TECHNIQUES TO ACCELERATE THE TRANSITION TO A NEW GENERATION OF TERRESTRIAL DIGITAL TV

Daniel Soto Álvarez

PhD THESIS UPF / 2020

Thesis supervisor

Dr. Josep Blat Gimeno

(Dept. of Information and Communication Technologies)

DEPT. OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Universitat Pompeu Fabra Barcelona

[Blank page]

Life is the most valuable gift
you've ever received.
— Φφ —

Abstract

This thesis explores the problem of compatibility between generations in the field of Digital Television, and proposes solutions to this problem. Since its inception, the DTV has been under development without taking into account the ability to continue receiving broadcasts with legacy equipment when a new generation is introduced. Therefore, when it comes time to introduce a new technology, a simulcast solution is usually used.

The study carried out in this thesis explains the reasons why a traditional simulcast is a problem. As a concrete study example, the case of the transition from SD MPEG-2 to HD H.264 broadcasts according to the DVB-T standard is analyzed. And how this problem unnecessarily extends the transition periods between generations. In this way it postulates that one way to promote the development of new generations of Digital Television is precisely to shorten transition times.

In order to facilitate the transition to a new generation, different technical solutions have been studied and developed that provide a certain degree of compatibility between generations. The results obtained conclude that it is not only feasible, but also desirable, to incorporate inter-generational compatibility solutions in Digital Television standards. Using the techniques described here, and others listed as future work, it would be possible for new standards and generations to incorporate this necessary capability.

Resumen

Esta tesis explora el problema de la compatibilidad entre generaciones en el área de la Televisión Digital, y propone soluciones a ese problema. Desde su inicio, la TVD se ha ido desarrollando sin tener en cuenta la capacidad de seguir recibiendo emisiones con equipos antiguos cuando se introduce una nueva generación. Por lo tanto, cuando llega el momento de introducir una nueva tecnología, se suele utilizar una solución de emisión simultánea o *simulcast*.

El estudio realizado en esta tesis explica las razones por las que un *simulcast* tradicional es un problema. Como ejemplo de estudio específico, se analiza el caso de la transición desde emisiones SD MPEG-2 a HD H.264 utilizando el estándar DVB-T. Y cómo esta dificultad alarga innecesariamente los períodos de transición entre generaciones. En esta línea se postula que una forma de promover la implantación de nuevas generaciones de Televisión Digital es precisamente acortando los tiempos de transición.

Para facilitar la transición a una nueva generación, se han estudiado y desarrollado diferentes soluciones técnicas que proporcionan un cierto grado de compatibilidad entre generaciones. Los resultados obtenidos concluyen que no sólo es factible, sino también deseable, incorporar soluciones de compatibilidad intergeneracional en los estándares de Televisión Digital. Utilizando las técnicas aquí descritas, y otras enumeradas como trabajos futuros, sería posible que los nuevos estándares y generaciones incorporaren esta característica necesaria.

Keywords: *Digital Television, Joint Simulcast, inter-generational compatibility, SD to HD migration, MPEG-2 to H.264, aggressive compression, independent encoding.*

Table of contents

[Blank page]

Chapter 1

INTRODUCTION

Digital Television or DTV[1] today can undoubtedly be considered a success. Currently 90-95% of the world's population can enjoy this service through some form of transmission (terrestrial, satellite or cable). However, during the transition from analogue to digital television, there were many doubts as to whether this would really be the case. Among the many questions that have been analysed, it is notable that the technical questions were the least questioned —which is actually logical because DTV standards are very robust—. Thus, socio-cultural and economic aspects, such as potential user resistance, the digital gap or equipment prices, were extensively analysed and discussed in the preparation of the migration plans. Although in the end, all the efforts and work resulted in a transition in which very few serious problems were reported, so the migration to Digital Television has finally not been questioned.

But despite its success, DTV also has some important counterparts. A very remarkable one, which is the basis of this thesis, is the lack of compatibility between generations. In the industry we usually talk about backwards and forwards compatibility. These terms are used when it is intended to give continuity to old or legacy equipment when new innovations are introduced. And depending on whether the innovations are compatible with the legacy equipment (backward compatibility) or whether the legacy equipment is compatible with the new technology (forward compatibility), then one term or the other is used. However, in the field of DTV it is more appropriate to talk about *compatibility between generations*[2]. This term describes the characteristic that older equipment can continue to function even if the technology used is replaced by a new one, or more precisely when a new technology has been introduced. And this is what happens when a new generation is deployed. This is why the term inter-generational compatibility is used.

And concerning this inter-generational compatibility none of the current DTV standards incorporates that feature. In other words, when a new generation is introduced, it is true that the new devices that are coming onto the market will have the capacity to receive old signals. But the old receivers are unable to do so. Therefore, an incompatibility gap is created between generations. However, it should be noted that despite this, DTV standards are still very flexible. Certain parameters such as the number of services, the error correction rate, the degree of compression, and many others can be varied over time to best suit the needs of the moment. And that capability was actually non-existent within the analogue television. Therefore, in itself this flexibility represents a major advance over previously existing technology. However, this flexibility is limited, and

[1] Even "DTV" or "Digital TV" can be considered a name, in particular related to the terrestrial broadcasts of digital television in USA, we use here DTV as the acronym of the generic concept of *Digital Television*.
[2] We will use the terms *"compatibility between generations"* or *"inter-generational compatibility"* interchangeably.

does not allow for inter-generational changes. This is why it can be said that one of the mistakes made in DTV is the lack of compatibility between generations. We will briefly analyze this statement in a little more detail in this introductory section.

1.1 Why the lack of Digital TV inter-generational compatibility is a pitfall?

When we say that the DTV has a fault —the lack of compatibility between generations— we are not saying that from a technical point of view it contains a bug. Rather, we are claiming that it lacks a technical capability that would be necessary. Whether or not to include a particular feature in any system is a design decision. And without a doubt the different standards of digital television have been meticulously designed and tested. However, from our point of view, the negative impact of the lack of compatibility between generations has been underestimated.

Like many other technological products, the DTV incorporates the common backward compatibility. In other words, new equipment is also compatible with previous generations, but not the opposite. This backward compatibility is easy to implement technically because it only requires that new devices do not remove capabilities. It is only necessary to design and add the latest features, and then configure the new devices to work indifferently with one or another technology, so the compatibility is guaranteed. Therefore, it is only after several generations that the obsolete capabilities are finally discarded, as they are no longer used. This is the common cycle of work in the industry when dealing with subsequent generations.

However, forward compatibility between generations is much more complex and difficult. One serious problem with it is that it has to be included in the design from the beginning. In other words, each new generation necessarily requires the ability to operate even when unknown functionalities are added. And that of course has an impact in terms of both complexity and costs. For this reason it is common that this capability is not taken into account when designing a new technology, unless it is clearly necessary. And that seems to be the reason why so far no digital television standard has incorporated compatibility between generations. But a brief analysis of the experience gained over the last years in the use of DTV suggests that this decision has at least been underestimated. This statement will now be discussed briefly.

When comparing digital television with digital mobile telephony, it can be observed that both technologies share certain parallels. When it comes to development and deployment, the model applied in both technologies is generally very similar. And this development model basically consists of designing and implementing a new generation from time to time, incorporating in each new features as well as numerous optimizations. The improvement is of course substantial, but with the particularity that each new generation implies a technological jump that breaks the compatibility. And yet this will not necessarily cause a problem, at least in the case of digital telephony. The reason is that when there is a high rate of renewal of the devices and a short life span of them, then the transition between generations is simple. Because it is understood that when the market is mature, then the deployment of the new technology will just start and gradually the equipment base will be renewed. In other words, the market is automatically adapted by the rapid replacement of the equipment. However, this progressive migration model, which works without problems with digital telephony, does not work in the same way with digital television. A possible explanation for this is given below.

In the digital telephony market the introduction of any new generation always follows the same pattern. Once the technical specifications have been completed and the standards have been approved, the frequencies to be used are offered at auction and the design of devices begins. Later, operators start to deploy their networks and the first terminals appear. Then progressively the networks will grow in extension and the number of compatible terminals will increase. That process continues for an indeterminate period until the vast majority of networks and terminals are compatible. And the process finally ends when support for obsolete generations is removed and legacy networks are switched off. Ultimately, the whole process may take two or three generations, but it always has the same end.

However, in the case of digital terrestrial television, although the technological evolution should more or less follow the same approach, it turns that it is not exactly working in the same way. Transition periods are much longer and migration from one generation to another is much more complex and costly. A probable explanation for this difference could be the following. Currently the greatest effort for the deployment of a new digital telephony technology seems to be related to the deployment of the new network. This is because a simple reason. After all, it is only when the operator has the necessary spectrum and begins to install the first antennas that compatible terminals arrive to the market. So there is no gap with the market and therefore the beginning of the deployment does not imply an excessive risk. On the other hand, the transition is made smoothly because the terminals are gradually renewed by new ones, which are always compatible with the new technology. And again there is no mismatch between the deployment of the networks and the terminal base. Everything seems to fit together smoothly.

But experience is showing that this is not really the case with how digital television is evolving. And the main reason for this may be a very different rate of equipment renewal. Consider this difference in detail. With regard to digital mobile telephony the average time between generations is generally close to a decade, and the period of renewal of the terminals is between 12 and 36 months. However, with regard to digital television, there is no clear horizon between generations. In fact, the transition periods are diffuse and diverse. There are even territories where one generation has skipped. And that is because the renewal rates of DTV receivers are very different. In this case the times are not measured in months, but in years. Currently, in the first world the average renewal time varies between 5 and 8 years. But this period is extended much further in other territories, reaching figures of between 12 and 15 years. This difference should certainly have an important impact on all of this. So in the next section, there is an analysis of the causes of this fact.

1.2 The migration troubles of the different Digital TV Generations

A plausible explanation for the mismatch in the evolution of DTV is the possible vicious circle that arises when a new generation appears. This vicious circle can be explained as follows. With the arrival of a new generation to the market, three main actors play: industry on the one hand, broadcasters on the other, and lastly users.

Assuming that the standards are fully finalized and tested, the introduction of the new generation should not cause major problems if the forces between the actors are balanced. But in the case of digital television there is usually a mutual interlock between the actors. On the one hand, broadcasters do not start broadcasts with the new technology for lack of spectrum and users, which makes new broadcasts unprofitable for them. On the other hand, users do not renew their equipment due to lack of emissions and high prices. And finally, the industry does not produce compatible equipment for mass use due to the lack of a market. And the vicious circle continues until at some point there is a radical change. For example, when an external actor such as a regulator imposes a new technology; when the industry floods the market with compatible equipment because the technology has gone outdated; or when new radio spectrum has been released to allow additional emissions with the new technology at low cost. In any case, it can be seen that the result of this vicious circle will be that the transition from one generation to another becomes complicated and difficult.

A simple review of the different generations of digital television can help to see the causes behind the difficulties of migration between generations. However, a first aspect to be clarified is what is understood by generation in the field of digital television. In this thesis, the focus of what is a generation of digital television concentrates on what users perceive as such. That is, they are the different generations of reception equipment. Or named otherwise, they are the multiple consumer generations. So, it is not the generations of transmission standards (DVB-T, DVB-T2, etc.), nor the generations of encoding formats (MPEG-2, H.264, H.265, etc.). Rather, it is the combinations of all of them that come to the market at a given time. For example, here we will consider the first generation the SD services in MPEG-2 with DVB-T, the second generation the HD services in H.264 with DVB-T, the third generation the HD with H.264 over DVB-T2, and the fourth generation the UHD H.265 over DVB-T2.

However, this division is incomplete. Basically, because there is no single standard for digital television. Currently there are four of them in use, and this classification only takes into account the DVB standard. Therefore, a brief analysis of what the different consumer generations are with each of the different standards will be presented. The standards currently used are: ATSC, DVB, ISDB and DTMB. Although some of them have different sub-standards for different transports (terrestrial, satellite, cable, etc.), the analysis focuses particularly on the terrestrial versions. Within the digital terrestrial television specifications, the following consumer generations can be distinguished for each of the different standards:

	First	**Second**	**Third**	**Fourth**
ATSC	SD/HD MPEG-2 MPEG-TS 8-VSB	HD H.264 MPEG-TS 8-VSB	HD/UHD HEVC MMT OFDM	
DVB-T	SD MPEG-2 MPEG-TS DVB-T	HD H.264 MPEG-TS DVB-T	HD H.264 T2-MI DVB-T2	UHD HEVC T2-MI DVB-T2
ISDB-T	SD/HD MPEG-2 BTS BST-OFDM	SD/HD H.264 BTS BST-OFDM	UHD/8K HEVC MMT BST-OFDM	

DTMB	SD MPEG-2 MPEG-TS TDS-OFDM	HD H.264 MPEG-TS TDS-OFDM		
Formats: SD / HD / UHD - Codecs: MPEG-2 / H.264 / HEVC Containers: MPEG-TS / BTS / T2-MI / MMT Modulations: 8-VSB / DVB-T / BST-OFDM / TDS-OFDM / OFDM/DVB-T2				

Table 1. The different generations of products within the diverse DTV standards

Although the most recent generations of some standards are still unfinished in some cases, it is noticeable that the number of generations is unequal. Moreover, it happens that the degree of implementation of each standard is even more variable. For example, while in some territories the switch from analogue to digital television has been made directly to a modern generation, in other cases what has been done is to start with an obsolete or almost obsolete generation. But there have also been cases where the switch was made at the beginning of digital television. And there are also even more scenarios. For example, when a certain generation has been skipped in order to avoid the transition to an already outdated generation. All this suggests that migration between generations is not an easy process when it comes to digital terrestrial television.

To analyze an interesting case, the implementation of DTT in Spain will be described. In this country, digital television broadcasting began at the end of the 1990s. These broadcasts started using the first generation DVB-T standard with the MPEG-2 codec and SD services. Today, at the end of the 2010s —twenty years later— it is still the most used standard in this territory. Although it is true that there have been since the beginning of the 2010s regular broadcasts using the second generation (HD services with H.264 codec), there is not even a transition plan and only a small number of channels are distributed in HD. And the most likely scenario at the moment is that for the early 2020s the first generation signals will be dropped, and the second generation will be adopted, while tests begin with the third generation. But this evolution can be considered very inefficient because: 1) the fourth generation already exists; 2) the second generation is virtually outdated; 3) the third generation is used in many other territories on a regular basis; 4) without a clear transition plan the migration to the third —or even the fourth if the third generation is skipped— could require another 15 years or more. All this shows with this example that there is actually a problem in addressing the switch from one generation to another.

Taking into account this fact, that it is not easy to move from one generation of DTV to another, the author of this thesis proposes a potential solution: it implies the reduction of transition times between generations by applying technical measures based on inter-generational compatibility. This means that since it is not possible to add to the standards an inter-generational compatibility that was not initially included, then perhaps it would be easier to make those migrations smoother by exploring alternatives that would at least reduce transition times. Of course, the most optimal solution would be if each standard natively incorporated forward compatibility. But that has not been done so far, although it would be potentially interesting. It is therefore appropriate to present solutions that address this problem and facilitate as much as possible the negative impact that the lack of compatibility generates on the migration process. The following section presents some ideas and contributions concerning the problem outlined.

1.3 Contributions to accelerate the migration between Digital TV Generations

As explained in the previous section, the difficulty in migrating from one generation of DTV to another is caused in part by the lack of forward compatibility of older receiving equipments. Therefore, it makes sense to assume that the migration process could be simplified if existing devices can continue to receive television programs in some way when a new generation is introduced. Based on this premise, the author proposes to add the following restriction regarding the migration process to achieve the objective: "When a new generation is introduced, it must be ensured that all services continue to be received using the legacy devices". This restriction, although it might not seem to make much difference, should be noted that it can be applied in any of the following two scenarios:

1. When the number of services broadcasted with the previous generation does not change, but new services are added using the new generation. In that case the new services must also be accessible with older equipment.
2. When all the services broadcasted with the previous generation are migrated to the new generation. In that case the reception of these services is preserved with legacy devices.

It is easy to see that the key factor in complying with the proposed restriction in any of the above scenarios is based on the basic principle of "do not eliminate services and only add new functionalities". And this goal can be achieved in two ways: through compatibility or by simulcast. The first solution involves to distribute the services in such a way that they can be processed by both legacy and emerging devices. While the second solution provides a way to distribute services using both technologies at the same time. Each of these two options has advantages and disadvantages. It is therefore necessary to analyse each of them.

The implementation of compatibility necessarily requires changes to the standards. If these do not include the ability to be forward compatible —which would be the optimal solution— then it is necessary to modify the new generation standard so that it will be partially compatible with the previous standard. This is not the same as backwards compatibility. In the latter, the new equipment can work with either of the two generations. But when partial compatibility is added to the new standard, what is raised is that the new equipment can take advantage of the signals of the previous generation at the same time that they work with the signals of the new generation. In that sense, part of the signal —from any source— is divided into two sections: the former and the modern. The old one is broadcasted in the usual way using the old standard, and the new one is processed in such a way that part of it is broadcasted using the new standard. The new receiving equipment will then reconstruct the entire signal by processing both sources. This solution can be called hybrid compatibility.

This first hybrid compatibility solution is considered by the author in the first of the papers included in this thesis. However, its practical application is partly difficult. Mainly because it is necessary to modify both transmitter and receiver equipment to accept hybrid broadcasts. And while designing and including such modifications is not as costly as defining a new standard, the steps required are actually the same: design, testing, development and deployment. It also requires that both industry and broadcasters, as well as in many cases regulators, agree to use such modifications. The result is that only by applying a consensus model is it possible to use this approach. However, on a technical level it is a functional solution, as demonstrated by this research.

The second option discussed is to use a full simulcast during the entire transition period. With regard to this solution, it should be noted that a full simulcast is not the same as a partial simulcast. If only a subset of the total services can be received using both technologies —the partial simulcast case— then an artificial gap is being created between users. And this is true, whether it is chosen to broadcast services only with the new technology, or to keep others only with the old one. In both cases, doing so will most likely slow down the transition to the new technology. The reasons are simple. If there are services that are only available with the old technology, then even if all receptors can reproduce 100% of the services, there will be no incentive for users to migrate to the new technology. On the other hand, if services are only accessible with the new technology, then broadcasters will see a limited penetration of the new services. And that may result in the broadcasting of these services become impractical due to low profitability.

Moreover, using terrestrial networks the available radio spectrum is currently very scarce. After successive frequency dividends, which have reduced the available bandwidth by 200MHz, the useful spectrum is small. In addition, taking into account the adjacent areas that occur in most territories, the effective number of channels can be reduced by up to 33% of the total. This is why in stable conditions, i.e. when only one technology is used, more than half of the existing channels are already being used. Therefore, the lack of sufficient free frequencies can seriously block the deployment of new simulcast frequencies. And yet there are still more potential problems.

A second problem with terrestrial broadcasting may be the high cost of coverage. With very large areas to cover and complex mountain terrains the number of transmitting antennas required can be very high. This can lead to prohibitive network expansion costs. Considering that a complete simulcast may require doubling the network capacity, the problem becomes very complicated in such cases. And progressive coverage will not always help, as the user base will be fragmented and that can make the transition as difficult as with a partial simulcast.

Finally, a third problem may also emerge. This is the reduction of signal robustness in the distribution network. When it is necessary to perform a full simulcast using the same number of frequencies, a workaround is to increase the network bandwidth by reducing redundancy. With this simple change in the network configuration it is possible to distribute more services without upgrading the network topology, but this can inevitably create reception problems. This is because some of the existing reception facilities may be at the limit of their performance, and any reduction in signal strength has the consequence that services may not be played back. For this reason it is generally assumed that decreasing redundancy will reduce network coverage. And the most obvious solution would then be to increase the number of antennas. But then the necessary action is almost the same as that required to include more frequencies, with the only exception of not requiring more spectrum. Therefore, this solution will also not be feasible in many cases.

To solve all these problems this research presents a viable option over the solution of a complete simulcast. This solution is based on distributing the same services in simulcast using the same bandwidth. To achieve this goal it is necessary to extremely compress and share elements between services. In the last two papers published by the author, two ways to achieve this objective are presented. Although this approach may seem risky at first, it has a clear advantage over any other. Since there is no need to modify the transmission network, costs are practically minimal. On the other hand, it also has other important advantages such as the rapid introduction of the new generation, which can greatly help to reduce the migration period. And in that case, the loss caused by the extreme compression will be diluted because in a short time it will be possible to abandon the legacy technology. It will then be possible to use the full capacity of the broadcasting network for the new technology, and thus extract the maximum quality from it.

1.4 Related research and similar projects

Although at the moment there does not appear to be any investigation focusing on intergenerational compatibility in DTV, this does not mean that there are no parallel lines of research. In fact, it would be inaccurate to say that Digital Television standards lacks any kind of backward compatibility at all. As discussed in the previous sections of this chapter, different DTV standards support at a minimum that new devices are compatible with older generations. In addition, all of them have adjustable parameters that allow changes to be made over time according to the circumstances. And even allowing for some improvements to be added later, which can sometimes be done more or less transparently. But outside of these general lines of research, there are other specific lines that are adjacent to the work of this research. Among others, the following can be listed as parallel lines: scalable compression, flexible modulation and some advanced processing and compression techniques. In the following paragraphs a brief review of these lines of research will be made in order to position the work carried out.

Scalable compression consists of dividing a sequence of images into two or more streams so that the different versions differ in their characteristics and are complementary. This means that one or more of the streams can be used to reconstruct the original sequence, thus obtaining approximations with more or fewer characteristics. The characteristics commonly used in scalable compression are: spatial resolution, temporal resolution and quality. For example, an HD sequence can be decomposed in spatial resolution into an SD version (lower resolution) and an extended version (the difference between the SD version and the HD version). In this way, the HD version can be reconstructed through the SD version and the extended version. Similarly, the temporal resolution can be decomposed, for example, by dividing a sequence at 50fps into two streams at 25fps. And the same is possible with the quality, by decomposing the sequence into a low quality basic version, and another extended version with higher definition. Thus, it would be possible to reconstruct the initial sequence in two ways: one with low quality, using only the basic version; and another with high quality, combining the basic version with the extended version.

In either case, the key of the scalable compression is to decompose the sequence into multiple layers and define a dependency relationship between them. And when such decomposition and subsequent compression are efficient, then this scalable compression makes sense. For example, if it is desired to migrate from HD to UHD services, using scalable compression it would be possible to use existing HD versions to reduce the bandwidth requirements needed for the UHD versions. In fact, scalable compression could be one of the pillars for a full compatibility between Digital Television generations. However, at the moment its use in DTV is negligible. Although there are many studies and developments in this field. For example, some standards such as H.264/SVC or H.265/SHVC are defined and could be used in DTV if desired. In any case, most of these lines focus only on two different areas: on increasing or extending the robustness of the broadcast network, or on its use in IPTV or Streaming environments. However, this concept of scalable compression was explored by the author and led to the idea of scalable compression with independent encoding (see paper 1 at chapter 2).

On the other hand the flexible modulation concentrates on the signal transport instead of the optimization of the encoding, as does the scalable compression. Signal modulation is the third major column of any DTV standard. The first column is encoding (A/V stream compression), the second is transport (packaging and signalling), and the third is modulation (RF signal encoding). A flexible modulation can be defined then as the ability to use different RF signal modulation modes, even unknown modes, without losing signal compatibility. This allows a modulated DTV signal using a particular standard to be combined with another digital signals within the same spectrum. In other words, flexible modulation can be used to reserve part of the bandwidth for other uses. These uses can be other digital television services, using the same or another technology, or totally different services.

And in order to achieve the goal of combining the DTV signal with another digital signal, both time division multiplexing and frequency division multiplexing can be used. But in any case it is indispensable that the DTV standard supports at least one of them in order to be able to combine different modulations. And not all of them do so. For example, the ATSC (1.0) and DVB-T standards completely exclude support for this capability. On the other hand, others such as ISDB-T, DTMB or DVB-T2 allow some combination of different modulations on the same frequency. This feature is usually used to assign different levels of robustness to different services distributed on the same channel. Thus, different modulations are used for each of the different services. This allows the reception performance of these services to be adapted to different uses: mobility vs. fixed, HD vs. SD, low power vs. regular performance, different distance/coverage, etc. However, each standard uses distinctive techniques, and only some of them allow the digital television signal to be combined with other completely different digital signals. That is why this area of research is partially under development. For example, regarding the group of DVB standards, there are studies combining different signals with DVB-T2 broadcasts (DVB-T2 Lite, DAB, etc). In these cases the combination is based on the Future Extension Frames (FEF) included in the DVB-T2 standard.

Using the combination of signals, and according to the experience acquired by the author of this thesis, flexible modulation could be a good basis for a complete compatibility between different generations of DTV. This argument is discussed in more detail in the chapter on future work (see chapter 3). This is because it would be much easier to build mixed transmission modes by simply incorporating into a standard the ability to combine the digital TV signal itself with any other type of digital signal. These new modes of hybrid compatibility would then consist of signals which combine old and new encodings. And between that point and full compatibility between generations there would be only a thin line that could be technically easy to achieve. This is why the author particularly advocates for this feature in future investigations into this line of research.

Finally, the third related line of research is the study of advanced processing and compression techniques. Undoubtedly, investigation into the improvement of audio and video codecs remains an open line today. However, leaving to one side the efforts to develop new standards, which are directly related to the development of new generations of Digital Television, there is the area of optimizing the compression. This is possible because the codecs used in DTV are defined in such a way that only specifications on how the signal should be decoded are dictated, and not really on how it should be encoded. In other words, the decoder is broadly described, but not the encoder. This allows the capabilities of the encoders to be improved and optimised even when the technology is already deployed. In fact, this feature has been used from the beginning in all DTV standards. And it is generally accepted that during the time of use of any particular generation, the level of compression will improve. In fact so much is so, that the area of research in the processing and optimization of both video and audio is very extensive.

And precisely because that is a very wide area of research, it is why the objective of this thesis is not competing in that area. In fact, our contributions to progress in this field could be classified as modest. But taking into account that the objective of this research is not to obtain better levels of compression, but to optimize the transition between generations, then it can be understood that the proposal is purely for innovation. And in that sense, to achieve a complete simulcast using the same available bandwidth —that the author postulates as the best option to accelerate the migration between generations and decrease the transition period—, it is necessary to apply the maximum possible innovation in the field of compression optimization. In fact, any technique that maximizes performance will be useful in this context. That is why even the simple aggressive compression proposed in the last document of this thesis (see paper 3 at chapter 2) is helpful to improve the results obtained.

As a conclusion, the author is confident that current and future developments in the listed areas will contribute to improve the distribution in the simulcast of DTV services. These improvements will certainly serve to improve compatibility between different generations of DTV, which as explained is a completely open field of research. It is especially expected that pitfalls that could stop the future progress and development of DTV will be avoided. Experience is already showing that generation transitions are difficult in DTV, so there is room for improvement in this area. Otherwise, it is more than likely that other services will end up replacing the space occupied by terrestrial DTV broadcasts.

Chapter 2

TECHNICAL DESCRIPTION

The content of this chapter describes the published research papers related to the research carried out in this thesis. Of the total work done, we present here only the part that has acquired visibility. This corresponds to three papers that describe different techniques that may be useful to achieve the objective proposed in the previous section. Which is no other than smoothing the transition from one generation to another in Digital Television. All the papers described in this section should be considered as a technical approach to potential solutions, and not as a theoretical study on improvements in aspects of the encoding field. This comment is relevant because the results of the research are geared towards innovation rather than pure research. Therefore, the results obtained are based on the design and implementation of technical tests that demonstrate the viability of the ideas proposed. And the validity of the results are then guaranteed by the publication of the papers.

The order in which these papers are presented is chronological, although the years of publication are not. This difference exists because the techniques resulting from the research have been developed over a considerable period of time. And the results have been published only when relevant results were obtained. However, the process has gone through different phases in which various ideas had been explored. Some of them were successful and others were not. Therefore, the vision of the author in this regard is to present the work according to a timeline. This makes it possible to observe the evolution of the ideas that have been developed throughout the research. And this is relevant because some techniques are based on previous results (both positive and negative).

More specifically, the research started from the concept of spatial scalability. The initial idea was to exploit this feature to make video compression more efficient. However, the idea was not new, as the MPEG-2 codec already provided for this feature. However, the number of research papers on the development of this concept using this codec was actually quite small. On the other hand, during the development of the H.264 codec, a lot of work was published in relation to scalability extensions. This resulted in the publication of these extensions under the H.264/SVC standard. At that time none of the results obtained in this research were at the level of the developments incorporated in that standard. However, much research on this concept could be reviewed and studied. This encouraged the author to consider other options, such as the possibility of spatial scalability using independent codecs.

This novel idea has consisted in generating two video streams in a scalable mode, but compressed completely independently. In other words, there was no relationship or shared data between the encoder and decoder for each of the streams. The advantage of this idea is that it could be applied to the migration of DTV broadcast. Basically because there was no need for elements that were not already available within the receiver equipment. Basically because there was no need for elements that were not already available within the receiver equipment. This third mode, different from the operating mode of the first generation and the operating mode of the second generation, would be to implement the scalable model at the beginning of the broadcasts with the new technology. In this way, the efficiency obtained by the application of the scalable model was intended to reduce the bandwidth requirements to provide enough free space to start emissions with the new generation. The results of this research are presented in the first of the papers described in this chapter.

On the basis of the results obtained, a second complementary technique was then proposed. So, if the first idea was focused on improving the efficiency of the video, the next one would concentrate on the audio. After exploring different alternatives, the idea of sharing the audio streams was considered. In fact, up to that point no research work had been done that addressed this option. Instead, the main line of research was focused on the development of new audio codecs that would be much more efficient at the bit-rate level. And at that time few of these codecs were based on scalable techniques. However, later on, with the emergence of a new generation of audio codecs —with advanced multitrack capabilities, lossless compression, and vector flows—, techniques based on complementary streams were developed for the audio encoding. The greatest exponent of this technique in the audio field is the DTS codec, which from its first version has included this concept in its design, and it was a sign of identity. Therefore, and due to the fact that there was already ongoing research in this line, other alternatives were explored. And this is how the idea of sharing audio streams emerged.

The technique of sharing audio streams in Digital Television is based on a simple idea: each DTV service is composed of N data streams. These streams can be of different types: video, audio, teletext, metadata, etc. And since a Transport Stream can contain different services, the proposal is to share audio streams between the services in the simulcast. Therefore, if two services of the same program in simulcast are included in the same MPTS, then instead of duplicating the audio streams, they would be shared. This allows for bandwidth savings that may not look very meaningful, as audio streams represent a small percentage compared to video streams. However, since digital television services are usually associated with two or three audio tracks —depending on the language in use— then the savings can be significant. Furthermore, this would be multiplied by the number of simulcast services within the same MPTS. Thus, when applied to a complete multiplex, this represents a not negligible gain. Accordingly, the second paper describes in detail a technique for sharing audio streams between DTV services.

Finally, the research of the author focused on the practical and realistic application of the work developed. Then the problem described in the previous chapter about the slow migration of digital terrestrial television services was identified. And it was observed that the lack of compatibility between generations was a critical factor. Then, reusing the results, it was considered the possibility of generalizing the advances obtained to facilitate this migration and thus reduce the transition times. Once the problem was identified, the possibility of optimizing the simulcast of services was considered. As discussed in the previous chapter, the ability to perform a complete simulcast is an effective way to facilitate the transition between generations. Based on this premise, the following strategy was developed: to optimize a simulcast, all elements should be shared and those that could not be shared should be aggressively compressed. The goal would then be to achieve the bandwidth required for the simulcast to be approximately the same as for broadcasting with only one of the two generations.

Following this idea in the third of the papers, a technical solution is described with this concept. The research shows how it is possible to share different elements in DTV services, and at the same time introduces techniques to aggressively compress video streams. The results show that this approach is valid and can be applied today in digital terrestrial television networks in production. It is therefore a viable solution to unlock the transition to newer technologies in some of the territories where this transition has become a problem. Consequently, the research presented in this thesis can be considered a current and viable line of research. It is also an open research area, as the argument that it is positive to consider the inter-generational compatibility in digital TV seems to be conclusive. The next chapter will argue this question by outlining potential new lines and ideas that can be included in future generations to minimize the problem of upcoming migrations.

The following publications were published in connection with the results of this thesis and are reproduced in full afterwards (listed in the order they are presented):

- Soto, Daniel. "Digital Television Backward Compatibility Based on Mixed Simulcast using Independent Scalable Video Coding."[3] *SMPTE Motion Imaging Journal* 125.9 (2016): 42-56.
- Soto, Daniel. "Digital TV simulcast with shared audio streams."[4] *2015 IEEE 5th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*. IEEE, 2015.
- Soto, Daniel. "Aggressive joint compression for DTV simulcast."[5] *Journal of Digital Media & Policy* (pending final publication). Intellect, 2020.

[3] https://ieeexplore.ieee.org/abstract/document/7803449
[4] https://ieeexplore.ieee.org/abstract/document/7391272
[5] https://www.ingentaconnect.com/content/intellect/jdmp/2020/00000011/00000002/art00004

[Blank page]

2.1 Paper 1 – Digital Television Backward Compatibility Based on Mixed Simulcast using Independent Scalable Video Coding (2016)

Scalable compression is not a new research area. However, the work published in this paper approaches this technique from a completely new perspective. It is a scalable but completely independent compression. This novel technique is based on performing scalable compression outside the encoder. In other words, by using standard encoders without any support for scalability, it is then possible to perform scalable compression with a certain gain in efficiency. This technique has been called by the author as *Independent Scalable Compression*.

The scalable decomposition can be done in several ways. The three possible ones are: spatial, temporal and in quality. And although it would be possible to approach any of the three with the independent compression technique, the study carried out in this work focuses exclusively on spatial scalability. In this regard, it should be noted that there are currently many studies on advanced techniques for optimizing spatial scalability, but this work is still novel in that it specifies a unique way to perform the coding/decoding process without having to modify the codecs used. The basic concept is the same as that used with standard scalable compression: to decompose the sequence into two streams: the base stream and the extended one. However, this technique analyzes the necessary steps to be able to compress each of the streams independently, so that the original sequence can be reconstructed without adding noise and obtaining some improvement in performance.

However, making such scalable decomposition without further action does not generate a valid reconstructed signal. When using lossy codecs, the direct application of the scalable model only introduces noise. This is why this paper proposes and analyzes a technique that really works. The key to this technique is based on two essential considerations: 1) to minimize or control the error introduced in the base stream; and 2) to preserve the efficiency of the compression in the extended stream. Based on this, the technique describes how to achieve both objectives in such a way that the reconstructed signal has an acceptable quality. And it applies the results to the specific case of the MPEG-2 codec for an SD base stream and the H.264 codec for an HD extended stream.

So far, this technique has not been applied to any particular product. However, the idea can be applied to new standards such as H.265 and others to increase the efficiency of UHD and UHDV services. This is possible because it does not require the scalable extensions that the industry does not include in current equipment. Therefore, it is still an open line of research where it is still possible to find new types of processing and decomposition that could increase the efficiency of the independent scalable compression.

This paper was published in[6]:

[6] Original result files are mirrored in:
https://dsaupf.github.io/publications/isvc2014/repo.html OR
https://drive.google.com/drive/folders/190HFrxdla8mXH_fzakVQBZ24i5Zqxjx-?usp=sharing OR
https://drive.google.com/drive/folders/1a8al8J80iBizlQ_kPlNWfEmU17LdQ14X?usp=sharing .

- Title: Digital Television Backward Compatibility Based on Mixed Simulcast using Independent Scalable Video Coding
- Journal: SMPTE Motion Imaging Journal (Volume: 125 , Issue: 9)
- Page(s): 42 - 56
- DOI: 10.5594/JMI.2016.2619460 [7]
- Publisher: SMPTE
- Date of Publication: Nov.-Dec. 2016
- Print ISSN: 1545-0279
- Electronic ISSN: 2160-2492

On the following pages it is reproduced in its original format.

[7] https://doi.org/10.5594/JMI.2016.2619460

Digital Television Backward Compatibility Based on Mixed Simulcast using Independent Scalable Video Coding

By Daniel Soto

Abstract

A significant problem in digital television broadcasting is the introduction of new coding techniques when new formats arrive. If the new codec is not fully compatible with the previous one, then backward compatibility is lost. Typically, when incompatible video coding technologies are used, the solution is to begin a simulcast model. This paper discusses one technique to circumvent this problem. The technique is based on the idea of a mixed simulcast. It is based on the concept of spatial scalability without sharing information between different encoding algorithms. This paper explains the application of this technique to the combination of an MPEG-2 video codec with SD signals and an H.264/AVC video codec with HD signals. While the focus here is on the introduction of HD in a DVB-T network without losing compatibility with legacy SD devices, the method described can also address the transition from HD to UHD.

Keywords

Backward compatibility, HD, independent scalable video coding, simulcast, UHD

Introduction

The switch to digital TV has several advantages over analog TV. In addition to the enhanced quality of audio and video, the digital broadcast allows more programs, multiple audio tracks, and supplementary services linked to the audio-visual content.[1] However, backward compatibility has been a significant problem in some early implementations of digital TV, unlike our experiences with improvements in analog TV such as the introduction of color, stereo sound, closed captions, and the like.[2–6]

This paper introduces an innovative solution to the problem of backward compatibility in digital TV based on the concept of scalable compression. A new method of independent scalability is proposed to achieve a mixed simulcast. Rather than the typical simulcast, where the two streams are fully independent, the new method creates a mixed simulcast, where one stream is based on the other to obtain better efficiency. It differs from other scalable solutions in the way the streams are compressed. Our solution relies on the use of two independent encoders that do not share internal data, so the streams are completely independent at the encoding level. The first encoder generates an SD stream from the original HD signal, which is fully compatible with legacy receivers. The second, using a new advanced codec, produces a stream that excludes information already included in the SD stream. Any legacy device can decode the SD stream, whereas new devices can reconstruct the original HD signal using both streams.

Rather than the typical simulcast, where the two streams are fully independent, the new method creates a mixed simulcast, where one stream is based on the other to obtain better efficiency.

While this paper focuses on a move forward from MPEG-2 576i@50 signals to H.264 1080i@50, the method described is equally applicable to transitions from HD to UHD and beyond. The study explains implementation of the proposed method and includes evaluation tests to verify its effectiveness. It describes a new algorithm to calculate the enhanced stream while maintaining the quality of the reconstruction.

Related Work

The idea of using multiple streams for video coding is not new. Different studies have explored this option,

Digital Object Identifier 10.5594/JMI.2016.2619460
Date of publication: XX November 2016

even in the analog domain.[7–10] Several techniques using multiple streams have been proposed since the beginning of MPEG digital compression.[11] Initially labeled *Hierarchical Coding* or *Multiresolution Coding*, they were forerunners of the modern concept of *Scalable Video Coding* (SVC onward). This concept divides a video sequence into a minimum of two-bit streams. The first, called *baseband layer*, is characterized by more limited properties (i.e., lower spatial or temporal resolution, or a lower bitrate). While the second, called *enhanced layer*, extends the properties of the previous stream. Thus, an additive relationship between the two layers is created in which it is possible to reconstruct a signal with advanced features combining the two streams. A receiver with limited performance will use only the baseband layer, while a more advanced one will use both layers. Each device receives a suitable signal according to their capabilities but without sending a specific complete signal to each one of them.

Different operating modes have been established within SVC: *Spatial Scalability* for resolution, *Temporal Scalability* for refresh rate, and *Quality Scalability* when fidelity is improved by either setting a precise Signal-to-Noise Ratio (SNR) or maintaining a particular bitrate. When the above scalability modes are applied together, it is possible to achieve almost any combination.

Despite the inherent complexity of SVC techniques, there have been implementations. Best known is the one presented by the *Joint Video Team* (JVT) as a scalable extension to the H.264/AVC standard. There have been numerous research articles related to scalability, but the number of commercial implementations is still very low, and none of them has been applied to digital television broadcasting yet. Nevertheless, there is interest in applying these techniques to digital TV broadcasting.[12]

Veteran video standards, such as MPEG-2, used in the first generation of digital TV already allow a scalable option within specific profiles.[13] The same approach was later used in standards H.263 and MPEG-4. However, these three attempts to standardize and promote the use of scalable techniques have not been commercially successful. For digital TV, a few research proposals exist[14,15] but have not materialized in practical implementations because of a lack of efficiency and increased complexity along with a lack of a real market need.

This is not the case with the H.264/AVC standard used in second generation digital TV. Here, a complete annex to H.264/AVC has been added.[16] It allows support for optional streams of temporal, spatial, and quality scalability, and describes how to maintain backward compatibility when the scalable extension is not supported, thus ensuring backward compatibility with existing devices. The publication of the *Joint Scalable Video Model* (JSVM)—*reference encoder*—software by the *Joint Video Team*[17] has been the basis of numerous subsequent studies. However, the number of implementations is still limited, but there has been recent renewed interest with hybrid TV solutions, such as Catch-Up services and VoD. For example, there are some efforts to combine the MPEG consortium's *Dynamic Adaptive Streaming Protocol over HTTP* (DASH)[18] with the scalable extensions of H.264/AVC.[19]

Other works have attempted to better exploit the potential of scalable coding. Some of them[20,21] discuss new methods to optimize the process of layering. They have explored the mathematical component of spatial scalability to reduce or eliminate redundant information between different layers. So instead of using the traditional pyramidal spatial decomposition,[22] these studies explore solutions employing mathematical minimizations of the components between layers. This results in less redundant data on extended layers, so the overall compression performance is increased.

In addition, *Hierarchical Modulation*[23] is very interesting for digital television. This allows splitting the spectrum of the digital signal—the physically transmitted wave—in two different layers. This way, standards such as DVB-T have provided support for more than one encoding scheme in the same multiplex. Two channels are defined using hierarchical modulation: the *Low-Priority channel* and the *High-Priority channel*, each with its own coding and modulation schemes. Both signals can be retrieved independently. It is possible to adjust the transmission parameters in accordance with the needs of each layer.

Coalescence of both scalable video coding and hierarchical transmission techniques allows transmission of the baseband video layer via a robust transmission layer and the enhanced video layer via a higher bitrate layer, thus ensuring basic reception under certain minimal conditions and under more favorable circumstances, reception with greater resolution and/or quality.

However, as of now, there are no systems—among the first and second generation—using the hierarchical mode. Probably, future video coding advances in new generations of digital TV will allow their application of both techniques together.

Fundamentals of Scalable Video Coding Without Sharing Information

Concept of Independent Scalable Video Coding

To achieve backward compatibility in digital television broadcasts, the concept of spatial scalability is proposed. Test results for this alternative model are included in Experimental Results. Next in this section, the fundamentals of this alternative model are presented.

Basic principles of scalability in video encoding have been extensively studied. For more detailed explanations, see Ohm.[11] **Figure 1** shows the general model of scalable coding using T layers. This is the model applied in MPEG-2, H.263, and MPEG-4 standards

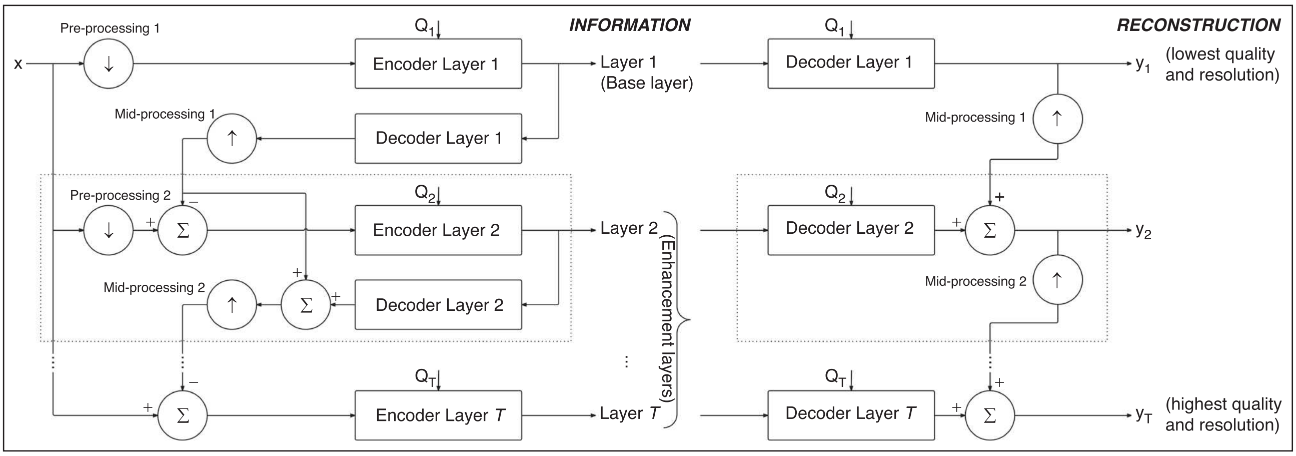

FIGURE 1. Principle of scalable coding using T layers, as described by Ohm.11

for scalability extensions. Complexity and efficiency were the reasons scalable profiles of these standards were underutilized. However, with the emergence of the H.264/SVC standard, new techniques were introduced into the scalable model, making it possible to reduce the complexity while increasing performance. For a more detailed explanation, see Schwarz *et al.*[16]

The common approach has been to incorporate the scalable model inside the encoder, that is, changing the encoding and decoding processes by adding extensions for scalable signals. This approach maximizes the advantages of scalability and improves efficiency. However, it prevents use with legacy decoders.

A different approach is to use the scalability model outside the encoder. This can be done without modifying the internal behavior of coding algorithms. Using this approach, it is possible to have a completely independent model that can be applied even when using different types of coding. This advantage comes at the cost of efficiency, since it is not able to use the codec's internal data from one layer to complete the encoding of another layer. It actually equals the model using full images. In fact, if the smallest processing unit is the full frame, then the scalability model outside the encoder corresponds to a scalability model of type *interlayer prediction with two-loop structure* (**Fig. 2**).

As has already been discussed, all models prior to the H.264/SVC standard were mainly oriented toward this model and did not provide a good compromise between efficiency and complexity.[24]

However, the approximation of the scalable model outside the encoder is useful under certain circumstances. Among them is the ability to create a stream that is an independent combination of two different codecs (SD and HD in this case) that is a necessary and sufficient condition to obtain backward compatibility.

To implement the proposed encoding model, we start by applying the classical Gaussian pyramid

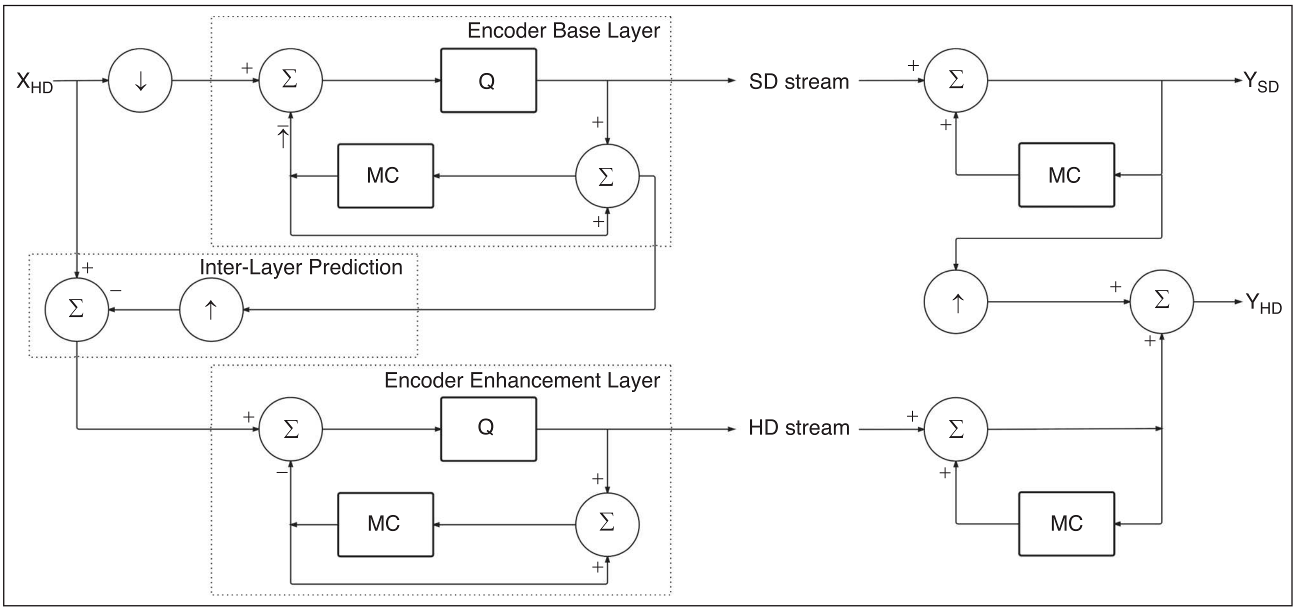

FIGURE 2. Structure of the hybrid coder with interlayer prediction and two-loop scalability.

decomposition[22] to individual HD frames or fields in case of interlaced images. The result is an SD version with lower resolution, obtained from filtering and decimating the original image, and an HD difference image with the same resolution of the original one, computed as a disparity image.

The main problem for spatial scalable compression outside the encoder is its possible negative efficiency. The use of an over-complete decomposition means that energy distributed between the two layers increases, penalizing the performance. Additionally, the lossy compression of current hybrid video compressors—like MPEG-2 and H.264/AVC—presents serious problems because the errors introduced in the baseband layer influence the extended layer. In a fully independent model, the compressor of the enhanced layer must compress not only the information in the differential signal but also the errors introduced in the base layer. Consider the following: In an image pyramid, the decomposition follows the expression

$$e = S_H - b \tag{1}$$

where S_H is the original image (with HD resolution), b is the resized image of the baseband layer (upsampled from SD to HD), and e is the component of the enhanced layer (in HD). Therefore, the ideal reconstruction process follows the expression

$$R = b + e \tag{2}$$

where R is the reconstructed signal, and b and e are the received streams (baseband and enhanced, respectively). But when lossy compression is applied, the reconstruction process obtained actually follows the expression

$$R' = (b + E_b) + (e + E_e) \tag{3}$$

where E_b and E_e denote the errors introduced by each compressor into its corresponding layer.

In this case, R and R' could be quite different according to the errors introduced. Because lossy compressors are based on factors such as consistency of the image, degree of human perception, and similar factors, when such compressors are used independently, it is not possible to find a direct correlation between E_b and E_e. What usually happens is that one type of error amplifies the other. On this basis, if no adjustments are applied, the signal R' will degrade heavily regarding R. However, with independent compression, such direct adjustment is not possible since the internal data of one compressor is not shared with the other.

Nonetheless, the distorting effect that the lossy compression on the baseband layer has over the enhanced layer can be countered by limiting the influence of the error E_b in the calculation of the signal e to ensure that signal R' is sufficiently close to the ideal signal R. It is also possible to restrict the direct effect of the error E_e over the signal R' by maintaining the internal integrity of the signal e. However, only by applying deterministic techniques, it is possible to maintain the independence of the compression algorithms. Therefore, we present a naive approach that relies on the addition of two new processes to the scalable computation. The application of these two new processes can improve the result of the reconstruction without sharing any internal data between the different compression algorithms.

The first process uses a filtered version of the baseband layer. Instead of directly using the decoded image b to compute the component e of the enhanced layer, the image b is preprocessed before calculating the difference. The objective of this preprocessing is to reduce the impact of the error E_b without compromising the information in the signal. The second process involves the use of a distinct version of the difference signal as the enhanced layer. Therefore, a simple mapping function of the signal from e to e' is proposed for this process. This mapping function must be invertible and also robust against lossy compression. In this way, it is possible to minimize the error E_e introduced by the encoder of the enhanced layer. The expected outcome is that E_b and E_e errors will be smaller, thereby increasing efficiency of the signal R'. Different ways to implement these additional processes are presented below.

Filtering the Baseband Layer

The objective of this process is to minimize the error introduced by the compression algorithm in the low-resolution layer. It is based on a simple deterministic decomposition, which relies on the nature of hybrid lossy compressors. Given that S_L is the original signal in low resolution and S_L' is that same signal compressed, then the following expression is applicable:

$$S_L = S'_L - E_b \tag{4}$$

where E_b is again the error introduced by the lossy compression in the baseband layer.

Considering that artifacts caused by lossy compression in a video signal particularly affect higher frequencies, this expression would be fulfilled

$$L(S_L) \approx L(S'_L) \tag{5}$$

when $L()$ is a low-pass filter, because $L(E_b) \rightarrow 0$. However, although it is not strictly true that all errors in a lossy compression affect only high-frequency components, it is true that use of a filter of this type will match the rule $E_b >> L(E_b)$.

Based on the above, the proposed method consists in filtering the decoded signal from the baseband layer before using it for calculating the enhanced layer. However, because it is necessary to keep independent operation between the compressors of both layers, the type of filtering to be applied must be deterministic. This means that no internal data of the compression process can be shared for only in this way will the two signals be fully independent. This allows the receiver to work with each signal in an autonomous way without relying on additional information, so it can decode each layer separately and then compute the reconstruction.

There are some issues regarding the particular type of filter to apply. If a very aggressive low-pass filter is used, there will be less noise, but the information given to the enhanced layer will be minimal. This implies that the pyramidal decomposition was very inefficient. If the filter is too inaccurate, then the noise level will be increased. That will negatively impact the compression process of the enhanced layer. Based on the outcomes of the tests performed, a filtering technique that has been successful for this purpose is presented below.

In a pyramidal image decomposition, much of the information in the difference component of two successive layers is concentrated at the edges of the objects. Since what is intended is to limit the negative effects of noise generated by the lossy compression in the low-resolution layer, it may be useful to apply this simple rule: employ the low-pass filter only in areas where there are no edges. This ensures that the noise is discriminated selectively, while a deterministic function is applied. This can be done without sharing internal data between encoders. It is sufficient to find the edges of objects in the decoded image. Whether artifacts appear around the objects edges or not, the enhanced layer always contains information for that part of the image. Therefore, the work of the enhanced layer compressor does not vary much because it was concentrated on that part of the image anyway. The purpose of this technique is to remove the noise present only in areas where there are no edges. It is computationally simple, does not depend on external information, and takes place intraframe.

For the computation of the enhanced layer discriminating the noise in the baseband layer, the following procedure is suggested. First, compute the upscaling of the decoded baseband image—the filtering is conducted on the image with noise—to match the resolution of the layer with high resolution. Second, perform the function of detecting the edges of objects over each component (Y,U,V) of the image. The result is then stored as a binary mask indicating whether an edge is found. Third, apply the low-pass filter to the image. Finally, combine the result of the low-pass filter with the rescaled image using the mask. The result is the component that will be used for calculating the image difference. The signal of the enhanced layer is finally computed by subtracting the original image from this calculated component. **Figure 3** shows an example of the use of this algorithm for the calculation of the enhanced layer.

Employing the above process, the high-resolution image used as the enhanced layer benefits in two ways. One, it has little noise in areas where there are no object edges. Two, it is visually consistent with what would have been expected if the original, uncompressed low-resolution image was used to calculate the image. Thus, the negative effects on errors introduced by lossy compressors are avoided for both layers.

Calculating the Enhanced Layer

As our objective is to compress baseband and enhanced layers with different and independent codecs, the problem to be solved is minimization of noise introduced by the compressor in the enhanced layer. As it has already been discussed, it is necessary to adapt the signal information to the characteristics of the compressor. This is achieved by maintaining the internal

a

b

c

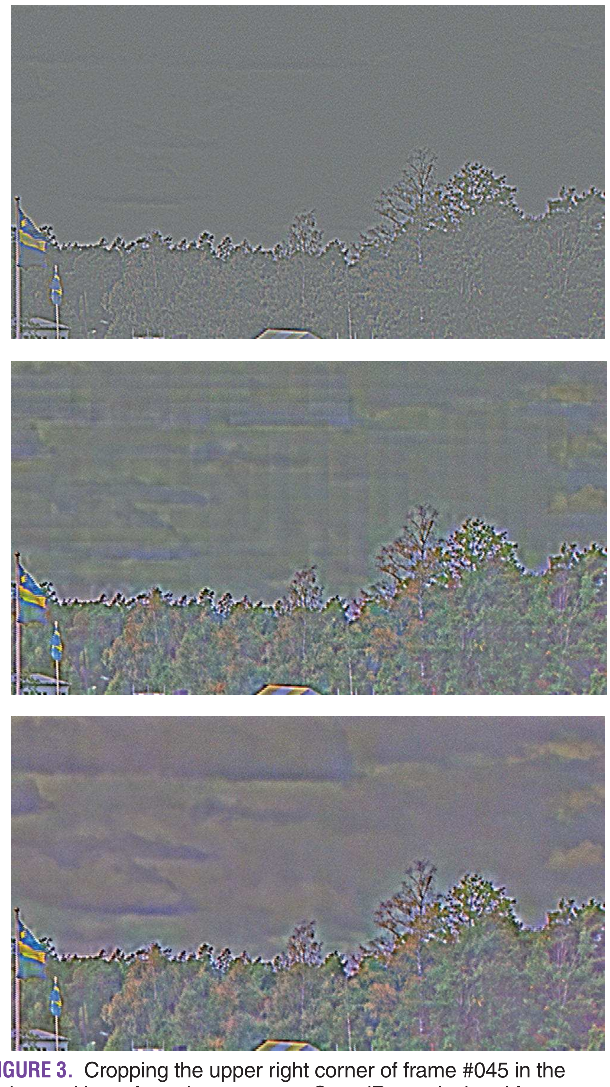

FIGURE 3. Cropping the upper right corner of frame #045 in the enhanced layer from the sequence CrowdRun calculated from: (a) original uncompressed SD image; (b) decoded SD image without filtering; (c) decoded SD image with filtering.

consistency of the image. This is the right way to keep the error introduced by the lossy compression within reasonable levels. Consequently, the exact function used to calculate the enhanced component becomes very relevant. Improving the calculation of such component is discussed below.

A classic function used in models of spatial scalability is the difference function. This function is used in a layering division with bilinear rescaling to compute the values of adjacent layers. Determining these values involves calculation of the difference between the original picture and the image with lower-resolution upscaled. This decomposition is known as spatial Gaussian pyramid. The enhanced component follows the expression

$$dH = S_H - rS_L \quad (6)$$

where dH is the image difference, S_H is the original image in high resolution, and rS_L is the upscaling of the low-resolution image S_L—the image S_L is obtained by downscaling the original image S_H. As already indicated, the decomposition calculated using this function does not remove all redundant information between the layers. However, this residual redundancy helps maintain the internal consistency of the image, which is good for the compressor of the enhanced layer. Besides, this function is deterministic and easy to calculate and therefore valid for use in the scalable model with independent encoders.

When using the difference to calculate the enhanced layer, the compressor must meet two requirements. The first is that the error introduced by the compression of the baseband layer is not amplified by the compression of the enhanced layer. The second is that the artifacts introduced by the compressor of the enhanced layer do not ruin the outcome of the reconstruction. Using this kind of compression, the motion compensation function loses efficiency when images vary considerably from one frame to another. The redundancy present in the pyramidal decomposition helps to keep the internal integrity of the images, thus contributing positively to the motion compensation function. Tests performed during this study indicate that suitable results can be obtained using the difference function and doing a slight requantization of the outcome.

The specific requantization we propose is a smooth mapping function. When the difference function is used, the pixel level values are in the range [−127, +127], which is transferred to the normalized range [0, 255]. It is easy to check that these values are often close to the center (value 127) of a Gaussian function. Also, the values will be closer to the center when the error in the baseband is lower. Based on this, a simple mapping function is proposed so that only values below a certain threshold are requantized. Such requantization may be a simple round to a single value within different intervals. With this mapping, it is possible to keep the high-frequency information provided by the enhanced layer while limiting the errors introduced by the baseband compressor. The result is improved compression of the enhanced layer, which results in a smaller error in the reconstruction. The downside of such mapping is that it will add a fixed error to values that are far from the values in the baseband layer. This is why the empirical results suggest using a smooth and nonaggressive mapping function.

It is also important that this mapping function is easily invertible and error resistant because lossy compressors can make the values calculated in the receiver become very different than what is expected. It is important that the compression artifacts present in the enhanced layer not generate anomalous values in the reconstruction. Using functions with transposed values should be avoided; rather, smoothing functions are more helpful. Empirical results suggest a better outcome with this type of function, especially for values that are far from the center of the Gaussian.

Architecture of Independent Scalable Video Coding for Retro-Compatibility

This section describes how to implement the spatial scalable model outside the encoder. A functional technical model is presented. The proposed approach avoids making changes to existing components; therefore, it is not necessary to change the video codecs already present in both generations and only add some extensions. This makes it much easier to implement the proposed model.

The aim of the new extensions is to allow a transition period in which the backward compatibility is guaranteed. Thus, instead of doing a full simulcast, it will be possible to perform a mixed simulcast—based on a scalability model—which is more efficient. These scalable extensions will be unnecessary after the transition period. When the legacy codec is discontinued, the new codec will be the only one in use, and the scalable extension can be disabled because the proposed technology does not limit future uses. It does not interfere with new developments but ensures backward compatibility as long as necessary.

Furthermore, the proposed solution does not excessively increase complexity. Although the proposed system uses the two-loop scalable model with full processing of two streams, this is not a real problem because second generation receivers already include support for both the legacy and new codecs. In that sense, the common solution is to implement two different decoders in the same device. The reality is that many implementations are already able to decode two streams in parallel. From this point, the overhead of the new solution is minimal because the only new task is the reconstruction of the HD signal combining the two streams. Therefore, the complexity of the solution is determined only by the cost of the operations to be performed for scalable reconstruction, which is fairly simple, and even less complex than other proposals.[25]

Model Design

The basic architecture of scalable backward compatibility is based on transmitting two video streams simultaneously. The encoding of these two streams is performed separately, and thus the scalability is applied outside the encoder. The first, *baseband stream,* carries only the low-resolution signal (SD) and is compressed to be fully compatible with the legacy standard. The second, *enhanced stream,* contains the information necessary to enable the receiver to reconstruct the HD signal. This second video stream is encoded independently by using the more advanced codec based on the new standard to be introduced on the market. However, a constraint between the HD and SD signals is required. No changes in refresh rate or interlaced mode are allowed; only changes in the vertical and horizontal resolution between the two versions are permitted. Any changes in width and height are allowed without imposing any restrictions as in some other studies.[26]

In the receiver, this technology will be commercially viable only when existing components are reused and complexity is kept to a minimum. Only when the cost is lower than the benefit obtained will the model be practical. We propose a design for the receiver in which only four new components are added: (1) decoding the second video stream, (2) upsampling the *baseband stream*, (3) mixing both video streams, and (4) synchronization control.

No changes are required for the audio because digital TV standards allow more than one audio stream in the same program. Therefore, it is necessary to add only additional audio tracks. All receivers can process the audio if at least one audio track was found encoded with a compatible codec. No troubles are generated when different audio codecs are mixed in the same program since all systems natively allow simulcast audio.

The same rule of simplicity should be applied to the transmitter. However, because only one transmitter exists in the production chain, the complexity of the encoder does not add significantly to the cost, so the design of the transmitter architecture is less restrictive. The transmitter design will keep the elements necessary to compress the signal with the legacy codec and add only new processes to generate the extended signal: (1) downscaling of the original HD signal to SD used in legacy systems, (2) the decoding of the *baseband stream*, to provide SD data for the enhanced stream calculation, (3) the calculation of the *enhanced stream*, (4) the encoding of the *enhanced stream* using the new codec, and (5) the resynchronization and multiplexing of the two streams.

Decoder Architecture

The architecture of the legacy SD decoder is not changed in any way. The new extended architecture of the HD decoder based on spatial scalability can be seen in **Fig. 4**. As shown, the design of this decoder is fairly simple. Using a receiver that supports two parallel streams requires only four new components.

The components included in the receiver are as follows:

- *SD Decoder*: This is the standard decoder of the low-resolution signal. It is exactly the same as used in any legacy receiver; however, it is important that no additional post-processing operations are performed, because any operation for image enhancement needs to be performed only at display level. This is necessary to maintain consistency between the data calculated by the sender and receiver.
- *HD Decoder*: This is the decoder of the high-resolution enhanced stream. Its implementation should follow the standard of the new codec. No modifications are required with respect to an ordinary decoder. However, as with the SD decoder, this component should not do any post-processing that is not described by the standard.
- *Upscaling*: This component is responsible for matching the resolution of the signals in SD and HD. It needs to take into account the characteristics of the formats used. For example, in a progressive signal, the rescaling must be applied at the frame level; while for an interlaced signal, it should be applied at field level. The algorithm used is free depending on the

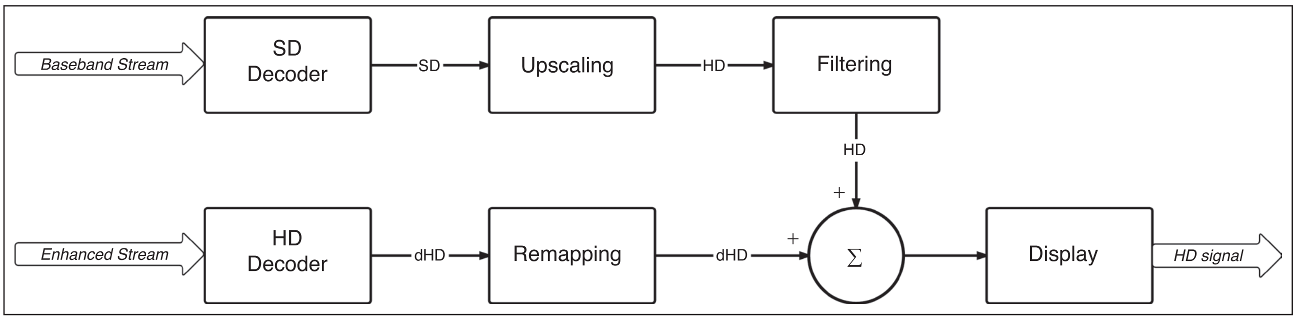

FIGURE 4. Architecture of independent scalable decoder.

implementation but must be the same in the encoder and decoder. This is mandatory because different algorithms generate different versions with different values, and the values must be identical on both ends for correct processing.

- *Filtering*: This component performs the task of filtering the image from the *baseband stream*. This filtering operation is applied to the image only after it has been *upscaled* to match the HD resolution. The algorithms used will vary but should always be the same in the encoder and decoder. Although there is no restriction regarding the complexity of this process, it is really necessary to use a deterministic process because no information is shared between codecs to maintain the independence between them.
- *Remapping*: If any requantization of the values in the *enhanced stream* was used during the encoding process, then this component must perform the inverse function. However, if this function is really simple, it could be done directly in the calculation of the reconstruction process.
- *Reconstruction*: The reconstruction process is usually the simple addition of the images of both layers. The data from the *baseband stream* after decoding, rescaling, and processing is used for the base layer. The signal of the *enhanced stream* uncompressed and restored (from the optional requantization) is used. Before applying this function, it is mandatory to check whether overflow values have arisen. This step is necessary because the artifacts caused by lossy compression can generate phantom values. Furthermore, any resynchronization process between both streams must take place at this time.
- *Display*: Finally, the reconstructed image is used as the source of the video to display. Here is where the post-processing algorithms can be executed to improve the image. Note that this does not include any internal filters related to the process of encoding/decoding but only the processing linked to the display procedure. Examples of these functions could be deinterlacing, antinoise filters, etc.

Encoder Architecture

The design of the video encoder is much more complex than the decoder. We describe the encoder after presenting the decoder because some components are the same. The architecture of the encoder is shown in **Fig. 5**.

The components included in the encoder are:

- *Downscaler*: The first task is to generate the low-resolution version from the original source in high resolution. After this point, the system will have two signals, one in high resolution (HD) and one in low resolution (SD). The quality of this component is critical to the ultimate quality of both signals but does not affect the process of the scalable compression. Therefore, it is possible to use any suitable downscaling algorithm. The recommendation is to use the best available system that preserves the highest quality signal.

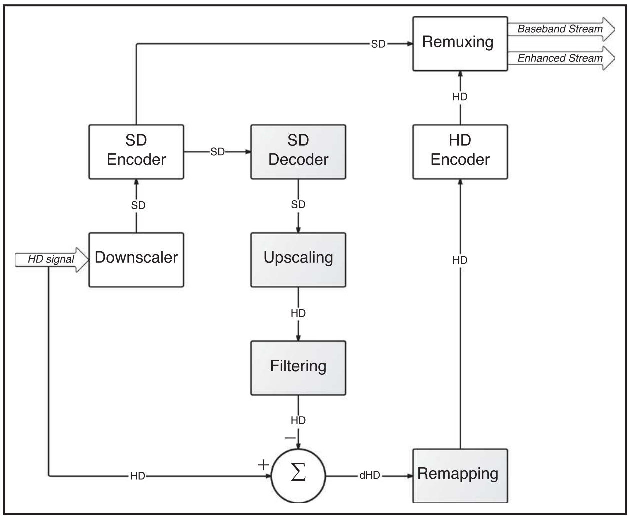

FIGURE 5. Architecture of independent scalable encoder.

- *SD Encoder*: The SD version of the signal feeds this component. Any encoder that maintains backward compatibility can be used. However, due to the negative impact of noise on the scalable compression, it is recommended that the parameters of the encoding process be set with values that can reduce compression artifacts. In particular, settings that preserve the low-frequency information and minimize the noise caused by the lossy compression are useful.
- *SD Decoder*: This component is the same as the one used in the receiver.
- *Upscaling*: This component is also the same as the one used in the receiver. The same algorithm and techniques should be used at both ends, as their output values need to be the same.
- *Filtering*: This component is also identical to that used in the receiver.
- *Difference*: This component calculates the difference between the signal of the base layer and the extended layer. It works in the domain of the high-resolution layer. Note that if the input signal is interlaced, then the function is applied at field level and not at frame level. Besides, this component is responsible for the synchronization of the streams because the base layer signal is always delayed in reference to the original signal.
- *Remapping*: This is the same component that exists in the receiver.
- *HD Encoder*: This component uses the new codec with which the *enhanced stream* is compressed. Its work is completely independent of the other components. However, as with the SD encoder, it is recommended that compression values be set so that the highest quality of the reconstructed signal is preserved. Filters should not be used if they will

degrade the signal. It is generally recommended not to activate any preprocessing of the signal at this point.

- *Remuxing*: This last component handles the remux of the two streams. The process must be done even if both signals use different transmission paths. This is required to maintain synchronization between the streams, which will facilitate the work of the receiver. If it is not done, there will be a long delay between the two signals, and it will be necessary to use a large buffer in the receiver. Taking into account the delays introduced by each component, the delay between the two signals can be very large. It is best to buffer this delay at the source to reduce complexity in the receiver.

Experimental Results

Several tests were performed to evaluate our proposal for maintaining backward compatibility by broadcasting the same content in SD and HD without having to perform a full simulcast. While the tests were geared to this specific scenario, the results can be extrapolated to a general model of a mixed simulcast based on spatial scalability with independent compression. The details of the tests and results are outlined below.

Scope of the Tests

These tests were performed to verify our proposal in as close to real market conditions as possible. Instead of using typical research tools, tools and commercial models that followed recommendations accepted by broadcasters—specifically those established by the EBU[27]—were used:

1) SD format: MPEG-2 576i@50
2) HD format: H.264 1080i@50
3) Encoder: FFmpeg (standard version)

It should be noted that the proposed model also applies to broadcasts in progressive format. However, since the natural transition from SD to HD keeps the interlaced format, only results related to interlaced formats are discussed. Also, the compressed MPEG-2 signal is always compressed as interlaced—whether it is originally interlaced or not. Results obtained by mixing interlaced and progressive formats are usually not suitable for spatial scalability.

The FFmpeg[28] encoder was chosen as the compression tool rather than the common *reference encoders* (JSVM software[17]) because, in current production environments, the encoders used to compress television programs have several levels of optimizations that should be taken into account during the tests. For example, the noise caused by lossy compression has a strong impact on the outcome of reconstruction. This is why we approximate the characteristics of real production systems as closely as possible. Therefore, the values used for the bitrate, PSNR, and others are those commonly used in commercial digital television broadcasts.

Synthetic Tests

The sequences used in the tests fall into two different groups. The first type was the *"SVT High Definition Multi Format Test Set"* provided by the EBU[i]. The second was a complete sequence of a soccer match provided by a broadcaster. The objective was to evaluate the behavior of both common research sequences and real television sequences. All sequences are in compliance with the SMPTE 274M format (using the common 1080i@50 mode) and the BT.709 color space and were converted to YUV 4:2:2 with 8bpp for the tests. **Table 1** shows the sequences used.

The *Soccer* sequence is the production signal of a soccer match summary. The original format is SMPTE VC-3 (aka Avid DN × HD intraframe compression) with a fixed bitrate of 120 Mbps, corresponding to the standard for HD signals in production environments. It was considered suitable for testing because the quality of the sequence is in accordance with current standards in the broadcast field. The sequence includes several scenes with graphics, camera changes, foreground images, wide shots, motion sequences, camera movements and zooms, and occassional slow-motion replays. It represents a significant and varied source of mainstream broadcast content.

All the operations performed in tests have been done using the AviSynth processing tool[29] Version 2.5.8. It is a versatile, open-source environment and allows programming very quickly in a Microsoft Windows environment. It includes support for plugins, scripts, and any file format. It also allows working at pixel level and supports fields and frames. Finally, it can be used transparently to the display tool, and in real time. This allowed multiple testing and easy changes on-the-fly.

To move the results obtained with the AviSynth processing tool to the FFmpeg encoding tool, we used MKV files. All sequences were stored as individual files using this container format, and the lossless compressor HuffYUV 4:2:2 has been used to reduce the space inside the files. All sequences compressed in MPEG-2 and H.264 are stored using the same MKV containers while keeping their own video codec.

Test Parameters

Two sets of parameters were selected for these tests. The first set were those used for Video and compression, and the second were the values used for scalability operations. Due to the specific field of application, compression values were chosen first and then results using different values for scalability were explored.

[i] They can be obtained from the address "ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/". Specifically, the sequences used are files unprocessed and uncompressed in the 1080i25_CgrLevels_SINC_FILTER_SVTdec05_directory.

TABLE 1. SEQUENCES USED IN TESTS.

CrowdRun_1080i25

DucksTakeOff_1080i25

InToTree_1080i25

OldTown Cross_1080i25

ParkJoy_1080i25

Soccer_1080i25

For video and compression, the following values were selected:

- SD Streams:
  - Codec: MPEG-2 interlaced 4:2:0
  - Resolution: 720 × 576 (16:9 anamorphic aspect ratio)
  - Frame rate: 25 (50 fields per second)
  - Pre-filtering: none
  - Bitrate: 6000 kbps
- HD Streams:
  - Codec: H.264 profile Main@L4.0 4:2:0
  - Resolution: 1920 × 1080 (16:9 native aspect ratio)
  - Frame rate: 25 (50 fields per second)
  - Pre-filtering: none
  - Bitrate: 10 500 kbps for full simulcast
  - Bitrate: 5500 kbps for scalable layer

Using these values, two broadcast modes were compared. One was the full simulcast mode, where one SD and one HD version of the same program are transmitted simultaneously. The other was the mixed simulcast mode, with a normal SD program and another one using the scalable technology proposed in this paper. **Figure 6** shows these two transmission modes. On the left, the full simulcast mode used a bitrate of 6 Mbps for the SD channel in MPEG-2 and H.264 with a bitrate of 10.5 Mbps for HD. On the right, in the mixed simulcast, the bitrate for the SD channel in MPEG-2 remains at 6 Mbps, but the scalable stream for reconstructing the HD version used only 5.5 Mbps. Using the second mode, it would be possible to broadcast an HD channel while

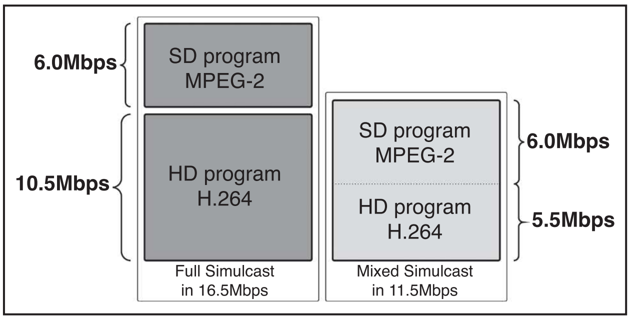

FIGURE 6. Complete simulcast vs. mixed simulcast.

maintaining backward compatibility with only a 10% increase of the overall bandwidth.

Once the compression parameters for the tests were decided, empirical analysis identified the following values for the scalable technique described in this paper:

- Baseband layer filtering function: 2D low-pass Laplacian filter, 3×3 kernel with coefficients defined as

$$C = \begin{bmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{bmatrix} \quad (7)$$

- Iterations of filtering function: four steps
- Edge detector function: Sobel filter
- Requantification of difference function: none

It should be noted that not applying requantization in the difference function does not invalidate the results. The previous tests revealed that the use of this requantization may improve the outcomes. However, for the

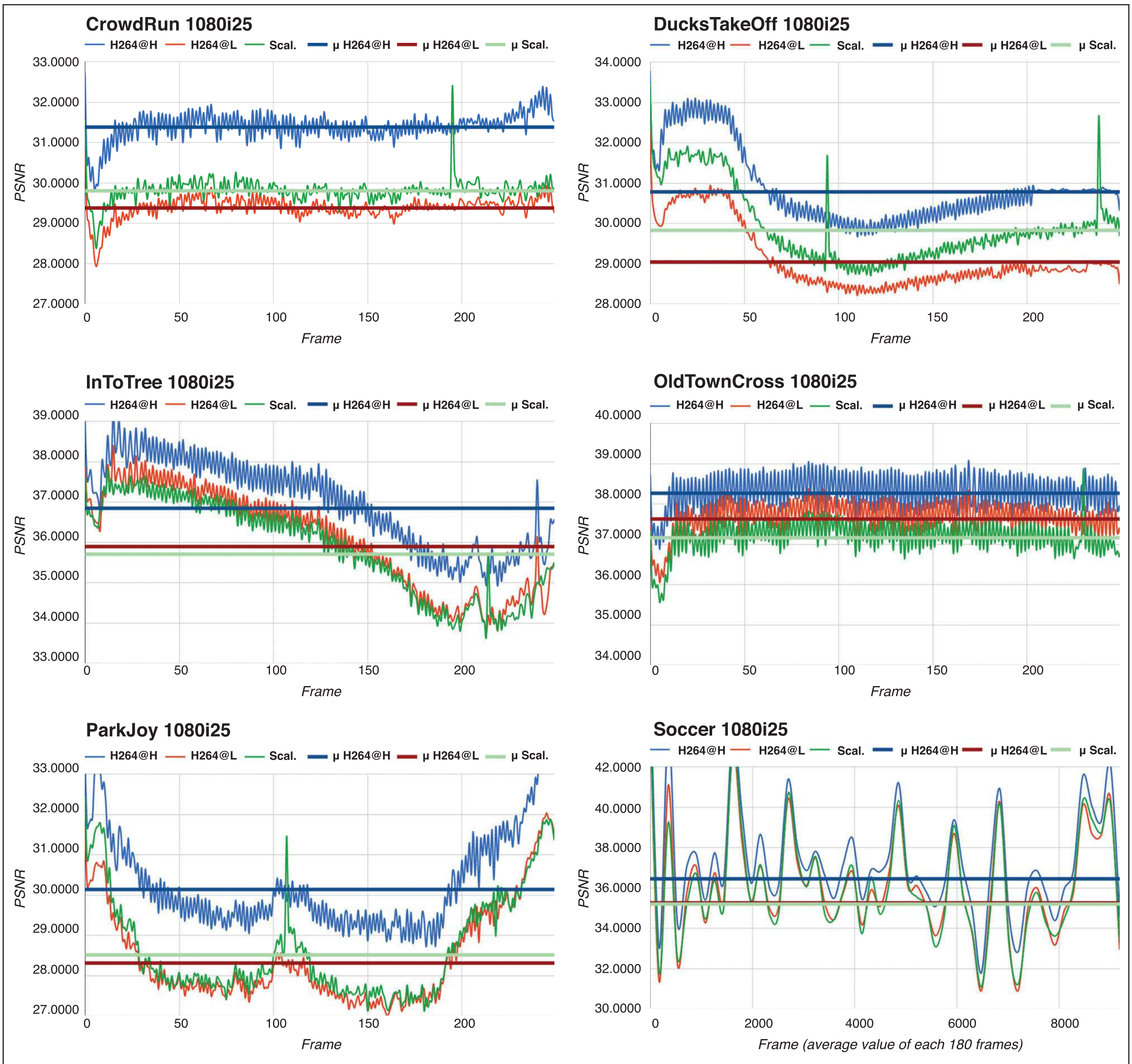

FIGURE 7. Graphs showing PSNR comparison for each sequence.

TABLE 2. Summary of results.

	CrowdRun	DucksTakeOff	InToTree	OldTownCross	ParkJoy	Soccer
Duration	10'	10''	10''	10''	10''	6' 10''
H.264 @10.5Mpbs Overall PSNR	31.38	30.78	36.85	38.27	30.14	36.45
H.264 @5.5Mpbs Overall PSNR	29.38	29.04	35.89	37.63	28.31	35.26
Reconstructed from MPEG-2 @6Mbps + H.264 @5.5Mbps Overall PSNR	29.80	29.83	35.71	37.16	28.51	35.20
Deviation	+0.42 dB	+0.79 dB	−0.18 dB	−0.47 dB	+0.20 dB	−0.06 dB

specific use that is pursued in this paper, we have obtained better results disabling this process. So the results presented here summarize only tests done without requantization.

For the iterative filtering of the baseband, the objective was to remove as much noise as possible for high frequencies. However, using lossy compression, artifacts can be easily generated not only in the high-frequency domain. The tests performed suggested that the repeated application of the filter function with a smaller window gave better results. Thus, the same filtering function was applied in multiple passes. The process was: rescale the image from the baseband layer to the size of the high-resolution image, apply the filtering function i times over the same image, calculate the mask with the edges of the objects on the *unfiltered* rescaled image, apply the resulting mask on the *filtered* rescaled image, and finally use the result to calculate the difference image. The result is that the image was filtered only where there are no edges, leaving the edges untouched.

Performance Analysis

The results of the performance tests are presented in **Fig.** 7. In the graphs, the *H264@H* signal represents the HD sequence compressed with H.264 codec using a bitrate of 10.5 Mbps. The *H264@L* signal represents the sequence in HD compressed with H.264 codec using a bitrate of 5.5 Mbps. And the *Scalable* signal represents the reconstructed sequence from an SD layer compressed in MPEG-2 with a bitrate of 6 Mbps plus an HD extended layer compressed with H.264 codec using a bitrate of 5.5 Mbps. In the case of the *Soccer* sequence graph, PSNR values are grouped on the average of 180 consecutive frames because of the length of the sequence.

From the above results, a number of conclusions can be drawn. See **Table 2** for a quantitative summary of the results. The first conclusion is that, with few exceptions, the values of PSNR of the *Scalable* version are between values of *H264@H* and *H264@L* versions. This means that, in general, the compression using the technique presented in this paper offers higher efficiency than a full signal compressed with H.264 codec using the same low bitrate, although without equating, in either case, the efficiency obtained by compressing the full signal with standard bitrate, that is, with a bitrate equal to the sum of the bitrate of the two scalable streams.

A second conclusion is that, in some cases, our scalable compression technique is not better than full simulcast. In the case of the *OldTownCross* sequence, the *Scalable* signal is ~0.5 dB, worse than *H264@L* signal. And in the case of the *InToTree* sequence, the *Scalable* signal is slightly below, with ~0.2 dB less. In both cases, this result is due to the low complexity of the signal in comparison with the bitrate used because of the high PSNR values in signals compressed with H.264 codec. Therefore, the worst performance of the scalable compression results from the low efficiency achieved by MPEG-2 compression in the baseband layer. So the error introduced into the SD signal impairs the outcome of the reconstructed signal because with high PSNR values the signal difference is slightly more difficult to compress. The conclusion from this result is that from a certain threshold the scalable technique presented here offers no advantages in absolute levels. The benefits of the scalable model are realized only below that threshold. Therefore, when applying this technique, it will be necessary to adjust all the parameters of the compressor, especially the control values corresponding to the MPEG-2 codec, because this signal needs to suffer a minimal degradation in relation to noise and artifacts.

A third conclusion is that the behavior of this compression technique is grouped into three different sets. The first group—type 1—would be represented by sequences with medium complexity, such as *CrowdRun* and *DucksTakeOff*. Compressing these sequences with the scalable model proposed here improves the overall efficiency. So for type 1 sequences, the performance of the scalable technique is as expected. In the second group—type 2—there are sequences with low complexity but much high-frequency information. In this group, *InToTree* and *OldTownCross* are examples that have large amounts of information but fluid movements and continuous camera traveling, which are easy to compress. In these cases, the higher efficiency of H.264

codec with respect to MPEG-2 codec is evident. So the compression of the entire HD signal is very efficient, whereas the compression of the SD signal has low-quality levels. Therefore, for type 2 sequences, the scalable technique has worse outcomes. Finally, there is the group—type 3—of high-complexity sequences as *ParkJoy* in which the scalable compression is at least as efficient as the full compression. In this latter case, the result is expected because there is no degradation in the reconstruction; the values obtained are at the same level as those obtained with the nonscalable compression.

The results of the *Soccer* sequence deserve a separate consideration because it is a real sequence with multiple scenes of varying complexity. The first result is that, in general, the PSNR values are very similar to those obtained by the compression with the H.264 codec at the same reduced bitrate. In fact, in the total sequence, the difference between the signals *H264@L* and *Scalable* is only 0.06 dB for the first. However, in subjective tests, a greater presence of visible artifacts has been observed in the *H264@L* signal, which is not seen in the *Scalable* signal. This is because many of the scenes of this sequence are type 2, where there is a little complexity, and PSNR values are high. Note that the average PSNR of the entire sequence is above 35 dB in all cases. Therefore, the increased efficiency of the scalable compression is manifested in only certain parts of the sequence. However, there is a visible improvement in subjective quality using the scalable technique. The overall conclusion of the analysis of this sequence is that the technique presented here is valid for the production signals described in the case study.

A final remarkable aspect of the analytical results is the presence of peaks in PSNR values. These rare peaks around +2 dB in the *Scalable* signal appear in individual frames occasionally. They seem to be related to the decoupling of the compressors, so there are times when some overlap of information occurs when a completely independent compression is used. One of these particular cases is when both encoders coincide on using an intraframe and assign it a high priority. In that specific case, the image quality is appreciably increased over adjacent frames. However, in such particular cases, there is an absence of visible artifacts, and therefore it is an irrelevant effect imperceptible to the viewer.

In addition to tests for the analytical calculation of PSNR, we have also tested the subjective perception of the sequences. These tests were conducted to evaluate the presence of annoying artifacts introduced by the different types of compression. The tests consisted of high-resolution sequences displayed in different modes: full-screen playback, comparative playback, search over individual frames, etc. The aim was to look for the presence of visible artifacts and evaluate their impact on the result.

Overall, the results of subjective analysis points to two conclusions. The first is that the quality of the model proposed in this paper improves results in most cases. In the rest, the quality does not degrade. The second conclusion is that artifacts introduced by the scalable compression are less noticeable than without it. That means that observable artifacts in compressed sequences with the *H264@L* mode are much more noticeable than those present in the sequences compressed with *Scalable* mode. For example, in the *CrowdRun* sequence, in the case of the *H264@L* signal, there is an annoying *floating effect* in the sky textures. Usually, temporal artifacts are better masked by the *Scalable* mode than the *H264@L* mode.

To summarize from the subjective point of view, the independent scalable solution is as acceptable as the full compression with H.264 codec at typical bitrate (10.5 Mbps in tests). But the advantage of the scalable solution proposed in this paper over the full simulcast is that it maintains backward compatibility with MPEG-2 decoders, which is the objective of this work.

Conclusion

The results of the tests performed prove that the model described in this paper is valid for application in the field of digital television. It is particularly suitable for maintaining backward compatibility with legacy devices during a transition from the SD format to the HD format. It allows a smoother transition by keeping emissions in both formats simultaneously during the entire transition period. This is possible due to the improved efficiency of the mixed simulcast compared to a full simulcast. The scenarios that best take advantage of this technology would be territories where digital terrestrial television is deployed in SD format with the MPEG-2 codec that now begin to introduce HD formats with the H.264 codec.

As for the best way to apply the technology, the following recommendations are suggested. First, the SD stream compressed with the old codec should maintain the highest possible quality. In the case of overcompression of the signal, artifacts introduced in SD have an adverse effect on the scalable compression. Thus, it may be desirable to reduce the effective resolution of such signal by using low-pass filters. With the application of low-pass prefiltering, it is possible to reduce artifacts resulting from lossy compression, at the cost of adding a slight blur effect to the image. Therefore, even by using less bandwidth for the SD stream, the image quality at low frequencies will remain. This ensures that the scalable system will not be penalized and will be able to recover all the high-frequency information during reconstruction.

Second, the signal quality of the enhanced layer must be preserved. For the reconstruction process to be effective, it is essential to ensure a minimum quality of the signal. If this threshold is not respected, the compressor of the enhanced layer cannot maintain the internal consistency of the frames. The result will be an excessive degradation of the reconstructed signal with the appearance of annoying artifacts. Considering, however, the increased compressor efficiency of the enhanced layer, the basic recommendation is to maintain a fixed bitrate for the base layer and use a variable bitrate for the enhanced layer, guaranteeing in the latter case a minimum quality. This solution could easily be implemented using statistical multiplexing of all the programs transmitted on the same multiplex.

Third, the compression parameters need to be readjusted over time during the transition period. A remarkable result drawn from the tests is that it is possible to adjust the settings to give more quality to one layer or another. Thus, at the beginning of the transition period, while the number of HD receivers is low, the SD stream has more bitrate and therefore more quality allocated. This gives more priority to the baseband layer and ensures fewer artifacts at a cost of less crispy HD images; whereas, at the end of the transition period, when the number of SD receivers is less, the quality of the SD stream can be reduced. This flexibility would allow a much smoother transition, and users can gradually migrate to the new technology. However, under any circumstances, the quality of the SD stream cannot be reduced to the point that it increases artifacts, because that would affect the quality of the reconstruction of the HD signal. The outcome of this progressive adjustment is that consumers with legacy devices are encouraged to make a more rapid replacement of their receivers as the quality of legacy emissions is falling.

Finally, the scalable model will not be necessary beyond the transition period. After broadcasts with the legacy codec switch off, the scalable model makes no sense, since the standard compression of the entire HD signal with the new codec has greater efficiency because the new receivers are fully compatible with the new codec. The scalability model is just an added module, so it is possible to simply disable the scalable reconstruction when it is no longer necessary. The technique does not in any way limit future developments.

Future Work

The efficiency of the spatial scalable model with independent compression outside the encoder could be improved in several ways. One possibility might be looking for new functions for the calculation of the enhanced layer. If it were possible to reduce the redundant information of the pyramidal decomposition while maintaining the internal consistency of the enhanced layer, then it would be possible to improve the results of reconstruction. Moreover, it would also be advantageous to find new configurations that allow compressing the enhanced layer with a higher efficiency. The high flexibility of H.264 codec allows the encoder to make decisions based on the signal content. So far, this way has not been explored for the scalable model outside the encoder; however, by including specific algorithms in the H.264 encoder, it may be possible to improve the results by maintaining the independence of encoding processes between layers. So as long as the encoders are not sharing data, the model remains valid.

Regardless of the possible technical improvements, an alternative pathway for future work would be to apply the same model using other video codecs. The recent emergence of a successor for the H.264/AVC standard, the new codec H.265/HEVC,[30] raises the possibility of using the same technique to maintain backward compatibility between these two codecs. This would allow a smooth migration from the second generation digital television (HD) to third generation (UHD). Possibly, the new levels of efficiency achieved with the H.265/HEVC codec will allow an even higher efficiency for the scalable model with independent compression. This would certainly contribute to faster adoption of the UHD format in the distribution of digital television signals.

Lastly, although the viability of the solution presented in this paper focuses on a very specific scenario, it would be interesting to explore solutions in other areas. An interesting scenario might be systems for distributing mixed signals. Systems of this type may be transmitting signals in low resolution using multicast to all receivers, while some receivers receive a signal

TABLE 3. Filenames of test sequences.

TYPE	FILENAME	CODEC
uncompressed original HD Version	*_HFYU-yuv422p.AVI	HuffYUV 4:2:2
MPEG-2 @6.0Mpbs SD Version	*.avs.SD-Video.MPG2.MKV	MPEG-2
H.264 @10.5Mpbs 'H264@High'	*.avs.HD-Video.H264.MKV	H.264
H.264 @5.5Mpbs 'H264@Low'	*.avs.HDlow-Video.H264.MKV	H.264
H.264 @5.5Mpbs 'Scalable'	*.avs.dHD-Video.H264.MKV	H.264
Reconstructed from MPEG-2@6Mbps + H.264@5.5Mbps	*.avs.dHDrec-Video.H264.MKV	HuffYUV 4:2:2

[ii] They Accessible from the address "http://www.dtic.upf.edu/~dsoto/publications/isvc2014/sequences/"

with higher resolution in a unicast mode. Given this case, if there were a technique with scalable compression, it would allow a more efficient use of the capabilities of each device. It would therefore be possible to generate different versions of the enhanced layer, which are sent in unicast mode to each device with HD support, while all of them receive the same baseband layer in multicast. Besides, it would not even be necessary for all enhanced receivers to use the same codec, or stream, but each one could use its own, thus providing greater flexibility to this solution.

Appendix: Test Sequences

The outcomes of the simulations described in this paper can be downloaded[ii] to verify the results. All test sequences are offered freely, without restricted permissions. Each sequence has six different versions, as listed in **Table 3**.

References

1. A. Brown and R. G. Picard (eds.), *Digital Terrestrial Television in Europe*, Mahwah, NJ: Lawrence Erlbaum, 2003.
2. M. S. Goldman, "Migrating from MPEG-2 Video to MPEG-4 AVC," *SMPTE Conf.*, vol. 10, 2010.
3. Y. Wu, et al., "Overview of Digital Television Development Worldwide," *Proc. IEEE,* 94(1):8–21, 2006.
4. F. Rancy, E. Zilles, and J-J. Guitot, "Transition to Digital TV and Digital Dividend," *IEEE 10th Int. Conf. Telecommun. Modern Satellite Cable Broadcast. Serv. (TELSIKS),* vol. 1, 2011.
5. M. T. G. Leiva, and M. Starks, "Digital Switchover Across the Globe: The Emergence of Complex Regional Patterns," *Media, Cult. Soc.,* 31(5):787–806, 2009.
6. K. McCann, et al., "Technical Evolution of the DTT Platform: An Independent Report by ZetaCast, Commissioned by Ofcom," 2012, http://stakeholders.ofcom.org.uk/binaries/consultations/uhf-strategy/zetacast.pdf. Accessed Oct.2014.
7. A. G. Toth, M. Tsinberg, and C. W. Rhodes, "NTSC Compatible High Definition Television Emission System," *IEEE Trans. Consum. Electron.,* 34(1):40–7, 1988.
8. K. M. Uz, M. Vetterli, and D. J. LeGall, "Interpolative Multiresolution Coding of Advance Television With Compatible Subchannels," *IEEE Trans. Circuits Syst. Video Technol.,* 1(1):86–99, 1991.
9. T. Naveen, et al., "Comments on Interpolative Multiresolution Coding of Advanced Television With Compatible Subchannels"(and reply and additional comments), *IEEE Trans. Circuits Syst. Video Technol.,* 2(1):95–100, 1992.
10. M. O. Polley, S. J. Wee, and W. F. Schreiber, "Hybrid Channel Coding for Multiresolution HDTV Terrestrial Broadcasting," *Proc. IEEE Int. Conf. Image Process., ICIP,* vol. 1, pp. 243–7, 1994.
11. J-R. Ohm, "Advances in Scalable Video Coding," *Proc. IEEE,* 93(1): 42–56, 2005.
12. A. Kouadio, M. Clare, L. Noblet, and V. Bottreau, "SCV - A Highly-Scalable Version of H.264/AVC" *EBU Technical Review*, Q2: 35–56, 2008.
13. S. Okubo, "Requirements for High Quality Video Coding Standards," *Signal Process. Image Commun.* 4:141–51, 1992.
14. T. Chiang and D. Anastassiou, "Hierarchical Coding of Digital Television," *IEEE Commun. Mag.*, 32(5):38–45, 1994.
15. S. Aign and K. Fazel, "Temporal and Spatial Error Concealment Techniques for Hierarchical MPEG-2 Video Codec," *IEEE Inter. Conf. Commun., ICC'95 Seattle, "Gateway to Globalization,"* vol. 3, 1995.
16. H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H. 264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.,* 17(9):1103–20, 2007.
17. JSVM reference software (2006), https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/image-video-coding/research-topics/svc-extension-of-h264avc/jsvm-reference-software.html. Accessed Oct. 2014.
18. ISO/IEC FCD 23001-6, "Part 6: Dynamic Adaptive Streaming Over HTTP (DASH)," *MPEG Requirements Group*, Jan. 2011.
19. Y. Sánchez et al., "Scalable Video Coding Based DASH for Efficient Usage of Network Resources," *Third W3C Web and TV Workshop*, 2011.
20. M. Flierl and P. Vandergheynst, "An Improved Pyramid for Spatially Scalable Video Coding," *IEEE Int. Conf. Image Process. ICIP.* vol. 2, 2005.
21. W. Yang, G. Rath, and C. Guillemot, "Scalable Video Coding With Interlayer Signal Decorrelation Techniques," *EURASIP J. Appl. Signal Process.*, 1: 216, 2007.
22. J. M. Ogden et al., "Pyramid-Based Computer Graphics," *RCA Eng.,* 30(5): 4–15, 1985.
23. S. O'Leary, "Hierarchical Transmission and COFDM Systems," *IEEE Trans. Broadcast,* 43(2):166–74, 1997.
24. C. A. Segall and G. J. Sullivan, "Spatial Scalability Within the H. 264/AVC Scalable Video Coding Extension," *IEEE Trans. Circuits Syst. Video Technol.,* 17(9):1121–35, 2007.
25. H. Schwarz et al., "Constrained Inter-Layer Prediction for Single-Loop Decoding in Spatial Scalability," *IEEE Int. Conf. Image Process., ICIP*, vol. 2, 2005.
26. G. Marquant et al., "Extended Spatial Scalability for NonDyadic Video Formats: From SDTV to HDTV," *Visual Commun. Image Process.*, International Society for Optics and Photonics, 2005.
27. EBU-TECH 3328, *Current Status of HDTV Delivery Technology*, Geneva, May 2008.
28. FFmpeg software tool (2000–2014), http://ffmpeg.org/. Accessed Oct. 2014.
29. AviSynth framework (2005), http://avisynth.nl/. Accessed Oct. 2014.
30. G. J. Sullivan et al., "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.,* 22(12):1649–68, 2012.

About the Author

Daniel Soto was born in Barcelona, Spain, in 1973. He received the BS and the MS in computer science and engineering from the University of Balearic Islands (UIB), Spain, in 1994 and 1999, respectively, and the MPhil in information and communication technologies from the Universitat Pompeu Fabra (UPF), Barcelona, Spain, in 2004. Since 2000, he has been a researcher and part-time professor with UPF. His research interests focus on digital TV and streaming media, Internet of Things, content delivery networks and software-defined networking, and virtual infrastructure technologies.

[Blank page]

2.2 Paper 2 – Digital TV simulcast with shared audio streams (2015)

Advances in audio compression have been significant in the last decade. Although traditionally considered legacy codecs, such as MP2 (MPEG-1 Audio Layer II) seem obsolete, the quality they provide is very good. However, they are not as good in terms of compression efficiency. This is why the most recent codecs achieve much higher compression rates while maintaining very similar quality. However, because the quality of the audio streams seems to have reached a peak, recent efforts seem to have concentrated on multi-channel and vector audio streams. And in many cases, this new type of encodings relies on the use of some sort of more or less scalable technique.

However, these advances in audio compression have not provided any significant advantage in the field of Digital Television. Newer codecs have simply been adopted, which basically do more or less the same job with slightly less bandwidth. For example, there is currently no research underway to apply scalability between multilingual audio tracks, which could be considered a major advance. Between the lack of work on audio optimization specifically for the field of DTV, and leaving aside the area of developing new codecs, the work presented in this paper offers a novel technique.

The technique developed allows the sharing of audio streams between different services. Although the idea seems simple, technically it is not easy to implement. None of the digital television standards directly address this functionality. Fortunately, however, they do not make it impossible at all. The problem, therefore, is to find a viable way to achieve the objective.

To achieve this objective, the paper focuses on identifying and defining a technique capable of generating transport streams in which the audio streams are shared between different DTV services. But not only to do that, but also to be able to perform that task without interfering with the encoding process. In other words, by simply adding an additional task after the encoder, it is possible to synchronize and share the audio streams between different services of the same MPTS. This represents a substantial advantage in the distribution of the simulcast services, since the efficiency obtained is equivalent to the total number of audio streams multiplied by the number of services in the simulcast Thus, the bandwidth saving can represent between 5 and 10% of the total space occupied by a multiplex.

This technique remains unique today, and also represents the basis of the work presented in the following paper. Basically what it demonstrates is that it is possible to share the same PES stream between different separate services within the same MPTS without disturbing the DTV devices.

This paper was published in[8]:

- Title: Digital TV simulcast with shared audio streams

[8] Original result files are mirrored in:
https://dsaupf.github.io/publications/ssas2015/repo.html OR
https://drive.google.com/drive/folders/1Q1pMwAhZhgIK3dvlvlC83jzNy4fs8MBB?usp=sharing OR
https://drive.google.com/drive/folders/1YQ2MTEIRB2Kk2hz5zQp-9_YUKy216nIJ?usp=sharing .

- Proceedings: 2015 IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)
- Date of Conference: 6-9 Sept. 2015
- DOI: 10.1109/ICCE-Berlin.2015.7391272 [9]
- Publisher: IEEE
- Conference Location: Berlin, Germany
- Electronic ISBN: 978-1-4799-8748-1
- DVD ISBN: 978-1-4799-8747-4

On the following pages it is reproduced in its original format.

[9] https://doi.org/10.1109/ICCE-Berlin.2015.7391272

Digital TV Simulcast with Shared Audio Streams

Daniel Soto
University Pompeu Fabra (UPF)
Barcelona, Spain
daniel.soto@upf.edu

***Abstract*—A common problem in the field of digital television is the transition to new encoding formats. The incompatibility between new and old codecs requires a simulcast over a transition period. In some scenarios —particularly when using terrestrial broadcasts— this simulcast represents a significant waste of bandwidth and raises many problems. This article explores the possibility of share audio streams between programs within the same simulcast and presents a technique to reduce the bandwidth required using the ASC (Audio Sharing Compactor). This solution has been validated through the implementation of a demonstrator capable of processing a Transport Stream in real-time before send it to the modulator transmitter.**

***Index Terms*— DTV, simulcast, audio stream sharing, multiple video streams sharing audio.**

I. Introduction

In the evolution of digital television systems, lack of backward compatibility is a common problem. Although digital TV systems have some flexibility, which allows for changes over time to adapt its operation to the needs of the moment, this flexibility has limits. That is why often it is chosen to keep broadcasting of duplicate content during a transition period. However, the simulcast over a long transition period can be a problem in some circumstances, usually because the available bandwidth is limited.

The idea presented in this article, instead of focusing on optimizing the video stream, is to share different audio streams between programs within the same simulcast. Thereby, if each program contains one or more audio tracks, the same streams shall be used by all versions of the same program. This ensures that when the simulcast is generated, only the video stream is duplicated, saving all the bandwidth consumed by repeated audio streams.

It is important to note that the solution presented here is completely transparent, as well as easy to implement within the production chain. Generally, the process involves the processing of the Transport Stream where the programs are multiplexed to filtering out duplicated audio tracks. This task is carried out just before the input to the modulator transmitter. Therefore, the synchronization of streams and modification of SI tables is performed after compression and packaging, thereby achieving a transparent solution. Fig. 1 shows a block diagram of this solution

This article describes the basis for completing this task of removing duplicate audio tracks while maintaining compatibility with all type of receptors. Specifically, it describes how to perform this optimization using the DVB-T standard for simulcasting a program in SD MPEG-2 and HD H.264; although it is possible to easily adapt this technique to other digital television standards.

II. Motivation

In some areas the distribution of digital television signals is mostly done using terrestrial broadcasts. In these regions is not uncommon to find troubles when migrating from one generation to another. One scenario in which this issue has reached huge proportions is the territories in which the transition to digital television was performed using the first generation (i.e. using the MPEG-2 codec and SD format). In places where this transition was completed some time ago, the migration to second generation is on the way (i.e. using the H.264 codec and HD format). The problem now is the very limited available bandwidth for terrestrial broadcasts.

Ignoring the option to go forward to the third generation, the simulcast is the only alternative. But the lack of sufficient space to deliver the content using the new technology is sometimes critical. Then the option may be to decrease the quality of content, broadcasting two versions using the old and the new formats. But that does not make sense in emissions using the new codecs, as consumers expect improvements with the new technology. The consequence is that users are less attracted to purchase new equipment and this is not helping the renewal of the legacy devices. It is within this scenario where the present study provides a novel solution that optimizes the available bandwidth for simulcast, thereby allowing the successful migration to a new generation of digital television.

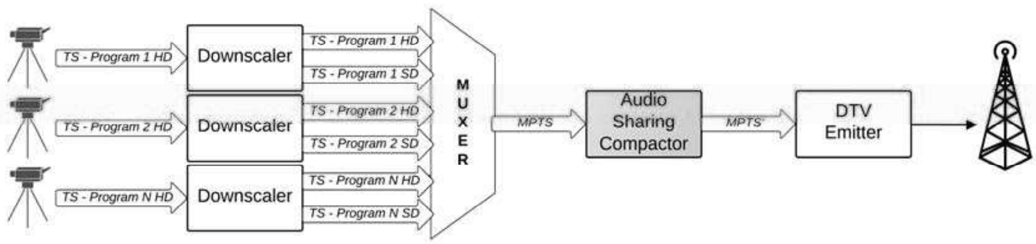

Fig. 1. Overview of the MPTS generation with shared audio streams

III. RELATED WORK

So far, there are very few studies dedicated to optimize the audio of a simulcast in digital television, and nearly all focus primarily on backward/forward compatibility with multichannel audio without making use of simulcast operation [1]. The reason seems to be the low interest until now for optimizing broadcasts with simulcast; as the migration to new digital television standards has been considered a necessity only recently. So the lack of solutions in this research line is more due to low interest from the industry than to potential technical difficulties in their application.

However, although the minimal work directly focused on the purpose of improving performance of simulcast broadcasts, that does not mean there are no solutions regarding the optimization of the audio signals on television. Besides the well-known advances in audio compression, such the HE-AAC [2] or E-AC3 [3] codecs included in the most recent generations of digital television, solutions that have helped to optimize the use of the audio tracks have been presented throughout history. For example, relating the analog television, audio technologies such as A2 Stereo (Zweiton) or digital NICAM [4] allowed a program to have two different independent audio tracks (dual sound) using the stereo channels. This was not a complex implementation, but proved to be very useful, which is why all devices were prepared to work in this mode.

On the other hand, with the move to digital systems new gates were opened for significant progress in audio optimization. Thus within the standards of Digital Video —which mark the foundations of DTV— it is possible to find solutions with some similarity to the idea presented in this work. One may be the support for multiple video tracks. At the beginning of digital video systems only a single video source was supported (with several audio tracks). Later, the support for multiple video sources in the same signal was added, in order to provide compatibility with multi-camera systems. Thus the multiplexing of different synchronized streams was born, which could be of any nature —audio, video or anything else—. This concept was then directly incorporated into the MPEG-2 Transport Stream [5], which represents the basis of all current digital TV standards. In fact, there are no large technical differences between a Transport Stream multiplexed with multiple television programs, and one with multiple camera views with different audio tracks. That is precisely the main concept on which is based this work of sharing audio tracks between programs of the same simulcast.

IV. FUNDAMENTALS

There are different ways to achieve the goal that two or more programs in the same Transport Stream share the audio streams. The most straightforward solution is to assign to the encoder the task of generating the different versions of the same program and produce the multiplexing. The advantage of this solution is that the resulting Transport Stream meets with the constraints established by the digital television standard used from the start.

However, this direct solution presents difficulties: The first is the complexity of the compression. Since the encoder will generate more than one stream of the video, the work that it must complete is larger as well as more complex. The second difficulty is the timing synchronization during multiplexing. This is due to the restrictions on size of the buffers and the delay between audio and video packets. This work is complex and usually managed by a multiplexer responsible for combining various Transport Streams —of SPTS type, each with a complete program— into a single Transport Stream —of MPTS type with multiple programs— [6]. Therefore, leaving this task within the encoder makes it much difficult to implement it. Finally, the third difficulty is that none of the current digital TV standards explicitly supports sharing audio tracks between different programs, even though no standard forbids it [5 *(C.6)*]. Although doing this does not directly represent any technical problem, the specifications recommend not sharing streams between programs. This is recommended to avoid errors from applications handling MPTS streams.

Due to the above stated difficulties, addressing the problem in the opposite direction is possible to simplify the process. The use of this approach with the application of the solution at the end of the production chain is much more effective and has several advantages: First, encoders can be independent and use different codecs, as in regular simulcasts. Second, the multiplexing process remains unchanged; which is highly important as the statistical multiplexing is critical for improving the efficiency of the simulcast and get it to be effective [7,8]. And last, all possible incompatibilities with the modified Transport Stream are eliminated while performing the process at the end.

However, the practical implementation of the proposed solution is not simple. There are two main problems to be solved which are the stream synchronization and the signalling of programs. Only when these two problems can be resolved satisfactorily, will the system work successfully maintaining full compatibility with digital TV standards. Furthermore, the processing must be done in real-time, because this approach requires modifying the stream at the end of the production chain. However, the complexity of the required tasks is simple enough to do so without difficulty.

V. ARCHITECTURE

One solution for a system capable of generating a multi-program Transport Stream in which the audio tracks are shared is the ASC (Audio Sharing Compactor). The objective of this system is simple: remove the redundancy introduced by the multiple versions of the same audio tracks without changing the structure of the Transport Stream. Therefore the final result must be a new Transport Stream that keeps the same characteristics, but in which the audio streams are mapped to more than one program. This compacting process then reduces the bandwidth consumed by the Transport Stream. Fig. 2 shows the proposed architecture to implement this technique.

Regarding the details of the architecture, it must be said that it follows the pipeline model. At functional level, the filters only need to process the Transport Stream as single packets. That means it is not necessary to unpack the PES data for each of the pids. Having divided the process into multiple sub-elements has added some flexibility. This allows adapting the processing to the needs of each case without modifying the entire system. Each of the three elements in the architecture will now be described in detail.

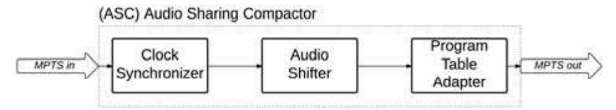

Fig. 2. Internal architecture of the (ASC) Audio Sharing Compactor

A. Clock Synchronizer

The purpose of this component is to synchronize the clock values of each of the programs that form the simulcast. Starting from a particular program the clock values are recalculated in the remaining programs in a way that all are synchronized with the first. Doing this operation makes sure that even if the programs are independent, all share the same time base. Timing should therefore be very precise, as timestamps must be indistinguishable in each of the programs. Although this task may at first seem complex, it is not at all. It will be necessary only to calculate differences between the timestamps in each program, and make the necessary changes as outlined below.

At low level, the task to be performed by this component is to rewrite PCR, PTS and DTS timestamps of audio and video streams. This process can be done by computing the variation of time bases of the programs, and using this value as a fixed offset to recalculate all timestamps of pids inside the same program. Therefore, a precise synchronization between all streams belonging to the same simulcast is achieved in a simple way. It should be noted that once the offset is calculated the synchronization will not be lost, because the source is the same, for it is a simulcast.

At functional level, this component only needs to process the Transport Stream as single packets. That means it is not necessary to unpack the PES data for each of the pids, as the PCR, PTS and DTS data are embedded in the packet headers, specifically within the optional field. Thus, each time a packet is received, comparing the pid value and timestamps is enough. This way, the internal structure of the Transport Stream is never modified, thus ensuring that synchronization remains unchanged.

The final result of this filtering technique is that the output Transport Stream is identical to the input one, with the only difference that secondary programs of the simulcast have advanced or delayed the timestamps. The outcome achieved is fully transparent, both at the Transport Stream level and at the processed programs level, because changes only affect timestamps of the same program. Of course, this filtering can be applied as many times as necessary, so it is not a problem to process more than one set of simulcasts in the same Transport Stream. Fig. 3 summarizes the outcome of this processing.

B. Audio Shifter

The purpose of this second component is to guarantee that audio streams meet the specifications about decoder buffers boundaries; as the multiplexer did not take into account the synchronization between the different programs of the simulcast when creating the MPTS stream. Therefore, packets from audio streams coming from other programs will probably not appear in the right time with respect to the video stream. That problem is not related to timestamps, but to the place where the packets are inside the multiplexed Transport Stream.

To avoid this problem, one possible solution is to change the position of the packets within the Transport Stream. But reordering packets of video streams involves recalculating all PCR timestamps. And this is not a simple task, as it is in fact equal to perform a new re-multiplexing. On the other hand, the audio tracks are much easier to move. The reason is that audio streams do not include PCR timestamps, only PTS marks, and consequently the packets can be advanced or delayed without modifying any timestamp. Also, as audio tracks are usually compressed using a constant bitrate (CBR) the audio packets can be relocated within the same PES audio stream without changing the structure of the Transport Stream.

The way in which audio packets must be relocated is based on the structure of each individual Transport Stream. As an example, consider a simulcast between an SD program compressed in MPEG-2, and another one in HD compressed with H.264, each one with one audio stream using the MPEG-1 Layer II codec. $PidV_{SD}$ and $PidA_{SD}$ are then the video and audio streams of SD program, and $PidV_{HD}$ and $PidA_{HD}$ the streams of the HD program. Then, from the point of view of the decoder, packets in the Transport Stream corresponding to such streams are ordered as follows based on equivalent PTS values. First are coming the packets from $PidA_{SD}$, because audio packets are usually slightly advanced to video. Secondly, the packets from the corresponding $PidV_{SD}$ appear —typically, the SD program has lower delay than the HD program due to its less complex compression; therefore the SD streams comes before the HD streams—. Then, in third place, with much more delay, the packets from $PidA_{HD}$ appear. Finally, the packets from $PidV_{HD}$ are found. Consequently, in an instant of time T, the PTS

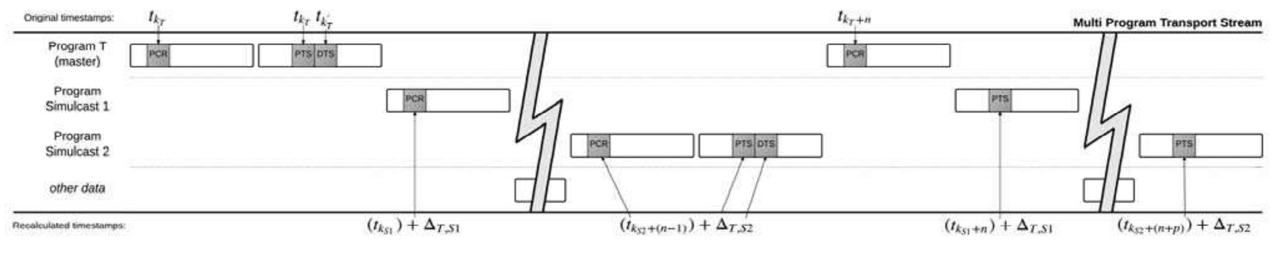

Fig. 3. Timestamps recalculation in subprograms of the same simulcast

values of the different streams follow the sequence $PidV_{HD}$ < $PidA_{HD}$ << $PidV_{SD}$ < $PidA_{SD}$. In this particular example, the problem is that the decoding buffer of the video stream from the SD program will overflow when playing coupled with the audio stream from the HD program. To solve this problem, a simple solution is to reallocate the audio packets from the HD program in a way that this sequence is satisfied: $PidV_{HD}$ << $PidA_{SD}$ < $PidA_{HD}$ < $PidV_{SD}$. With this rearrangement the audio packets from the HD program come earlier and, consequently, the SD video buffer will not overflow. However, this change involves that now the HD audio stream is advanced ahead with respect to the HD video stream. Usually this is not a problem because the size of the audio buffer —which is usually never filled at all— can easily compensate this delay in the video stream without any buffer overflow. In an analogous way, the same is true for the SD audio stream when playing coupled with the HD video stream. In this case, the audio packets arrive very similar in time to packets from the SD audio stream. The end result is that any of the four possible combinations can be played without buffer troubles. In Fig. 4 is illustrated this audio packets rearrangement over time inside the MPTS stream.

At technical level this component performs the function of a delayer. And this objective often consists in *advancing* the arrival of the audio stream during the decoding of the Transport Stream. This ensures that at any given moment in time, the packets of a particular stream inside the Transport Stream have higher PTS timestamps. To achieve this goal there is no other alternative but to delay the rest of the streams. This is because since the filtering works in real-time, it is not possible to manage packets that have not yet been received. Therefore, the only feasible solution is delaying all the other packets.

In addition to the discussed above, it is necessary to note that advancing one stream an arbitrary number of packets is not useful. This is because of the structure generated by the multiplexer. In order to fulfil with the restrictions about the size of the decoding buffers, the multiplexer selects very precisely the location of each packet inside the Transport Stream. Therefore, selecting carefully the number of packets to be advanced is important. An option to avoid problems is calculating the number of existing packets between two consecutive *Payload Unit Start Indicator* (PUSI) marks inside the PES stream. As audio streams are using a constant bitrate (CBR), this number does not change over time, and one PUSI mark appears every N packets of the stream. Consequently, it can be considered that all packets between those marks are in the same group. Therefore, replacing a set of N packets by another set from the same stream does not alter the structure of the Transport Stream, and thereby the transparency of the process is guaranteed.

C. Program Table Adapter

The objective of this component is to modify the PMT table of each program so that all programs of the same simulcast have the same audio streams. This can be done in various ways depending on the way in which audio streams were generated prior to the merging process. For example, if each program duplicates all audio tracks at encoding, then what needs to be done is changing the tables of secondary programs to point to the audio streams of the main program. On the other hand, if each program carries only one audio stream and what is wanted is that every program is able to use any of them, then what needs to be done is adding the remaining streams to each program.

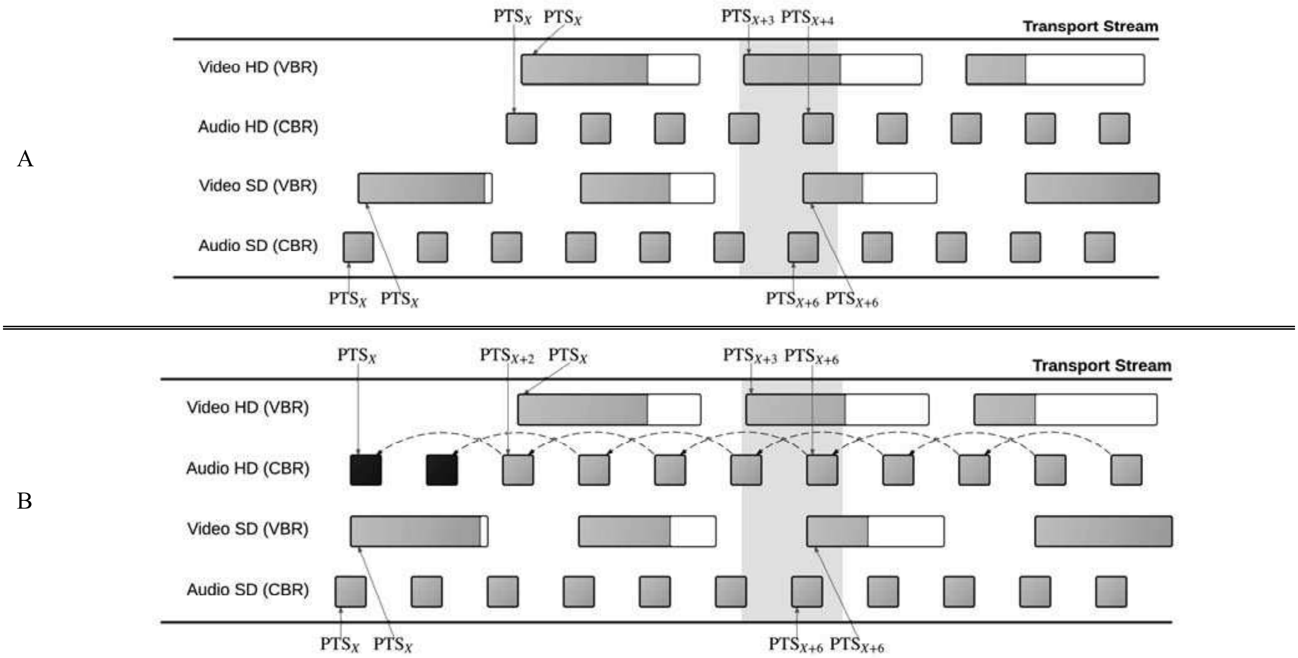

Fig. 4. Time position over the TS of audio and video PES packets from SD and HD programs in a simulcast: a) original position before reallocating; b) new position after time shift

At functional level, this process can be done easily by monitoring all PMT tables present in the Transport Stream and modifying them when necessary. Because these tables are repeated from time to time, but they do not change, the process is really simplified. Therefore, only when a change is detected in one of the tables pertaining to a program belonging to the simulcast, then, and only then, is when the table needs to be rewritten. From that moment on, every occurrence of this particular table only needs to be replaced with the rewritten version. Moreover, if the rewritten table size is the same then the structure of the Transport Stream does not even need to be modified.

At low level, is necessary to note that in the case that each program is encoded using all audio tracks in the source, it will also be necessary to remove redundant data from the audio streams discarded. To do this without changing the structure of the Transport Stream, replacing with padding all packets member of the streams that must be removed is sufficient. These stuffing packets can then be reused prior to broadcasting the signal to reinsert new information in the Transport Stream.

VI. Experimental tests

In order to validate the technique disclosed in this paper a test system that follows the architecture presented has been implemented. This experimental environment is capable of processing in real-time one Transport Stream to complete all the tasks described in Section V. The process feeds a full Transport Stream —ready to be sent to a distribution network of digital TV—, and modifies it in a way that the resulting output is the same Transport Stream but with programs in simulcast sharing the audio tracks.

This demo implementation runs on Linux X86 platform, although the code is easily portable to other platforms. The computational requirements are minimal, considering that the processing of the full Transport Stream is done in real-time. In order to facilitate the development, the package *OpenCaster* [9] has been used as a basis. The source code of some tools in this package has proved useful for implementing the skeleton of new tools. Therefore, without much effort it was possible to implement a complete system capable of performing the processing required for the testing. The binaries of the resulting tools are available for external review[1].

Among the various tests performed, a demonstrator is included, which receives the full Transport Stream of commercial DVB-T broadcasts using a standard tuner. This testing includes the tune of frequencies that include programs in simulcast (SD MPEG-2 and HD H.264). The demo consists of processing the received Transport Stream for removing any duplicate audio streams. The result is identical to the original Transport Stream, but with shared audio pids belonging to the simulcast programs. The whole process is done in real-time and the outcome can be reproduced in any standard digital TV receiver without any trouble. In each program of the simulcast, different audio tracks are selectable and all of them maintain the synchronization during the playback smoothly. Some captures of these testing Transport Stream are available for review[2].

VII. Discussion and conclusion

Although different digital TV standards do not directly provide support for sharing audio over different programs in a simulcast, it can be done while maintaining compatibility. This allows a saving in the bandwidth required for simulcasting programs, as it is no longer necessary to replicate the audio streams. Problems regarding synchronization and signalling can be satisfactorily resolved using the technique described in this paper. The saved space may be used to improve the quality of the programs within the same frequency. This allows the simulcast to be more efficient, and enables more programs to be broadcasted in simulcast. In turn, this will allow completing the transition from one format to another more quickly and smoothly. Particularly in the case of the territories where SD MPEG-2 programs using DVB-T are being broadcasted, increasing the number of HD H.264 programs using the proposed technique will be possible. This should be translated into a faster acceptance of new formats, which will lead to a more rapid renewal of the receiving devices. This eventually involves discontinuing the use of legacy standards of digital TV earlier. It is therefore a major improvement to carry on.

References

[1] Breebaart, Jeroen, et al. "MPEG spatial audio coding/MPEG surround: Overview and current status." *Audio Engineering Society Convention 119*. Audio Engineering Society, 2005.

[2] Wolters, Martin, et al. "A closer look into MPEG-4 High Efficiency AAC." *Audio Engineering Society Convention 115*. Audio Engineering Society, 2003.

[3] Andersen, Robert L., et al. "Introduction to Dolby digital plus, an enhancement to the Dolby digital coding system." *Audio Engineering Society Convention 117*. Audio Engineering Society, 2004.

[4] ETSI/EBU: ETSI EN 300 163, V1.2.1 — *Television systems; NICAM 728: transmission of two-channel digital sound with terrestrial television systems B,G,H,I,K1 and L*; Mar. 1998.

[5] Rec, I. T. U. T. "H. 222.0| ISO/IEC 13818-1." *Information Technology—Generic Coding of Moving Pictures and Associated Audio Information: Systems, ITU-T/ISO/IEC* (2007).

[6] Rao, K. R., Z. Bojković, and D. Milovanović. "Joint video encoding in DVB/ATSC multi-program transmission: MPEG-2 vs. H. 264/AVC bitrate control." *Telecommunication in Modern Satellite, Cable, and Broadcasting Services, 2009. TELSIKS'09. 9th International Conference on*. IEEE, 2009

[7] Wang, Limin, and Andre Vincent. "Bit allocation and constraints for joint coding of multiple video programs." *Circuits and Systems for Video Technology, IEEE Transactions on* 9.6 (1999): 949-959.

[8] He, Zhihai, and Dapeng Oliver Wu. "Linear rate control and optimum statistical multiplexing for H. 264 video broadcast." *Multimedia, IEEE Transactions on* 10.7 (2008): 1237-1249.

[9] Avalpa Digital Engineering s.r.l. OpenCaster (2008). <http://www.avalpa.com/the-key-values/15-free-software/33-opencaster>. Accessed online on June-2015

[1] Accessible at "http://www.dtic.upf.edu/~dsoto/publications/ssas2015/tools/"

[2] Accessible at "http://www.dtic.upf.edu/~dsoto/publications/ssas2015/tests/"

[Blank page]

2.2 Paper 3 – Aggressive Joint Compression for DTV Simulcast (2020)

The distribution of television services in simulcast is currently quite high. The reason is that we are immersed in several transition processes. Therefore, it is not uncommon to find services that are currently distributed in up to three different versions: SD, HD and UHD. The reason behind this is that in some cases it is not yet possible to completely abandon SD services, and on the other hand the demand for UHD content is growing.

In the context of such scenario with multiple versions of the same content, the work presented in this third paper identifies a very interesting and technically viable solution. It involves nothing more and nothing less than sharing all possible elements between the different versions of a simulcast. This inevitably implies that services are distributed together, which makes sense in many scenarios. On the other hand, the paper also discusses how to achieve a complete simulcast of all services using the same bandwidth. This option can be very beneficial in areas where bandwidth is very limited, and where it is not possible to expand the distribution network to include all programs in the simulcast. This is the case, for example, with the Digital Terrestrial Television networks, on which the study is focused.

Furthermore, this published research is entirely new in terms of providing a viable solution to the problem of the inter-generational migration in DTV. So far there are no technical papers focusing on simplifying or facilitating this migration. However, it is becoming more and more evident that transition times between generations are becoming longer. Therefore, the research provided in this line of work is completely up to date.

But in addition, the same paper also suggests the opportunity to explore similar solutions for the transition from HD to UHD. Although there is still some time for this migration to take place, in fact it will not be necessary to wait too long. There are already regular terrestrial broadcasts of UHD in certain territories, and more and more are being made using other distribution systems (satellite and cable, in addition to IPTV). Therefore, it is urgent to find solutions to minimize the impact of the future switch to UHD. In that sense, this work can be a good starting point.

This paper was published in[10]:

- Title: Aggressive Joint Compression for DTV Simulcast
- Journal: Journal of Digital Media & Policy (previous published as Journal of Digital Television) (Volume: 11 , Number 2 , 1 June 2020)
- Page(s): 151-174(24)
- DOI: 10.1386/JDMP.00021.1 [11]
- Publisher: Intellect
- Print ISSN: 2040-4182

[10] Original result files are mirrored in:
https://dsaupf.github.io/publications/jdmp2019/repo.html OR
https://drive.google.com/drive/folders/1XwLxMmWSwYNcHAn-Hz_dM5VCGbXw1Tsr?usp=sharing
OR https://drive.google.com/drive/folders/15l2_ITLnEcSKGtePcff4rKTcIDxy-_Dt?usp=sharing .
[11] https://doi.org/10.1386/jdmp_00021_1

- Electronic ISSN: 2040-4190

On the following pages it is reproduced in its original format.

Journal of Digital Media & Policy
Volume 11 Number 2
 English language. https://doi.org/10.1386/jdmp_00021_1
Received 16 September 2019; Accepted 6 March 2020

DANIEL SOTO
Universitat Pompeu Fabra

Aggressive joint compression for DTV simulcast

ABSTRACT

A complete simulcast on digital TV (DTV) can be complex to implement in some scenarios, such as when the network bandwidth has many limitations. A common solution in such scenarios is to limit the number of duplicate programmes in the simulcast to a minimum. However, this practice slows down the migration to a new generation. This article analyses the causes of this delay and proposes a technique to improve the efficiency of the simulcast. Focusing on the migration from MPEG-2 SD to H.264 HD broadcasts in DVB-T networks, the article presents an efficient solution that can accelerate the transition period.

KEYWORDS

DTV simulcast
SD–HD migration
MPEG-2 to H.264 transition
DVB-T joint simulcast
joint compression
aggressive compression

INTRODUCTION

Since the migration from analogue to digital television (DTV) was completed, several generations of consumer products[1] have been introduced into DTV over the past twenty years. The transition from one consumer generation to another is not straightforward, as DTV standards do not provide precise support for it. This means that a transition period is necessary to migrate from one to another. During that period, the common solution is to broadcast a programme using two services that use different technologies, what is known as simulcast. But that solution requires more bandwidth, so improving the efficiency of simulcast could shorten the transition period.

1. As for the products on the market, the first generation is DVB-T MPEG-2 SD, the second generation is DVB-T H.264 HD, the third generation is DVB-T2 H.264 HD and the fourth generation is DVB-T2 H.265 UHD.

One way to improve simulcast efficiency is to consider the linking of the services in both formats, simulcasting the same programme as a single set rather than as separate services. Using some simple techniques, it is possible to reduce the bandwidth required to broadcast a TV programme in two different formats, for example, by sharing some parts such as audio and other data. In addition, it is also possible to adjust the compression parameters of video streams to minimize the footprint. By making intelligent use of these two techniques, and adjusting them over time, it is possible to broadcast a full package of TV programmes in simulcast using around the same space without entirely duplicating all sources. This would result in faster and less stressful transitions.

BACKGROUND

Terrestrial broadcasts of DTV have specific challenges for simulcast compared to other distribution channels. In satellite or cable distribution, more bandwidth usually exists or can be easily incorporated. So in these cases, the migration process is not directly limited by bandwidth constraints. In contrast, when terrestrial networks are used, bandwidth is the first limiting factor. For example, at the beginning of the DTV, the spectrum used was taken from that of analogue television. Thus, during the transition period, until the analogue switched off, the bandwidth available for digital broadcasts was very limited. However, after the switch-off, the scene did not improve much. The two digital dividends in the 800 and 700 bands have reduced the spectrum available for DTV by limiting it to only the 470–694 MHz bands IV/V. Therefore, the total number of available UHF channels is less than 28 multiplexes when 8-MHz-wide channels are used (channels 21–48). But, taking into account adjacent territories and the overlap of national and regional networks, the channels actually usable for a particular network usually range from 9 to 14 multiplexes, all depending on the border agreements and layer coverage (ITU-R 2014). Therefore, the use of sixteen or more RF channels for a simulcast exceeds the capacity of most terrestrial networks. Consequently, the simulcast using terrestrial DTV broadcasts is very complex, and the widespread solution is to perform it for only a very small subset of the total programmes. But this often delays migration and the transition period becomes longer.

We maintain that by allowing simultaneous transmission of all programmes, an efficient simulcast could alleviate the above drawbacks. And for this reason, we present a technical solution to achieve this. The argument we adopted is as follows. If the transition between generations is easier when the introduction of the new generation is not disruptive, then it would be easier to finish the transition if all programmes are accessible with both technologies. There are three key points in arguing this. The first one has to do with consumers: when the new generation of receivers becomes available, the early adopters are more likely to upgrade their devices as soon as the new signals are available, even if the quality of the new signals is not the best possible. This could help because the migration process will always accelerate as soon as the critical number of consumers is reached. Secondly, the need to improve the distribution network is much lower for network operators with an efficient simulcast. As the bandwidth requirements remain the same, no new emitters are required, so the low initial investment lowers the potential barriers to the introduction of the new technology. Therefore, the new emissions could be in operation as soon as the new generation arrives on the market. The third key point concerns the broadcasters. When the network can simulcast all programmes, the return on investment in the new technology could be quickly recovered, because there will be enough audience, even if it is necessary to stay ahead of competition in quality during the transition period. As can be seen, one factor that could be critical in accelerating the migration is the improvement of simulcast solutions.

JOINT VS. INDEPENDENT SIMULCAST

The common *independent simulcast* of a programme involves the production of two independent services. So, even if the programme is only produced once, the output is invariably two different and independent services. Each service can then be transmitted throughout different or in the same multiplex. It is a straightforward solution, except for the serious problem of the bandwidth needed. In fact, all parts of the programme need to be duplicated. And this may be very inefficient, and becomes a barrier as discussed earlier. Our solution to the problem is a *joint simulcast*.

The idea of *joint simulcast* is conceptually simple: broadcast a programme using two services within the same container – that is, within the same Multi-Programme Transport Stream (MPTS) – while sharing the maximum number of elements between them. But the implementation of this concept may be complex. This article proposes a specific approach to implement it, and shows that it is technically feasible when using the same transmission standard. We first analyse the elements of the programmes that can be shared or combined. Next, we discuss how all these elements can be brought together. Finally, we describe how to optimize these shared elements so that the bitrate used will be as low as possible. In short, we provide a technical solution that allows operators to provide a full simulcast, which can be used to greatly facilitate the transition from one generation to another when the transmission standard is assumed to be constant throughout the simulcast period.

A typical example is the transition from MPEG-2 broadcasts in SD to H.264 in HD using DVB-T signals. In this scenario, our proposed solution is valid when using commonly operated DVB-T multiplexes. Typical recommendations limit the number of programmes to four when using a bandwidth of 20 Mbps, while five is an usual number using 22–24 Mbps. Fortunately, our approach also allows to vary the level of quality between SD/HD versions. Due to the need of compressing both signals to the maximum, the simulcast quality is worse compared to *independent simulcast*. However, through a preference factor, the quality of one version can be prioritized. This preference factor provides a relevant feature: the possibility of changing this quality parameter over time. Thus, during the transition period, it would be possible to modify the quality depending on the penetration of new receivers. For example, the migration could start with the highest priority to SD, meaning that the HD version will have limited quality. But later, as the number of new receivers grows, the quality of HD version may increase, while the SD quality decreases. At the end, there could be a period when the SD version becomes legacy, with its quality reduced to an essential minimum. The result could be a shorter and less stressful transition period, which is the ultimate goal of this proposal.

TECHNICAL DESCRIPTION

This section describes a technological solution that enables a full simulcast without increasing the required bandwidth. As a practical example, we focus on the objective of a simultaneous transmission of SD and HD versions of the same programmes in DVB-T networks, with MPEG-2 and H.264 codecs.

Our proposed approach is based on five aspects related to common techniques to optimize bandwidth. The novelty is to apply them at the same time to maximize efficiency in a specific way for a joint simulcast. They are:

- simulcasting within the same multiplex;
- track sharing between services;
- HD aggressive compression;
- SD video quality tweaking;
- combined statistical multiplexing.

Figure 1 illustrates the differences between a regular independent simulcast and our joint simulcast. As can be seen, there are basically two general differences: (1) With the exception of video, all other elements are included only once in the joint simulcast, while they are repeated in the standalone solution. (2) The video streams in the joint simulcast solution are significantly more compressed than in the standalone solution, with the aim of occupying the space equivalent to a single service. Due to the latter, the final quality is lower than that achieved with an independent simulcast. But this loss of quality can be evaluated. And consequently for a particular DTT network, it would be sufficient to quantify it to determine at what threshold the solution can be considered valid. We call this solution *Aggressive Joint Simulcast*, since the compression factor is extremely high in order to be able to *de facto* double the number of services in the distribution network using the same number of RF channels.

Sharing elements in an MPTS

Sharing items between different services in DTV is not new. The different standards already include some elements that are shared by services. For example, those that provide the Programme Guide data (EIT) include information for several services (Benoit 2002). But there are other more significant examples, such as teletext services. It is typically shared by services from the same provider when they are in the same multiplex. In this case, a global teletext service is built to be used by a group of similar services. This can be achieved because there is no technical limitation in the standards that prevents a given PES from being referenced by more than one service. In this way, it is acceptable that two or more PMT tables refer to the same PID. However, this sharing requires special measures. For instance, when the teletext stream is shared and it carries subtitles, then it is necessary to include these subtitles in the same container and identify them correctly. This is achieved by inserting each one on a different page, and then adding the corresponding teletext descriptor to the PMT pointing to the matching page in the subtitle description section (ETSI 2017). This example illustrates what needs to be done to share an item across different services of the same MPTS: the element must be constructed in such a way that it can be shared, and then a descriptor must be added to each service that uses it.

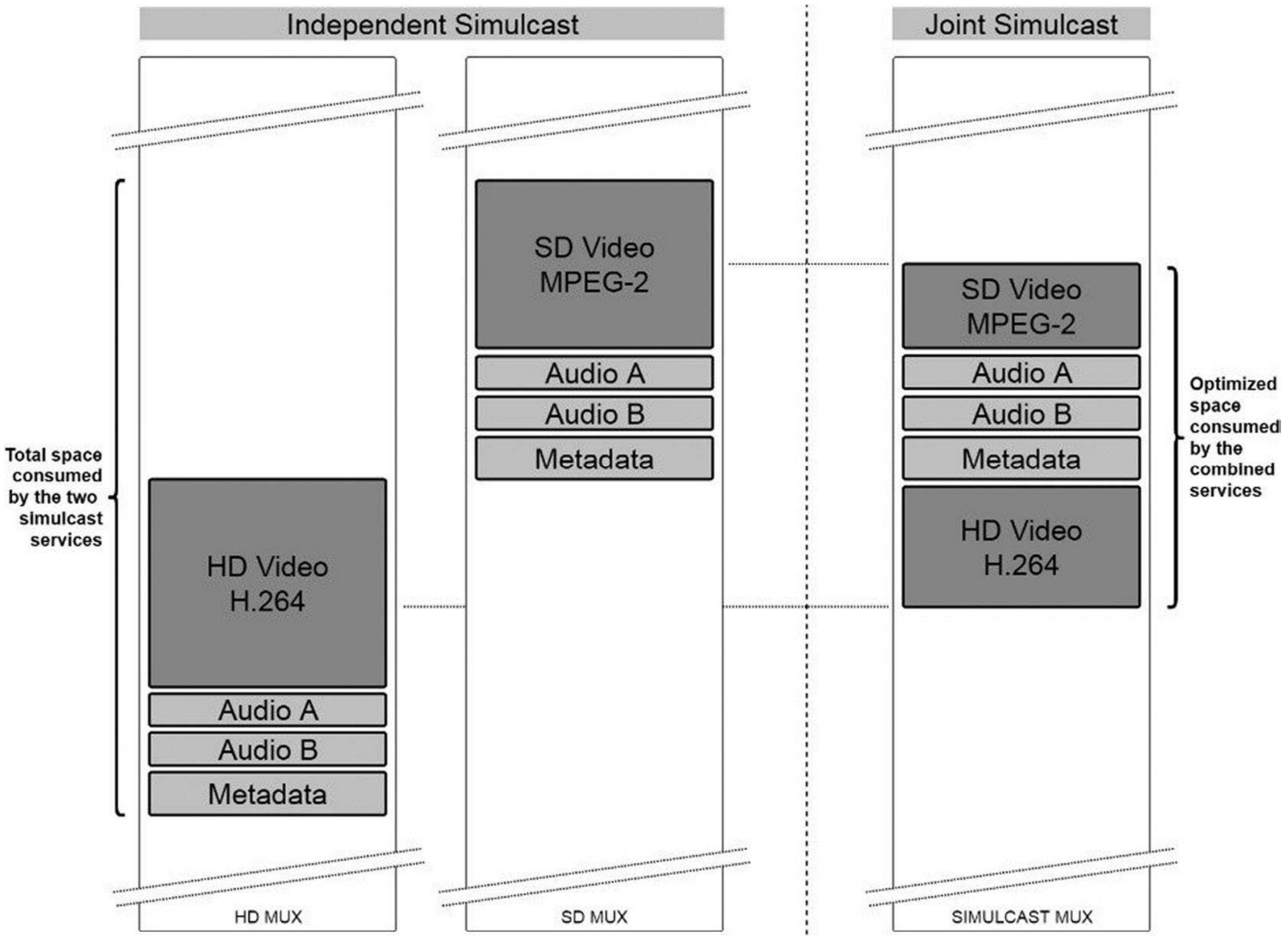

Figure 1: Independent simulcast vs. aggressive joint simulcast.

To make this possible in the simultaneous transmission of programmes in SD and HD, an essential requirement must be fulfilled: the services must always be within the same Transport Stream. This forces the use of only one modulation technology without being able to upgrade to a new one. As a result, although the DVB-T2 standard provides a significant increase in bandwidth performance compared to DVB-T, there is no choice but to continue using the latter. And, even though it is possible to increase the bitrate by 5–10 per cent using the same standard just by adjusting some network parameters, this can be counterproductive. The reason is that it reduces the robustness of the signal, and consequently the coverage is lowered. So, it is better not to change the network parameters. Therefore, sharing elements using the same transport network is shown as the best possible solution.

Starting from that idea, the different elements that compose a service and that could be shared are: teletext, audio, metadata, additional services (Carousel, HbbTV, etc.), conditional access data (EMM and ECM) and so on. In fact, anything can be shared, with the sole exception of the video, since it is the differentiating element. Audio is not a singular element because services can natively include several audio tracks encoded using different codecs. Therefore, the inclusion of multiple audio tracks does not cause troubles, as players discard unknown formats. So, only by including one audio stream encoded with a legacy codec then all receivers will be able to play the service (Soto 2015). However, it is mandatory to at least include this legacy audio track that is compatible with the previous generation, as without it the legacy devices will not be able to play any audio. Therefore, it will probably be more advantageous to simply use only previous-generation audio codecs for all tracks.

However, to make this functional, the video in both versions must be perfectly synchronized. This means that since all other shared elements are the same, it is necessary that the timestamps be identical for both video sequences. Only then the time references for the rest of the elements will be valid regardless of the video version, and then there will be no synchronization problems in the reproduction. However, in order to apply this technique correctly, it is still necessary to perform other optimization tasks. The following subsections explain other steps necessary to achieve the goal of a complete simulcast in the same MPTS.

Video simulcast synchronization

Each DTV service must have a master clock to achieve proper synchronization of all elements pertaining to that service. Without a master clock, it may be impossible to use the PTS/DTS marks to synchronize the different PES streams. The master clock signal is transmitted by including PCR timestamps at short intervals on the packets of the PES stream that serves as the master clock. Usually, in a TV service, the PCR is then included in the video stream. However, when there are two video streams, only one of them can carry the master clock signal, so a master video stream and a slave video stream are needed. This is the case, for example, with multi-camera services or when using 3D content compressed using independent views, also called simulcast of multiview video (Vetro et al. 2011). And a similar approach is used in the Blu-ray specification, where the BDAV format uses a fixed PID (0x1001) with empty payload that carries the PCR marks. In all those cases, there is simply a video slave synchronized with a master video.

Using the same solution it may be also possible to generate two simulcast services sharing the same time base. Figure 2 summarizes the process of generating an MPTS stream with synchronized simulcast services.

The process begins with the production of the source programme (in HD format). But, instead of compressing the video in the usual way, a second version of the video is generated by downsampling the original. The compression is then performed in simulcast. This means that each version is compressed using a different codec, but sharing the same time base. This would basically be the equivalent of generating a multi-camera video but using different codecs. At the same time, all other elements (audio, teletext, etc.) are compressed in the usual way.

The next step is to add PCR timestamps to both video streams. In theory, this is beyond DTV standards, because only one stream can be the master clock. However, this does not raise any technical issue at all. This is because PCR timestamps are linked to packets of a given PES. Therefore, including more PCR timestamps into other streams does not interfere with existing timestamps. This is true because the service description in the PMT table will continue to point to the master video stream, and its PCR timestamps will be present. And any player will just ignore the timestamps present in other streams. Consequently, the resulting Transport Stream will remain valid. However, it is still necessary that the PCR timestamps added to the slave video stream be perfectly synchronized with those in the master video stream.

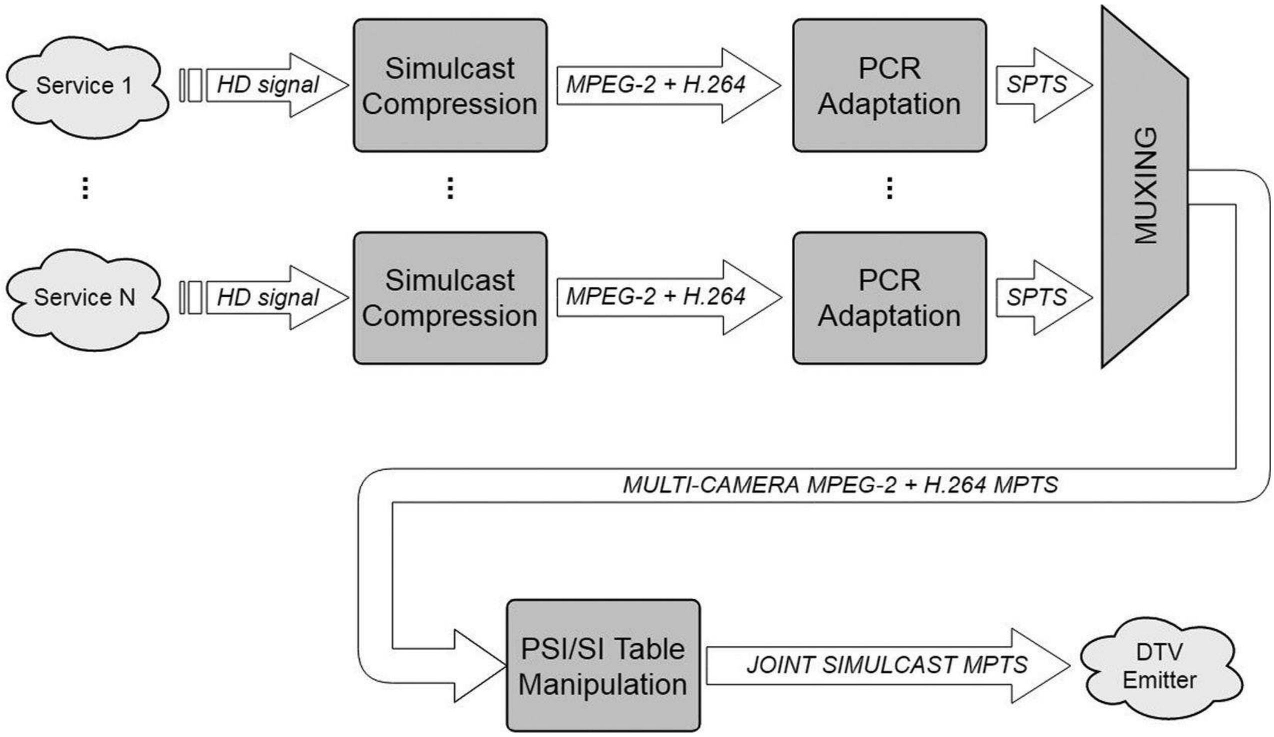

Figure 2: Workflow for synchronizing a joint simulcast.

A single restriction is that the repetition intervals of PCR timestamps of both videos must be independent.

The third step is to multiplex the service into the final MPTS. In this case, the procedure does not differ from the usual. So, after generating the service, it merely has to be merged into the outgoing MPTS. The only difference being that the multiplexer must maintain the PCR timestamps in both video streams. This is a problem that can be easily addressed. For all the packets of the same service, multiplexation is performed using the PCR values of the master video, and without taking into account the timestamps of the slave video, taking care to not remove the additional PCR timestamps present in the slave video packets. So, only a simple re-stamping of such PCR values in the slave video stream will be necessary without needing to move their position. With all the above, although the PCR timestamps seem to be independent, synchronization of both streams occurs within the same clock.

Although this solution would certainly work, there is another alternative – the one applied to Blu-ray discs. This involves adding a third track with empty payload, which only contains PCR timestamps. This additional stream is then shared between the two simulcast services. In this way, both video streams simply do not carry PCR marks and the master clock is associated with this extra stream. The final result is exactly the same as in the previous case, so either solution can be applied indistinctly.

The fourth and last step is performed with the final MPTS. In it, the PSI/SI tables are modified to divide the generated service into two different services. This task consists of performing two things: first, de-linking the slave video stream from the service; and second, creating a new service with that slave video stream and the rest of the shared elements. For example, for a master video in HD, the following four actions would be sufficient to complete the process: (1) copy the PMT packets of the service into a new free PID (which will then be a new PMT table); (2) modify the original PMT table of the service by removing the reference to the secondary video of the SD stream; (3) reconstruct the copied PMT table to assign it a new identifier; point the PCR PID to the SD video stream; and remove the reference to the HD video stream. Note that when using the third empty stream with PCR marks, additional steps are required: first create the extra stream, then copy the PCR marks from the master video, then remove the PCR marks in the master video and finally change both PMT service tables so that they point the PCR to this extra stream; and (4) make changes in the PSI/SI tables to add the new SD service (in this case, identifying the service as SD instead of HD). Later it is possible to make additional changes by manipulating other PSI/SI tables of the MPTS, but all these modifications will be complementary.

The outcome is then an MPTS with two services in simulcast that share elements. The signalling of each of these two services is independent. But the synchronization between elements is shared and accurate. Therefore, any receiver can play back either service without problems. When decoding the HD service, the original PCR timestamps will be in use. And when the SD service is being played, the PCR timestamps will be the ones associated with the slave video. But, there might be no synchronization problems because the other elements are synchronized with the master video and, at the same time, the slave video is still synchronized with it. Therefore, the PCR timestamps of the slave video can operate as the master clock without troubles. And when the extra PCR-only stream is used, there would be no problem either, as the playback uses shared PCR marks.

However, it is relevant to mention here that after the splitting process, it will no longer be possible to re-multiplex again. Any modification to the MPTS implying a remuxing process will lose the synchronization among both services. For this reason, the process of dividing up the services must always be made at the end of the production chain. Before reaching that point, it is possible to operate as usual while maintaining a multi-camera service or keeping the PCR timestamps shared. Therefore, after splitting the services, it will only be possible to make changes to the MPTS so far as that it would not impact synchronization. However, it is obvious that this constraint is not effective when using a re-multiplexer that is aware of the synchronization between different simulcast services. In that case, this restriction could be removed.

Aggressive HD compression

The distribution of HD services can be done using a wide variety of compression parameters. In fact, all DTV standards are flexible and allow to choose different parameters depending on the compromise between different factors (e.g. number of channels, total bandwidth, amount of redundancy, quality of service) (Benoit 2002; Duarte and Eldar 2011). And this could be done in any of the different layers: transmission (RF), encapsulation (MPEG2-TS) and compression (A/V). This is how broadcasters adapt their services to the transport network. However, in the case of a joint simulcast, the fundamental parameter to be controlled is the level of video compression. It is therefore critical to optimize the compression of video streams to the extreme. And this is true for both SD and HD versions. However, each requires its own processing, so we will first discuss the case of HD video.

Regarding the video, it will be necessary to use what we can call *aggressive compression*. It basically means using the most extreme values allowed by the codec to obtain the lowest possible bitrate. Although these extreme values are not used in typical broadcasts, it is essential here. This is because the goal is to bundle two versions of the video into the space normally occupied by a single service. Consequently, the negative impact of using such extreme compression outweighs the benefit of the required bandwidth savings. Among the parameters that can be adjusted to increase the compression level of video are the following:

- *GOP length:* A fundamental principle is that the greater the distance between two I-frames, the greater the compression efficiency (Huszák and Imre 2010). Therefore, a longer GOP will improve compression efficiency when controlling possible degradation from scene changes. However, in DTV, the duration of GOP sequences cannot be very long, as this would disrupt the start time of playback during zapping. This is why values close to 1000 ms or less are often used, for a typical half-second cadence of channel change. However, it is possible to slightly increase these values close to 2000 ms. And since the current specifications of DTV receivers allow for these large values, this change will only impact the response time when a channel is tuned. And in fact during the transition period, this is not a determining factor.
- *Look ahead:* The efficiency of inter-frame compression is significantly improved when future frames are taken into account (Goldman 2009). Such frames will provide additional information about the usage of each macro-block, assigning different priorities to them depending on the degree of influence on the prediction of future frames. That is why common DTV compressors operating in real time and not using information from near future cannot achieve the same efficiency as those working with streaming content. Therefore, by introducing a certain delay into the stream, it is possible to take advantage of this feature also when operating in real time. This delay is not desirable in the distribution of DTV services, but as there will be no delay between the different versions of the simulcast, on account of synchronization, the total delay will not be noticed directly by the users. Therefore, this will not be a determining factor during the transition period either.
- *Minimal bitrate:* The use of very low bitrates is not recommended for DTV broadcasts either. The main reason is potential buffer overflow or underflow, as well as more complex multiplexing (Wang and Vincent 1999). This is why when the complexity of a sequence is very low, the encoders commonly fill the bitrate to a minimal threshold, even if the quality setting is above the target value. However, by relaxing this restriction, it is possible to take advantage of the space available to improve the performance of joint simulcast. But, although technically this adjustment is possible, it is necessary to do it very carefully in order to not generate a Transport Stream with buffer errors.
- *Optimization of spatial resolution:* Another basic principle is that the smaller the spatial resolution, the more efficient will be the compression. That is why the video pre-processing allows to reach higher efficiency rates (Bruckstein et al. 2003). But, in addition to doing that, it is also possible to directly reduce the spatial resolution. It is important to note that the pre-processing filters used in DTV are generally adjusted to preserve as much spatial information as possible. But, since in the transition period it is not strictly necessary to work with the maximum resolution, this restriction

can also be relaxed. And that will allow to increase the compression factor even without producing appreciable artefacts. This may be especially important at the time of simulcast launch, because due to the minimum number of receivers installed, the presence of HD services is much more important than its effective quality.

- *Other compression optimizations:* Because the available bandwidth during the time of a simulcast is very scarce, it is convenient to apply all available optimizations within the new codec. But doing so may not be a very common measure when introducing a new generation of DTV. The reason is that the technology may not be mature enough, and therefore there may be problems with certain receivers. But this is just a standardization problem. By making the necessary compatibility tests, it is possible to solve this problem. In this way, from the first moment of the transition period, it should be chosen to use the maximum possible efficiency of the new codec.
- *Lower QP target:* The last and possibly most critical parameter that can be adjusted is the final quality of the video, controlled by the quantization parameter (QP). All lossy codecs have an objective factor that they try to achieve in order to provide the best possible visual quality. To avoid potential artefacts in DTV broadcasting, conservative values are used. However, during the transition period, as the simulcast is only viable using an aggressive compression, the restriction of potential artefacts must be relaxed. It should be noted, however, that the efficiency of the new codec is substantially higher than that of the previous one. Therefore, adjusting the QP value to a somewhat lower value for the HD version may allow a significant part of the bandwidth to be available, which can then be used by the much-less-efficient SD version.

Finally, it should be noted that video compression parameters may be adjusted over time. As a result, different periods could be defined during the transition period. These would depend on the circumstances, such as the number of receivers installed, and different values could be used in them. For example, in the early stages of the transition, HD services could use very extreme settings. These values could then be progressively normalized as the transition period comes to its end. This might not be difficult because technically no problems would arise and it might only have an impact on the final quality.

SD video quality fine-tuning

The processing of SD video differs from that of HD video. There are several reasons for this. On the one hand, the maturity of SD technology is already consolidated. Therefore, the efficiency level of the old codec will not increase much. On the other hand, the transition period should serve to turn the old codec into a legacy, until it finally is no longer in use. In this sense, as the new technology is adopted, it makes sense to reduce the quality of SD version as much as possible. To do that, some of the adjustments that can be made are the following:

- *Temporal resolution reduction:* The first generation of DTV is characterized by maintaining the formats of analogue broadcasts. This is especially true when it comes to the use of interlaced footage. In recent generations of

DTV, interlaced formats have been abandoned in favour of only progressive formats (Sugawara et al. 2014). This is a clear step forward, and therefore interlaced broadcasts are now meaningless. This fact can be used to improve the compression efficiency of legacy SD video. In this way, by switching from interlaced compression to progressive compression with half of the temporal resolution, it is possible to offer a similar quality by significantly reducing the required bitrate. For example, one SD service mastered at 720 × 576/50i can be transformed into an SD 720 × 576/25p service. This change will not cause troubles with regular DTV receivers, as nearly all MPEG-2 decoding chips have been compatible with this progressive format for a long time (ETSI 2019). The only loss with this change is temporal resolution, but in progressive receivers the post-processing can partially compensate for this loss.

- *Spatial resolution adjustment:* Similar to what happens with the HD version, reducing spatial resolution can increase compression efficiency. Although it is possible to use the capability of lowering spatial resolution, as the standards support different sizes, this should only be done in the later stages of the transition period. One reason might be that low-band-pass filters can also be applied effectively, as the content is progressive. Using either option, it is then possible to reduce the QP target value without causing significant artefacts in the video, and this might result in another significant bandwidth savings.
- *Other compression optimizations:* Finally, it is also desirable to make other minor adjustments to the compression values. The GOP length could be increased, although in this case not much can be done with the HD version. This is because the specifications of the MPEG-2 codec for DTV are usually much more restrictive than for the H.264 codec. Even so, this parameter can still be extended a bit beyond the usual. It is also possible to activate other optimizations, but it is important not to violate the guidelines of a specific DTV standard used. It is necessary to keep in mind that not all extensions will be available on installed receivers, and a significant change beyond the standard could create unexpected problems.

Narrow statistical multiplexing

When statistical multiplexing is applied in an MPTS, the gain in video quality is noticeable (He and Wu 2008; Rao et al. 2009). When using VBR instead of CBR, complex sequences can be encoded without degradation using higher bitrates, and simpler sequences then require lower bitrates. This fact can be exploited to increase the effective number of services delivered through a multiplex. However, this implies a new challenge when simulcasting SD- and HD-linked services in the same MPTS. As the HD and SD streams are directly correlated, the same complex sequences appear at the same time. The challenge is then how to apply a simple statistical multiplexing without increasing the complexity of encoders.

Since statistical compression is a complex problem, there are several strategies to make use of it. But regardless of the algorithm used, in order not to further increase complexity when using joint simulcasting, it is worthwhile to apply a simple simplification. It consists of grouping the bitrates of both video streams as if they were a single one. For example, multiplexing N programmes in simulcast implies the presence of $2 \times N$ services in the MPTS. So, instead of computing for a total of $2 \times N$ streams, it is enough to do it for the N sets of

SD+HD. This simplification may be effective because the sequences that are part of one simulcast would be identical. So, if the HD version needs more bitrate, the same would be true with the SD version. Therefore, it will be necessary to reduce the bitrate of the remaining streams to assign it to this set.

However, when two different codecs are used, this correlation may not always be the same. Because the new codec is more efficient than the old one, the efficiency of the HD version may be much higher in certain circumstances. It is nevertheless possible to adopt a simple solution to take advantage of this fact. It consists of using only the quality values of SD versions for all computations while always maintaining the sum of the two versions as the value of the resulting bitrate. This simplification might work in most cases, except where the HD version contains a large amount of high-frequency information that is not present in the SD version. However, such cases are unlikely to appear over a long period of time, so there is no need to be worried about it.

As for how to distribute the available bandwidth between the two versions of the same simulcast, it is also possible to apply another straightforward strategy. This is to first compress the HD version reserving a specific percentage for the SD version. Because the efficiency of the HD codec is superior, the resulting bitrate may possibly be below the target limit of the HD+SD ensemble. This allows that extra space to be allocated to the SD version, which is then compressed based on the space finally available. However, the bitrate assigned to the SD version may not be high enough to guarantee a minimum quality. In that case, the tuning algorithm would progressively increase the allocation to the SD+HD ensemble, thereby compensating for the lack of bitrate at any given time. With a simple deterministic reservation of bitrate for the SD version, it is possible to easily distribute the available bandwidth between the two versions without needing complex computations. In the same way, this strategy allows to adjust the bandwidth reserved for the SD version during the transition period.

A latter factor could also be considered to further increase the efficiency of statistical multiplexing. Since the two versions of the simulcast are strongly correlated, it is not appropriate that GOP beginnings overlap at the same time. This is due to the fact that the I-frames use more space than the rest. It is therefore convenient to distribute GOP beginnings in such a way that they overlap but not coincide. This can be achieved by using GOP lengths that are multiples of each other, and by including a fixed distance between the start of one and the other. This would reduce bitrate peaks for the simulcast, and improve the overall efficiency of the entire MPTS.

EXPERIMENTAL RESULTS

This article presents the results obtained by applying the concepts with a demonstrator. We conducted two different tests with this demonstrator: Test A and Test B. The purpose of the first was to prove the feasibility of using an aggressive joint simulcast in terrestrial broadcasts. The second test focused on showing the video quality obtained using this technique.

To build the demonstrator, we used an open source software. The well-known *FFmpeg* package (Bellard 2019) was used as the compression tool. In order to achieve quality levels similar to professional compressors, GPU hardware was employed. In particular, the *NVidia encoder* (Nvidia 2019) was used for encoding with the H.264 codec, and the *Intel encoder* (Intel 2019) for MPEG-2 compression. On the other hand, the *TSDuck toolkit* (Lelegard

2019) was used to manipulate the Transport Stream. This tool could perform any kind of transformation on the MPTS, so its use made it possible to separate, combine and manipulate the services in the tests. We developed specific patches for booth packages to complete certain functionalities not available by default that are necessary to complete the demonstrator. In addition, particularly for the second test, we developed a naive implementation of a statistical multiplexer.

The workflow illustrated in Figure 3 was followed for the tests. Although the tasks in a production environment must be executed in parallel, an offline processing was chosen. In this way, each of the intermediate results was stored in temporary files. The last of the resulting files is an MPTS that meets DVB-T specifications and can be transmitted using a modulator, to be played back by standard DTV receivers.

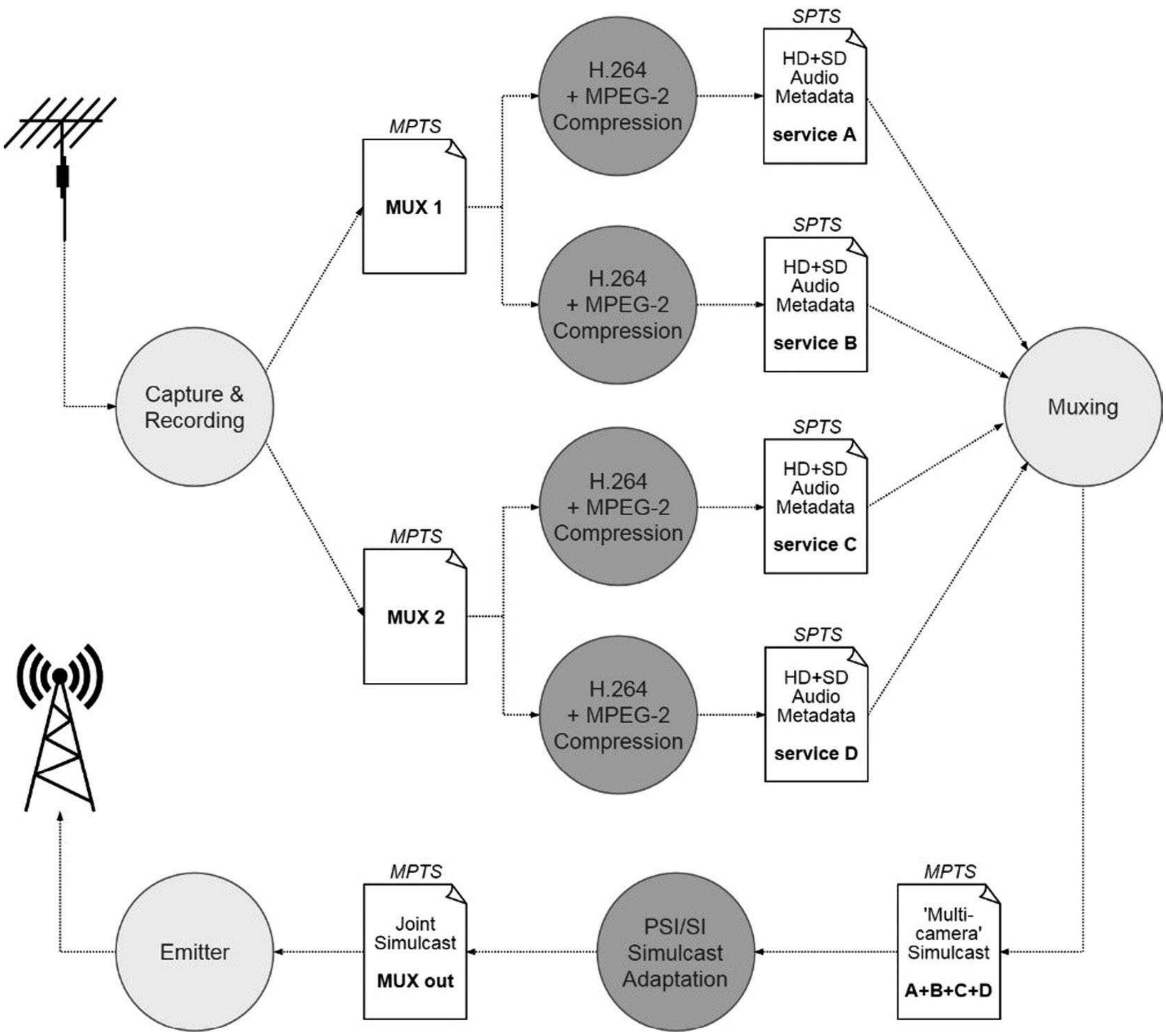

Figure 3: Workflow of our demonstrator.

First test

In a simplified way, Test A captures regular DVB-T broadcasts and transforms the content into a new MPTS using the aggressive joint simulcast technique. The result mimics all aspects of the original source, preserving all metadata attached to the services. This includes multiple audio tracks, subtitles, teletext, HbbTV, etc. The goal is then to verify that sharing elements between services of the same programme in simulcast would not interrupt current and legacy players, and that all the original capabilities are preserved at the same time. Therefore, the objective was to validate the functionality of this technique in a production environment.

Technically, the test consists of capturing several multiplexes of the Spanish DTT. These digital broadcasts currently comprise only a few services in HD, which are also independently simulcasted in SD and complemented with other services exclusively in SD. Using these services in HD as source programmes, implementing the workflow resulted in a single MPTS containing the simulcast of all programmes. All original information was preserved. The resulting MPTS can be broadcast via a DVB-T modulator. For validation, different types of receivers were fed with the emitted signals to evaluate compatibility.

Because a professional multiplexer with stat-mux capabilities was not used in this test, we consider five different cases. Each test case generates a CBR MPTS with a different bitrate, matching fully different standard DVB-T operating modes. Every case uses the same four different programmes in full HD resolution (1080/50i), with audio tracks compressed in MP2, and generates one MPTS with eight different services. These are four pairs of SD (MPEG-2) and HD (H.264) linked services in a full simulcast. Therefore, the main difference in each case is the compression of the video. This makes it possible to compare different levels of aggressive compression with the independent simulcast. Especially in cases where almost the same bandwidth is used as in the captured broadcasts. This proof-of-concept demonstrates the feasibility of a full simulcast using the same bandwidth, although with some loss in the final quality.

For video compression, we applied a naive approach between CBR and VBR with statistical multiplexing. It used a constant quality rate control that generated VBR video streams. Compared to real statistical multiplexing, the quality obtained was constant for all streams throughout the entire sequence. Although this was suboptimal, it was sufficient to evaluate the test results. In fact, for each test case, we manually selected the target QP and used this value for the whole sequence. The calculation of this target value was based on different checks to obtain the maximum bitrate below the nominal bitrate of the MPTS.

Here is a summary of the technical parameters of source inputs:

- MPTS nominal bitrate: 19.90 Mbps.
- DVB-T modulation: 8 MHz bandwidth, 8K mode, 64QAM constellation, 2/3 code rate, 1/4 guard interval.
- Services: four HD (+ two SD independent simulcast + two SD additional programmes).
- Video streams HD: 1920 × 1080/50i, H.264 VBR (2~8 Mbps).

- Video streams SD: 720 × 576/50i, MPEG-2 VBR (2~6 Mbps).
- Audio streams: 2/3 tracks, MPEG-2 or AC-3 CBR (128~192 Kbps).
- Metadata: PSI/SI, EIT, HbbTV, teletext.

The content of the services selected as source is as follows:

- A3-HD (SID:149): news programme, medium complexity.
- L6-HD (SID:151): sports programme, high complexity.
- T5-HD (SID:190): quiz show (first) and news (after), medium/high complexity.
- C4-HD (SID:191): reality programme, low complexity.

With these sources we performed the tests by combining four HD programmes in full simulcast into a single MPTS. That is, a total of four services in HD and four in SD, compared to the original of two services in HD and four in SD. The resulting MPTS is fully compatible with the DVB-T standard and retains all the characteristics of the original services, as well as the management tables of the broadcasting network. In addition, all MPTS generated were tested using an analyser in compliance with the ETSI TR 101 290 (ETSI 2014) recommendation. Each one of them passed all priority 1, 2 and 3 tests.[2] The five test cases conducted are summarized as follows:

- *Case 0 (Independent-Simulcast)*: A simple re-mux of the HD services found in sources plus the original SD versions. It represents a complete independent simulcast. It shows how much bandwidth is wasted with a regular simulcast.
- *Case 1 (Joint1)*: A joint simulcast of the previous test without any re-encoding. It represents the most simple joint simulcast, sharing only elements. It confirms that shared elements can be played without violating the standard.
- *Case 2 (Joint2)*: A joint simulcast with re-compressing only the SD video. It shows the quality level achievable when using the progressive resolution with aggressive MPEG-2 encoding.
- *Case 3 (Joint3)*: A joint simulcast of all video streams aggressively recompressed. It shows the quality obtained after aggressive H.264 encoding.
- *Case 4 (Joint4)*: A joint simulcast with the minimum bandwidth. This test increases the compression level until it matches the original bitrate. It proves the feasibility to achieve a full simulcast using the same bandwidth.

The results are summarized in Table 1, including the parameters of the streams used as well as the performance obtained. Data are grouped according to each test case, showing for each video stream the format used, the bitrate consumed (minimum/average/maximum), the QP target value used when encoding and the resulting PSNR (minimum/average/maximum). Special rows are included to describe the complete Transport Stream and the sum of all the video streams within that Transport Stream.

Overall the results of Test A validate the proposal in two different areas. First, our approach did not generate problems at the compatibility level. In all the tests performed, no problems were detected when playing the resulting

2. With the exception of some particular warnings for Test 3.5 about errors in the EIT tables. They are insignificant errors relating to the repetition intervals of the programmes in emission.

Test	Service	Format	BR.min	BR.avg	BR.max	QP (I,B,P)	PSNR.min	PSNR.avg	PSNR.max
Source	*_MPTS_*			1,99,05,882					
Indep.	*_MPTS_*			3,16,68,449					
Joint1	*_MPTS_*			3,01,60,427					
Source, Indep. & Joint1	A3-HD	1080/50i	1.999	3.383	4.502		∞	∞	∞
	A3-SD	576/50i	1.005	2.314	3.269		∞	∞	∞
	L6-HD	1080/50i	2.300	3.842	5.569		∞	∞	∞
	L6-SD	576/50i	1.628	2.670	4.372		∞	∞	∞
	T5-HD	1080/50i	2.577	3.750	5.636		∞	∞	∞
	T5-SD	576/50i	1.907	3.211	4.227		∞	∞	∞
	C4-HD	1080/50i	2.499	2.870	3.759		∞	∞	∞
	C4-SD	576/50i	1.159	2.045	3.225		∞	∞	∞
Ind. & J1	*_subtotal_*			24.085					
Joint2	*_MPTS_*			2,71,44,385					
	A3-HD	1080/50i	1.999	3.383	4.502		∞	∞	∞
	A3-SD	576/25p	0.319	1.593	3.585	5 / 5 / 9	36.76	41.89	54.25
	L6-HD	1080/50i	2.300	3.842	5.569		∞	∞	∞
	L6-SD	576/25p	0.707	2.010	4.605	5 / 5 / 9	36.36	41.76	50.09
	T5-HD	1080/50i	2.577	3.750	5.636		∞	∞	∞
	T5-SD	576/25p	0.670	2.048	4.029	5 / 5 / 9	36.39	41.45	59.30
	C4-HD	1080/50i	2.499	2.870	3.759		∞	∞	∞
	C4-SD	576/25p	0.500	1.243	2.038	5 / 5 / 9	38.72	42.65	48.05
	*_subtotal_*			20.739					

(*Continued*)

Test	Service	Format	BR.min	BR.avg	BR.max	QP (I,B,P)	PSNR.min	PSNR.avg	PSNR.max
Joint3	*_MPTS_*			2,21,17,647					
	A3-HD	1080/50i	0.212	1.999	4.917	17 / 18 / 25	6.90	19.16	40.47
	A3-SD	576/25p	0.278	1.173	2.664	7 / 7 / 11	34.80	39.98	53.50
	L6-HD	1080/50i	0.662	2.614	5.479	18 / 19 / 25	9.06	17.33	39.06
	L6-SD	576/25p	0.553	1.483	3.369	7 / 7 / 11	34.93	39.91	48.54
	T5-HD	1080/50i	0.327	2.473	5.489	18 / 19 / 25	3.58	17.95	40.64
	T5-SD	576/25p	0.511	1.516	2.952	7 / 7 / 11	34.87	39.56	56.64
	C4-HD	1080/50i	0.560	1.596	2.526	17 / 18 / 25	8.87	20.25	41.22
	C4-SD	576/25p	0.419	0.924	1.439	7 / 7 / 11	37.36	40.83	46.60
	*_subtotal_*			13.778					
Joint4	*_MPTS_*			1,99,05,882					
	A3-HD	1080/50i	0.169	1.554	4.022	19 / 20 / 26	6.90	19.14	39.45
	A3-SD	576/25p	0.270	1.105	2.492	8 / 8 / 12	34.05	39.41	52.90
	L6-HD	1080/50i	0.590	2.350	5.000	19 / 20 / 26	9.06	17.33	38.48
	L6-SD	576/25p	0.526	1.392	3.127	8 / 8 / 12	34.42	39.35	47.80
	T5-HD	1080/50i	0.303	2.237	4.991	19 / 20 / 26	3.58	17.94	39.91
	T5-SD	576/25p	0.484	1.422	2.748	8 / 8 / 12	34.30	38.98	55.85
	C4-HD	1080/50i	0.418	1.250	2.061	19 / 20 / 26	8.87	20.26	39.81
	C4-SD	576/25p	0.403	0.872	1.343	8 / 8 / 12	36.92	40.31	46.12
	*_subtotal_*			12.182					

Table 1: Test A results.

Test	Service	Format	BR.min	BR.avg	BR.max	QP.avg	PSNR.min	PSNR.avg	PSNR.max
Statistical multiplex with four programmes in joint simulcast 4×HD+4×SD	*_MPTS_*			1,99,05,882					
	RTVE-HD	1080/50i	0.285	1.687	4.093	30.32	30.24	43.18	67.46
	RTVE-SD	576/25p	0.312	1.119	2.869	13.08	8.41	26.53	51.56
	SES-HD	1080/50i	0.106	2.721	5.557	26.72	8.03	23.81	60.76
	SES-SD	576/25p	0.322	1.109	3.068	13.18	8.27	23.99	59.06
	HB4K-HD	1080/50i	0.117	2.761	5.491	29.20	5.01	19.44	51.41
	HB4K-SD	576/25p	0.186	1.794	4.105	13.47	5.19	22.78	52.52
	FTV-HD	1080/50i	1.042	3.703	8.319	28.16	4.11	16.71	46.35
	FTV-SD	576/25p	0.420	1.984	3.123	13.49	4.33	17.35	47.12
	unused		0.219	1.137	3.129				
	*_subtotal_*			18.015					
Statistical multiplex with five programmes in joint simulcast 5×HD+5×SD	*_MPTS_*			2,34,18,685					
	RTVE-HD	1080/50i	0.277	1.645	4.469	30.55	30.05	43.05	69.68
	RTVE-SD	576/25p	0.315	1.128	2.875	13.09	8.41	26.51	51.56
	SES-HD	1080/50i	0.736	2.688	5.880	26.80	8.03	23.81	60.76
	SES-SD	576/25p	0.315	1.104	3.049	13.22	8.27	24.00	59.06
	HB4K-HD	1080/50i	0.119	2.756	6.047	29.41	5.00	19.90	51.41
	HB4K-SD	576/25p	0.180	1.809	4.109	13.40	5.19	22.14	50.66
	FTV-HD	1080/50i	0.957	3.634	8.750	28.31	4.11	16.93	47.91
	FTV-SD	576/25p	0.448	1.989	3.044	13.47	4.33	17.35	47.12
	QVC-HD	1080/50i	0.139	1.743	3.592	27.81	8.98	26.56	51.98
	QVC-SD	576/25p	0.268	1.243	2.901	13.00	9.30	26.71	48.71
	unused		0.300	1.384	3.441				
	*_subtotal_*			21.124					

Table 2: Test B results.

streams with receivers of different types. Second, it was possible to increase the compression level of the video to the degree that the joint simulcast uses the same bandwidth as non-simulcast broadcasts. Although the quality was low, it was sufficient for making a full simulcast during a certain transition period. Furthermore, only by sharing all the data in the simulcast was it possible to increase the efficiency by 5 per cent without affecting the video quality.

Second test

Test B focused on maximizing the video quality of the aggressive joint simulcast. In this case, different source programmes in UHD resolution were taken to generate a full HD/SD joint simulcast. Only the video stream and one audio track were used. A relevant difference with the previous test is that statistical multiplexing was used. The objective was to verify that the quality of the video should be in line with that obtained with a traditional simulcast.

For this test, four different UHD programmes were captured from satellite broadcasts plus another from the Spanish DTT. These programmes were down-converted and interlaced in high quality to remaster the HD version, and similarly for the SD version. In this way, contribution feeds were simulated with the same level of quality as used by broadcasters. Using these sources, two different MPTS versions were generated: one version with four programmes (eight services) and another with five programmes (ten services). The first one is in pair with DVB-T networks that broadcast using bitrates of around 20 Mbps (e.g. in Spain); and the second one targets DVB-T networks using 22–24 Mbps (e.g. in Italy or France).

In this case, statistical multiplexing was used with a custom implementation. Our implementation was based in a naive forward algorithm restricted to closed GOPs. Due to its purely deterministic nature, the algorithm can be executed in real time. However, this requires a cluster of different GPU processors running in parallel. Therefore, for our tests, we ran each part individually on the same computer. The algorithm divided the processing into four steps: pre-processing, analysis, compression and multiplexing. Each of these tasks was executed at the GOP boundaries. Therefore, for executing in real time, it is necessary to finish in less time than the size of GOP. The total delay in real time is then equal to four times that value. The algorithm is time-limited since it pre-calculates the target size of each resulting GOP, selecting the value for each service based on the analysis. Thus, the compression task simply uses predetermined values that maximize the global quality when processing each GOP. This ensures that the maximum bitrate will not be exceeded in the multiplexing phase.

Here are the technical parameters of the input videos:

- *Stream UHD type 1*: 3840 × 2160/50p 8 bit, HEVC CBR (25 Mbps).
- *Stream UHD type 2*: 3840 × 2160/50p 10 bit, HEVC HLG CBR (18 Mbps).
- *Stream UHD type 3*: 3840 × 2160/25p 10 bit, HEVC HLG CBR (21 Mbps).
- *Stream UHD type 4*: 3840 × 2160/25p 8 bit, HEVC CBR (10 Mbps).
- *Stream UHD type 5*: 3840 × 2160/50p 8 bit, HEVC CBR (20 Mbps).

3. With the exception of some particular warnings for Test 3.5 about errors not relevant in the EIT tables.

The content of the services selected as source is as follows:

- *RTVE-UHD*: documental programme, medium complexity.
- *SES-UHD*: musical demo programme (first) and sports (after), medium/high complexity.
- *HB4K1-UHD*: travel programme, medium/high complexity.
- *FTV-UHD*: fashion programme, high complexity.
- *QVC-UHD*: commercials , low/medium complexity.

With these sources, we generated two resulting MPTS that include eight or ten simulcast services. Both video versions (HD and SD) were pre-processed with a high-quality 3D noise reduction filter, plus a fine gaussian blur. Then they were compressed using aggressive parameters. As for the audio track, it was compressed with the MPEG-1 Layer II codec and shared, as well as the EIT data. In addition, all basic PSI/SI tables were included to meet with the DVB-T standard. The unused space was between 1.1 and 1.4 Mbps on average, and could be used to include other additional data such as a full EPG. As in the previous test, the generated MPTS was tested using an analyser in compliance with the ETSI TR 101 290 (ETSI 2014) recommendation. Both passed all priority 1, 2 and 3 tests.[3] The results are summarized in Table 2.

A detailed analysis of the outcomes of this second test could lead to some conclusions. First, by bringing together all the available techniques, it is feasible to reduce the total bitrate of the MPTS to achieve the goal of a full simulcast. By combining the five aspects listed in the previous section, the resulting quality may be sufficient to apply this technique. It should be noted that our solution addresses a specific scenario – terrestrial broadcasts – where a full simulcast may be impossible in other ways. Then, comparing our solution with a regular independent simulcast, the improvements were evident.

A second conclusion concerns possible efficiency gains. Since our statistical multiplexer is suboptimal, improvements could be obtained using a professional equipment. Extreme compression values used in our tests showed that it is possible to use them without problems, while it is also possible to share data through services. Then, the potential efficiency gain using all together would be about 15 per cent maintaining similar compression distortion. This assumption would be based on studies indicating that statistical compression provides approximately a 10 per cent increase in efficiency; an additional 5 per cent of data sharing would yield the indicated 15 per cent.

CONCLUSIONS

Our tests showed that it is possible to optimize the simulcast of DTV programmes by sharing elements. It was also demonstrated that by reducing somewhat the quality it is possible to achieve a complete simulcast of all programmes using an aggressive compression. Doing so may be beneficial for the transition between different generations of DTV sharing the same transmission standard. One case where this may be worthwhile is in the transition from SD MPEG-2 to HD H.264 broadcasts using the DVB-T standard. As explained before, in the case of terrestrial networks, there are often multiple factors delaying the migration to the new technology. Therefore, in such situations the application of this technology could accelerate the migration compared to a mere independent simulcast.

In addition, the quality of the simulcast can be adjusted. It is explained here how an aggressive compression of each element can be adjusted so that one or the other version is given more priority or less quality. Thus, during the transition period when the simulcast is being used, several phases could be established in which different qualities can be applied. Then, with sufficient consumer acceptance and with the support of broadcasters, the migration could be accomplished gradually and steadily, even without disturbing the most neglected users. In fact, neither user will be forced to upgrade until the migration is fully completed – that is, at the moment when the simulcasts are finally removed.

Therefore, the technology described in this article could potentially accelerate the migration between specific generations of DTV. The cost of doing so is minimized by limiting the investment to the equipment needed to perform the proposed simulcast with aggressive joint compression, while the distribution network does not require any modification.

4. https://drive.google.com/drive/folders/1XwLxMmWSwYNcHAnHz_dM5VCGbXw1Tsr?usp=sharing. Accessed 29 February 2020.

FUTURE WORK

The work presented here, although using only freely accessible solutions and custom implementations, shows the potential improvement that can be achieved using an aggressive joint simulcast. Therefore, it is expected that using professional equipment adapted to this technique might provide superior results. This is the next step we want to take.

However, as for the next transition from HD to UHD, it also seems feasible to apply the idea of sharing elements between services in order to achieve a more efficient simulcast. However, although the principles remain the same, the present study only offers a solution when the transmission standard is the same. Therefore, for a transition from DVB-T to DVB-T2, this technology might not be directly applicable. However, it is possible that new proposals based on FEF frames will allow sharing part of the multiplex using different coding technologies. In that case, it would be possible to apply the same concept and then develop an HD+UHD simulcast. To this end, our idea is to explore the use of H.265 scalable extensions to provide a layered joint simulcast. However, it would then be necessary to study which elements should be optimized and which extreme compression values to use to obtain a viable simulcast that uses approximately the same bandwidth.

APPENDIX 1

The results of simulations described in this article can be downloaded[4] for verification. All the test sequences are offered freely, but with restrictions since the original material is copyrighted and permission from the author must be obtained. In our case, the use of short extracts under fair use policy applies. The files used are listed in Table 3 for Test A and in Table 4 for Test B.

Filename	Description	Bitrate	Modulation
mpe2-2018-09-20_21_14.ts	Captured from Spanish DTT (4SD+2HD ind. simulcast)	19.91 Mbps	64QAM 2/3 1/4
mpe3-2018-09-20_21_13.ts	Captured from Spanish DTT (4SD+2HD ind. simulcast)	19.91 Mbps	64QAM 2/3 1/4
mpe2_3-indep-simul.ts	Independent Simulcast {basic remux} (4SD+4HD)	31.67 Mbps	64QAM 7/8 1/32
mpe2_3-joint1-simul.ts	Joint Simulcast type1 {no video transcoding}4SD+4HD)	30.16 Mbps	64QAM 5/6 1/32
mpe2_3-joint2-simul.ts	Joint Simulcast type2 {SD transcoding only} 4SD+4HD)	27.14 Mbps	64QAM 3/4 1/32
mpe2_3-joint3-simul.ts	Joint Simulcast type3 {full transcoding} (4SD+4HD)	22.12 Mbps	64QAM 2/3 1/8
mpe2_3-joint4-simul.ts	Joint Simulcast type4 {extreme transcoding}(4SD+4HD)	19.91 Mbps	64QAM 2/3 1/4
*.analyze.txt	Summary of the characteristics of each Transport Stream	–	–
*.psnr.log	Log files of the PSNR analysis	–	–

Table 3: List of Test A files.

Filename	Description	Bitrate	Modulation
RTVE-4K_2019-12-09_14-30.ts	Captured from Spanish DTT (1×4K 2160/50p)	VBR	–
SES-UHD1_2019-*.ts	Captured from Astra 19.2E (1×4K 2160/50p-HLG)	VBR	–
HotBird-4K1_2019-*.ts	Captured from HotBird 13E (1×4K 2160/25p-HLG)	VBR	–
FTV-UHD_2019-*.ts	Captured from HotBird 13E (1×4K 2160/25p)	VBR	–
QVC-UHD_2019-*.ts	Captured from Astra 19.2E (1×4K 2160/50p)	VBR	–
*-master.HD-remaster.ts	Remastered HD version 1080/50i	VBR	–
*-master.SD-remaster.ts	Remastered SD version 576/25p	VBR	–
Simulcast.statmux.p4.ts	Joint simulcast with STAT-MUX (4SD+4HD)	19.91Mbps	64QAM 2/3 1/4
Simulcast.statmux.p5.ts	Joint simulcast with STAT-MUX (5SD+5HD)	23.42Mbps	64QAM 2/3 1/16
*.analyze.txt	Summary of the characteristics of each Transport Stream	–	–
*.ldjson	Log files of the QP values	–	–
*.psnr.log	Log files of the PSNR analysis	–	–

Table 4: List of Test B files.

REFERENCES

Bellard, F. (2019), 'FFmpeg tool', FFmpeg: The Leading Multimedia Framework, 15 September, http://www.ffmpeg.org/. Accessed 29 February 2020.

Benoit, H. (2002), *Digital Television: MPEG-1, MPEG-2 and Principles of the DVB System*, New York: Routledge.

Bruckstein, A. M., Elad, M. and Kimmel, R. (2003), 'Down-scaling for better transform compression', *IEEE Transactions on Image Processing*, 12:9, pp. 1132–44.

Duarte, M. F. and Eldar, Y. C. (2011), 'Structured compressed sensing: From theory to applications', *IEEE Transactions on Signal Processing*, 59:9, pp. 4053–85.

ETSI (European Telecommunications Standards Institute) (2014), *Digital Video Broadcasting (DVB); Measurement Guidelines for DVB Systems*, TR 101 290 v1.3.1 (2014-07), Sophia Antipolis: ETSI/EBU.

——— (2017), *Digital Video Broadcasting (DVB); Specification for Conveying ITU-R System B Teletext in DVB Bitstreams*, EN 300 472 v1.4.1 (2017-01), Sophia Antipolis: ETSI/EBU.

——— (2019), *Digital Video Broadcasting (DVB); Specification for the Use of Video and Audio Coding in Broadcast and Broadband Applications*, TS 101 154 v2.5.1 (2019-01), Sophia Antipolis: ETSI/EBU.

Goldman, M. S. (2009), '"It's not dead yet!": MPEG-2 video coding efficiency improvements', *proceedings of the NAB Broadcast Engineering Conference*, Las Vegas, NV, 18–23 April.

He, Z. and Wu, D. O. (2008), 'Linear rate control and optimum statistical multiplexing for H. 264 video broadcast', *IEEE Transactions on Multimedia*, 10:7, pp. 1237–49.

Huszák, Á. and Imre, S. (2010), 'Analysing GOP structure and packet loss effects on error propagation in MPEG-4 video streams', *2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, Limassol, Cyprus, 3–5 March, IEEE, pp. 1–5.

Intel Corporation (2019), 'Intel quick sync video SDK', Intel Media SDK, 15 September, http://software.intel.com/en-us/media-sdk/. Accessed 29 February 2020.

ITU-R (International Telecommunication Union, Radiocommunication Sector) (2014), *Spectrum Requirements for Terrestrial Television Broadcasting in the UHF Frequency Band in Region 1 and the Islamic Republic of Iran*, ITU-R BT.2302-0, Geneva: ITU.

Lelegard, T. (2019), 'TSDuck toolkit', TSDuck: The MPEG Transport Stream Toolkit, 15 September, http://tsduck.io/. Accessed 29 February 2020.

Nvidia Corporation (2019), 'Nvidia video codec SDK: NVEncode', NVIDIA Video Codec SDK, 15 September, http://developer.nvidia.com/nvidia-video-codec-sdk/. Accessed 29 February 2020.

Rao, K. R., Bojković, Z. and Milovanović, D. (2009), 'Joint video encoding in DVB/ATSC multi-program transmission: MPEG-2 vs. H. 264/AVC bitrate control', *9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services*, Nis, Serbia, 7–9 October, IEEE, pp. 3–12.

Soto, D. (2015), 'Digital TV simulcast with shared audio streams', *IEEE 5th International Conference on Consumer Electronics, Berlin (ICCE-Berlin)*, Berlin, Germany, 6–9 September, pp. 330–34.

Sugawara, M., Choi, S.Y. and Wood, D. (2014), 'Ultra-high-definition television (Rec. ITU-R BT. 2020): A generational leap in the evolution of television (standards in a nutshell)', *IEEE Signal Processing Magazine*, 31:3, pp. 170–74.

Vetro, A., Tourapis, A. M., Muller, K. and Chen, T. (2011), '3D-TV content storage and transmission', *IEEE Transactions on Broadcasting*, 57:2, pp. 384–94.

Wang, L. and Vincent, A. (1999), 'Bit allocation and constraints for joint coding of multiple video programs', *IEEE Transactions on Circuits and Systems for Video Technology*, 9:6, pp. 949–59.

SUGGESTED CITATION

Soto, Daniel (2020), 'Aggressive joint compression for DTV simulcast', *Journal of Digital Media & Policy*, 11:2, pp. 151–174, doi: https://doi.org/10.1386/jdmp_00021_1

CONTRIBUTOR DETAILS

Daniel Soto was born in Barcelona, Spain, in 1973. He received the BS and MS in Computer Science and Engineering from the University of Balearic Islands, Spain, in 1994 and 1999, respectively, and the M.Phil. in information and communication technologies from the Universitat Pompeu Fabra (UPF), Barcelona, Spain, in 2004. Since 2000, he has been a researcher and part-time professor with UPF. His research interests focus on DTV and streaming media.

Contact: Department of Information and Communication Technologies, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain.
E-mail: daniel.soto@upf.edu; daniel.soto.upf.edu@gmail.com

Chapter 3

CONCLUSIONS AND FUTURE WORK

As described in the previous sections, DTV is a good technology. However, the lack of inter-generational compatibility can be considered a pitfall. Certainly not in all scenarios this deficiency has a significant impact. However, as discussed above in some cases — such as the broadcast of terrestrial television services— it can become the main cause of large and unnecessary migration periods between generations. This is why it is important to work to ensure that there is a sufficient degree of compatibility in the future. The aim of this research has been to work in that direction. But that goal still needs much more work to be fully implemented.

So there are still many areas of research open in the field of Digital Television. In addition to ongoing research into the development of new audio and video compression techniques, which are evolving and are likely to increase the efficiency of DTV broadcasts, there are also other topics that require attention. The reason is due to the development model that has been used so far in DTV. In summary, the problem is that each new generation is based on the adoption of new technical solutions that improve one or more of the three areas on which DTV is based: encoding, packetization and transport. However, as previously discussed this purely forward model is problematic. However, as discussed above, this purely forward model is problematic. And the potential risk is that at some point the gap between one generation and another will be large enough to disrupt the evolution of Digital Television.

With regard to that risk, it is now clear that DTV is a technology that competes with many others. Over the past four or five decades the regular consumption of television as a form of entertainment has been tremendous. But since the birth of DTV, consumption habits have changed substantially. And the cause seems to be more related to the introduction of new services that compete with TV than to any other factor. The problem is that the broadcasting model used in TV, although very efficient for mass consumption, is very slow to absorb the changes. And therefore, compared to other services with which TV competes, if the speed of technological evolution is not accelerated, there is a real risk that consumption will decrease significantly due to a lack of innovation. And the final consequence would then be that DTV could become a marginal service.

To avoid this dark scenario, one element that can help is the research and development of technical solutions that focus on smoothing the evolution between the different generations. Among some of the potential solutions, the following new technical challenges currently faced by DTV can be considered as examples: the future coding standards, the jump to the next generation, and the optimization of the broadcasts using the current standards. Therefore, in this chapter we will try to discuss these issues from the point of view of future work.

3.1 Future Coding Standards and the Hybrid Scalable Compression

The work carried out during this research has concentrated on the MPEG-2 and H.264 video codecs because they are the standards used in the first and second consumer generations of DTV under the DVB-T standard. However, the standardization of the HEVC or H.265 video codec has been completed some time ago, so the industry has already started using this codec for a while. It is therefore urgent to start defining migration plans towards the use of this codec. Because if not properly planned, the same problem of incompatibility between generations will arise again. In fact, it is already possible to cite a specific example of this problem. The ATSC 3.0 standard —soon to be ratified— uses this video codec by default. And as has happened before, this revision of the standard does not incorporate any kind of inter-generational compatibility. Therefore, it will be necessary to replace all the devices in order to make them compatible with this standard. And that will probably make the migration process using this standard very slow for the reasons explained in this thesis.

In addition, and to make the problem even worse, the next generation of video codecs is already in development. Work is currently underway to complete the specifications of the codecs that will replace the H.265/HEVC. Candidates include AV1 and H.266/FVC/VCC. Regardless of the characteristics of each one, as well as which one is ultimately chosen by the industry for use in future DTV standards, it should be noted that at this time there is no work focused on incorporating compatibility between standards from different generations. Therefore, it can already be assumed that broadcasts with these codecs will not be interoperable with current devices. And that will once again raise the same challenge of migration between generations. However, this may not necessarily be the case. After all, once a problem has been identified, solutions can be found. Based on this, and in accordance with the criteria of the author that the migration between generations in DTV is a true problem, the following are some possible lines of research to overcome this problem.

The first outstanding challenge is the compatibility between different video codecs. When it is the case that the codecs are incompatible with each other, i.e. one is not a subset of the other, as is currently the case, then specific techniques must be applied to obtain a certain degree of compatibility. In this sense, the author proposes to use the concepts of scalability already exposed in this thesis. This would facilitate the transition from the H.264 codec to the H.265/HEVC. Some open lines of research based on this idea are presented below.

One potential technical solution is to use the scalable native extensions of some codecs. The H.264 codec has an extension that provides scalability, it is the codec called H.264/SVC. The efficiency of this extension is good enough to be used in production environments. Although it is true that so far the industry seems to have shown little interest in adopting such extension. While it is possible that the cause is the lack of environments where its use is really productive. However, from a technical point of view such extension is very robust. Furthermore, the H.264 codec standards require that all decoders must be transparent to the use of this. And that is a very helpful feature. Because when a video stream is encoded using the H.264/SVC codec, it is guaranteed that any decoder, even if it does not support the scalable extensions, will be able to decode the base stream. And this simple requirement can be very valuable for implementing compatibility between different codecs. For this reason, the author proposes the *hybrid scalability* as a potential line of investigation. Below is a brief explanation of what this concept consists of and how to apply it.

A first model where it is possible to apply some hybrid scalability is when the two codecs to be used support scalability extensions. That is, when both the old codec in use and the new codec to be introduced are capable of operating in scalable mode. When this happens, it will always be possible to decompose the video stream into two streams, regardless of the codec used, and make them to work together. This is already possible using the H.264/SVC standard in combination with the more modern H.265/HEVC standard, since the specifications of the scalable extensions that were concluded not long ago in the latter, under the name SHVC, provide precisely support for that case. This means that a scalable stream in which the base stream is encoded with H.264 and the extension is encoded with H.265/HEVC, is then processable by a H.265/SHVC decoder. Therefore, since it is then technically feasible for both codecs to work in scalable mode, the simplest strategy would be to compress the video stream in scalable mode, but compress each of the streams using a different codec. The only problem, however, is that at present the development of this form of scalable hybrid encoding remains a line of research that needs to be further explored.

Therefore, as a future work based on this specific line of research, the author proposes the following similar approach to be applied to the DTV. It would consist in compressing the base stream with the old codec —a codec without scalable support—, and the extended stream with the new codec, but using as base stream the one encoded with the old codec. This solution, which we can call *translated hybrid scalability*, compared to the *native hybrid scalability* mentioned above, is in general simple but technically more complex. The reason is that no direct support is included in the codecs in use, so some degree of adaptation is required. For example, just with regard of the encoder, a new challenge comes up. Since the codecs are different and no scalable extensions are supported, the encoder needs to generate both streams together and do the decomposition at the same time. However, such type of encoder does not currently exist. Another, and more important problem, is the compatibility with decoders. In order to support this translated hybrid scalable mode, certain changes are required inside the decoders. Such new changes need to be developed, tested, implemented and verified. Nevertheless, this should not be a huge problem, as the industry has not yet deployed the equipment for the latest generation involved, so there is still enough time.

At the technical level, this proposal for a translated hybrid scalable coding is based on two new concepts: on the one hand, on the transparent encapsulation of the extended stream; and on the other hand, on the scalable decoding using different codecs. The first is necessary to avoid creating incompatibilities with legacy devices. The problem here is that if the extended stream is compressed directly using the new codec, this could cause problems with unadapted decoders that do not support hybrid scalability. A simple solution would be to encapsulate the stream as if it were a scalable stream using the old codec, but adding some small change so that it would be discarded by the decoders of the old codec with scalable support. Then there would be no problem either with the old decoders or with the new ones, whether they have scalable support or not, because they would never use the extended stream. Only the new decoders with support for this kind of translated hybrid scalability would use both streams to reconstruct the sequence from these streams. The end result would be a seamless operation when it comes to unsupported devices.

On the other hand, as for the decoders compatible with this new format, where changes should be made, these could be reduced to the following. First, support for scalable extensions should be mandatory. Obviously, without this support it is impossible to implement any type of hybrid scalability solution. But as already mentioned, for the HEVC standard such extensions are defined and standardized, so the only requirement is that the industry implements them in DTV devices. Second, support for scalable hybrid streams should also be mandatory. And on this last requirement there are at least two possible technical ways to implement this functionality: native support or direct transcoding.

The native support of scalable decoding consists simply in that a scalable decoder is able to determine with which codec each scalable stream is coded and work indistinctly with one or the other. Thus, if the base stream is encoded using the old codec, simply use that other codec, instead of the native codec. This as a technical solution is very easy to implement, as it is common for new equipment to incorporate support for both new and old codecs. So support for the non-native codec is already present, and it would then be very easy to add such an extension. Actually it would only be necessary to add the support to interpret the transparent encapsulation of the extended stream, which is trivial to be done. The only variation then with regard to a regular scalable decoding is that the base stream will use a different codec. Doing this should not limit the reconstruction results in terms of quality, although it would affect the efficiency that can be achieved. The reason is that the efficiency of the old codec to encode the base stream will be lower than that which would be obtained using the new codec. So as can be seen, this option is technically feasible and easy to implement. It is for that reason that as discussed above the SHVC extensions of the H.265/HEVC codec are supporting this mode of operation with a base stream in H.264. However, for a general implementation it would be necessary that all codecs used now and in the future in DTV have that particular feature.

With regard to the other implementation option, the direct transcoding solution, it is a bit more complex but also feasible. In this option, the scalable decoder does not require any modification, but the video stream is transformed before being injected into the decoder. This transformation is done in real time and is divided into two different tasks. The first, and simplest, is to decapsulate the extended stream and remove the wrapper that masks the data as if it were an extended stream encoded with the old codec (but note that this stream is transparent to any decoder of the old codec with scalable support). After this simple process this extended stream will 100% compliant with the scalable extensions of the new codec standard. The second task is therefore to transcode the base stream to be compatible with the new codec. This task, although it may seem costly, can be done in a simple way. On the one hand, the resulting stream does not have to be transmitted, but will be consumed directly. Therefore, in theory there are no major bandwidth restrictions. So it would be feasible to apply a mere pure I-frame coding only, which is very fast and requires few resources. This stream, created in real time with a fast transcoding, will then be used to feed the standard scalable decoder that uses the new codec. With this approach, the only essential prerequisite is that the transcoding performed must be absolutely lossless. Otherwise, the reconstruction will include errors.

An interesting advantage of the latter solution of direct transcoding is that it does not require technical changes of any kind in the decoders. And this is a relevant key point. Of course, it requires a specific encoder and the transcoding capability implemented in the receivers. But there is still the option that none of that processing has to be done inside the devices. Using external equipment, such as CI cards, it is possible to carry out this translation. So the decoders will not be modified in any way. This makes the solution easy to implement, and for this reason it is therefore the option recommended by the author when a native hybrid scalability is unavailable.

In sum, it can be seen that there are solutions to the problem of future compatibility between generations. And although these solutions have not yet been explored outside of the work done in this thesis, it has been possible to see a small overview of how they could be implemented. Thus, a very interesting line of future work is being suggested to enable future DTV broadcasts to be compatible between different generations. And as already mentioned, it is already possible to natively combine HD streams with the H.264 codec along with UHD extended streams using the H.265/HEVC codec with the SHVC extensions. This would undoubtedly facilitate the transition between generations of Digital Television using these codecs. Therefore, it would be very interesting if the industry would be encouraged to include and promote the use of scalable extensions in DTV devices and standards.

3.2 GPU based Advanced Compression

Although the main objective of this thesis, as already mentioned, is not the optimization of video compression, it has been possible to explore certain related areas in which it would be useful to be able to continue working. One of these fields is the use of GPU hardware for video encoding. Along this line it is necessary to comment that in the research on aggressive compression the author has been able to know this technology closely, and the results are very interesting. If until now it was necessary to use professional equipment to compress DTV streams in order to achieve acceptable quality, now with GPU assisted encoders this requirement no longer seems necessary. The price of encoding equipment for Digital Television has for a long time exceeded by several units of magnitude the economic capabilities of both individual users and modest companies and researchers. What has led during all this time to limited results around this line of research outside large groups specifically dedicated to the area of DTV.

However, with the popularization of GPUs it is now possible to remove this barrier. Since video compression on the GPU can now be considered a commodity, it is possible to generate fully compatible DTV streams with non-professional equipment. This opens up the opportunity to investigate new optimization methods, such as statistical multiplexing. So far, this type of multiplexing only exists in professional equipment when it comes to real time. However, a more innovative technical implementation could now be possible. For example, it would be possible to develop new open source tools — which do not yet exist—, that could generate MPTS streams using this technology. This would allow the techniques presented here, the aggressive compression and the joint simulcast, to be explored further. And as the results published in this thesis show, a simple working statistical multiplexer based on GPU compression could be built. Although this prototype does not work at the moment in real time, it allows to see that such implementations are possible. So its use could well be used to find more effective techniques to improve significantly the efficiency of simulcasts. This would perhaps attract more attention to this line of research, and more studies could be done. And perhaps that could eventually encourage the industry to take a greater interest in the inter-generational compatibility in the Digital Television.

Specifically speaking, the advantages of GPU hardware assisted encoding are many. On the one hand, it is possible to work in real time using a modest computer. On the other hand, it is possible to achieve efficiency and quality rates that are not possible with current software algorithms. And the fact that GPUs are now a mass market product means that it is possible to use several GPUs at the same time. This opens up the possibility of simultaneously processing multiple streams, and even applying multiple filtering processes simultaneously to the same stream. In this way, using clusters of GPU equipment, large-scale processing tasks that would otherwise be much more difficult can be addressed.

However, it should be noted that video compression on the GPU is not intended to replace professional DTV equipment. It is not even intended to be used in any other type of Digital Television environments. This is because all existing implementations are based on proprietary drivers. And the GPU manufacturing industry has incorporated the ability to compress video for other reasons. These reasons are basically two: hardware video decoding and video streaming. To understand the limitations of this technology, these two reasons are detailed below.

Today, whatever device the user uses —no matter if it is a PC, a smartphone, a laptop or any other electronic device— the ability to play video content is considered inherent. However, the increasing use of video in electronic devices means that this capability needs to be added to the hardware. The reason is that the bandwidth required for uncompressed video exceeds by far the transmission capabilities of the networks. This means that compressed video is required. And because better quality is increasingly desired, but taking up the same space, it is necessary to use codecs with high efficiency rates. However, to achieve this goal, massive computational efforts are required. For this reason it is necessary that the devices have dedicated video decoding hardware. Otherwise, they will not be able to play back compressed video streams smoothly.

So the primary use case for GPU decoders is for conventional multimedia playback. This often excludes formats commonly used in DTV, such as interlaced video. Fortunately, the implementations provided by manufacturers often support interlaced formats as well. As do existing open source tools used by GPU decoding libraries. This is very useful, because in order to apply complex compression processes it is often necessary to process a stream in multiple stages. And to do so, the simplest way is to recompress the same stream several times. Therefore, to retain high quality while being fast enough to operate in real time, it is useful to rely on specialized hardware for decoding. For this reason, decoding on the GPU is extremely useful, even though the work involved is the compression of DTV streams.

When it comes to GPU hardware encoding, the reason why the industry supports this feature is completely different. There are basically two reasons that can be identified, although they are both related. The first is the ability to capture sessions in video games and the second is the ability to project the desktop onto external displays. The first use case is very significant because, in fact, much of the success of GPUs is in their use for video games —they are actually their niche market—. And the second use case is notable because it is not limited to desktop devices. Today, the screen projection on other displays is being introduced strongly in the smartphone and notebook environments. There the aim is to provide simple solutions for the user to connect additional displays. And one way to do this is to capture the screen as a video stream and then send it to an external device. This use case works very well as long as dedicated hardware is available.

Thus, following the needs of both use cases described, the industry has now turned video encoding/decoding on the GPU into a commodity. However, because the focus is completely different from that of Digital Television, there are many limitations to how GPUs can be used in this field. In essence, it must be considered that the support provided by manufacturers for video encoding using the GPU will not provide the expected support for DTV broadcasts. At least two reasons have been identified during the development of this thesis. First, the settings that would be common when compressing a television stream are not suitable for what are considered common usages in the GPU market. On the other hand, bitstreams generated with a GPU often do not meet the requirements for DTV broadcasts. The reasons for this are explained below.

With regard to the first point, the settings used in the video encoding, it should be noted that in the case of synthetic images, such as in video games or computer monitors, the final quality can be greatly compromised if specific settings are not used. For example, when using lossy codecs it is essential that certain types of noise are not introduced into the video sequence. As well as not losing certain details. However, the video sequences typically used in Digital Television are in many cases relatively immune to the effect of both of these issues. Therefore, it should be understood that GPU video encoders may have quality drawbacks when compared to professional DTV equipment. A simple solution would be to test and select settings that will give similar results to those obtained by dedicated equipment. A future line of research could then be to explore these configuration parameters and try to improve the results.

And as for the second point, the compliance of the bitstreams generated by the GPU video encoders with the DTV standards, here the problem is to adhere to the specifications. The standards are very strict in this regard in terms of how the stream should be in order to be played back smoothly by DTV decoders. The work done during this thesis has highlighted this problem, as these bitstreams often do not meet the strict specifications imposed by the standards. These limitations have been shown when using analysers to verify the streams and regular DTV boxes for playback. However, it has also been found that it is actually viable to adapt the generated bitstreams to meet the specifications. To achieve this, two techniques were used which have proven to be effective: modifying the streams on-the-fly and forcing certain compression parameters. By doing that the streams that are obtained can be perfectly used for DTV broadcasts without any trouble. However, it should be highlighted that this limitation is very important, so it would be necessary to work on improving the compatibility in order to allow the compression in the GPU to be used in the field of DTV in a more trouble-free way.

And it should also be mentioned that in addition to video encoding/decoding on the GPU, this dedicated high-performance hardware can also be used for other computational processes. Because video pre-processing can be used to increase compression efficiency, dedicated hardware is very useful for performing these filtering tasks. In fact, this functionality is already present in many GPU drivers. So it is easy to take advantage of this feature. And since the entire video stream is already stored in GPU memory, it is faster and more efficient to do all the filtering on the GPU than to move the video stream into RAM for processing it by the CPU, and then copy it back to the GPU memory. Therefore, using standard libraries such as OpenCL, it is possible to apply any type of filtering to video streams, which can certainly help improving efficiency. So another area for further research would be to use GPU processing to optimize the compression performance.

And in close connection with that, there is one last point to consider about the use of the GPU. In recent years, there have been lines of research on AI and BigData that use GPUs. But so far there are few studies that apply these techniques to video processing, although this is likely to be a very promising area as well. If these techniques and knowledge could be applied to joint compression within a DTV simulcast, then it might be possible to make the correlation between the streams in the simulcast even further exploitable. This would be possible by using a platform where all the processing and coding can be done on a GPU cluster, which would then be able to address the investigation of the application of these AI and BigData techniques to the sequence to be processed.

3.3 Channel Bonding and Mixed Modulation Broadcasts

One of the areas of research that has not been addressed during this thesis concerns the modulation of RF signals. As mentioned earlier, this area —the transport layer— is the third basic component of Digital Television (the other two are encoding layer and packetization layer). Therefore, all the results presented here are based on the premise of using a single modulation mode to broadcast all the data within a single multiplex on a single frequency. This means that all simulcast services are within the same MPTS. However, it is also possible to go further and improve intergenerational compatibility by making changes in the transport layer.

Aside from the improvements made to the various standards in terms of modulation modes —which is an open research area, as is audio and video coding— improvements focused on adding compatibility between different broadcast signals are potentially possible. Among the possible approaches, the author proposes at least two in this area: the channel bonding and the mixed modulation.

The bonding technique is a simple solution to the problem of combining two modulations, one new and one old, which also adequately complements the joint simulcast. Bonding in general involves the use of two paths to send a single stream. This technique is widely used in other areas of telecommunications, such as in networking, where by using the transport or link layers it is possible to combine the bandwidth provided by multiple transport connections. But this technique could also be applied to the broadcasting of DTV signals. In fact, some standards already support this technique to some degree. For example, DVB-C2 incorporates support for a *Channel Bonding* extension; also the future ATSC 3.0 will support the *Channel Bonding* functionality of two channels; and DVB-S2X also incorporates an option that makes it possible. In all of them, the channel bonding solution basically involves the use of two frequencies (or more, depending on the standard used) to broadcast a single MPTS. Thus, the packets of the Transport Stream are distributed among the different physical channels. And the initial order of the packets is reconstructed in the decoder to obtain the original MPTS. In this way it is possible to increase the effective bandwidth by combining more than one frequency and achieve higher bit-rates.

But this technique in itself only increases the bandwidth. It does so mainly by aggregating the bandwidth of the channels. Although also by reducing the overhead of the channel spacing when the two are adjacent and the modulation used allows this gap between the two frequencies to be exploited. However, it has one drawback: it requires the use of more than one tuner (or using one with a wider channel width, if the frequencies are adjacent). But in return there is a very large bandwidth that can be exploited for better efficiency from statistical multiplexing. And this will increase the total number of services that can be broadcasted, which in itself is a clearly significant improvement, especially when the services occupy a large part of the MPTS. For example, by joining three frequencies it is easy to go from 9 services distributed in three groups in 3 frequencies, to 11 (or more) services distributed in a bond of those same frequencies. And the best part is that all of this is using exactly the same RF spectrum and maintaining the same quality of services.

But better still, this technique could also be used for the purpose of allowing inter-generational compatibility. To do so a combination of two (or more) frequencies would be required, but using different modulations in each. In this way it would be possible to build a WTPS that shares parts of the services, as explained in this thesis, but instead of using a single channel to broadcast it, several channels are used. Thus, one of the channels could still use the old modulation, which is compatible with legacy devices. And the others could use the new, more advanced modulation, which will be compatible with the new devices. Then, using the channel bonding, the video streams using the old codec, as well as the rest of the service data, would be delivered through the channel with the old modulation. And video streams using the new codec would be delivered on the channel(s) with the new modulation. This would maintain compatibility with legacy devices, but allow the new devices to receive the new services by reusing some of the data broadcasted using the old technology.

However, this mixed bonding mode is completely new and is not covered by any standards. So further work on its implementation would therefore be very beneficial. And due to the great advantage that its use could bring to the terrestrial distribution of DTV broadcasts, its incorporation as an extension of the DVB-T2 standard would be very attractive. However, it should be pointed out that the type of bonding proposed implies the use of different modulations. Therefore, some form of encapsulation would be necessary to make the use of this technique on the old frequency —the one that uses the old modulation and carries the services using the old codecs—. This is because usually the use of the bonding technique implies that all channels are used to receive the original stream. However, in the proposed scenario the bonding would be asymmetrical. In other words, one channel would operate on its own as if there were no merging channels at all, while the other would operate as an extension. Therefore, there would actually be two modes in operation: the first is the basic mode (the legacy mode) which uses only one channel, and the second is the extended mode (the new mode) in which both channels are combined. This operating method could be called *Asymmetric*

Channel Bonding with Multiple Modulations. And this is one of the developments that the author is interested in continuing to work on.

The other technique that could be investigated further in this area is the *Mixed Modulation*. This technique involves modulating more than one digital signal within the same channel on the same frequency. And in the case of a DTV broadcasting, it is equivalent to the fact that part of the occupied spectrum is used to transmit other digital signals. In a nutshell, it involves making a Time Division Multiplexing (TDM), or a Frequency Division Multiplexing (FDM), but using completely different digital signals. Both modes of signal multiplexing are commonly used in many areas of telecommunications. It is therefore not surprising that certain of the DTV standards incorporate particular modes of operation that are based on some of these two techniques. For example, the ISDB-T standard divides the spectrum of each channel into different parts and it may use different modulations in each part. In other words, using FDM, each television channel is subdivided into smaller portions. And each of these sections can be used for a different purpose. For example, single frequency broadcasting services can be provided with different capabilities in relation to error correction and other similar characteristics. Or each of these sections can be used for completely different services. Even so, the standard only allows the use of specific modulations within the channel, and always for broadcasting DTV services.

And as for TDM, a combination using time division is also possible. For example, the DVB-T2 standard, with the incorporation of the Super-frames and supporting the Future Extension Frames (FEF), is able to allocate part of the bandwidth to other digital signals. This feature allows dividing the symbols of the signal carriers into frames, and signalling and identifying each of them. The frames can then be of different types, even unknown types. Thus, it is possible to send any digital signal within unused frames. And using this feature it is possible to modulate DVB-T2 Lite (radio) or LTE (mobile phone) signals together with the DVB-T2 signal on the same channel.

However, so far there is no solution to overlay two DTV signals using different modulations. Although it would clearly be useful to develop some extension that allows a partial modulation mode that reuses an old modulation. For example, it would be useful to explore how to encapsulate a DVB-T signal into a DVB-T2 multiplex. One way to do this could be to increase the bandwidth of the DVB-T2 channel to be wider than that of the DVB-T signal, and then place the entire spectrum of the DVB-T signal inside the DVB-T2 channel. To do this it would be necessary to develop a new extension that allows to define larger channels in DVB-T2 (the current maximum is 10MHz, which would be insufficient to contain a DVB-T channel using 6, 7 or 8MHz), and then use a new FDM mode that allows to combine both signals. This approach would be necessary because the problem cannot be solved by using TDM and super-frames, as the DVB-T signal does not allow TDM with other signals. However, it should be mentioned that in case of using a fixed mode of TDM with wider channels, it would be functionally equivalent to the bonding solution, and then there would be no advantage in using it. Therefore, the Mixed Modulation would only make sense if it could be used on the same frequency and be transparent to the old modulation.

In any case, if any of the above solutions could be used to combine transport signals, it would certainly facilitate the migration between different generations of DTV. The main reason is that the *Joint Simulcast* technique presented here could be applied much more efficiently. Instead of just sharing elements at the packetization level, this sharing capability would also extend it to the transport layer. And by combining this with the results of the *Hybrid Scalable Compression* an optimal and high-efficiency model for inter-generational compatibility could be achieved. The following lines describe how this would be possible.

The basis of this new highly efficient model would be to include all simulcast services within the same MPTS. That MPTS could be identified as the *complete MPTS* (the reason is explained below). In this MPTS the services would be distributed in a joint simulcast solution, where each service would be composed of a shared part (audio and other data streams) and two different video streams. And these two versions of the video will be scalable, so there will be one basic video and one extended video. All elements would then be compressed using the old generation codecs, except for the extended scalable video stream that is compressed using the new codec. Finally, the MPTS is divided into two parts, one compatible and one extended. And the compatible part is broadcasted on one channel using the old generation modulation. The extended stream, along with some data required for the reconstruction, is broadcasted on another channel using a new modulation. In this way, any legacy device will be able to reproduce all services by tuning to the old channel. And the new devices will be able to play the services in simulcast using the improved video quality in conjunction with the other shared elements. To achieve this, they must additionally tune into the extended channel and rebuild the entire MPTS. Then, using the entire reconstructed MPTS, the new devices will only need to decode the enhanced quality video using the scalable hybrid compression. So, since the base video stream is encoded using the old codec, and all the base service streams are located in the shared part of the MPTS broadcasted in the channel with the old modulation, then the process for the legacy devices is completely transparent. Therefore the three parts into which the Transport Stream is divided are: the *complete MPTS* which is the original MPTS, the *basic MPTS* for the compatible part shared with legacy devices, and the *extended MPTS* which is the scalable part for new devices only.

With this model it would be possible to achieve the maximum level of efficiency in a Joint Simulcast. And although here two channels are used to broadcast the entire simulcast, this could actually be done by reusing pre-existing frequencies. In other words, from a set of N frequencies, N-X (where $X<N$) could be reserved for the old technology (using the old modulation carrying the basic MPTS) and then use the remaining X frequencies for the new technology (which will use the new modulation to transport the scalable extended stream compressed with the new codec). The main advantage is that if this allocation can be done transparently, which would be possible by adding the necessary extensions to the current standards, then this would also be transparent to users. And then it would be possible to vary the distribution of the bit-rate between the old and the new technology. This would be a very interesting functionality to be used during the transition period, as described in this thesis. However, there would be the limitation that the basic versions can only go on the frequency with the old modulation. This should not be a problem, though, as this restriction does not apply to the extended versions.

Therefore, as described above, it is actually feasible to implement an optimal model for distributing DTV services in such a way that it allows smooth migration between different generations of Digital Television. And that is the ultimate goal of this research.

So based on the results obtained, and the future work that has been outlined, it is demonstrated that it is certainly viable to develop the technical resources necessary for a real inter-generational compatibility in the standards of DTV. This would undoubtedly be very useful, and is the work for which the author has been fighting since the beginning of this investigation.

3.4 Specific recommendations for the Spanish DTT

To conclude this document, the author would like to include a number of recommendations for a particular scenario. During the research of this doctoral thesis, an specific scenario has always been used as a target example: the DTT in Spain. And the reason why these final comments are not included as an appendix is because the results can be directly applied in that scenario. In fact, they can be considered future work, as the author is currently working on being able to make this objective a reality. Consequently, it is valuable at this point to conclude this document with an attempt to apply to that case the experience acquired during the realization of this research. To this end, the author presents a list of recommendations based solely on his own experience.

As a first general comment, it can be asserted that the introduction of DTT in Spain in substitution of the traditional analogue television must be considered like a success. Among the different reasons that support this argument there are several, but there are two in particular that stand out apart from the obvious improvement in quality: the universal access to services, and the increase of the television offer. With the use of analogue technology, access to television services was closely linked to the place where the audience resided. On the one hand, the quality in the reception of analogue signals was deficient in places distant from large population centres. On the other hand, advanced services such as dual and stereo audio, were only active in a few broadcasting centres (basically in only two —Torrespaña and Collserola— despite there being thousands of total emitters). And the extended services were reduced exclusively to Teletext. This created a technological gap between two population groups: those living in the most populated parts of the territory, and all the others.

Moreover, the television offer was certainly very small with the analogue signals. And it varied greatly according to the location of the population. This was due to the fact that in some places multiple local and regional broadcasts were received, while in others the offer was exclusively the minimum of the services with national coverage. Also the number of providers was limited exclusively to the public broadcasters and a set of three or four private broadcasters. All this made up a rather limited offer of television services which again divided the population into groups according to their location.

On this basis, it can be concluded that the substitution of analogue television by DTT in Spain led to a democratisation of terrestrial television in that territory. On the one hand, all services became identical. And all of them were then accessible under equal conditions in all the places with coverage. On the other hand, digital television being able to offer extended services such as multiple audio tracks, subtitles in multiple languages, audio description, etc. And as a result, it was possible to ensure that the entire population could now have access to all extended services. Furthermore, due to the significant increase in the number of broadcasted services (in an order of magnitude of 5 or 6 times more, due to the multiplexing capabilities of DTV), the public was able to start consuming content that was otherwise considered not cost-effective. This also further democratised the television experience, as a large part of the population was no longer directly subject to the rules of the small group of media operators.

But this apparent success, not always perceived as such by the people, is not free of problems. However, it is necessary to say that many of these problems are not due to a technical issue, but rather to other external factors; such as the effects of regulations imposed by legislators, the economic objectives of operators, and the failed proposals of some services. But even so, it must be noted that there are some technical deficiencies that could have been solved otherwise. In fact, the basis of this thesis shows that this is true: since almost twenty years after the first digital terrestrial television broadcasts in Spain, the same generation of digital television is still being used. And this despite the fact that the technology used is already obsolete and there are already a few generations that could replace it. So the current situation is that everything seems to have stalled for the reasons explained throughout this thesis.

For this reason, and although the author does not discuss the success of the DTT in Spain, but rather values its success, he also wants to give visibility to some technical issues that could be improved. The purpose of this is to provide greater technical knowledge, as well as innovative solutions that could improve the DTT service. This is particularly significant when it is taken into account that some of the aspects that could be improved can be enforced based on the guidelines established by the regulator. In fact, these are viable solutions that only require a commitment from the regulator to implement them quickly with some limited investment in the improvement of the distribution networks.

Among the aspects that could be improved, the following four are detailed:

1. Design of the distribution network.
2. Packaging of the broadcasted services.
3. Multiplex management.
4. Rules for service migration.

We will analyze in a very short detail what each of these aspects implies.

3.4.1.1 Design of the distribution network

The distribution network used in Spain employs slightly different values from those used in other neighbouring countries (mainly France, Italy and the United Kingdom). Basically the problem is that the effective bit-rate is lower than used elsewhere. The reason for this difference has to do with the roots of the implementation of DTT in Spain. The initial configuration was planned to be an SFN network in the whole territory. The advantages of a network of this type are obvious, because for the user is much simpler to use. In addition, the effective coverage is usually greater, since receivers can be fed with the highest quality signal (usually the strongest) when several antennas are emitting the same signal. However, at the time of designing the network it was decided to use the most extreme value of the Guard Interval in order to maximize coverage using the minimum number of broadcasting centres. But that implies a notable reduction in the effective bit-rate that can be used on each frequency, leaving the value at 19.9Mbps. And this is below the most common 22-24Mbps used in other territories. So now in order to achieve those much more streamlined values it is necessary to make changes. And only two options are possible: either the guard interval is reduced, or the redundancy is reduced.

But altering the second parameter seems impractical because in that case the coverage decreases. In addition, many of the installations currently in use, which may be at their operating limit, could suffer signal drops. In fact, the author agrees that the current redundancy value (FEC 2/3) is correct. However, the guard interval should be changed. And there is a good reason to do so. Many of the broadcasting centres in use today are not SFN networks covering large areas. This is because some time ago changes were made in the distribution network to dismantle the use of SFNs at the national level and move to regional/local signal distribution. So currently in many cases a few, or even a single emitter, are covering one area. And in such circumstances it is very unusual to exceed the maximum distance between transmitters by using a greater value of the guard interval. Therefore, with a minor reconfiguration of the network and a minimal investment in a few additional transmitter centres, 10-20% more bandwidth could be gained. Which would be an important advantage that would be worth considering.

3.4.1.2 Packaging of the broadcasted services

From the inception of digital television specifications, the concept of *Service Bouquets* has been part of the standards. These bouquets or packages are usually linked to distribution networks in order to group services according to operators. This feature is essential in the operation of satellite broadcasts, where many operators share a single orbital position. But it is also widely used in cable and terrestrial networks for other reasons. Basically the grouping of services allows extended functionalities that are otherwise not available, such as channel management (LCN), shared programme guides (EPGs), and other similar ones.

For example, the use of LCN descriptors greatly simplifies the use of digital television for users. One of the problems most criticized by DTT users in Spain is the arrangement of channels. At present, when a search is performed, services are stored without any logic. And the complex user interfaces of both TVs and STBs are often useless for sorting channels. This was in many cases the biggest problem detected by users during the initial DTV transition. But exactly the same problem has been repeated in subsequent re-tunes in the receivers due to changes made in the distribution networks. And in fact it is not an exaggeration to say that from the point of view of the user experience, this is the biggest problem of digital television in that territory. However, technically this problem does not exist, as all receivers are able to sort services according to a broadcasted list. And this is being used without any problem in other territories such as France or the United Kingdom, where users are not aware of such problems. But the Spanish regulator does not enforce the use of this capability. We can assume that the reason for this is undoubtedly the lack of consensus at political and commercial level among the operators. From any other point of view, however, it is a serious mistake not to use this capacity, which unnecessarily penalises users. And for the same reason in other territories this feature is always used.

Another similar case, although this is already a minor problem, concerns program guide services (EPGs). Currently, EIT data from EPGs are sent only at the level of the same multiplex. The reason for this again is the lack of regulation on coordination. And this simplistic multiplex coordination model unnecessarily limits user accessibility. Now in order for a user to have all the EPG information available on his device, s/he must navigate through all the received frequencies. But this is a fact that most users are unaware of. They misunderstand why the EPG service does not work properly on their terminals. And this is artificially creating an access barrier for users to use this enhanced service; and the worst thing is that on a technical level the problem does not exist. The feature of sharing EIT data between different frequencies is perfectly defined and easy to implement. When deployed, it is sufficient to tune into any television channel so that all data from all services are available. This greatly improves user accessibility and should therefore be a priority for the regulator.

3.4.1.3 Multiplex management

Since the initial deployment of the DTT in Spain, the regulator imposed that the different broadcasters operating the same multiplex should agree on its technical operation. However, this approach prevents maximising the capacity utilisation of a multiplex when it is not managed by a single operator (which is the case for at least half of the multiplexes in use). This implies that, for example, at the technical level, all the services distributed in the Spanish DTT are CBR; with the only exception of those which are broadcast in a multiplex where all the services are from the same operator. Only in those cases the statistical multiplexing is enabled. This results in insufficient video quality for some services, when it would not be the case if statistical multiplexing were enforced on all multiplexes.

In addition to the abovementioned, the model in which the regulator establishes digital TV licenses is not very well suited. The current model establishes a license to the exploitation of 25% of the space of a multiplex for one service, and then imposes some restrictions to its use. From a technical point of view, this approach does not help to optimize the use of the multiplex. The reasons are simple to explain. On the one hand, bandwidth is wasted on non-essential services (data, radios, etc.) because it is permissible to do so (up to 10% of the assigned space). On the other hand, the empty space used for padding after multiplexing services is around 5-10% of the total. This leads to a significant waste of the multiplex, which can be easily verified when comparing the multiplexes managed by a single operator with the other multiplexes. Therefore, there is no doubt that better management would be beneficial for users, since, for example, it would be possible to simulcast services that would otherwise only be accessible with a single technology.

3.4.1.4 Rules for service migration

Another consequence of the implantation of a non-centralized multiplex management model is that service migration is very complex. For example, simple operations such as moving a service from one frequency to another, adding a new service or restructuring part of the network are such complex tasks that they require a long and carefully planned procedure. This contrasts with the ease that a satellite operator, for example, can add and remove services, make changes to the channel network, add or remove frequencies, etc. All these tasks, which are only technical adjustments that are fully covered by DTV standards, can be performed by the operator without user intervention.

However, in the case of Spanish DTT, such changes are very difficult or even impossible to carry out in a transparent way. And this is not logical when the necessary technical tools exist to ensure that users only have to turn on their devices and receive automatic updates of the configurations.

Furthermore, the problem is aggravated when it is proposed to initiate a transition process to another generation. In that case it is convenient to be able to move some services to new frequencies, in order to free up space for new services that use the new technology. But although it could be done in a way that is transparent to the user, it is now impossible. In addition, by using technical solutions already implemented it is possible to further simplify the migration. For example, when the HD Simulcast LCN identifier is activated then the transition will be even easier for the users. This identifier allows receivers with support for decoding HD signals to automatically replace the SD channel with the HD version, if this version is available. This means that users will start seeing the new services with better quality from day one if their equipment accepts it, and all without any manual intervention.

3.4.2 Analysis of the proposed solutions

These listed above, are the four most notable areas in which the author has detected deficiencies throughout the realization of this thesis. It should be noted, however, that the points listed above are only some of the problems that exist. The author does not claim that this is a complete list of all problematic areas; nor does he intend to suggest that they are the only ones that require attention or can be improved. These are simply the ones that generate relatively significant problems and yet could easily be improved.

In this regard, a number of potential improvements are detailed below. These are improvements that from a technical point of view could be implemented without major problems. However, they fundamentally require changes at the regulatory level. In fact, virtually none of them can be implemented without the Spanish regulator making changes to its recommendations. However, this is far from being a problem, but rather an advantage. The reason is that the regulator bases its decisions partly on the recommendations of the technical working groups that collaborate with it[12]. And some of these problems have already been identified. Therefore, it is feasible to consider the recommendations analyzed here within these forums. This could perhaps encourage the regulator to work in that direction.

Entering into this analysis, among the particular improvement actions suggested by the author, the following five are listed:

- Migration to HD H.264 DVB-T using the Joint Simulcast.
- Changes in the management model of the multiplexes.
- Modifications to the technical parameters of Multiplexes.
- Joint planning for upcoming migrations (DVB-T2 and UHD).
- Alternative user access from other transport networks (TDT-SAT).

[12] See the documents of the *Technical Forum of the Digital Television group of the Spanish Government / Foro Técnico de grupo de Televisión Digital del Gobierno de España*.

Each of these points is now analysed in detail.

3.4.2.1 Migration to HD H.264 DVB-T using the Joint Simulcast

Currently in the Spanish DTT, an independent simulcast is used to broadcast only some services in HD. This partial simulcast has been running for many years, in fact for almost a decade. So actually the Spanish DTT has been in a migration process for a long time. But this situation have no sense. For example, in other countries the migration to the next generation, using HD H.264 services with DVB-T, is complete or nearly complete. A notable case is the neighbouring country of France. In that territory, from 2016 all broadcasts are made using only the H.264 codec. So the fact that the first generation is still being used in Spain today, when in fact digital TV services started 20 years ago, makes it certainly a rather peculiar case. But worst of all, this suboptimal situation is likely to persist for a few more years. According to the recently published plans for the second digital dividend[13], it seems that SD signals in MPEG-2 will not be eliminated until 2022[14]. However, this decision may not be very successful as it will prevent the full deployment of all HD services for another two years. The underlying problem will be that until that date, without direct action to provide a solution, about half of the television services will continue to be accessible exclusively in SD. This will certainly can further increase the serious delay in the development of digital terrestrial television suffered in this area. And this delay is more evident when comparing with other countries, such as the United Kingdom or Germany, where third generation services (H.264 HD with DVB-T2) are operating. So unless something is done, the problem will get worse. Being anchored in a completely obsolete technology is not good, so any solution addressed to solve this problem could compensate for the lost time. In that sense, the technology of the Joint Simulcast proposed in this thesis can be very positive to accelerate the migration to services only in HD. The key is that using this solution right now it would be possible to easily switch all services to HD, without having to turn off SD services yet. That would most likely move the public to upgrade their television equipments, and when the switch-off date comes, there will be no problem in completing the switch-off the MPEG-2 signals.

At this point the opportunity factor for implementing such solution is significantly relevant. The reason is that currently the number of legacy receivers that only support MPEG-2 SD is still significant, so eliminating such broadcasts at this time would create social alarm. However, it is equally true that all products currently on the market are all compatible with one of the next two generations, that is all with H.264 HD, but with DVB-T or DVB-T2. It is therefore reasonable to implement as soon as possible a solution that allows all services to be received in HD. In fact, the only reason why this is not being done at the moment, as studied in this thesis, is because the shortage of available bandwidth. But by applying the Joint Simulcast solution it is expected that the migration of users to HD will be rapidly accelerated. Leaving only a residual amount of them for the 2022 date. All this without forcing anyone to complete the migration, and allowing for a non-traumatic change. With the added advantage of being in a better position to start the next migration on the date of the SD MPEG-2 codec removal.

---

[13] See the document https://www.televisiondigital.gob.es/2DD-5G/Documents/plan-actuaciones-2DD5G.pdf .

[14] See the document https://avancedigital.gob.es/es-es/Participacion/Documents/Proyecto-RD-TDT-segundo-Dividendo-Digital/Proyecto-RD-%20Plan-Tecnico-TDT.pdf .

In addition, it must be taken into account that the economic costs of implementing the Joint Simulcast solution are very low. This is because the changes are only at the headend level, specifically in the coding equipment, so the investment is minimal compared to the benefits. Furthermore, if the application of these changes was made to overlap with other re-tuning processes, it would make their application practically transparent. Therefore, on the basis of the opportunity cost and the benefits obtained, it does not seem reasonable to refuse to apply this option as soon as possible.

3.4.2.2 Changes in the management model of the multiplexes

If the Joint Simulcast solution is selected, it would be useful to make an additional change to make the result as efficient as possible. This change would involve a modification in the management of the Spanish DTT multiplexes. Instead of continuing to use the current uncoordinated management model, results would be improved by opting for centralized management. Basically, the new model, which is used in other countries, consists of replicating the operations of a network operated by a payment broadcaster on a vertical platform. The objective is not to establish a pay-per-use scheme, but to avoid the current ecosystem in which each of the multiplexes is managed as a separate unit. In the new model, all television programs are delivered to a single network operator by contribution feeds. The network operator then packages the programs into digital television services and distributes them through the corresponding networks (national and regional). This allows services to be aggregated on a purely technical basis to improve efficiency; rather than simply allocating 1/4 of a multiplex for each of the broadcasting licences. Furthermore, this opens up the possibility of using the technical features for auto-configuration of the receivers. This would greatly simplify actions such as switching services from one frequency to another. The only requirement to be able to work in this way is to define the figure of the network operator and to assign the licenses according to a broadcasting right of a digital television service.

An interesting fact is that making this change would not be difficult. At present, there is already an operator who operates nationwide and is responsible for the transport of all signals with national coverage. It manages the entire distribution network for these signals. Therefore, the necessary infrastructure for such centralized management is already de facto in place. It would therefore only be necessary to define the figure of that network operator at the regulatory level and then directly begin to carry out the management in a coordinated way.

And from the point of view of the users, this change of model has many advantages. On the one hand, certain aspects of the configuration of the broadcasts would be fully automatic: the ordered list of the services (main defect perceived by the users of the Spanish DTT), the automatic re-tuning of programs in case of changes, the global EPG service, the automatic selection of the HD versions in case of being present and compatible with the receiver, etc. On the other hand, it would be possible to maximize the full use of all network capabilities, which would result in the delivery of better quality to all users. This in turn would alleviate the pressure of the challenges associated with the migration of services to future generations when a new technology becomes available. This would be true because the necessary changes in the broadcasting network could in many cases be virtually transparent to users. And additionally that would make it much easier to free up space, or make other relevant changes, to start emissions using new standards. Essentially, all this would be possible simply because the services, by being centralized, would be able to move between frequencies and the parameters within the network would change transparently. However, nothing of this is possible under the current regulation, so a change would be very necessary.

3.4.2.3 Modifications to the technical parameters of Multiplexes

The change proposed above in the management model would allow an easy modification of the parameters of the broadcasting network without affecting the users. This would open the possibility to plan ways to readjust the technical parameters of the network —always in a transparent way for the users— that would allow to increase the effective bandwidth of the different multiplexers. The objective would be to obtain more free space to facilitate the inclusion of the simulcast of all the services currently present in the DTT network. This is technically feasible because the values currently used were established more than 20 years ago with a particular objective that is not being met now (basically the goal was to use nationwide SFN networks). This increase in free space could now be achieved by changing certain parameters such as the guard interval, which would not reduce the robustness of the signal, and therefore the same coverage levels would be achieved. It should be mentioned, however, that making this particular change could mean in some specific cases having to incorporate some additional transmitting tower, due to the decrease in the effective separation distance between the antennas. But this would surely involve a perfectly acceptable cost. But it should still be taken into account that making changes in the parameters of the broadcasting network without disturbing the users necessarily implies that the management model has been modified before, in any other case such changes would be very difficult to carry out. Moreover, if this were not done, it is very likely that the holders of broadcasting licences would be clearly penalised.

Thus, in order to achieve this readjustment in the broadcasting network, it would be necessary to start first with the frequencies using the DVB-T standard (currently all of them); but then this would serve to initiate a smooth migration to the DVB-T2 standard (instead of doing a full or partial block-based migration to establish this transport standard). The increase in bandwidth proposed by the author would be based on a small change in the guard interval from the current 1/4 value to a value of 1/8. This would increase the effective bit-rate to 22.12Mbps from the current 19.90Mbps. The only problem with this modification is that the maximum distance between the transmitters would change from the current 67.2 Km to only 33.6 Km, because the maximum time between symbols would be reduced from 224µs to 112µs. And as explained above, this could imply the need to install some additional transmitting antennas. However, once a centralized broadcast network is in operation, it is relatively easy to schedule readjustments without affecting users. A simple study of the impact of this change on the current network could then quantify the total cost. However, it is quite easy to anticipate that since the network is currently divided into 75 geographical areas, each of which forms its own SFN, the expected technical and economic effort would be relatively low.

As for the free space gain in each multiplex (which would represent a significant 10% increase in bandwidth) it could then be used in different ways. On the one hand, some of the currently distributed services could be packaged more efficiently by decoupling them from the current 25% usage model within a multiplex. That is, services could be distributed in such a way that there would be 5 or more within a single multiplex (not including the simulcast versions). This would facilitate the introduction of the simulcast versions of the total of all services, since the Joint Simulcast used would not need to be very aggressive regarding the compression values. In addition, this would even free up space for other uses. For example, being the case that currently 7 multiplexes operated at 20Mbps are used with national coverage, then the total bandwidth available is 140Mbps. So by switching to multiplexes with sizes of 22Mbps only 6 of them would be needed to reach almost the same bandwidth (132Mbps). Thus, the released frequency could be reused for other uses, such as being used to initiate broadcasts using the DVB-T2 standard with national coverage from the first moment (i.e., without adding new broadcast equipment).

3.4.2.4 Joint planning for upcoming migrations

The potential flexibility offered by a centralized management of the broadcasting network is a very important factor to take into account. The change to a centralized model not only helps to improve efficiency in the current scenario, but can also bring benefits in the future. For example, in addition to the mentioned above, when this type of management is introduced, migration to future generations can be improved. The main reason behind this is that it provides simpler and more effective means of reducing transition periods, which will undoubtedly benefit in accelerating the adoption of the new technologies. Thus, taking the two upcoming migrations as examples, it is possible to analyse how these could be carried out. Note that the first of these future migrations will be the move from using the current DVB-T transmission standard to the more efficient DVB-T2, but maintaining the H.264 codec for video encoding. While the next will be to jump to the UHD using the H.265 codec and leaving the use of HD services encoded with the H.264 codec. See below how these future changes could be more easily accomplished.

The first aspect to consider is that if nothing is done to simplify and/or accelerate the current transition —the switch from SD MPEG-2 to H.264 HD— the scenario is potentially very dark. On the one hand, by not doing anything the end of the transition period will be extended in time. This is likely to delay the introduction of the next generations, which would result in a slowdown in technological development and could have a negative impact on reducing the use of digital terrestrial television among users. However, if some changes, such as those suggested in the previous points, are intelligently implemented, it is possible that future changes can be implemented more quickly. The main reason for this is that once the efficiency in the use of the broadcast network has been maximized (either by a better packaging of the services, or by the increase of the bandwidth, and also by the feature of being able to make changes in the network parameters in a transparent way for the user), then it is possible to make the introduction of any new technology much easier. This is obviously so, because by minimizing the difficulties it is much easier to bring forward the time when changes are introduced, and therefore transition times can be expected to be much shorter. So below we see the possible specific scenarios that could occur with regards to the Spanish DTT.

A first scenario to discuss would be to have implemented in the near future all the suggestions proposed by the author. Thus, in an ideal scenario, it would be possible to effectively eliminate the compatibility with the first generation at the expected date of 2022. That is, all SD services with MPEG-2 would be switched off by that date. This could be done in a non-traumatic way by applying the joint simulcast described in this thesis. This would be because the implementation of this solution would possibly result in that the number of devices that are not compatible with H.264 HD services would be virtually residual by that time. The reason that could be expected is because all services will already be available in their HD version, and since the quality of the SD versions has been reduced below the average, then users would be driven to upgrade their equipment, although not forced to do so. Then, and as long there were no changes in the DTT broadcasting licences at national level, the Spanish DTT would be expected to have the following characteristics:

- 7 nationwide multiplexes in DVB-T with a bandwidth of 22.1Mbps.
- 26 programs in full HD delivered only with H.264 codec.
- Centralised management of all services within the broadcasting network.

The last of these points is actually very important because, as explained above, it is a necessary condition for a smooth migration. From this scenario, some assumptions can then be formulated in order of planning the next migration to the use of DVB-T2 in the broadcasting network. First, there would be new free space in the network due to the switch-off of SD services. But because the use of the joint simulcast eliminates the need to duplicate the shared parts of the services, then in fact only the space occupied by the SD video stream is released. In addition, when aggressive compression is used, the space occupied by SD services is relatively small. And since it is also the case that the quality of HD services will be slightly below the common standards, the reality will be that not so much space will be released. Therefore, the first option when removing the SD services would be merely to occupy all the space with the HD services at maximum quality. That option, although attractive in principle, might not really be the best option. We should look at the reasons why it would not be.

If it is decided to keep only HD services with H.264 codec after the SD services switch-off, the migration to the next generation of digital TV could be seriously delayed. In fact, if the introduction of the DVB-T2 standard is not started quickly, the move to UHD services could be very difficult. The most important reason is that there will be no incentive for either users or broadcasters to embrace such technology, no matter how much compatible equipment is offered by the industry. Therefore, it seems much more appropriate to focus on a quick transition to the DVB-T2 standard in order to prepare the existing base of receivers for the use of UHD services. In this way, the migration periods could be greatly shortened because the changes towards future generations actually overlap.

Based on this idea, a much more ambitious roadmap would consist of the following phases for future migrations:

- Phase 1: Reallocating of H.264 HD services with DVB-T.
- Phase 2: Progressive migration to DVB-T2 with HD/UHD simulcast.
- Phase 3: Services only in UHD with H.265 and DVB-T2.

As can be seen, the ultimate goal is to migrate all services completely to the most modern generation, i.e. to the use the DVB-T2 standard to broadcast all programs in

UHD quality. Doing this would represent a recovery of all the time lost in Spanish DTT, as the last phase could be reached in a relatively short time. However, achieving this goal in a non-traumatic way and with the agreement of users and broadcasters is not directly achievable. Therefore, the author proposes a different approach to this transition so that the time needed to complete it is not excessively long.

For the proposal to be technically feasible, the above-mentioned roadmap should then be divided into the following sub-phases:

- Phase 1.0: Reallocating of H.264 HD services with DVB-T.
- Phase 2.1: First partial migration to DVB-T2 and starting of HD/UHD simulcast.
- Phase 2.2: Second partial migration to DVB-T2 with added HD/UHD simulcast.
- Phase 3.1: Total migration to DVB-T2 only and complete UHD/HD simulcast.
- Phase 3.2: Shutdown of H.264 HD services and H.265 UHD only with DVB-T2.

It should be noted that the duration of each of these phases need not be predetermined in principle. Possibly the degree of deployment of the receiving equipments would be the best way to establish the maximum and minimum duration. However, the technical actions to be implemented would be well established. These would be the following.

3.4.2.4.1 PHASE 1.0

The objective of this first phase, once the MPEG-2 SD services have been switched-off, is simply to rearrange the existing multiplexes to accommodate the maximum number of services in the minimum space. This will release entire frequencies that could be reallocated for DVB-T2 migration. In other words, this phase would only be the prelude to the implementation of the next phase. And it must be noted that doing this so quickly would not be a problem for the majority of the users because the centralised management of the network would allow the changes to be made in such a way that the receiving equipment would be self-configured to the new distribution of channels. Therefore, as there are no changes in the frequencies used, it would only be a process of thin channels search. The result of applying this change would configure the network as follows:

- 4 multiplexes in DVB-T at 22.1Mbps.
- 3 empty multiplexes.
- 26 programs in total distributed as follows:
  - 2 multiplexes with 6 services in full HD (all in H.264).
  - 2 multiplexes with 6 services in full HD, plus 1 HD in 720/25p (H.264).

The most important fact is that with this change the total number of services would remain the same, i.e. 26. And as for quality, with the exception of two services with reduced vertical and temporal resolution (which could correspond to two public services, one of 24h news and another exclusively of cartoons, which would therefore be minimally affected by this change in resolution), all the others would be in full HD resolution. It is then understood that the change would simply be in the location of the services. However, the level of compression would be more aggressive than usual, since for a multiplex of that size the usual number of programs is 5. And this should in principle have a negative impact on the quality of services. But the reality is that until this moment the space of the multiplexes —with that same size— was being shared with the SD services, which implied that the level of compression was already bigger than usual. Therefore, in practice, the differences in quality by making this change would be minimal. Only two of the multiplexes would require additional space for the two low-resolution programs. However, on a technical level, applying the aggressive compression techniques presented in this thesis, it would be perfectly feasible to handle this distribution of services.

3.4.2.4.2 PHASE 2.1

The next phase, immediately following the previous one, which means that they would actually be carried out together —although separated here for better understanding—, would consist of launching the simulcast of some programs in UHD format. And while this may seem premature, it would make perfect sense. We will explain the reason. This would be done by implementing H.265 encoded UHD services distributed over multiplexes using the DVB-T2 standard. This would be done by reusing the free frequencies left over from the previous phase. Therefore, no significant additional network investment would be required to migrate part of the network, as no new frequencies would need to be added. Moreover, this would not be a problem for users either, as no adaptation would have to be made to the antennas. Thus, assuming that a configuration of the multiplexes with a nominal capacity of 33.3Mbps was chosen for the operation of the network (the configuration corresponding to DVB-T2 more similar to that used then in the other DVB-T multiplexes), then the following organization would be obtained for the Spanish DTT:

- 4 multiplexes in DVB-T at 22.1Mbps.
- 3 multiplexes in DVB-T2 at 33.3Mbps.
- 26 total programs, 26 HD services and 12 UHD services, distributed as follows:
  - 2 multiplexes DVB-T with 6 services in full HD (H.264).
  - 2 multiplexes DVB-T with 6 services in full HD, plus 1 HD in 720/25p (H.264).
  - 3 multiplexes DVB-T2 with 4 services in UHD 2160/25p (with H.265).

This change allows that after SD services are disconnected, then the broadcasting can be started directly with the latest generation available. Thus overlapping the migration between the next two generations. Which would certainly be very interesting. A fundamental aspect is that the 12 new services are not new programs, but only a simulcast of some of the current services. And in fact these would represent almost half of all distributed programs. So there would be a significant advantage. And this feature is expected to attract both users and broadcasters. The users because they will be able to directly have the services already deployed with more quality (even if it is not with the full UHD quality). And to broadcasters because it will not force them to update their production channels, as only those interested in making the move to UHD production may have the opportunity to do so, which will be possibly for the programs with higher audiences. Consequently, change is likely to be very attractive to everyone.

However, since this is a transition period, the quality with which these UHD services are offered will not be at their maximum. In fact, the concepts of aggressive compression such as those proposed in this thesis for the SD/HD simulcast will have to be applied. It must be taken into account that the current recommendations for the use of UHD technology indicate a bit-rate of no less than 10Mbps. Therefore, for the suggested size for the DVB-T2 multiplexes the number of services is being increased by 1 more in each one. But this should not be a problem. It is more than likely that the early introduction of UHD services will attract more users than if there were no such services at all. This could translate into a larger installed base of compatible equipment with the latest standards (H.265 and DVB-T2), which as explained in this thesis. As a consequence, this reduction in initial quality, which can be compensated through an aggressive compression, is not expected to be a problem at this stage. Especially since none of the HD services have been removed, and all users continue to have access to all programs.

3.4.2.4.3 PHASE 2.2

At a later undetermined stage, when the installed base of UHD and DVB-T2 compatible devices is sufficiently large, and at the same time the efficiency of H.264 compression has already reached its maximum, a second phase of service migration could be initiated. Note that from time to time the efficiency of professional compression equipment usually improves, and the H.264 codec is still capable of higher efficiency rates than the present ones. This phase would then consist of a simple change in the use of one of the multiplexes used with DVB-T to DVB-T2. The advantage of making this change is that it would increase the number of services in the UHD simulcast to 16 (4 more services). But at the same time it would mean relocating the 26 services in HD in only 3 DVB-T multiplexes. Then, the configuration of the Spanish DTT will be as follows:

- 3 multiplexes in DVB-T at 22.1Mbps.
- 4 multiplexes in DVB-T2 at 33.3Mbps.
- 26 total programs, 26 HD services and 16 UHD services, distributed as follows:
    - 1 multiplex DVB-T with 8 services in full HD (H.264).
    - 2 multiplexes DVB-T with 9 services in full HD (H.264).
    - 4 multiplexes DVB-T2 with 4 services in UHD 2160/25p (with H.265).

A relevant factor about this new reallocation is that it should not be problematic at all for the following reasons. On the one hand, again the movement of services between frequencies would be almost transparent for users. On the other hand, as there is a much larger pool of users ready to receive UHD services, the potential impact on quality for the HD services would not affect all users. And finally, as the efficiency of H.264 coding has increased, performance should not be significantly compromised. In fact, if necessary, and if the time from the previous phase is very long and the number of UHD compatible equipment is very high, the spatial and temporal resolutions of HD programs could be reduced to avoid compression problems. Therefore, the setting of 8 and 9 services in a single DVB-T multiplex would be perfectly tolerable. And furthermore, with regard to UHD services, based on future technology development, it may also be possible to move them from 2160/25p to 2160/50p.

3.4.2.4.4 PHASE 3.1

This new phase solves the problem of the over-compression of HD services with the H.264 codec. Although users will not be required to migrate to the new technology, it will be true that priority will be given to UHD services. Therefore, from that moment the HD services in DVB-T will start to be residual. Therefore, the loss of quality of these services due to over-compression should not be a problem. However, despite this, the above phase should not be kept active for very long time. The reason is that a significant number of users without UHD compatible devices may be able to access DVB-T2 signals, which will then only be used for UHD simulcast. But technically it would be possible to distribute the same services in HD over those frequencies. Furthermore, it is very possible that by the time this phase is introduced the number of receivers without DVB-T2 support will be very low. It is therefore clear that the introduction of this new phase is focused on eliminating the second generation equipment, i.e. only those that support HD H.264 and DVB-T. That way the Spanish DTT configuration would become in this phase less complex:

- 7 multiplexes in DVB-T2 at 33.3Mbps.
- 26 programs in full HD/UHD simulcast, distributed as follows:
    - 5 multiplexes DVB-T2, 4 services in UHD (H.265) and 4 in HD (H.264).
    - 1 multiplex DVB-T2, 3 services in UHD (H.265) and 5 in HD (H.264).
    - 1 multiplex DVB-T2, 3 services in UHD (H.265) and 6 in HD (H.264).

From this moment it can be seen that the complete package of programs will be accessible through the most modern technology. In addition, the quality of UHD services could be improved by increasing the temporal resolution to the common rate of 50fps, if that has not been done before. And at this point, a joint simulcast could be applied to all services to maximize the network capacity and reduce HD services to the video stream only. This will also maximize overall efficiency, allowing this configuration to be maintained over time as long as necessary until the next phase is reached. However, it should be noted that by applying a joint simulcast it will also be possible to vary the priority in the quality between the HD and UHD services. Therefore, at the beginning of this phase, it would be possible to maintain the HD services with high quality, without increasing the quality of the UHD services, and adjusting the values until the HD services become residual. This would make a lot of sense as it may not be necessary at first to make any changes in order to increase the quality of the UHD versions compared to the previous phase.

3.4.2.4.5 PHASE 3.2

Finally, a last phase would be reached in which HD services with H.264 codec would be completely removed. The configuration of the Spanish DTT would then be:

- 7 multiplexes in DVB-T2 at 33.3Mbps.
- 26 programs in total distributed as follows:
    - 5 multiplexes DVB-T2 with 4 services in UHD (H.265).
    - 2 multiplexes DVB-T2 with 3 services in UHD (H.265).

The quality of the services would then be maximum, since all the space could be used for them and the capacities of the H.265 compressors would have been improved. And all this will possibly have been achieved in a relatively short time, perhaps as early as 2030. However, it is necessary to comment that migrations do not necessarily have to end at this point. In fact it is quite possible that the space released now could actually be used for other purposes. Perhaps to deploy the simulcast of the services with new available standards. In fact, the opportunity to count with free space would open the possibility to migrate in a smooth way to the next new generations of digital television that will be developed in the future. Furthermore, in the case of being able to have some intergenerational compatibility in the new standards, which would certainly be desirable from the point of view of the author, then it would be even easier to be able to make those changes.

In any case and as a final summary the following list shows the distribution of the available bandwidths in the proposed phases, which helps to better understand the development along these transition phases:

- Phase 0 (current): **155Mbps** (23 SD MPEG-2 + 11 HD H.264).
- Phase 0.1 (joint simulcast): **155Mbps** (26 SD MPEG-2 + 26 HD H.264).
- Phase 1.0 (SD swtich-off): **88Mbps** (26 HD H.264) + **100Mbps** (free space).
- Phase 2.1 (UHD sim I): **88Mbps** (26 HD H.264) + **100Mbps** (12 UHD H.265).
- Phase 2.2 (UHD sim II): **66Mbps** (26 HD H.264) + **133Mbps** (16 UHD H.265).
- Phase 3.1 (DVB-T2 only): **233Mbps** (26 HD H.264 + 26 UHD H.265).
- Phase 3.2 (UHD only): **233Mbps** (26 UHD H.265).

And as can be noticed, it is a fairly smooth and technically feasible roadmap. The author trusts in this way because it does not present major drawbacks. It should only be taken into account that it is essential to make a centralized control of the broadcasting network. This implies changing the regulation to make it possible. However, if the regulator decides to take this route, which it can do without legislative constraints, then there would be no major problems in carrying it out. And from the perspective of the author it would then be an outstanding way to make use of the knowledge acquired throughout the realization of this thesis.

3.4.2.5 Alternative user access from other transport networks

A last but not least important section is related with the coverage of the Spanish DTT. To date, the coverage of the DTT in Spain is quite good. However, it does not reach 100% of the territory. The solution adopted to cover the shaded areas is somewhat particular. It is based on using satellite distribution to provide coverage to users in that shaded areas. However, the signals received by satellite are not specific to that service, but rather are reused. In other words, they are the link feeds that are used to feed the broadcasting towers. The nature of these feeds makes it impossible to give open access to them, so the system adopted requires the use of equipment specifically designed for the purpose of receiving such signals. However, this has many drawbacks. The first is that users cannot access this service freely. Only by requesting the installation of official installers is it possible to obtain the necessary equipment to receive these signals. And the installers must first certify that there is no terrestrial coverage, and then submit a request to the regulator, who will approve or reject it. However, such equipment is simple, low quality set-top-boxes that only provide a basic functionality, and only connected to a single television device. It is therefore obvious that such model is far from offering an universal access to digital television services. Which unnecessarily creates a gap between citizens who have coverage and those who do not.

A potential solution to this problem is however actually possible. The proposal that the author suggests is based on following the same model applied in other surrounding countries. For example, in France - as in other territories - it is possible to have free access to equipment that technically provides access to the DTT service via satellite. The basic requirement for purchasing the equipment is simply to be a citizen of the country and to be able to prove it. It is thus possible to acquire any of the equipment certified to receive the signals, which is based on the encryption of the signals as carried out by digital pay-TV operators. In other words, simply by introducing a state-owned digital pay-TV operator offering DTT's signals by satellite access, it is possible to guarantee access to the entire population. This model has many advantages. On the one hand, nobody is discriminated against with regard to access to this universal service because of their area of residence. On the other hand, it ensures that the digital television service is technologically neutral, as it does not require the use of any specific and limited equipment. It is therefore possible to use any of the equipment available on the market in accordance with the users' preferences. And finally, access management can be easily controlled by using the conditional access solutions available on the market, which are well tested. Furthermore, this also increases protection against potential abuses against the signal.

However, nothing of the described is possible with the current DTT-by-SAT service in Spain. Therefore the author suggests the following change based on the model described above and already implemented in other countries. Assuming that the above suggestions have already been implemented, at least the one concerning the centralised management of the broadcasting network, then the figure of the network operator will be operational. It would then only be necessary for this operator to distribute the signal via the satellite and protect it as if it were a pay-TV carrier. But as explained, such operator already exists and the signals are already being distributed via satellite. Therefore, the change would basically imply changing the current system of protection from an ad-hoc solution, which there is evidence of having been compromised, to a more standard model of conditional access television system. For the application of this new model, and to ensure greater flexibility, we recommend the use of a multi-CAM conditional access, which allows the use for any model of device regardless of the manufacturer. This would be much more economical and beneficial to users. And that would also eliminate the dependence on the technology used with the access service. This means that if the conditional service is embedded in the receiver, then the signals must be specifically compatible with them, so it will not be possible to migrate to a new generation when it becomes available. However, by using an external conditional access system then it will be possible to de-link the service from the access equipment. In this way the use of this satellite transport technology will be more similar to terrestrial access, as users will be able to upgrade their equipment freely to access to the features provided by new generation services. All without worrying about conditional access, which will be compatible with all generations.

This last point is certainly critical to ensure that any new movement towards a next generation is not disrupted by the use of DTT-by-SAT. For any public service with universal access, it is necessary to ensure that all citizens have access to it, as the regulation does not allow discrimination per se of any citizen. Therefore, the only way to ensure that the complementary access to DTT via satellite is always compatible with future technological changes that are likely to be incorporated into the terrestrial distribution network, is precisely to make this access as neutral as possible. And this can only be guaranteed by using standard conditional access not linked to the equipment being used. Commercially this solution already exists, and it is the previously mentioned multi-CAM model. It is therefore very important that in addition to the changes suggested in the previous sections, this model is also adopted for the DTT-by-SAT service for the benefit of all.

3.5 Final comments

The author of this thesis wishes to personally thank the support received by his thesis director during all the time he has been developing the hard research work that this thesis has entailed. In addition, he asks any potential reader of the thesis to contact him if they would like to collaborate in the development of solutions for intergenerational compatibility in the field of digital television. Personally, the author is already working on two future lines of research: the HD/UHD simulcast based on the scalable extensions of the H.265 codec, i.e. with the SHVC codec (scalable HEVC); and the DVB-T/DVB-T2 mixed modulation based on the use of FEF and PLP. The first one seems to be the most promising line of research, since the support for a combination of a main stream in H.264 together with an extended stream in SHVC, is a pre-existing option already included in the SHVC standard. However, as with the Aggressive Joint Simulcast solution developed in this thesis, it is still necessary to analyze how to deploy such solutions in a way that is transparent and commercially viable.

Bibliography

Paper 1 references:

[1] Brown, Allan, and Robert G. Picard. *Digital terrestrial television in Europe*. Routledge, 2004.

[2] Goldman, Matthew S. "Migrating from MPEG-2 Video to MPEG-4 AVC." *SMPTE Annual Tech Conference & Expo, 2010*. SMPTE, 2010.

[3] Wu, Yiyan, et al. "Overview of digital television development worldwide." *Proceedings of the IEEE* 94.1 (2006): 8-21.

[4] Rancy, Francois, Elmar Zilles, and Jean-Jacques Guitot. "Transition to digital TV and digital dividend." *2011 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS)*. Vol. 1. IEEE, 2011.

[5] Garcia Leiva, Maria Trinidad, and Michael Starks. "Digital switchover across the globe: the emergence of complex regional patterns." *Media, Culture & Society* 31.5 (2009): 787-806.

[6] McCann, K., et al. "Technical evolution of the DTT platform: An independent report by ZetaCast, commissioned by Ofcom." (2012).

[7] Toth, Arpad G., M. Tsinberg, and C. W. Rhodes. "NTSC compatible high definition television emission system." *IEEE transactions on consumer electronics* 34.1 (1988): 40-47.

[8] Uz, Kamil Metin, Martin Vetterli, and Didier LeGall. "Interpolative multiresolution coding of advanced television with compatible subchannels." *IEEE Transactions on Circuits and Systems for Video Technology* 1.ARTICLE (1991): 86-99.

[9] Naveen, Thumpudi, et al. "Comments on" Interpolative multiresolution coding of advanced television with compatible subchannels"(and reply and additional comments)." *IEEE Transactions on circuits and systems for video technology* 2.1 (1992): 95-100.

[10] Polley, Michael O., Susie J. Wee, and William F. Schreiber. "Hybrid channel coding for multiresolution HDTV terrestrial broadcasting." *Proceedings of 1st International Conference on Image Processing*. Vol. 1. IEEE, 1994.

[11] Ohm, J-R. "Advances in scalable video coding." *Proceedings of the IEEE* 93.1 (2005): 42-56.

[12] Kouadio, Adi, et al. "SVC–A highly-scalable version of H. 264/AVC." *European Broadcasting Union, EBU Technical Review, Edition* (2008): Q2.

[13] Okubo, Sakae. "Requirements for high quality video coding standards." *Signal Processing: Image Communication* 4.2 (1992): 141-151.

[14] Chiang, Tihao, and Dimitris Anastassiou. "Hierarchical coding of digital television." *IEEE Communications Magazine* 32.5 (1994): 38-45.

[15] Aign, Susanna, and Khaled Fazel. "Temporal and spatial error concealment techniques for hierarchical MPEG-2 video codec." *Proceedings IEEE International Conference on Communications ICC'95*. Vol. 3. IEEE, 1995.

[16] Schwarz, Heiko, Detlev Marpe, and Thomas Wiegand. "Overview of the scalable video coding extension of the H. 264/AVC standard." *IEEE Transactions on circuits and systems for video technology* 17.9 (2007): 1103-1120.

[17] Fraunhofer, “JSVM Reference Software (2006).” *JSVM*, Fraunhofer Institute for Telecommunications—HHI, Oct 2014, https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/image-video-coding/research-topics/svc-extension-of-h264avc/jsvm-reference-software.html.

[18] Stockhammer, Thomas, et al. "Information technology—MPEG systems technologies—Part 6: Dynamic adaptive streaming over HTTP (DASH)." *ISO/IEC, MPEG Draft International Standard* (2011).

[19] Sánchez, Yago, et al. "Scalable Video Coding based DASH for efficient usage of network resources." *Third W3C Web and TV Workshop*. 2011.

[20] Flierl, Markus, and Pierre Vandergheynst. "An improved pyramid for spatially scalable video coding." *IEEE International Conference on Image Processing 2005*. Vol. 2. IEEE, 2005.

[21] Yang, Wenxian, Gagan Rath, and Christine Guillemot. "Scalable video coding with interlayer signal decorrelation techniques." *EURASIP Journal on Advances in Signal Processing* 2007.1 (2007): 054342.

[22] Ogden, Joan M., et al. "Pyramid-based computer graphics." *RCA engineer* 30.5 (1985): 4-15.

[23] O'Leary, Seamus. "Hierarchical transmission and COFDM systems." *IEEE Transactions on Broadcasting* 43.2 (1997): 166-174.

[24] Segall, C. Andrew, and Gary J. Sullivan. "Spatial scalability within the H. 264/AVC scalable video coding extension." *IEEE Transactions on Circuits and Systems for Video Technology* 17.9 (2007): 1121-1135.

[25] Schwarz, Heiko, et al. "Constrained inter-layer prediction for single-loop decoding in spatial scalability." *IEEE International Conference on Image Processing* 2005. Vol. 2. IEEE, 2005.

[26] Marquant, Gwenaelle, et al. "Extended spatial scalability for non dyadic video formats: from SDTV to HDTV." *Visual Communications and Image Processing 2005*. Vol. 5960. International Society for Optics and Photonics, 2006.

[27] EBU-UER, “Current Status of High Definition Television Delivery Technology”, EBU–TECH 3328, May 2008, Geneva (2008).

[28] *see reference [54*]

[29] “AviSynth framework.” *AviSynth*, Oct 2014, http://avisynth.nl/.

[30] Sullivan, Gary J., et al. "Overview of the high efficiency video coding (HEVC) standard." *IEEE Transactions on circuits and systems for video technology* 22.12 (2012): 1649-1668.

Paper 2 references:

[31] Breebaart, Jeroen, et al. “MPEG spatial audio coding/MPEG surround: Overview and current status.” Audio Engineering Society Convention 119. Audio Engineering Society, 2005.

[32] Wolters, Martin, et al. "A closer look into MPEG-4 High Efficiency AAC." Audio Engineering Society Convention 115. Audio Engineering Society, 2003.

[33] Andersen, Robert L., et al. "Introduction to Dolby digital plus, an enhancement to the Dolby digital coding system." Audio Engineering Society Convention 117. Audio Engineering Society, 2004.

[34] ETSI/EBU, “Television systems; NICAM 728: transmission of two-channel digital sound with terrestrial television systems B GH LKI and L”, ETSI EN 300 163 V1.2.1 (1998-03), Mar 1998, Sophia Antipolis (1998).

[35] ISO/IEC JTC 1, “Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems — Part 1”, ISO/IEC 13818-1:2007 / Rec. H.222.0, Apr 2007, Geneva (2007).

[36] Rao, K. R., Z. Bojković, and D. Milovanović. "Joint video encoding in DVB/ATSC multi-program transmission: MPEG-2 vs. H. 264/AVC bitrate control." 2009 9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services. IEEE, 2009.

[37] Wang, Limin, and Andre Vincent. "Bit allocation and constraints for joint coding of multiple video programs." IEEE Transactions on Circuits and Systems for video Technology 9.6 (1999): 949-959.

[38] He, Zhihai, and Dapeng Oliver Wu. "Linear rate control and optimum statistical multiplexing for H. 264 video broadcast." IEEE Transactions on Multimedia 10.7 (2008): 1237-1249.

[39] Avalpa. “OpenCaster software” OPENCASTER, Avalpa Digital Engineering, Jun 2015, http://www.avalpa.com/the-key-values/15-free-software/33-opencaster/.

Paper 3 references:

[40] ITU-R, “Spectrum requirements for terrestrial television broadcasting in the UHF frequency band in Region 1 and the Islamic Republic of Iran”, ITU-R BT.2302-0 (04/2014), Apr 2014, Geneva (2014).

[41] Benoit, Hervé. Digital television: MPEG-1, MPEG-2 and principles of the DVB system. Routledge, 2002.

[42] ETSI/EBU, “Digital Video Broadcasting (DVB); Specification for conveying ITU-R System B Teletext in DVB bitstreams”, ETSI EN 300 472 V1.4.1 (2017-04), Apr 2017, Sophia Antipolis (2017).

[43] Soto, Daniel. "Digital TV simulcast with shared audio streams." 2015 IEEE 5th International Conference on Consumer Electronics-Berlin (ICCE-Berlin). IEEE, 2015.

[44] Vetro, Anthony, et al. "3D-TV content storage and transmission." IEEE Transactions on Broadcasting 57.2 (2011): 384-394.

[45] Duarte, Marco F., and Yonina C. Eldar. "Structured compressed sensing: From theory to applications." IEEE Transactions on signal processing 59.9 (2011): 4053-4085.

[46] Huszák, Árpád, and Sándor Imre. "Analysing GOP structure and packet loss effects on error propagation in MPEG-4 video streams." 2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP). IEEE, 2010.

[47] Goldman, Matthew S. "“It’s Not Dead Yet!”-MPEG-2 Video Coding Efficiency Improvements." Proceedings of the NAB Broadcast Engineering Conference, 2009.

[48] Wang, Limin, and Andre Vincent. "Bit allocation and constraints for joint coding of multiple video programs." IEEE Transactions on Circuits and Systems for video Technology 9.6 (1999): 949-959.

[49] Bruckstein, Alfred M., Michael Elad, and Ron Kimmel. "Down-scaling for better transform compression." IEEE Transactions on Image Processing 12.9 (2003): 1132-1144.

[50] Sugawara, Masayuki, Seo-Young Choi, and David Wood. "Ultra-high-definition television (Rec. ITU-R BT. 2020): A generational leap in the evolution of television [standards in a nutshell]." IEEE Signal Processing Magazine 31.3 (2014): 170-174.

[51] ETSI/EBU, “Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications”, ETSI TS 101 154 v2.5.1 (2019-01), Jan 2019, Sophia Antipolis (2019).

[52] He, Zhihai, and Dapeng Oliver Wu. "Linear rate control and optimum statistical multiplexing for H. 264 video broadcast." IEEE Transactions on Multimedia 10.7 (2008): 1237-1249.

[53] Rao, K. R., Z. Bojković, and D. Milovanović. "Joint video encoding in DVB/ATSC multi-program transmission: MPEG-2 vs. H. 264/AVC bitrate control." 2009 9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services. IEEE, 2009.

[54] Bellard, Fabrice. “FFmpeg tool.” FFmpeg, Sep 2019, http://www.ffmpeg.org/.

[55] Nvidia. “Nvidia Video Codec SDK - NVEncode” NVIDIA VIDEO CODEC SDK, NVIDIA Corp., Sep 2019, http://developer.nvidia.com/nvidia-video-codec-sdk/.

[56] Intel. “Intel Quick Sync Video SDK” INTEL MEDIA SDK, INTEL Corp., Sep 2019, http://software.intel.com/en-us/media-sdk/.

[57] Lelégard, Thierry. “TSDuck Toolkit.” TSDuck, Sep 2019, https://tsduck.io/.