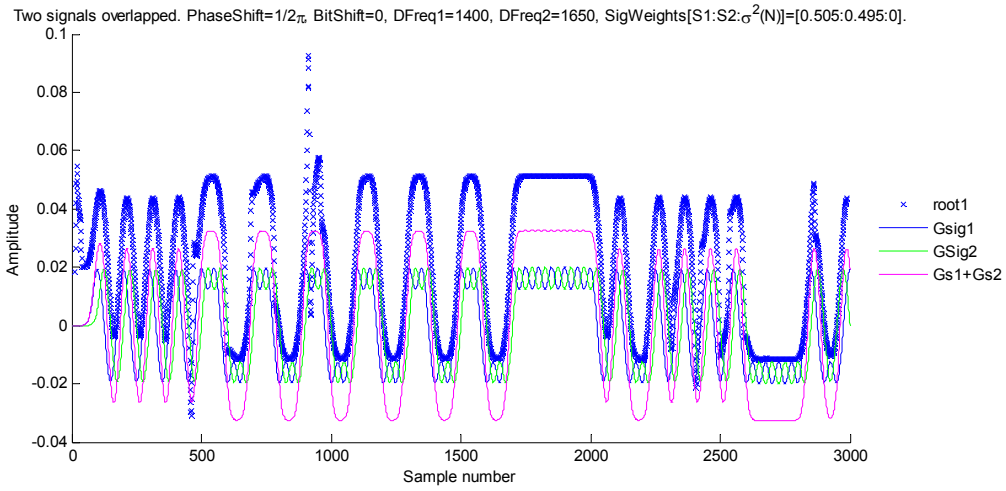
In Fig. 89 one can observe that the sharp down-falls from Fig. 89 a) turned to more and more saturated and smoother lines as the amplitude difference is increased Fig. 89 b), finally leading to a scenario in which the blue bold curve has recognizable characteristics of Signal1 and no Signal2 characteristics, Fig. 89 c).

The figures Fig. 87 to Fig. 89 lead to a conclusion that in the case of two signals overlapping with very near amplitudes, the first root of the MBD method can well reflect the Gaussian sum of the two underlying curves. When the amplitude difference between the two overlapping signals grows, sharp edges emerge, which are expected to disturb the recognizable characteristics of the summed curve. However, when the difference between the overlapping signals' amplitude is above some certain level, the first root curve well reflects the strongest contained signal in the IQ data, as it is well recognizable in Fig. 89 c). The signal amplitude levels at which the transition from two signals to single signal occurs will be analysed in Section 5.5.3 for different signal overlap scenarios.

5.5.2 Joint Gaussian sum demodulation for two overlapping signals (JGSUM)

To demodulate a single signal from MBD extracted roots, the demodulation process would require shifting roots signal to zero Doppler frequency offset, scaling it to 1, finding the bit clock and finally performing slicing to find the modulated bits. For the JGSUM algorithm the approach is very similar, with the except that the waveform composed of the extracted roots contains two Gaussian waveforms to be recognized and thus two bit clocks to be jointly sliced from the single waveform. In this chapter we will take a step by step look at the approach.

Let us start the JGSUM algorithm investigation by considering the extracted roots from an input IQ signal. As the first root is extracted from a single message IQ signal, it needs to be frequency shifted so that the maximum and minimum values are at equal distances from zero, thus compensating for the Doppler shift. Similarly in root extraction case from two overlapped signals, the root signal contains two signals with two different Doppler shifts and finding the mean frequency offset means finding the upper and lower bounds of the amplitudes contained within signal roots. Shifting the signal so that the upper and lower amplitude bounds are equidistant from zero and of unity amplitude would accomplish the preparation work. In the case of two overlapping signals, the required frequency offset does not point to a precise Doppler shift of the first or the second signal. Fig. 90 a) and b) below present the results of the single root extraction from two overlapping signals and the frequency offset compensation shift respectively.



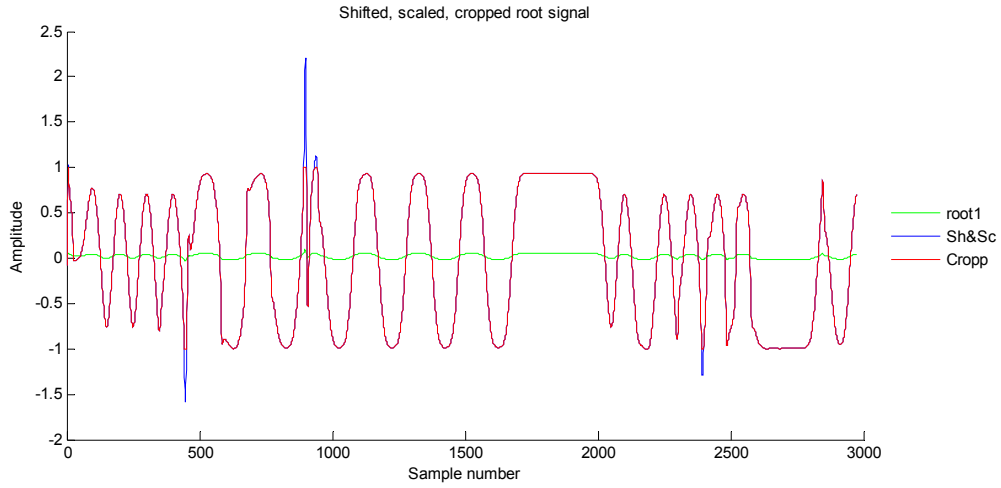


Fig. 90. a) Extracted root signal from IQ signal with two components b) extracted root signal (green), shifted and scaled (blue), cropped (red).

As it can be seen in Fig. 90 a), the extracted root signal oscillates in range of 0.015 to 0.055, which in Fig. 90 b) is then adjusted to be equidistant from zero and normalized to unity amplitude with ambiguous peaks cropped. The used approach to get the red curve in Fig. 90 b) from the blue roots curve in Fig. 90 a) consists of three following steps:

- Calculate the mean value of the roots signal and subtract it from the initial signal.
- Find the amplitude of signal peaks. Scale the signal to unity amplitude.
- Crop to 1 all the sharp edge peaks with amplitudes greater than 1.

Further improvements could be made to handle sharp edge peaks better than just cropping and also to perform signal amplitude search in a more advanced way than taking amplitude values from a smooth signal range. That would help to more precisely estimate the waveform adjusting parameters and increase robustness for automated applications. However, the used approach is well suited for demonstration purposes and completes the needed signal preparation for Gaussian shape search.

The next step is to find the symbol clocks of the bits contained within the signal. The optimal approach for that depends on the properties of how both waveforms overlap. For now let us assume that the bit clock values of the both signals have been precisely derived and use the provided values from the initial signal modulation. Knowing the bit clocks, joint shape reconstruction can be carried out similarly to the approach used in ShBD as described in Chapter 4. However, there is a difference in the type of shapes searched. In the case of the ShBD there were used shapes which mimic the IQ data of the signal like ones in Fig. 44. Whereas in case of processing the extracted signal roots, which reflect the frequencies contained within the signal, a new type of shapes, here called Gaussian shapes or shortened as GShapes, should be introduced, incorporating the signal's frequency properties. These shapes are derived from Gaussian-filtered bits, shown in Fig. 43 a), prior to passing the filtered signal to integration and modulation for building the IQ data. The basic Gaussian signal shapes for Signal1 and Signal2 are shown below in figures Fig. 91 and Fig. 92 respectively.

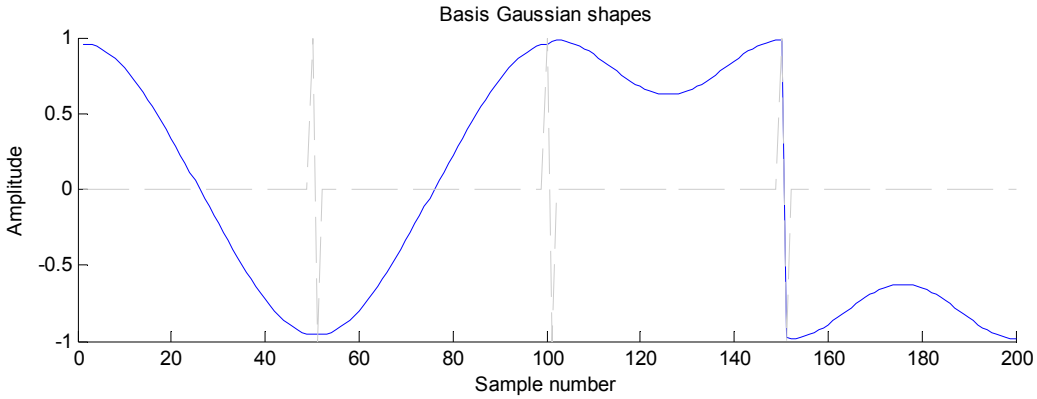


Fig. 91. Basis shapes of Gaussian signal

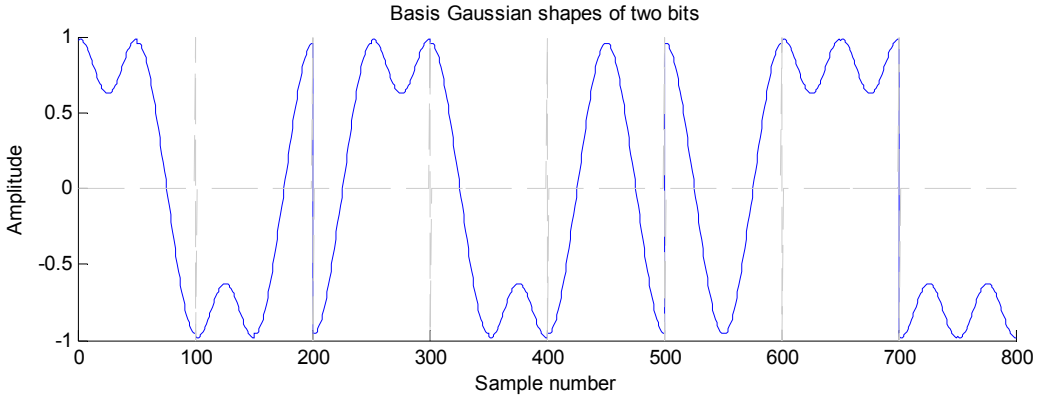


Fig. 92. Basis shapes of Gaussian signal for phase shifted second signal.

Similarly to the approach with the IQ shape construction for JShBD described in Section 4.2.2, the shape cuts for the Gaussian basic shapes are performed exactly at bit clock of Signal1 and named GShapes1. Also similarly to procedure outline for building Shapes2 in Section 4.2.3, the phase shifted version for shapes to approximate roots of Signal2 are constructed from valid GShapes1 pairs, and named GShapes2. Finally the variations of both basis waveforms are summed in each bit-step to perform joint signal approximation in sense of smallest L^2 difference between the GShapes1 and GShapes2 sum and the input roots signal. Since the reconstructed signal should be continuous, the Viterbi decoder as introduced and used previously in Section 4.1 can also be applied to correct discontinuity errors [39].

Let us here demonstrate a demodulation example. Let us consider the extracted root signal as shown in Fig. 90. Around the sample 900 there is a noticeable, sharply oscillating disturbance. When trying to recover the signal using the JGSUM algorithm, the misleading disturbance in the roots signal gives a rise to a misleading signal reconstruction with the signal shapes. Using the Viterbi algorithm this can be corrected as demonstrated in Fig. 93 to Fig. 95 below.

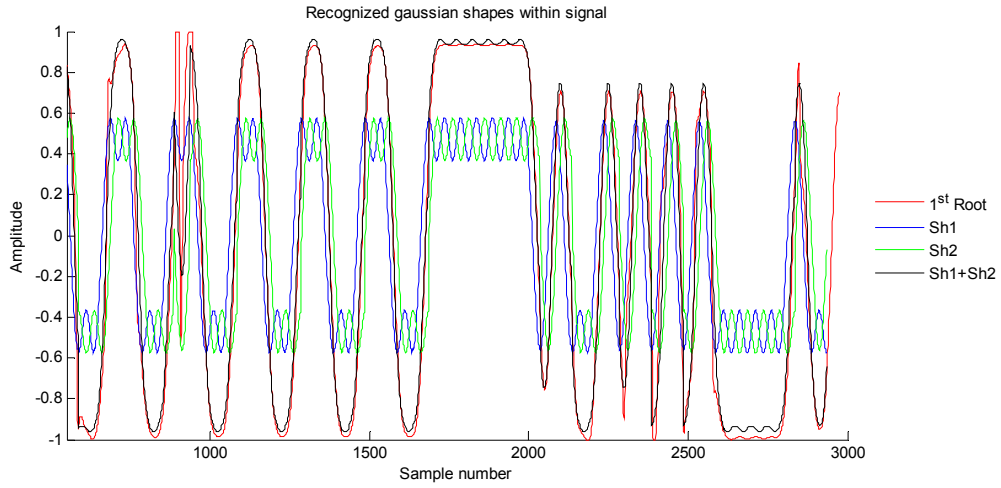


Fig. 93. JGSUM shape recognition applied to signal, without VD (Viterbi Decoder).

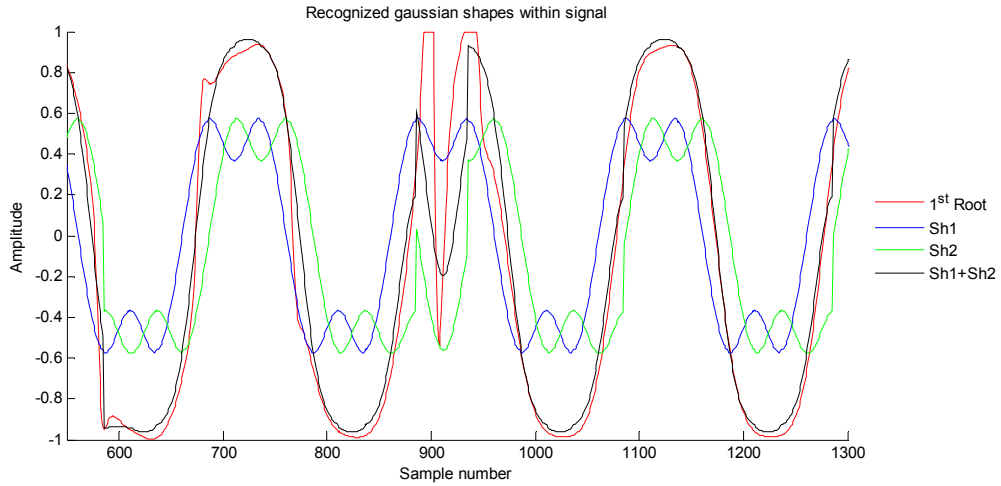


Fig. 94. JGSUM shape recognition applied to signal, without VD, zoomed around sample 900.

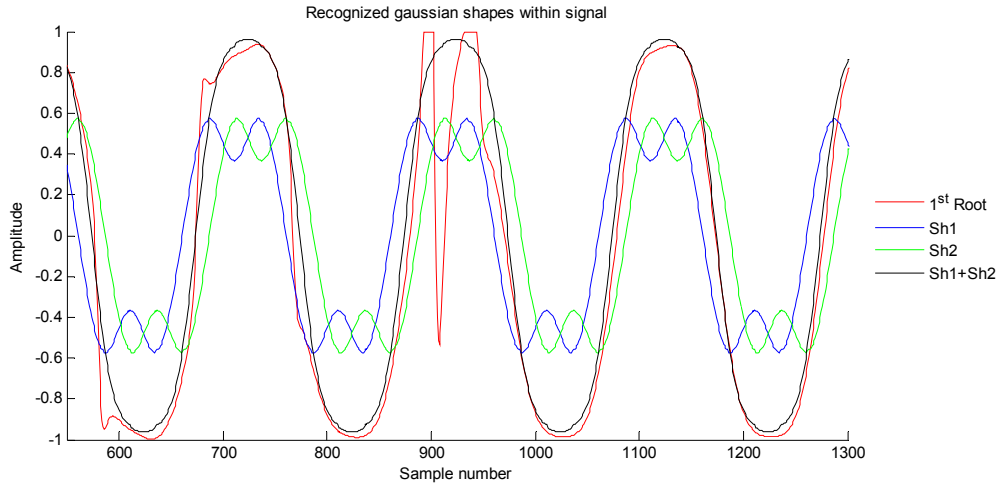


Fig. 95. JGSUM shape recognition applied to signal, with VD, zoomed around sample 900.

As it can be seen in Fig. 93, Fig. 94 and Fig. 95, the JGSUM algorithm can provide a good approximation of the two overlapping GMSK signal curves in a noiseless environment, while VD can

provide an important improvement for shape recognition, when discontinuities in extracted root signal are met.

The Viterbi algorithm can be used with different path lengths. Evaluating shapes by taking the L^2 norm distance over 2 previous bits and a single new one, gives a total shape length of 3 bits, which is an additional improvement compared to the trellis weight graph calculated from the shapes of a just single bit length. After finding shape weights, the VD was applied separately to construct the minimum weight path through trellis diagram for Signal1, repeating the same also for Signal2. Meaning that during the estimate of one of the signals, the shape values of the second signal were fixed. Joint Viterbi application could offer another step of further improvement.

For the Gaussian shape recognition algorithm application, there is a need to perform a priori estimate of the amplitudes of the two overlapping signals. The ratio 0.5:0.5 is a good selection when the overlapping signals are of equal or almost equal amplitudes. However, for the cases when the amplitude ratio of Signal1 to Signal2 is more than 0.6:0.4, the shape amplitudes ratio should also reflect that one signal being stronger than the other one. Using an empirical method there were found the best suiting shape amplitudes for different weights of the input overlapping signals. These are shown in the Matlab code below. The `S1_WEIGHT` and `S1_WEIGHT` stands for tested Signal1 and Signal2 weights respectively and the `SH_WEIGHT` reflects the found optimum shape weight ratio of GShapes1 to GShapes2 to reconstruct the roots signal. It can be read as follows: for the 0.6:0.4 input signal amplitude case, the shapes with weights 0.8:0.2, multiplied by 1.1 were chosen for simulation purposes.

```
S1_WEIGHT = [0.5, 0.5025, 0.5238, 0.5556, 0.6, 0.666, 0.8, 0.9];
S2_WEIGHT = 1 - S1_WEIGHT;
SH_WEIGHT =[[0.5, 0.5]*1.1;
            [0.6, 0.4]*1.1;
            [0.7, 0.3]*1.1;
            [0.8, 0.2]*1.1;
            [0.9, 0.1]*1.1];
map_SigW2ShW = [1, 1, 3, 3, 4, 4, 5, 5];
```

Both signals proved to be jointly recognizable from the roots signal produced by MBD method if their amplitudes were similar or slightly deviating, up to a difference of 0.525:0.475, meaning 0.87dB. For cases with more amplitude difference between the composed signals, only the strongest signal could be extracted, leaving the smallest signal extraction for consecutive signal cancellation algorithm implementation. Further analysis of the JGSUM algorithm is revealed in the next section.

5.5.3 Analysis of JGSUM demodulation capability

Let us consider having IQ data provided, with two overlapping signals contained. Let us further assume that the bit clocks of these are known. Let us analyse how the GSUM algorithm performs for different weight, Doppler shift, phase shift and noise ratios.

To generate IQ test data of overlapping signals for the analysis, two AIS messages of type 1 were chosen, each having a 256 bit length. The following settings were used for the simulations:

- $OSR = 50$, oversampling rate for bits,
- $N = 30$, number of samples used for FBLP,
- $L = 5$, number of roots searched,

- $M_{SVD} = 1$, number of roots selected.

5.5.3.1 Resistance to Doppler shifts

To analyse the Doppler shift impact on GSUM demodulation capability, one message signal with a static Doppler shift of 45Hz and a second signal with the Doppler shift varying in the range of 45Hz to 1600Hz was constructed. The analyses were performed for different weight ratios between the signals, with and without Viterbi coder. Phase shift for the case was assumed to be constant with ½ of bit length.

Table 9 below reports simulation results using the similar table structure as previously revealed in Section 4.5.2.3. In the left grey area there are indicated signal weights in different units: unity amplitude, amplitude difference in presents, signal to interference ratio and signal to interference ratio in dB. Within the grey area at the lower part of the table the variable parameter is identified, which in this analysis case is the ratio of Doppler frequencies of the first and the second signals along with the differences between them in Hz. The e1 and e2 columns are each for number of demodulation bit errors in Signal1 and Signal2 respectively. The vertical bar separates demodulation results with or without a Viterbi decoder in use. Thus, one can read the table below as follows: at $SIR = 0.83dB$ between the two overlapping signals, at 5Hz difference, with ½-bit phase shift, the first signal could be demodulated correctly whereas the second had 15 bit errors for approach with a Viterbi decoder and 20 bit errors otherwise. This cell is outlined with a thick orange line in Table 9. Fields where demodulation succeeded with zero errors, using or excluding Viterbi application, are marked in green.

The dB ratio of the signal to interferer was calculated using 0.7 as a constant for standard deviation value of the AIS signal at unity amplitude. For SIR and SIR_{dB} calculation formulas please refer to (25) and (26) respectively.

SIR_{dB}	SIR	Diff %	AMPL1	AMPL2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	
19.08	81.00	900%	0.900	0.100	0|0	84|101	0|0	90|101	0|0	106|105	0|0	83|84	0|0	107|110	0|0	102|100	0|0	109|107	
12.04	16.00	400%	0.800	0.200	0|0	65|66	0|0	68|72	0|0	80|80	0|0	63|69	0|0	76|80	0|0	82|82	0|0	96|86	
6.02	4.00	200%	0.667	0.333	0|0	59|64	0|0	54|55	0|0	53|54	0|0	48|47	0|0	63|62	0|0	68|65	0|0	68|68	
3.52	2.25	150%	0.600	0.400	0|0	53|56	0|0	42|43	0|0	63|56	0|0	37|38	0|0	48|50	0|0	49|54	1|1	71|72	
1.94	1.56	125%	0.556	0.444	0|0	41|42	0|0	31|31	0|0	39|35	0|1	27|28	0|0	31|37	0|0	40|43	3|2	43|43	
0.83	1.21	110%	0.524	0.476	0|0	0|0	0|0	15|20	0|0	9|12	0|0	8|14	0|0	16|19	0|0	20|21	2|2	9|13	
0.09	1.02	101%	0.502	0.498	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
0.00	1.00	100%	0.500	0.500	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
					45:45		45:50		45:100		45:200		45:400		45:800		45:1600		DF1:DF2
					0		5		55		155		355		755		1600-45		Δ Diff [Hz]

Table 9. Doppler shift analysis for two overlapping signals with fixed 45Hz Doppler shift for the first signal.

From the results in Table 9, there can be concluded two important GSUM demodulator properties in noiseless environment:

- For cases of signal amplitudes differing by less than 1% of the strongest signal amplitude, both signals can be always demodulated;
- The strongest signal can always be demodulated if the highest Doppler of both signals is lower or equal to 800Hz.

There was carried out also another simulation run with different Doppler frequencies, where the first signal had a static 400Hz Doppler shift, while the second signal had a Doppler varying from 450 to 2400Hz. These results of the demodulation capability analysis are shown in Table 10 below.

SIR_{dB}	SIR	Diff %	AMPL1	AMPL2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	
19.08	81.00	900%	0.900	0.100	0|0	84|85	0|0	107|108	0|0	108|106	0|0	101|105	0|0	95|98	0|0	94|93	0|0	104|110	
12.04	16.00	400%	0.800	0.200	0|0	61|62	0|0	79|83	0|0	87|87	0|0	79|80	0|0	74|77	0|0	92|84	0|0	93|94	
6.02	4.00	200%	0.667	0.333	0|0	51|53	0|0	59|64	0|0	67|70	0|0	58|62	0|0	60|60	0|0	79|82	0|0	71|64	
3.52	2.25	150%	0.600	0.400	0|0	41|44	0|0	44|45	0|0	49|53	0|0	46|48	0|0	44|45	0|0	73|73	0|2	58|55	
1.94	1.56	125%	0.556	0.444	0|0	31|32	0|0	34|40	0|0	39|40	0|0	36|36	2|2	32|37	3|12	70|96	6|21	56|86	
0.83	1.21	110%	0.524	0.476	0|0	15|19	0|0	19|24	0|0	21|24	1|1	15|17	1|1	14|15	5|19	56|90	11|23	57|82	
0.09	1.02	101%	0.502	0.498	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
0.00	1.00	100%	0.500	0.500	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	0|0	
					400:450		400:650		400:900		400:1150		400:1400		400:1900		400:2400		DF1:DF2
					50		250		500		750		1000		1500		2000		Δ Diff [Hz]

Table 10. Doppler shift analysis for two overlapping signals with fixed 400Hz Doppler shift for the first signal.

As it can be observed, at 400Hz static Doppler the two properties concluded from Table 9 also hold true.

5.5.3.2 Resistance to phase shift

Very similarly to the previous analytical approach with Doppler shift, the same amplitude variation ranges of Signal1 to Signal2 were used. The Doppler shifts of the first and the second signals were fixed at 400 and 650Hz while the overlapping phase of the second signal was varying. The phase variation was tested from ½ phase to zero phase overlap, referenced respectively as 50 to 0% in the bottom line of Table 11 below. Meaning, that the distance of the symbol clocks between the both signals range from maximum distinct to zero.

SIR_{dB}	SIR	Diff %	AMPL1	AMPL2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	e1	e2	
19.08	81.00	900%	0.900	0.100	0|0	107|108	0|0	108|103	0|0	86|97	0|0	70|83	0|0	65|69	0|0	66|61	
12.04	16.00	400%	0.800	0.200	0|0	79|83	0|0	83|82	0|0	79|83	0|0	70|78	0|0	56|64	0|0	47|53	
6.02	4.00	200%	0.667	0.333	0|0	59|64	0|0	57|65	0|0	53|63	0|0	46|59	0|0	36|50	0|0	32|42	
3.52	2.25	150%	0.600	0.400	0|0	44|45	0|0	44|47	0|0	41|46	0|0	32|42	0|0	26|38	0|0	20|35	
1.94	1.56	125%	0.556	0.444	0|0	34|40	0|0	33|36	0|0	24|34	0|1	18|26	1|1	9|24	0|1	7|25	
0.83	1.21	110%	0.524	0.476	0|0	19|24	0|0	17|19	0|2	5|17	0|17	4|23	1|14	1|21	2|8	16|21	
0.09	1.02	101%	0.502	0.498	0|0	0|0	0|0	0|0	0|1	0|1	0|3	0|4	3|15	2|15	52|72	52|72	
0.00	1.00	100%	0.500	0.500	0|0	0|0	0|0	0|0	0|0	0|0	0|2	0|2	4|10	4|10	52|55	52|55	
					50%		40%		30%		20%		10%		0%		Bit phase shift

Table 11. Phase shift analysis for two overlapping signals with 400 and 650Hz fixed Doppler shifts.

Since the time distance between the bit clocks of the two signals is symmetric, it is enough to test the phase shift from 50 to 0% and to reference the observed indicators also for the range of 50 to 100% phase shift.

From the results reported above, the following two additional important properties of the JGSUM demodulator can be concluded:

- Joint demodulation of both signals works only when the phases of both signals differ by at least 20% of their bit clock width.
- For cases with a phase shift below 20%, the strongest signal can be identified only if the amplitude of the weaker signal is below $\frac{2}{3}$ of the amplitude the strongest signal (the amplitude ratio 0.6:0.4).

From the first property above the conclusion can be derived, that for two similar amplitude signals overlapping the JGSUM algorithm can do the demodulation only if the phase shift of the overlapping bits ranges of 20% to 80%. Assuming a uniform phase shift distribution for the overlapping signals at satellite, this would mean that 60% of the equal amplitude overlapping messages can be demodulated.

5.5.3.3 Resistance to noise

The tests for the JShBD demodulation capability of noisy signals were constructed similarly to the two previous analysis approaches. Again the Doppler shift of the first and the second signals were fixed at 450 and 600Hz. Also the phase shift was fixed at ½ bit. A white Gaussian noise was generated with varied amplitudes within a range from $\sigma(N) = 0.005$ to $\sigma(N) = 0.5$. The results produced by this simulation are shown in Table 12. The strongest signal to noise ratio expressed in dB is provided in the yellow column, on the right of the error bit e1 and e2 columns. For the SNR and SNR_{dB} calculation formulas please refer to (27) and (28) respectively.

SIR_{dB}	SIR	% diff	AMPL1	AMPL2	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR
19.08	81.00	900%	0.900	0.100	0|0	105|107	42	0|0	103|100	28	0|0	100|99	22	0|0	88|85	16	0|0	113|111	9.99	0|0	104|100	6.47	0|0	106|105	3.97	4|6	110|119	2.03
12.04	16.00	400%	0.800	0.200	0|0	80|83	41	0|0	80|82	27	0|0	89|79	21	0|0	113|109	15	0|0	100|104	8.96	0|0	95|99	5.44	4|8	108|103	2.94	9|17	94|100	1.01
6.02	4.00	200%	0.667	0.333	0|0	59|65	39	0|0	60|64	25	0|0	63|67	19	0|0	88|83	13	0|1	88|81	7.38	8|10	84|79	3.86	16|24	93|91	1.36	29|44	112|101	-0.58
3.52	2.25	150%	0.600	0.400	0|0	44|44	39	0|0	44|46	25	0|0	43|47	19	0|0	53|48	12	5|10	65|70	6.47	15|19	72|74	2.94	25|40	96|101	0.44	49|59	107|122	-1.49
1.94	1.56	125%	0.556	0.444	0|0	34|40	38	0|0	34|38	24	0|0	32|38	18	4|4	34|41	12	17|26	76|88	5.80	24|34	77|99	2.28	42|58	105|134	-0.22	83|83	117|135	-2.16
0.83	1.21	110%	0.524	0.476	0|0	18|23	37	0|0	20|25	23	0|1	21|21	17	11|11	26|27	11	27|45	79|115	5.29	42|56	89|120	1.76	66|71	106|134	-0.73	87|87	112|130	-2.67
0.09	1.02	101%	0.502	0.498	0|0	0|1	37	0|3	3|3	23	1|10	5|12	17	14|18	16|19	11	29|38	34|35	4.92	87|94	93|65	1.40	97|88	97|86	-1.10	100|98	93|79	-3.03
0.00	1.00	100%	0.500	0.500	0|1	0|0	37	0|5	2|0	23	3|6	4|5	17	5|22	10|18	11	40|39	35|32	4.88	86|70	86|59	1.36	93|80	93|85	-1.14	95|93	93|93	-3.08
				σ(N)	0.005			0.025			0.050			0.100			0.200			0.300			0.400			0.500		

Table 12. Noise analysis for two overlapping signals with 400 and 650Hz fixed Doppler shifts.

The table above demonstrates two additional very important properties of the GSUM algorithm:

- Joint demodulation can succeed only when the overlapping signals have equal amplitude and are 23dB above the noise level or have 1% in amplitude difference and lay 37dB above the noise level.
- When the strongest signal is 19dB over its interferer, 4dB of SNR is enough for it to be demodulated.

The first of the above properties denotes very restricting consequences for the application of joint signal extraction with JGSUM algorithm to real world signals, since there is a low probability of observing two signals with amplitudes differing by less than 1%. Also the required noise threshold is very high. This shows the strong necessity of sophisticated noise filtering needed to clean up the input signal prior to using the GSUM algorithm.

However, on the other hand, from the second concluded property it follows that the method can extract the strongest signal alone when it is 4dB over the noise in absence of any interferers or interferers being below 19dB.

To demonstrate improved resistance against noise, a pre-filtering process could be applied to the signal. However, this is expected to disturb the recognition of joint signals and week signals. This can be observed in the corresponding simulation results reported in Table 13, with the same simulation carried out as before, but using pre-filtering of IQ data, smoothed with 10 sample averaging filter.

SIR_{dB}	SIR	% diff	AMPL1	AMPL2	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR
19.08	81.00	900%	0.900	0.100	0|0	97|85	42	0|0	96|86	28	0|0	95|88	22	0|0	100|90	16	0|0	105|98	9.99	0|0	103|99	6.47	0|0	102|102	3.97	0|0	100|96	2.03
12.04	16.00	400%	0.800	0.200	0|0	98|89	41	0|0	99|90	27	0|0	102|92	21	0|0	100|94	15	0|0	101|94	8.96	0|0	104|102	5.44	0|0	104|107	2.94	0|1	106|106	1.01
6.02	4.00	200%	0.667	0.333	0|0	69|66	39	0|0	69|69	25	0|0	68|68	19	0|0	72|67	13	0|0	72|71	7.38	1|0	85|84	3.86	1|0	92|101	1.36	8|10	95|104	-0.58
3.52	2.25	150%	0.600	0.400	0|0	62|61	39	0|0	61|61	25	0|0	61|62	19	0|0	65|64	12	1|1	63|63	6.47	4|3	80|83	2.94	10|9	81|90	0.44	19|19	80|91	-1.49
1.94	1.56	125%	0.556	0.444	0|0	32|44	38	0|0	34|42	24	0|0	32|42	18	0|0	29|37	12	5|3	33|39	5.80	15|16	71|83	2.28	27|30	82|93	-0.22	29|32	86|94	-2.16
0.83	1.21	110%	0.524	0.476	0|0	20|24	37	0|0	18|22	23	1|2	21|29	17	4|6	21|37	11	16|17	48|55	5.29	22|23	64|78	1.76	37|35	84|100	-0.73	41|53	79|116	-2.67
0.09	1.02	101%	0.502	0.498	2|3	4|8	37	3|6	5|13	23	6|11	9|14	17	4|11	11|16	11	25|39	29|31	4.92	40|58	40|38	1.40	67|81	71|54	-1.10	61|66	68|56	-3.03
0.00	1.00	100%	0.500	0.500	3|4	2|5	37	4|5	7|8	23	5|12	5|15	17	13|21	14|18	11	33|40	33|30	4.88	48|49	46|38	1.36	63|70	67|55	-1.14	70|76	72|59	-3.08
				σ(N)	0.005			0.025			0.050			0.100			0.200			0.300			0.400			0.500		

Table 13. Noise analysis for two overlapping signals with 400 and 650Hz fixed Doppler shifts. An averaging filter was applied to the IQ data prior to extraction of frequency roots.

In the table above it can be observed that the error number has increased in cases where both signal amplitudes were equal or almost equal, as compared to results reported in Table 12 without the filtering option used. On the other hand, the positive side is the capability to recognize the strongest signal within a lower SNR environment, in this case down to SNR=1.36dB for SIR>6dB.

5.5.3.4 Strongest-signal and equal-signal modes

As previously mentioned, an initial estimate of Gaussian shape amplitudes is needed to perform the shape search within the roots signal. This could be provided either by some algorithm or by testing few appropriate sets of amplitudes ratios, as shown in Matlab code at the end of Section 5.5.2. A good choice showed to be a test of the four following Signal1 to Signal2 amplitude ratios: 0.5:0.5, 0.7:0.3, 0.8:0.2 and 0.9:0.1. The amplitude ratio 0.6:0.4 is not present due to empirical tests showing that the signal tends to be better demodulated using shape pairs of either strongly similar or strongly differing amplitudes.

Therefore to facilitate the usage of JGSUM demodulation method there were derived two extreme JGSUM demodulation application cases: the strongest-signal mode and equal-signal mode.

The strongest-signal mode of JGSUM algorithm is the approach of using MBD with the single root signal extraction and performing a strongest signal search with the shape amplitudes ratio set to 0.9:0.1. If this is the case, the JGSUM method will find a single signal if it is strong enough over noise level and or interferers. The results of this simulation approach for various noise levels are reported in Table 14 below.

SIR_{dB}	SIR (dB)	% diff	AMPL1	AMPL2	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR
19.08	81.00	900%	0.900	0.100	0|0	99|88	42	0|0	97|86	28	0|0	98|92	22	0|0	102|96	16	0|0	106|108	9.99	0|0	100|103	6.47	0|0	103|99	3.97	6|2	106|112	2.03
12.04	16.00	400%	0.800	0.200	0|0	100|92	41	0|0	101|92	27	0|0	97|91	21	0|0	101|96	15	0|0	103|102	8.96	0|0	106|114	5.44	0|3	107|107	2.94	11|15	107|107	1.01
6.02	4.00	200%	0.667	0.333	0|0	94|91	39	0|0	93|90	25	0|0	92|91	19	0|0	94|91	13	0|0	92|92	7.38	5|6	97|101	3.86	25|22	100|100	1.36	22|31	102|104	-0.58
3.52	2.25	150%	0.600	0.400	0|0	76|77	39	0|0	79|80	25	0|0	77|75	19	0|0	73|76	12	5|5	76|90	6.47	10|8	83|106	2.94	30|33	98|106	0.44	38|38	102|103	-1.49
1.94	1.56	125%	0.556	0.444	0|0	58|88	38	0|0	60|88	24	0|0	64|96	18	1|0	66|92	12	17|18	91|108	5.80	35|30	86|99	2.28	56|45	108|105	-0.22	46|58	98|105	-2.16
0.83	1.21	110%	0.524	0.476	4|0	37|79	37	8|0	47|86	23	9|0	55|95	17	13|10	64|103	11	29|31	100|110	5.29	45|42	88|96	1.76	39|46	106|107	-0.73	71|70	97|105	-2.67
0.09	1.02	101%	0.502	0.498	24|0	43|83	37	24|6	42|91	23	31|7	76|122	17	26|13	71|103	11	43|26	93|102	4.92	51|55	98|94	1.40	55|54	90|95	-1.10	71|69	108|103	-3.03
0.00	1.00	100%	0.500	0.500	27|6	40|84	37	28|6	45|84	23	26|8	74|123	17	35|21	70|94	11	46|40	90|102	4.88	60|49	89|104	1.36	57|51	106|94	-1.14	77|84	93|94	-3.08
				σ(N)	0.005			0.025			0.050			0.100			0.200			0.300			0.400			0.500		

Table 14. Noise analysis for two overlapping signals with 400 and 650Hz fixed Doppler shifts. The JGSUM algorithm is applied with strongest-signal mode.

As expected, the results of the top three rows in Table 14 are very similar to the ones of general noise tests in Table 13, since the amplitude set 0.9:0.1 was already used there. However, in the lower rows the demodulation results are more disturbed, since Table 13 results were calculated using attuned shape amplitudes.

The equal-signal mode, which is the opposite of the strongest-signal mode, is the approach of GSUM algorithm application searched shape amplitude ratio of 0.5:0.5 for equal amplitude signal demodulation. Table 15 summarizes the demodulation results of such an approach.

SIR_{dB}	SIR (dB)	% diff	AMPL1	AMPL2	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR	e1	e2	SNR
19.08	81.00	900%	0.900	0.100	0|0	105|102	42	0|0	105|104	28	0|0	105|102	22	0|0	109|98	16	0|0	101|101	9.99	3|2	112|115	6.47	9|11	105|107	3.97	73|92	106|101	2.03
12.04	16.00	400%	0.800	0.200	0|0	108|98	41	0|0	109|98	27	0|0	108|98	21	0|0	101|96	15	0|0	94|99	8.96	6|7	94|95	5.44	15|19	94|101	2.94	80|91	91|90	1.01
6.02	4.00	200%	0.667	0.333	5|6	75|75	39	4|4	77|78	25	5|4	77|75	19	4|5	67|72	13	6|4	69|73	7.38	29|31	91|94	3.86	57|58	85|79	1.36	93|114	101|97	-0.58
3.52	2.25	150%	0.600	0.400	4|9	55|60	39	4|8	56|62	25	4|9	56|60	19	4|5	58|56	12	10|15	59|61	6.47	66|61	112|96	2.94	67|69	93|81	0.44	84|89	103|98	-1.49
1.94	1.56	125%	0.556	0.444	1|4	38|43	38	1|7	40|47	24	1|6	40|45	18	0|13	32|48	12	70|68	96|77	5.80	76|75	95|85	2.28	94|95	96|84	-0.22	94|98	78|87	-2.16
0.83	1.21	110%	0.524	0.476	0|7	23|32	37	0|6	18|30	23	0|9	22|34	17	26|27	34|35	11	77|66	88|73	5.29	111|102	102|79	1.76	91|86	93|84	-0.73	102|95	105|96	-2.67
0.09	1.02	101%	0.502	0.498	0|0	0|1	37	0|1	3|2	23	2|4	3|6	17	52|51	54|19	11	87|91	89|56	4.92	107|97	94|68	1.40	92|89	95|79	-1.10	93|98	108|98	-3.03
0.00	1.00	100%	0.500	0.500	0|0	0|0	37	1|5	3|4	23	4|9	3|5	17	21|27	31|21	11	94|86	91|52	4.88	105|102	97|76	1.36	98|94	96|86	-1.14	90|97	90|83	-3.08
				σ(N)	0.005			0.025			0.050			0.100			0.200			0.300			0.400			0.500		

Table 15. Noise analysis for two overlapping signals with 400 and 650Hz fixed Doppler shifts. The JGSUM algorithm is applied with equal-signal mode.

The two tables Table 14 and Table 15 above lead to concluding the following additional property of the JGSUM algorithm:

- The strongest-signal mode with the search shape amplitude ratio 0.9:0.1 and equal-signal mode with the search shape amplitude ratio 0.5:0.5 are the two sufficient modes to be evaluated. The both modes together cover the same demodulation range of the more complex approach with attuned shape amplitudes.

This concludes the analysis of the JGSUM algorithm with the MBD application for overlapped AIS signal demodulation using frequency analysis.

The next chapter will summarize the results derived in this and previous chapters to design an innovative algorithm for use in SAT-AIS applications.

6 Proposed approach for demodulation of SAT-AIS signals

6.1 Summarizing and linking demodulation capabilities to message overlapping scenarios

In Chapter 3 we reviewed different packet collision scenarios. In Chapter 4 and 5 we observed approaches for extracting overlapped signals along with simulations to estimate the properties of the demodulation methods.

The new idea, derived by author of this paper from investigations made and knowledge gained during the research period of this topic, is to focus the joint signal demodulation efforts on the CC1 and CC2 overlapping scenarios and approach the cases as follows:

1) To solve CC1, use JGSUM algorithm for joint frequency extraction of overlapping signals with very close amplitudes.
2) To solve CC2, use MBD to find the 4 characteristic parameters for each of the two overlapping signals by demodulating the message parts which are interference free. Then apply the JShBDD to demodulate the joint parts of the signals.

The overall principle of AIS signal demodulation, incorporating the new developed algorithms, is proposed by author of this work as follows. It is also schematically presented in the diagram in Fig. 96 below.

1) (A0:Single) Analyse the input signal, check if it contains only single transmission by trying to demodulate it. If this is the case (a signal is present, it is at least 4dB over noise and it looks to be composed of single transmission), proceed with single signal demodulation.
2) (A1:Strongest) Signal contains multiple users, check if one of them is at least 4dB stronger than the rest. If this is the case (the strongest signal can be independently demodulated alone), proceed with the single strongest signal demodulation.
    i. In case the HDLC flags seem to be buried within overlaps, check if the signal has enough overlap free zone to contain the 168 message data and 16 CRC bits. If it does, use MBD for frequency extraction and check data bits with a sliding check sum calculator (CC3).
3) (A2:CC2) Signal contains multiple users and the strongest is not significantly stronger than the rest and there are significant parts of messages overlapped. Check if the overlap is just partial with an overlap-free zone of 5 bits at least. If this is the case (there are parts of the two signals which overlap and at least 5 bits of overlap free zone), estimate the 4 characteristic parameters of each signal in its overlap-free zones and perform joint demodulation with JShBDD.
4) (A3:CC1) Signal contains two users, none is significantly stronger, and there are no single parts available. Check if the signal is well above noise (SNR=18dB). If it is, perform demodulation using JGSUM algorithm.
5) (A4:CSC loop) If the extraction of one or multiple AIS messages were successful, modulate the messages back, subtract from the initial signal and repeatedly apply the whole algorithm to the rest of the signal again.

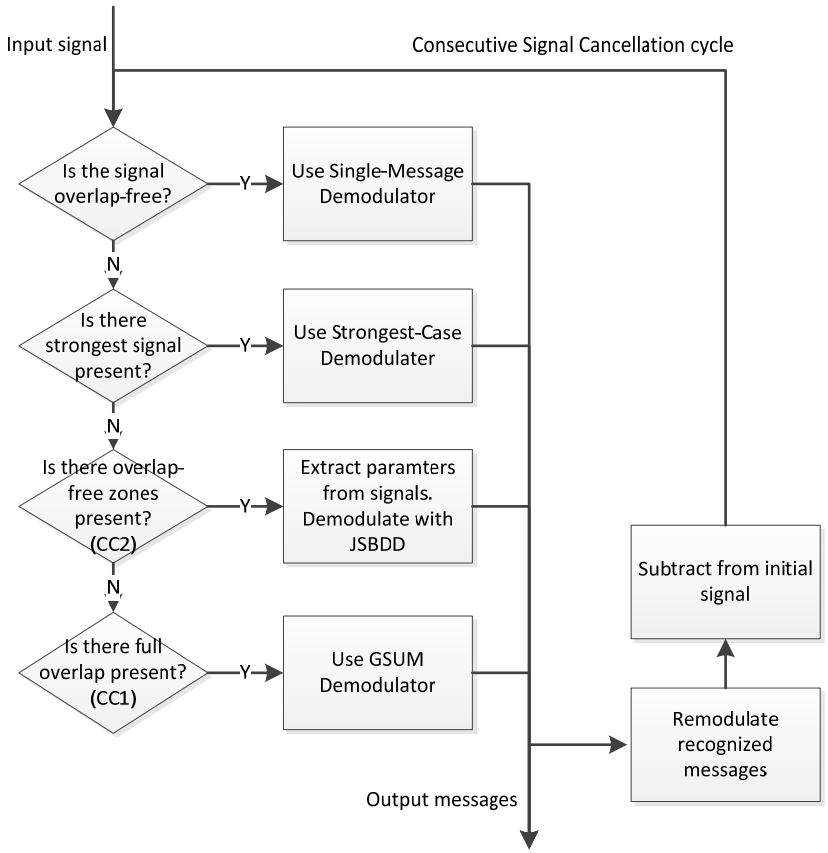


Fig. 96. The algorithm for signal demodulation, sequentially findng the most appropriate demodulation method.

6.2 Testing of algorithms

A demodulation simulator was developed, which processes a list of IQ signals built from generated AIS messages, considering every message and its overlap properties and which assigns an appropriate demodulation approach, if available. Thus every message passes through the decision tree revealed above in Fig. 96. As soon as message is identified to be recognizable by one of the demodulation algorithms in use, it is marked with the demodulation algorithm's number and directed to output.

The following table lists all of the parameters used by the demodulation simulator to decide about message extraction ability. The values for the parameters noted below were chosen based on simulation results described in previous chapters or publications referenced.

Parameter	Explanation
`noise_level = -143; %[dB]`	Noise level of receiver, in [dB]
`SC1_SNR =4;  %[dB]`	Strongest Case 1 (signal is overlap-free within the entire time slot). SNR - if the signal is more than 4dB above noise level and alone,

	it is considered to be demodulated.
SC2_SNRS=6; %[dB] SC2_SIR =4; %[dB]	Strongest Case 2 (signal is the strongest within the considered time slot). SNRS – minimal signal to noise ratio for the strongest signal. SIR – minimal signal to interference ratio for the strongest to weaker signal. If the strongest signal is more than 6dB above the noise level and it is at least 4dB above other interfering signals, it can be demodulated. In CSC loop it will also mean that another signal can be demodulated only if it has at least 4dB above and 4dB below free of interferers and it is 6dB over noise. Additionally, for the CSC to be applicable, the overlapping signals which are stronger than the one under consideration, should be successfully demodulated.
CC2_SNR =4; %[dB] CC2_SNRP=3; %[dB] CC2_SIR =0; %[dB] CC2_SIRI=4; %[dB] CC2_SIRJ=19; %[dB] CC2_syncb=5; %[bits]	Collision Case 2 (two or more signals are overlapping, signal parameters can be identified). The meanings of the parameters are following. SNR – the minimal boundary of signal to noise ratio in dB necessary for it to be identified for processing. SNRP – needed dB above noise for the four signal parameters (Doppler frequency, Doppler phase, bit clock, amplitude) to be estimated. SIR – minimum gap in dB for signals to be capable of joint demodulation. Here “0” means that signals can be of the same amplitude. SIRI – up to how many dB the difference between signal amplitudes account to be disturbing, requiring joint demodulation. SIRJ – maximum power level distance of two signals for them to be demodulated jointly. Syncb – minimum number of bits needed to perform the four signal parameter estimation.
CC1_SNRS=17; %[dB] CC1_SIRJ=0.1;%[dB] CC1_phase=0.2;%[bits]	Collision Case 1 (two signals are very close to each other in amplitude and they overlap completely, with no overlap-free zones). SNRS – the signal to the noise ratio of the strongest signal for the demodulation to take place. SIRJ – the signal to interferer ratio, below which the joint demodulation should take place (JGSUM). Otherwise the strongest signal is extracted first and then the interferer using JGUSM in strongest signal mode and applying CSC afterwards. Phase – minimum clock1 to clock2 distance in bits for the signals to be JGSUM jointly processed.

Table 16. Description of parameters used for demodulation analysis.

6.3 Simulation of satellite pass over ships

To test the contribution of each of the demodulation algorithm approaches, a realistic scenario was constructed for simulating satellite pass over an area of ships, which were modelled to transmit position reports as per ITU standard [1]. The simulation was built with the following parameters.

- A testing area for cruising ships was selected within a geographical rectangle of [-30..0] degrees longitude and [-30..0] degrees latitude. The area was filled with randomly distributed simulated ships with random voyage directions and speeds of applicable range. The chosen area is big enough to contain the observation area of a 1 minute satellite pass with a wide angle antenna. The total area of the chosen selected test field covers 10'626'000 km^2.
- Three different ship density scenarios were prepared over the area:
    - With 5'313 ships, meaning 0.5 ships per 1000 km^2.
    - With 10'626 ships, meaning 1 ship per 1000 km^2.
    - With 21'252 ships, meaning 2 ships per 1000 km^2.
- Satellite orbit was chosen to be at 700km altitude, similar to the common choice for LEO satellites carrying AIS payload. Two different antenna configurations were prepared for simulation:
    - Wide angle antenna with a 120° opening angle. From the particular satellite orbit configuration this antenna has footprint diameter of 3'108 km and covers approximately 7'588'000 km^2 area.
    - Narrow angle antenna, with a 38° opening angle. For the chosen satellite orbit configuration, this antenna has footprint diameter of 485 km and covers approximately 185'000 km^2 area.
- The pass of the satellite over the ship region was simulated for 1 minute, with the satellite flying over the middle of the ship covered area, as shown in Fig. 97 below. Thus, constructing circumstances which represent the signal reception on satellite when observing an area of uniformly distributed ships over the chosen time period.
- Reporting interval of all ships was set to 10 seconds. The selection of the AIS message reporting slots was subject to principles of SOTDMA cells operation, as described in ITU [1].

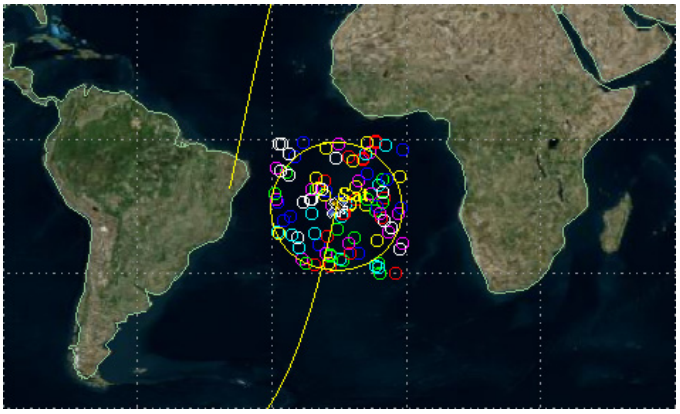


Fig. 97. The FOV of a satellite at 700km altitude with an 120° opening angle antenna passing over an area with randomly distributed ships.

The above outlined parameters were used as input to the Ship Generator application described earlier in Section 3.2 and shown in Fig. 29, to generated required satellite passage scenarios over the

ship region. The scenarios were passed to Signal Modulator for building corresponding IQ waveforms, determining which signals overlap and with which amplitude, time delay and Doppler shift properties.

6.4 Analysis of the algorithm performances

The results of the simulation strategies described above are shown below in pie charts for narrow and wide angle antennas in Fig. 98 and Fig. 99 respectively. The Table 17 and Table 18 report the simulation results in a table format respectively.

For the both figures there are two columns. The left one represents detection performance of single messages while the one on the right does the same for ships. Since it is enough to demodulate a single message from a ship per satellite pass to identify the ship as observed, the charts with ship detection rates identify higher detection performances.

In each column there are five rows which correspond to ship densities of 0.125, 0.25, 0.5, 1.0 and 2.0 ships per 1000 km^2 within the observation area.

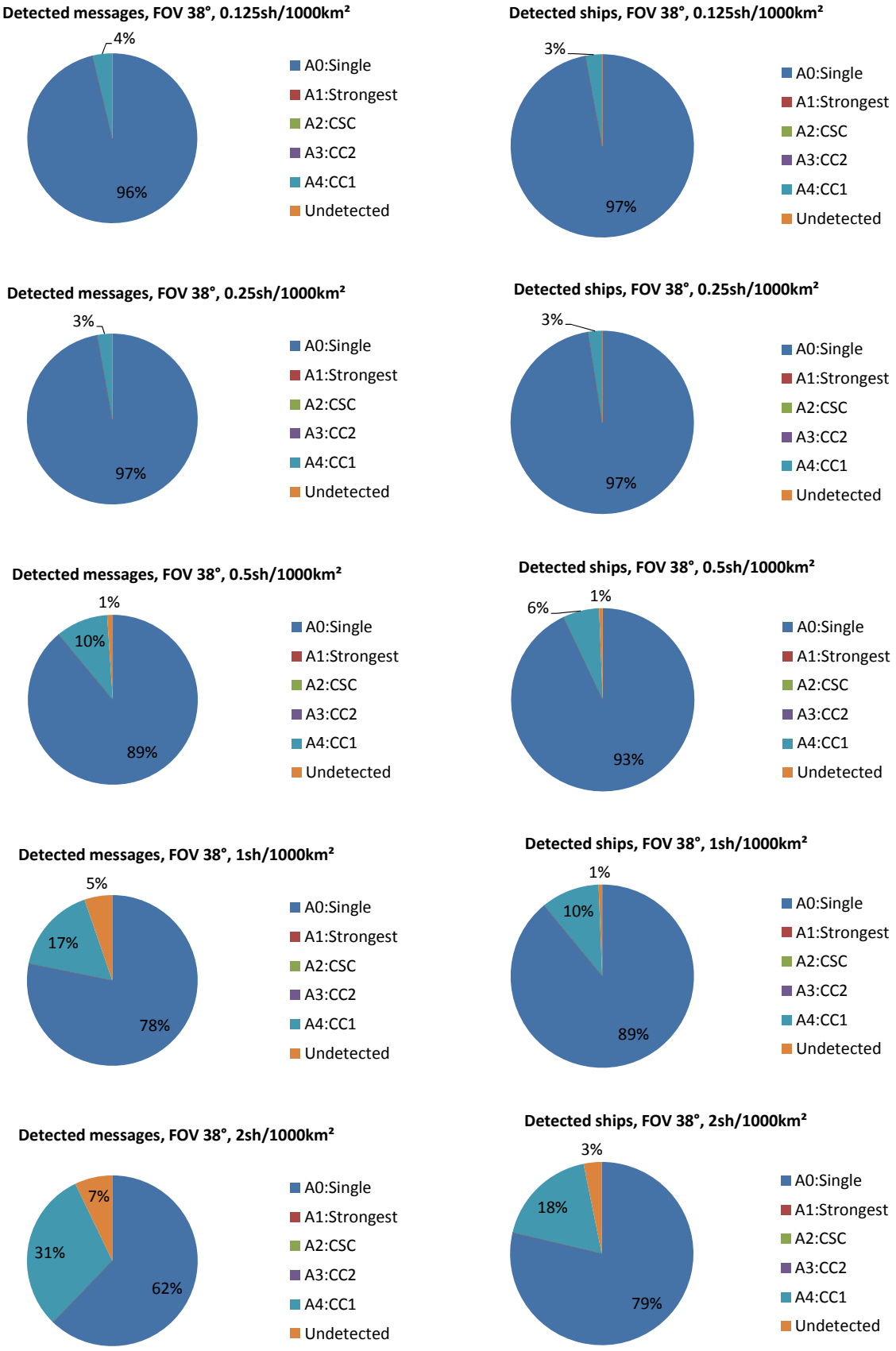


Fig. 98. Simulation results of AIS signal demodulation for narrow FOV antenna with 38° opening angle.

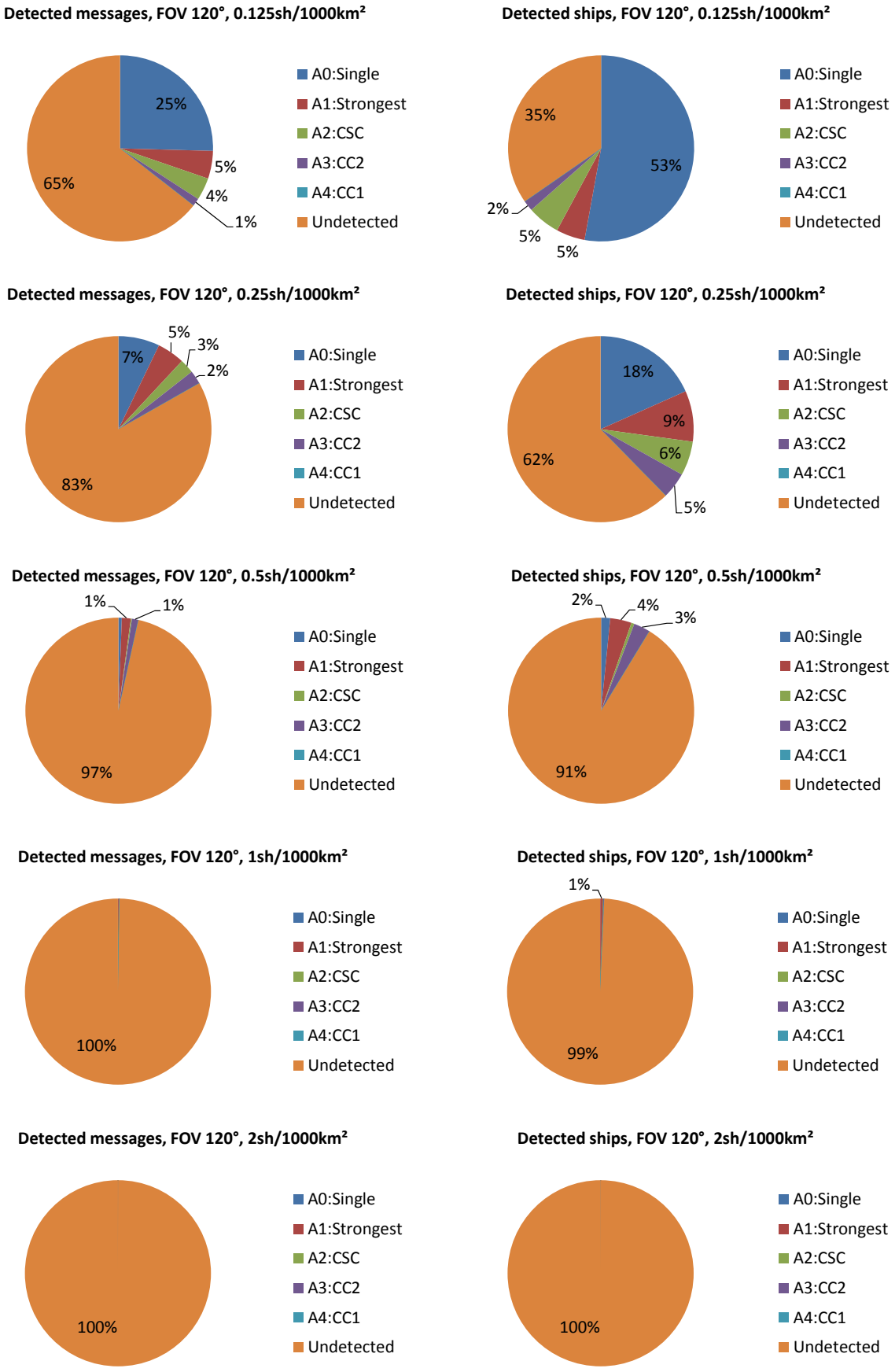


Fig. 99. Simulation results of AIS signal demodulation for wide FOV antenna with 120° opening angle.

	A0:Single	A1:Strongest	A2:CSC	A3:CC2	A4:CC1	Undetected	FOV	Given	A:SUM	A:SUM %
Detected messages, FOV 38°, 0.125sh/1000km²	51	0	0	0	2	0	53	3983	53	100%
Detected ships, FOV 38°, 0.125sh/1000km²	33	0	0	0	1	0	34	1328	34	100%
Detected messages, FOV 38°, 0.25sh/1000km²	137	0	0	0	4	0	141	7965	141	100%
Detected ships, FOV 38°, 0.25sh/1000km²	78	0	0	0	2	0	80	2656	80	100%
Detected messages, FOV 38°, 0.5sh/1000km²	250	0	0	0	28	3	281	15939	278	99%
Detected ships, FOV 38°, 0.5sh/1000km²	144	0	0	0	10	1	155	5313	154	99%
Detected messages, FOV 38°, 1sh/1000km²	444	0	0	0	94	30	568	31878	538	95%
Detected ships, FOV 38°, 1sh/1000km²	276	0	0	0	32	2	310	10626	308	99%
Detected messages, FOV 38°, 2sh/1000km²	689	0	0	0	340	79	1108	63756	1029	93%
Detected ships, FOV 38°, 2sh/1000km²	495	0	0	0	114	20	629	21252	609	97%

Table 17. Simulation results of AIS signal demodulation for narrow FOV antenna with 38° opening angle. Source data for Fig. 98.

	A0:Single	A1:Strongest	A2:CSC	A3:CC2	A4:CC1	Undetected	FOV	Given	A:SUM	A:SUM %
Detected messages, FOV 120°, 0.125sh/1000km²	739	140	113	40	0	1872	2904	3983	1032	36%
Detected ships, FOV 120°, 0.125sh/1000km²	572	54	61	20	0	375	1082	1328	707	65%
Detected messages, FOV 120°, 0.25sh/1000km²	409	273	142	134	0	4754	5712	7967	958	17%
Detected ships, FOV 120°, 0.25sh/1000km²	386	186	125	96	0	1312	2105	2656	793	38%
Detected messages, FOV 120°, 0.5sh/1000km²	67	177	23	129	0	11106	11502	15939	396	3%
Detected ships, FOV 120°, 0.5sh/1000km²	67	156	23	121	0	3856	4223	5313	367	9%
Detected messages, FOV 120°, 1sh/1000km²	2	45	0	16	0	23051	23114	31878	63	0%
Detected ships, FOV 120°, 1sh/1000km²	2	45	0	14	0	8423	8484	10626	61	1%
Detected messages, FOV 120°, 2sh/1000km²	1	3	0	0	0	46088	46092	63756	4	0%
Detected ships, FOV 120°, 2sh/1000km²	1	3	0	0	0	16950	16954	21252	4	0%

Table 18. Simulation results of AIS signal demodulation for wide FOV antenna with 120° opening angle. Source data for Fig. 99.

The two tables above report number of detected messages or detected ships for each of the demodulation algorithms introduced in Section 6.1 with the following meaning of its columns:

- *A0:Single* – the cases in which a single message signal was present, strong enough over the noise level to be demodulated (corresponding to configuration parameters for SC1 case within Table 16).
- *A1:Strongest* – the cases in which the strongest signal could be extracted (parameters for SC2)
- *A2:CSC* – the cases of underlying signal extraction, which could be identified after successful detection of the strongest signals and its subtractions using the consecutive signal cancellation (CSC) cycle (parameters for SC2).
- *A3:CC2* – the cases of partially overlapped signals extraction, when the overlapping signals possess the chance of individual signal parameters' estimation and JShBDD application (parameters for CC2).
- *A4:CC1* – the cases of fully overlapped signal extraction, when two overlapping signals have very alike amplitudes and are well over noise level for JGSUM application (parameters for CC1).

- *Undetected* – number of undetected messages or ships. It is calculated subtracting the *A:SUM* from *Given*.
- *FOV* denote number of messages or ships observed within the field of view of the satellite pass.
- *Given* denote number of messages or ships located within the whole rectangle area of simulation.
- *A:SUM* denote number of messages or ships processed by algorithms shown in the first four columns.
- *A:SUM%* denote the same what A:SUM and is expressed in percent from the total observed, shown in Given.

The next two sections present a deeper analysis into the results of demodulation characteristics reported in Fig. 98 and Fig. 99.

6.4.1 Case with 38° opening angle

Let us consider the simulation results in more detail and start with looking at the first column of the pie charts in Fig. 98, which represent message reception rates with narrow angle antenna. Since the blue fill of the pie charts in the figure represents a single message demodulation case A0-SC1, this tells us that for the density of 0.125 ships per 1000km^2, which is presented in the first pie diagram, 96% of the AIS message slots are found to contain with just a single transmission. Also there are about 4% of the messages being demodulated by an A4:CC1 – telling us that there are present overlaps of two messages with very similar amplitudes. Since the gain of the narrow angle antenna is relatively high with a setting of 12dB, which corresponds to the narrow angle antenna for AISat1, it helps for the colliding messages to overcome the high SNR requirement, set with a CC1_SIRJ parameter.

As the ship density increases, shown from top to bottom in Fig. 98, the number of single-slot messages decreases, down to 62%. Meanwhile, the rate of overlapping messages increases, leading to more CC1 algorithm demodulation cases, as much as 34%. Also, the rate of undetected messages shows an increase of up to 7%. The undetected messages correspond to the cases, when three or more overlaps are present, since the A4:CC1 algorithm was set to process a maximum of 2 signal overlaps.

The demodulation with A3:CC2 algorithm was not observed here, since a significant time shift among messages is needed for CC2 packet collision scenario to be observed. For the particular case with the antenna of 38° opening angle, the maximum bit delay from messages within FOV is 1.45 bits, which is below the 5-bit threshold, as set in CC2 parameters.

Also the demodulation with neither the A2:CSC, neither the A1:Strongest was observed, since these two algorithms process only the cases when overlapping messages have significant received power differences. The applied antenna had a narrow angle opening, therefore the difference in attenuation of the received signals from the ship transmitting through shortest path from nadir direction with 700km distance and the outer most ship of satellite's FOV transmitting at 19° angle from nadir with 745km signal travel distance, is 0.45dB. This is much too less than the required parameters of SC2_SIR and SC1_SNR for the algorithms A2:CSC and A1:Strongest correspondingly.

The three pie charts in the second column of Fig. 98 represent the same scenario as the first column, except that analysis is made for the rates of the received ships instead of messages. For a ship to count as received, it is enough that one message from it gets detected within the demodulation process during a satellite pass. Here the ships were assigned a demodulation algorithm corresponding to the order of the processing algorithms. Meaning, if the particular ship was received once by A0:Single and 3 times by A4:CC1, it will be counted as detected with A0:Single algorithm. The rate of the not received ships is expected to be lower than the rate of the undetected messages. Since a single ship is transmitting 6 times within the 60 second simulation time interval, which means that a missed information sent from a particular ship would likely get received by an another message at another time slot, while the ship is being observed within the satellite's FOV.

Similar to the analysis of the rate of received messages, as the density of ships in the observation area increases, so does the number of undetected ships.

6.4.2 Case with 120° opening angle

Let us next consider the simulations results reported in Fig. 99. These represent the 120° antenna opening case. Observing detected messages at a ship density of 0.25ships/1000km^2, the following can be noted: the algorithm A0:Single processes 7%, the A1:Strongest – 5%, the A2:CSC – 3%, the A3:CC2 – 2% of the total messages observed. These combined represent 17% successfully processed messages from the total observed ones.

Compared to the case of narrow angle antenna, here it can be noted successful application of the A1:Strongest, A2:CSC and A3:CC2. One would expect this, due to the fact that with the wide antenna opening angle some of the overlapping messages have enough difference in arrived signal amplitudes to be processed by A1:Strongest and A2:CSC and also enough time shift to identify individual characteristics of overlapping signals to process them with A3:CC2.

On the other hand, compared to the case of narrow angle antenna, here is no A4:CC1 demodulation observed. This is due to the fact that the wide angle antenna does not provide the necessary reception gain for the processing, by not meeting the parameter CC1_SNRS property of the signal strength.

In the 0.5ships/1000m^2 ship density conditions the A3-CC2 algorithm indicates the best performance for the weaker interfering messages extraction, although it makes only 1% of overall observed messages. There are only 3% of all received messages being recognized by all the algorithms together.

As the ship density increases, a strong reduction in the number of successfully processed messages can be observed. Increasing the ship density to 1ship/1000km^2, the CC2 algorithm gives 25% of all the demodulated messages, but corresponds to only about 0.1% of the total number of messages observed by the satellite's antenna. A further increase of ship density to 2ships/1000km^2 leads to a situation where none of the algorithms under consideration is capable of demodulating significant amounts of messages.

Observing the right column of the pie charts in Fig. 99, which represent the rate of received ships for the 120° opening satellite antenna, a very similar situation unveils. Although the percentage of

detected ships is higher than the message detection rate, the rate of detected ships significantly diminishes down to 0% for the case of the 2ships/1000km^2 ship density.

6.4.3 Conclusions from the 38° and 120° cases comparison

The following differences are observed while comparing the narrow and wide angle message detection simulation scenarios:

- The A3:CC2 demodulation in the narrow angle antenna case was not observed, while in the wide angle case the major contribution to demodulation capability was demonstrated, up to 33% of the total detected messages.
- The A4:CC1 demodulation gave a significant message detection improvement for the narrow angle antenna case up to 33% of the total detected messages, but was not applicable for the wide angle reception scenario.
- The A2:CSC demodulation showed its positive performance in the wide angle application, where messages overlap with more diverse signal amplitudes than in narrow angle case, giving up to 15% of the total detected messages.

6.4.4 Productivity evaluation for the demodulation algorithm groups

The classification of the message demodulation algorithms used in Fig. 98, Fig. 99, Table 17 and Table 18 can also be organised into three following groups, characterising AIS receiver readiness for space applications, used to represent results in Fig. 100 and Table 19 below.

- Group1: Basic AIS receivers, incorporating demodulation capabilities for A0 and A1 cases.
- Group2: Advanced AIS receivers with CSC capability, offering demodulation capability for A2 case.
- Group3: Sophisticated AIS receivers for processing signals of similar magnitude, offering demodulation capability for A3 and A4 cases. These require a selection of specific algorithms and approaches, formulated by author of this document in the previous two chapters.

The AIS message receiving equipment, which by its processing algorithms is similar to AIS receiving hardware on ships, could be assigned to belong to Group1. The Space-dedicated AIS receivers incorporating CSC capabilities could be assigned to belong to Group2. Finally, the Space-dedicated AIS receivers with a capability to process similar amplitude overlapping signals using various algorithms, could be assigned to belong to Group3.

In contrast to the basic terrestrial AIS receivers, which belong to Group1, the Space-dedicated AIS receivers belonging to Group2 and/or Group3 can offer significant performance improvements in space applications.

The following figure and table summarizes the successfully processed messages per AIS demodulation algorithm group in percent, with 100% representing the total successfully processed messages.

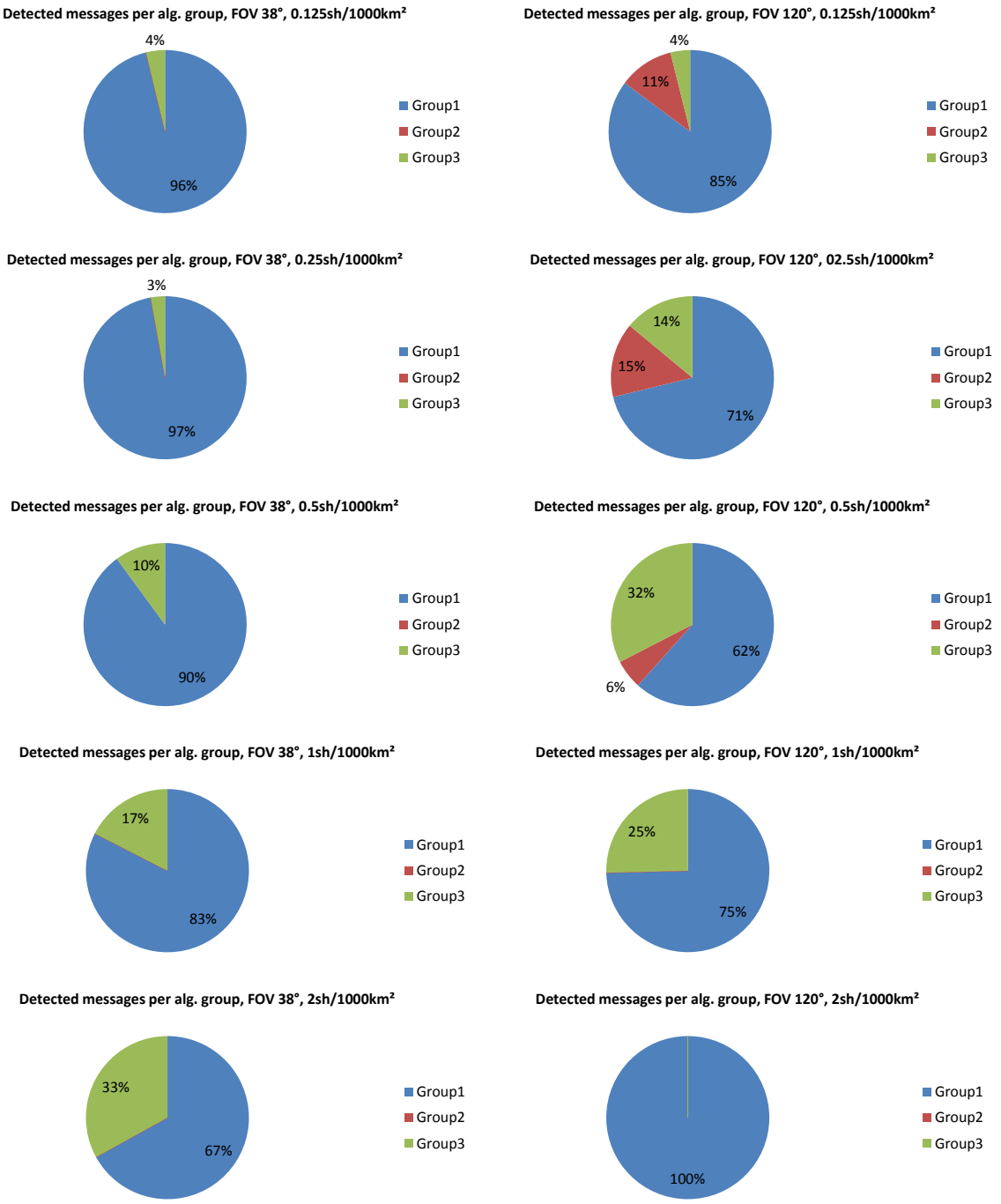


Fig. 100. Processed messages per AIS demodulation algorithm group for narrow and wide angle antenna scenarios.

	Group1	Group2	Group3
Detected messages per alg. group, FOV 38°, 0.125sh/1000km²	96.23%	0.00%	3.77%
Detected messages per alg. group, FOV 120°, 0.125sh/1000km²	85.17%	10.95%	3.88%

Detected messages per alg. group, FOV 38°, 0.25sh/1000km²	97.16%	0.00%	2.84%
Detected messages per alg. group, FOV 120°, 02.5sh/1000km²	71.19%	14.82%	13.99%
Detected messages per alg. group, FOV 38°, 0.5sh/1000km²	89.93%	0.00%	10.07%
Detected messages per alg. group, FOV 120°, 0.5sh/1000km²	61.62%	5.81%	32.58%
Detected messages per alg. group, FOV 38°, 1sh/1000km²	82.53%	0.00%	17.47%
Detected messages per alg. group, FOV 120°, 1sh/1000km²	74.60%	0.00%	25.40%
Detected messages per alg. group, FOV 38°, 2sh/1000km²	66.96%	0.00%	33.04%
Detected messages per alg. group, FOV 120°, 2sh/1000km²	100.00%	0.00%	0.00%

Table 19. Processed messages per AIS demodulation algorithm group for narrow and wide angle antenna scenarios. Source data for Fig. 100.

As it can be learned from Table 19 above, for the main demodulation work is performed by Group1 algorithms. The Group2 performs well in wide angle circumstances, unless too many signals are present. The Group3 offers up to 33% demodulation work for the narrow and wide angle antenna demodulation scenarios.

This demonstrates and concludes the significant contribution of the Group3 algorithms, which were developed and proposed by author of this document, to the demodulation process of AIS signals received by satellite in low earth orbit.

6.4.5 Analysis of parameter sensitivity

An interesting question could be raised by asking, where are the places of the bottlenecks for the demodulation processors? The question could be motivated by an interest to search for improved parameter configurations to reach higher demodulation rates of the AIS demodulation algorithms.

To carry out such an analysis, let us choose slightly relaxed parameter conditions for the demodulation processors than the ones provided in Table 16 and apply them for the wide angle antenna case for the varied ship densities as before.

For the first simulation case let us assume that for all the 4dB signal amplitude distance requirements in the strongest signal or joint AIS signal processors a 3dB margin would suffice. This relaxed condition is further referenced as RC1. That would lead to following parameter settings.

```
%RC1, relaxed conditions with SNR=3dB, FOV=120°
noise_level = -143 [dB]
SC1_SNR =3;  %[dB]
SC2_SNRS=6;  %[dB]
SC2_SIR =3;  %[dB]
CC1_SNRS=17; %[dB]
CC1_SIRJ=0.1;%[dB]
CC2_SNR =3;  %[dB]
CC2_SIR =0;  %[dB]
CC2_SIRI=3;  %[dB]
CC2_SIRJ=19; %[dB]
CC2_SNRP=3;  %[dB]
CC2_syncb=5; %[bits]
```

And let us introduce also another relaxed condition dedicated for A3:CC2 algorithm, assuming that the interferer margin is can be lowered to 2dB. This relaxed condition is further referenced as RC2.

```
%RC2, relaxed conditions with SNR=3dB and CC2_SIRI=2dB, FOV=120°
noise_level = -143 [dB]
SC1_SNR =3;  %[dB]
SC2_SNRS=6;  %[dB]
SC2_SIR =3;  %[dB]
CC1_SNRS=17; %[dB]
CC1_SIRJ=0.1;%[dB]
CC2_SNR =3;  %[dB]
CC2_SIR =0;  %[dB]
CC2_SIRI=2;  %[dB]
CC2_SIRJ=19; %[dB]
CC2_SNRP=3;  %[dB]
CC2_syncb=5; %[bits]
```

The message detection rates of the first and the second relaxed condition simulation approaches are reported in the second and third column of Fig. 101 respectively, with the demodulated message numbers shown in Table 20 below.

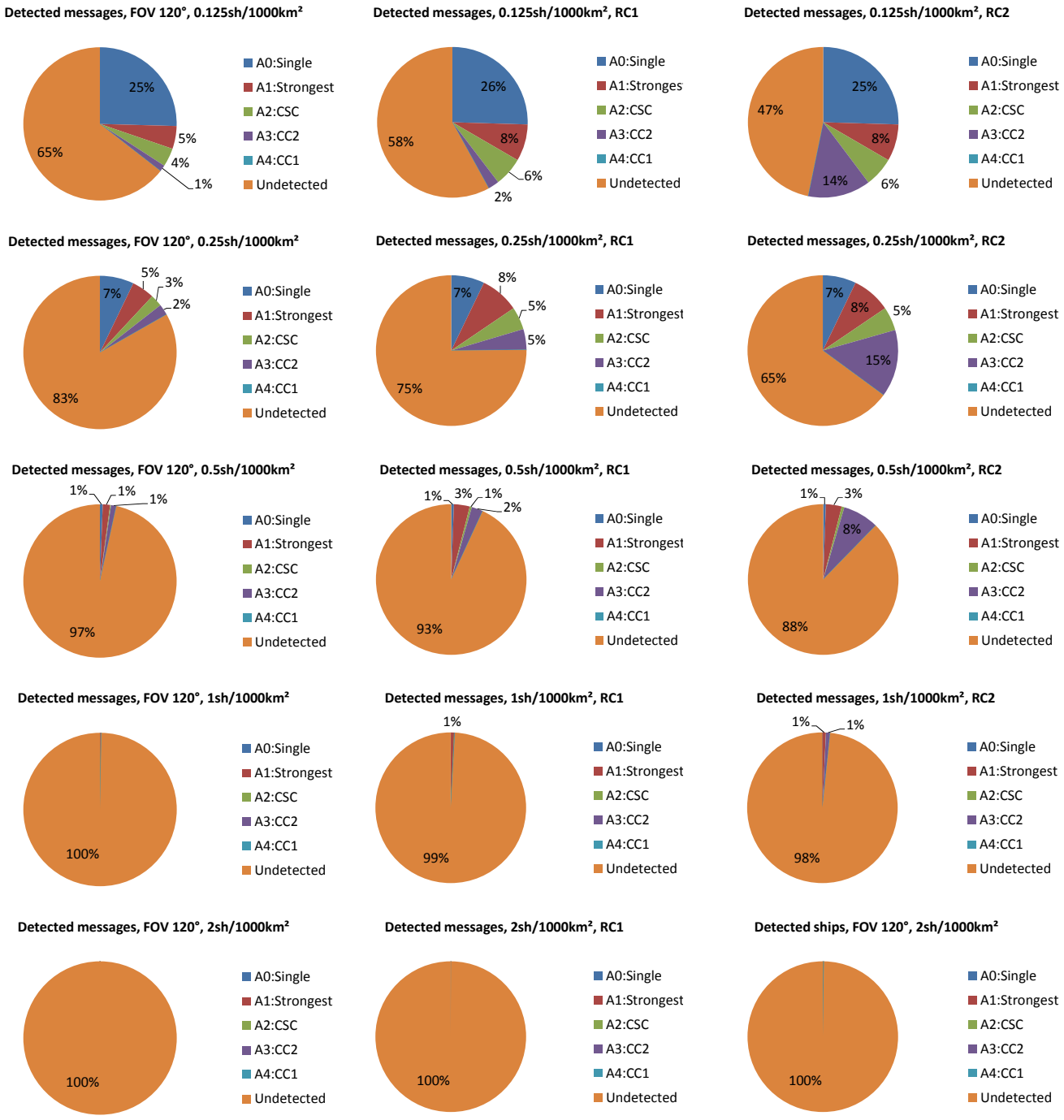


Fig. 101. Simulation results of AIS signal demodulation for wide FOV antenna with 120° opening angle with relaxed parameters. The first column is initial results as in Fig. 99. The second column presents results for the RC1 and the third column for the RC2 case.

	A0:Single	A1:Strongest	A2:CSC	A3:CC2	A4:CC1	Undetected	FoV	Given	A:SUM	A:SUM %
Detected messages, FOV 120°, 0.125sh/1000km²	739	140	113	42	0	1870	2904	3983	1034	36%
Detected messages, FOV 120°, 0.25sh/1000km²	409	273	142	144	0	4744	5712	7967	968	17%
Detected messages, FOV 120°, 0.5sh/1000km²	67	177	23	235	0	11000	11502	15939	502	4%
Detected messages, FOV 120°, 1sh/1000km²	2	45	0	136	0	22931	23114	31878	183	1%
Detected messages, FOV 120°, 2sh/1000km²	1	3	0	24	0	46064	46092	63756	28	0%
Detected messages, 0.125sh/1000km², RC1	739	230	181	69	0	1685	2904	3983	1219	42%
Detected messages, 0.25sh/1000km², RC1	409	477	278	255	0	4293	5712	7967	1419	25%
Detected messages, 0.5sh/1000km², RC1	67	390	54	278	0	10713	11502	15939	789	7%
Detected messages, 1sh/1000km², RC1	2	174	1	28	0	22909	23114	31878	205	1%
Detected messages, 2sh/1000km², RC1	1	20	0	0	0	46071	46092	63756	21	0%
Detected messages, 0.125sh/1000km², RC2	739	230	185	392	0	1358	2904	3983	1546	53%
Detected messages, 0.25sh/1000km², RC2	409	477	290	828	0	3708	5712	7967	2004	35%
Detected messages, 0.5sh/1000km², RC2	67	390	66	891	0	10088	11502	15939	1414	12%
Detected messages, 1sh/1000km², RC2	2	174	2	204	0	22732	23114	31878	382	2%
Detected messages, 2sh/1000km², RC2	1	20	0	20	0	46051	46092	63756	41	0%

Table 20. Source data table for the charts of Fig. 101.

Comparing results between the first two columns of Fig. 101, one can note a significant increase in message demodulation rate from left to right. For the ship density case of 0.25 ships per 1000km^2, the rates from initial demodulation results of 7%, 5%, 3% and 2% for the demodulation algorithms A0:Single, A1:Strongest, A2:CSC and A3:CC2 respectively increase to 7%, 8%, 5%, and 5%. As expected, the RC1 condition does not influence the number of single messages per slot found, thus the results for A0:Single algorithm stayed unchanged, whereas all the others showed a significant increase. Especially the A3:CC2, which more than doubled its rate of processed messages.

Comparing results further between the second and the third columns, one can note a significant increase in number of demodulated messages, introduced by the A3:CC2 algorithm. For the 0.25 and 0.5 ships per 1000km^2 densities there can be observed increase from 5% to 15% and 2% to 8% in the algorithm's demodulation rates, showing a dominance in a demodulation performance over other algorithms.

Comparing to results shown in Table 17, where it was 6, 1, and 0% rates reported for recognized messages of 0.5, 1, and 2 ships per 1000km^2 densities respectively, here the rates are 11%, 4% and 1%, which is about double that of the previous demodulation approach.

The relaxed demodulation condition analysis shown in Fig. 101 along with the algorithm performance analysis reported in Fig. 98 and Fig. 99 allow to derive a conclusion that an improved capability of packet separation in a densely interfering signal environment, especially the four signal characterising parameter estimation (the Doppler frequency and phase, amplitude and bit-clock), can significantly enhance an overlapped signal demodulation ability and using a joint overlapped signal demodulation capability can offer a remarkable benefit for AIS signal demodulation in space applications. The analysis also show the high benefit of the A3:CC1 and A4:CC2 algorithm's contribution to the collided message demodulation and points to promising further benefits, once the obstacles of high signal power requirement for A4:CC1 and easier signal parameter estimation for A3:CC2 could be solved.

7 Outlook

The learning curve

During the 5 work period in which the AIS topic was researched within DLR, the author of this document has learned and accomplished many AIS application related tasks. The work was started with developing a Software Defined Radio (SDR) based AIS transmitter to ensure a correct understanding of AIS modulation principles. After ascertaining that one's own modulated AIS messages can be received by a casual maritime receiver, a basic receiver was also implemented. Then, with entering into the developer team for the first DLR AIS satellite AISat1, an on-board AIS data compressing and analysis algorithms were developed for the satellite, which led to deeper knowledge about AIS data and its processing characteristics.

Also motivated by the needs of AISat1, an SDR based AIS-Testbed was developed to derive or prove performance characteristics for different AIS receivers. This included the analysis of a receiver's resistivity to noise and Doppler shift variations as well as the capability to extract overlapped AIS messages with various overlapping properties. Some part of the testbed development work was published in ASMS/SPSC2012 conference [49], which hosted a dedicated session for SAT-AIS topic.

The work accomplished

The main target for the author's PhD thesis was to develop algorithms which can be used for extracting overlapped messages from signals received by a single sensor antenna on satellites. Although there were met many unexpected complications during the pathway, the goal was accomplished, showing demodulation gains for different existing and newly proposed demodulation methods applied to overlapped AIS signals. From existing frequency extraction based methods, the JRRMBD was implemented and considered for more in-depth AIS applications, also the CSC, which is already used in space-AIS receivers and was just shortly mentioned within this work, has been implemented. Two new methods were derived: the JShBDD for joint IQ data estimation based on IQ signal shape analysis and the JGSUM method, for joint Gaussian sum estimation based on signal frequency analysis with the MBD algorithm.

The further works

The JGSUM algorithm in strongest-signal mode has already been successfully applied to real world IQ data samples thanks to raw frequency data recording availability from airplane pass over the German Bight and also AAUSAT-3 satellite mission [46]. All the demodulation algorithms still need to be further adjusted for use with real IQ data samples from space. Including customizations and optimisations needed for a continuous and real time data processing capability and implementation on FPGA chips for future real-world application on satellites.

For further analysis and implementation are also awaiting the two filled patents [Pat.1] and [Pat.2], which propose improvements for more efficient and more secure satellite-AIS channel use. As ship density along coastal zones increases and the SAT-AIS standard widespread, it would be of a significant benefit for SAT-AIS operators to have the two patents imposed internationally.

References

[1] *Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band*, Radiocommunication Sector of International Telecommunication Union, Recommendation ITU-R M.1371-4, 2010.

[2] "Carriage requirements for shipborne navigational systems and equipment," *International Convention for the Safety of Life at Sea (SOLAS)*, Chapter V.

[3] *Adoption of new and amended performance standards*, International Maritime Organisation, Resolution MSC.74(69), 1998.

[4] *SPO, Select AIS Device's Processing and Output, Command*, National Marine Electronics Association, Standard NMEA 0183, ver. 4.00, 2008.

[5] "AIS transponders ", International Maritime Organization, http://www.imo.org/OurWork/ Safety/navigation/pages/ais.aspx. (Accessed: 2 September 2014).

[6] G. Krebs, "Orbcomm 37 - 41", Gunter's Space Page, http://space.skyrocket.de/doc_sdat/ orbcomm-ql.htm. (Accessed: 2 September 2014).

[7] G. Krebs, "Rubin 9", Gunter's Space Page, http://space.skyrocket.de/doc_sdat/rubin-9-ais.htm. (Accessed: 2 September 2014).

[8] G. Krebs, "Rubin 7-AIS", Gunter's Space Page, http://space.skyrocket.de/doc_sdat/rubin-7-ais.htm. (Accessed: 2 September 2014).

[9] "exactAIS®", exactEarth, http://www.exactearth.com/products/exactais. (Accessed: 2 September 2014).

[10] *Performance Standards and Functional Requirements for the Long-Range Identification and Tracking of Ships*, International Maritime Organisation, Resolution MSC.210(81), 2006.

[11] *Adoption of Amendments to the International Convention for the Safety of Life at Sea*, International Maritime Organisation, Resolution MSC.202(81), 2006.

[12] *Revised Performance Standards and Functional Requirements for the Long-Range Identification and Tracking of Ships*, International Maritime Organisation, Resolution MSC.263(84), 2008.

[13] M.A. Cervera, A. Ginesi, K. Eckstein, "Satellite-based vessel Automatic Identification System: A feasibility and performance analysis", *International Journal of Satellite Communications and Networking*, vol. 29, pp. 117-142, March-April 2011.

[14] T. Eriksen, G. Hoye., B. Narheim, B.J. Meland, "Maritime Traffic monitoring using a space-based AIS receiver", *Acta Astronautica*, vol. 58, pp. 537–549, 2006.

[15] F. Hennepe, R. Rinaldo, A. Ginesi, C. Tobehn, M. Wieser, Ø. Olsen, Ø. Helleren, R. Challamel, F. Storesund, "Space-Based Detection of AIS Signals," in *5th Advanced satellite multimedia systems conference (ASMA) and the 11th signal processing for space communications workshop (SPSC)*, 2010, pp. 17-24.

[16] P. Burzigotti, A. Ginesi, G. Colavolpe, "Advanced Receiver Design for Satellite-Based AIS Signal Detection," in *5th Advanced satellite multimedia systems conference (ASMA) and the 11th signal processing for space communications workshop (SPSC)*, 2010, pp. 1-8.

[17] O.F.H. Dahl, "Space-Based AIS Receiver for Maritime Traffic Monitoring Using Interference Cancellation", M.S. thesis, Norwegian University of Science and Technology, Department of Electronics and Telecommunications, 2006.

[18] G.Maral, M. Bousquet, *Satellite Communications Systems*, Wiley, 5th Edition, 2010.

[19] G. Eiden, R. Goldsmith "Performance of AIS sensors in space - PASTA-MARE project final report executive summary", European Comission, 2010.

[20] R. Challamel, T. Calmettes, C.N. Gigot, “An European Hybrid High Performance Satellite-AIS System,” *6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications*, 2012, pp. 246-252.

[21] “Amver Density Plot Display”, United States Coast Guard, http://www.amver.com/density.asp. (Accessed: 10 June 2014).

[22] “Maritime traffic density - results of PASTA MARE project”, European Comission, 2010.

[23] J.A. Larsen, J.F.D. Nielsen, H.P. Mortesen, U.W. Rasmussen, T. Laursen, J.L. Pedersen, “Evaluation of AIS Reception in Arctic Regions From Space by using a Stratospheric Balloon Flight”, Department of Electronic Systems in Aalbourg University, 2012.

[24] T. Eriksen, A.N. Skauen, B. Narheim, O. Helleren, Ø. Olsen, “Tracking Ship Traffic with Space-Based AIS: Experience Gained in First Months of Operations”, International Waterside Security Conference (WSS), 2010.

[25] M. Picard, M.R. Oularbi, G. Flandin, S. Houcke, “An Adaptive Multi-User Multi-Antenna Receiver for Satellite-Based AIS Detection”, *6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC)*, 2012, pp. 273-280.

[26] M. Zhou, A.J. Veen, R. Leuken, “Multi-User LEO-Satellite Receiver for Robust Space Detection of AIS Messages”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.

[27] B.T. Narheim, R. Norsworthy, “AIS Modeling and a Satellite for AIS Observations in the High North, Draft New ITU-R Report Improved Satellite Detection of AIS”, in Radiocommunication Sector of International Telecommunication Union, ITU-R Working Party 5B, 2008.

[28] K.D. Kammeyer, *Nachrichtenübertragung*, Vieweg+Teubner, 2011.

[29] J. Proakis, M. Salehi, *Digital Communications*, Irwin/Mcgraw Hill, 5th edition, 2007.

[30] G. Hoye, “Observation modelling and detection probability for space-based AIS reception – Extended observation area,” Norwegian Defense Research Establishment, FFI/RAPPORT-2004/04390, 2004.

[31] *Improved satellite detection of AIS*, Radiocommunication Sector of International Telecommunication Union, Report ITU-R M.2169, 2009.

[32] J.K.E. Tunaley, “The Performance of a Space-Based AIS System”, London Research and Development Corporation (LRDC), 2011.

[33] J.K.E. Tunaley, “Space-Based AIS Performance,” London Research and Development Corporation (LRDC), 2011.

[34] J.K.E. Tnunaley, “A Stochastic Model for Space-Borne AIS,“ London Research and Development Corporation (LRDC), 2005.

[35] R. Kumaresan, "Estimating the Parameters of Exponentially Damped or Undamped Sinusoidal Signals in Noise", Ph.D. Dissertation, University of Rhode Island, 1982.

[36] D.W. Tufts, R. Kumaresan, “Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood,” *IEEE Proceedings*, vol. 70, 1982, pp. 975-989.

[37] D.W. Tufts, R. Kumaresan, “Improved Spectral Resolution I-III”, *IEEE Proceedings*, vol. 68, 1980, pp. 419-420, 592-597, 1354-1355.

[38] S.M. Kay, S.L. Marple, “Spectrum Analysis – A Modern Perspective”, *IEEE Proceedings*, Vol.69, 1981, pp. 1380-1419.

[39] B. Sklar, “How I Learned to Love the Trellis”, IEEE Signal Processing Magazine, vol. 20, 2003, pp. 87-102.

[40] A. Linz, “Efficient implementation of an I-Q GMSK modulator”, *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 43, pp. 14-23, 1996.

[41] L.J. Ippolito, *Satellite Communications Systems Engineering: Atmospheric Effects, Satellite Link Design and System Performance*, John Wiley & Sons, 2008.

[42] S. Tirro, *Satellite Communication Systems Design*, Springer, 1993.

[43] T. Turletti, “GMSK in a nutshell”, Telemedia Networks and Systems Group of Massachussets Institute of Technology, Technical Report, 1996.

[44] R. Stöckli, “July, Blue Marble Next Generation w/ Topography”, Visible Earth - A catalog of NASA images and animations of our home planet, NASA Earth Observatory, http://visibleearth.nasa.gov/view.php?id=74393. (Accessed: 4 October 2014).

[45] C.A. Balanis, *Modern Antenna Handbook*, John Wiley & Sons, pp. 59-61, 2008.

[46] „We are now offering a couple of raw AIS samples from space“, AAUSAT-3, http://www.space.aau.dk/aausat3/index.php?n=Main.Getdata. (Accessed: 15 September 2014).

[47] S. Sud, “A Robust Joint Model-Based Demodulator for Continuous Phase Modulation Signals in an Unknown Environment,” *IEEE Global Telecommunications Conference (IEEE GLOBECOM)*, 2008, pp. 1-5.

[48] J.S. Goldstein, I.S. Reed, L.L. Scharf, “A multistage representation of the Wiener filter based on orthogonal projections,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2943-2959, 1998.

[49] A. Dembovskis, “Testbed for Performance Evaluation of SAT-AIS Receivers,” *6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC)*, 2012, pp. 253-257.

[50] A. Dembovskis, “Decoding Performance Testbed for Space-AIS Receivers,” Collaboration in Space for International Global Maritime Awareness (C-SIGMA), poster presented in Frascati, Italy, June 2011.

Patents

[Pat.1] J. Behrens, A. Dembovskis, “AIS ship's transceiver” U.S. Patent US20140128098 A1, issued on May 08, 2014. “AIS ship transceiver”, European patent EP 2728564 A3, issued May 07, 2014. “AIS-Schiffstransceiver” German Patent DE 102012110540 A1, issued on May 8, 2014.

[Pat.2] A. Dembovskis, J. Behrens, “AIS-Schiffstransceiver”, U.S. Patent US 20140127990 A1, issued on May 08, 2014. “AIS marine transceiver detects a rogue base station”, European Patent EP 2728920 A1, issued May 07, 2014. “AIS-Schiffstransceiver”, German Patent DE 102012110541 A1, issued on May 8, 2014.

Glossary

ADC – Analog to Digital signal Converter

AGI/STK – Systems Toolkit from AGI

AIS – Automatic Identification System

AIS-SART – AIS Search And Rescue Transponders

APP – a posteriori probability

AtoN – Aids to Navigation

AWGN – Additive White Gaussian Noise

BLP – Backward Linear Prediction

CC1 – Collision Case 1

CC2 – Collision Case 2

CC3 – Collision Case 3

CC4 – Collision Case 4

CSC – Consecutive Signal Cancellation

CSTDMA – Carrier Sensitive TDMA

DAC – Digital to Analog signal Converter

DLR – German Aerospace Center (Deutsches Luft- und Raumfahrtagentur)

DSP – Digital Signal Processing

ESA – European Space Agency

FBLP – Forward Backward Linear Prediction

FFI - Norwegian Defence Research Establishment (Forvarets Forskningsinstitutt)

FLP – Forward linear Prediction

FOV – Field of View

FPGA – Field Programmable Gate Array

FSM - Final State Machine

GMSK – Gaussian Minimum Shift Keying

GSUM – Gaussian Sums algorithm

GUI – Graphical User Interface

HTZ – High Traffic Zones

IMO – International Maritime Organisation

IQ data – In phase (real part) and Quadrature (imaginary part) of sampled radio frequency data

ISI - Inter Symbol Interference

ITDMA – Incremental TDMA

ITUR – International Telecommunication Union Regulation

JGSUM – Joint GSUM algorithm

JShBD – Joint Shape Based Demodulator

LEO – Low Earth Orbit

LPF – Low Pass Filter

LRIT - Long-Range Identification and Tracking system

LS - Least Squares

LTZ – Low Traffic Zones

ML - Maximum Likelihood

MSG123 – AIS position report messages of type 1, 2 or 3

MTZ – Medium Traffic Zones

nm – nautical miles

NMEA – National Marine Electronics Association

NRZI – Non Return to Zero Inverter

OSR – Oversampling Rate, expressed in spb

PDF - Probability Density Function

RATDMA – Random Access TDMA

RC1 – Relaxed Condition 1

RC2 – Relaxed Condition 2

RF – Radio Frequency

SAT-AIS – Satellite based AIS

SDR – Software Defined Radio

SOLAS – Safety Of Life At Sea

SOTDMA – Self Organized TDMA

spb – samples per bit

sps – samples per second

TDMA – Time Division Multiple Access

TLE – Two Line Elements

VDA - Viterbi Decoding Algorithm

VE - Viterbi Equalizer

Acknowledgments

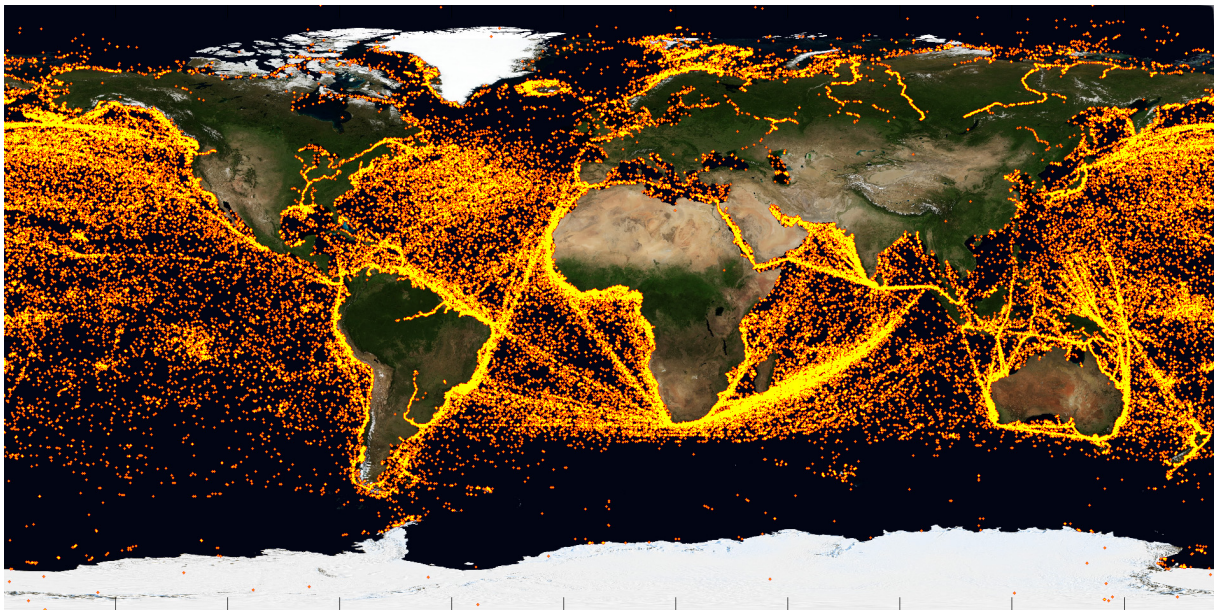
I would like to express my deepest gratitude to the DLR for enabling the realisation of this 5 year research work. Thanks to all of my great colleagues from the Satellite Systems department of DLR Bremen, especially to Mr. Jörg Behrens as head of the department for supporting and promoting this topic and to Dr. Birgit Suhr for the financial support during the last year of this work. Warm thanks to the whole AISat1 satellite team for the great collaborative work and for the involvement of me in the unique experience of the satellite development. Especially thanks to the colleagues Falk Nohka, Martin Drobczyk and Michael Jetzschmann. I learned a lot from your insights and was pleased to deliver AIS payload supporting software for the AISat1 project.

Similarly great thanks go to the Faculty of Mathematics of University of Bremen, to the colleagues in my working group and my doctorate supervisor Prof. Marc Kesseböhmer. Thank you for input, consultations and inspirations.

Also many thanks to the Faculty of Communications of University of Bremen, for helping me to gain familiarity with signal processing theories and techniques, especially a great personal thanks to Dr. Henning Paul for the highly valuable consultations.

Finally my best thanks to all of my friends and supporters for enabling and facilitating me to manifest this work, especially my parents and dance teachers. Thank you.

For a representative picture of my Satellite-AIS involvements, I have created a world map showing AIS position data, which the AISat1 team has been able to gather with the satellite's mission during the first 6 months of in-orbit activity. I wish that everybody could have the experience of seeing the world through one's own eyes and thanks the AISat1 team for giving me this opportunity.



Received AIS position reports by AISat1 mission, as of 2015.03.17. Source of the world map is [44].