Volume 6 | ISSN 2041-5923
2014

IET Journals
The Institution of Engineering and Technology

The Best of IET and IBC 2014

INSIDE Papers and articles on electronic media technology from IBC2014 presented with selected papers from the IET's flagship publication *Electronics Letters.*

Published by the Institution of **Engineering and Technology**

www.**theiet**.org/journals

Contents

An electronic version of this publication and previous volumes can be found at www.theiet.org/ibc or by scanning the QR code

ISSN 2041-5923

Introduction

Welcome to *The Best of IET and IBC*, 2014. This is the sixth volume of an annual joint publication between the International Broadcasting Convention and the Institution of Engineering and Technology.

The IET is a formal member of the IBC's partnership board but, beyond this, it has a long-standing and close relationship with the organisation, through which they together encourage and promote professional excellence in the field of media technology. Nowhere is this relationship more strongly reflected than in the pages of this publication, which celebrates the very best technical media papers from this year's *IBC Proceedings* and the IET's flagship journal, *Electronics Letters.*

This year, our editorial takes a look at IBC's Future Zone – one of the most exciting areas of exhibition space where the world's most hi-tech media companies and research organisations proudly demonstrate some of their very latest concepts and experimental technologies. Here, you can not only *see* tomorrow's media but you have the opportunity to *experience* it personally, leaving impressions that will remain with you long after IBC.

We then present eight papers chosen as the best contributions to IBC2014 by the IBC Technical Papers Committee and the executive team of the IET Multimedia Communications Network. These include the overall winner of IBC's award for the Best Conference Paper 'DVB-T2 Lite: Exploiting Broadcast HDTV Networks for Services to Mobiles' and papers representing the hot topics of 2014: UHDTV (Ultra High Definition Television), (HDR) High Dynamic Range video, HEVC (High Efficiency Video Coding), next generation studios and new directions in stop-motion film making.

We are also pleased to present personal interviews, both with an individual and with a research team, whose significant work appears in this volume. First *Gino Alberico* of RAI in Italy, the principal author of IBC2014's best paper, and then with the team at the Fraunhofer Institute for Integrated Circuits who have used advanced signal processing to inject novelty into the traditional skill of stop-motion animation. Find out the background on these individuals. What motivates them? What aspect of the work did they enjoy most? And if budgets were not a constraint, where would they take the future of their work?

From *Electronics Letters* this year we have three papers chosen from those published since IBC 2013.

Electronics Letters has a very broad scope, covering the whole range of electronics-related research and the three papers included this year deal both with research that may have a great impact on media technology in the long term as well as research aimed at addressing the immediate issues of expanding services with existing infrastructure.

The IBC papers you can read here represent the cream of almost 300 synopsis submissions originally received at our offices early in the year – this is a figure that has remained remarkably constant for well over a decade. We know that the writing and submission of papers for IBC is less often driven at the corporate level, but usually represents the motivation of individual researchers and teams who choose to write about their novel ideas, experiments, prototypes and trials, and to share these in our conference forums. We are extremely grateful and proud that so many media professionals continue to choose IBC for the publication of their technical work and as a forum for discussion with their fellow workers. This journal is a tribute to all those individuals who submitted synopses this year, whether successful or not. If you are inspired by the papers and stories presented here and would like to tell us about your own research, then please look out for our call for papers in January each year. And if your synopsis was unsuccessful this year, then please try again – we work hard to accommodate as many papers and posters as we possibly can.

I hope that you enjoy reading this collection of the best papers as much as I and my committee of specialists and peer reviewers. We would like to convey our thanks to everyone involved in the creation of this year's volume both at the IET and at IBC and to extend our best wishes for a successful and exciting IBC 2014.

Dr Nicolas Lodge
Chairman, IBC Technical Papers Committee

Who we are

IBC

IBC is committed to staging the world's best event for professionals involved in content creation, management and delivery for multimedia and entertainment services. IBC's key values are quality, efficiency, innovation, and respect for the industry it serves. IBC brings the industry together in a professional and supportive environment to learn, discuss and promote current and future developments that are shaping the media world through a highly respected peer-reviewed conference, a comprehensive exhibition, plus demonstrations of cutting edge and disruptive technologies. In particular, the IBC conference offers delegates an exciting range of events and networking opportunities, to stimulate new business and momentum in our industry. The IBC conference committee continues to craft an engaging programme in response to a strong message from the industry that this is an exciting period for revolutionary technologies and evolving business models.

The IET

The IET is one of the world's leading professional societies for the engineering and technology community, with more than 150,000 members in 127 countries and offices in Europe, North America and Asia-Pacific. It is also a publisher whose portfolio includes a suite of 27 internationally renowned peer-reviewed journals covering the entire spectrum of electronic and electrical engineering and technology. Many of the innovative products that find their way into the exhibition halls of IBC will have originated from research published in IET titles, with more than a third of the IET's journals covering topics relevant to the IBC community (e.g. *IET: Image Processing*; *Computer Vision*; *Communications*; *Information Security*; *Microwave Antennas & Propagation*; *Optoelectronics*, *Circuits & Systems* and *Signal Processing*). The IET Letters contained in this publication come from the IET's flagship journal, *Electronics Letters*, which embraces all aspects of electronic engineering and technology. *Electronics Letters* has a unique nature, combining a wide interdisciplinary readership with a short paper format and very rapid publication, produced fortnightly in print and online. Many authors choose to publish their preliminary results in *Electronics Letters* even before presenting their results at conference, because of the journal's reputation for quality and speed. In January 2010 *Electronics Letters* was given a fresh new look, bringing its readers even more information about the research through a colour news section that includes author interviews and feature articles expanding on selected work from each issue.

Working closely with the IET Journals team are the IET Communities team. The communities exist to act as a natural home for people who share a common interest in a topic area (regardless of geography); foster a community feeling of belonging and support dialogue between registrants, the IET and each other. Assisting each community is an executive team, made up of willing volunteers from that community who bring together their unique experience and expertise for the benefit of the group. Members of the Multimedia Communications Community executive team play an essential role in the creation of this publication in reviewing, suggesting and helping to select content. They contribute their industry perspectives and understanding to ensure a relevant and insightful publication for the broad community represented at IBC, showing the key part volunteers have to play in developing the reach and influence of the IET in its aim to share and advance knowledge throughout the global science, engineering and technology community.

ISSN 2041-5923

Editorial

Visions of Future Broadcasting at IBC2014

The Future Zone has become the jewel in the crown of the IBC Show. It is a fantastic collection of the latest cutting-edge projects and prototypes from leading R&D labs around the World, brought together into a single exhibition area in the Park Foyer, next to Hall 8, in the RAI Centre.

In selecting exhibits for the Future Zone, we try to select ideas and technologies from all sizes of companies – from large multi-national organisations to the smallest start-ups – and we start from the premise that the craziest proposals may well be the best. There are only really four guidelines for the selection Committee:

- the exhibit must be interesting, innovative and valid.

- it must not be a product that is already on the market – these can be shown on company stands in the other Exhibition Halls.

- it should consist of a practical demonstration, with hands-on interaction where possible.

- its presentation should have a 'wow-factor' (measurable on the Richter Scale), to encourage passing visitors to stop and discover more.

This year's Future Zone exhibitors are truly international, with representatives from a number of Japanese technology initiatives in both broadcasting and communications, MPEG and European collaborative projects, Chinese broadcast science projects, North American industry developments and Korean media research, amongst others. There are incredible exhibits from some of the major players at the forefront of our industry – like NHK, BBC, Cisco, VRT.

The practical demonstrations arraigned on the stands in the Future Zone this year, include: personalisation and enhancement of the viewer experience, highlights of the FIFA World Cup in 8K UHDTV at 120 fps, 4K UHDTV Cloud services, future Digital Radio broadcasting, a day in our future 'Connected Life', new technologies for AV archive restoration, 'smart' convergent media, 3608 viewable 3D displays, and more. A portal to view our future directions, rather than just a glimpse of the future!

The Future Zone is also home to IBC 'Posters'. These are the highly respected, rigorously peer-reviewed ideas chosen by the IBC Technical Papers Committee for their relevance to the subjects of IBC's Conference Programme and their ability to startle your imagination. The Posters are displayed as two A0-size 'snap-shots' of innovation, and this year we have created a great new viewing area with more of a modern 'art gallery' presentation style. This will enable visitors to view the displays more conveniently, and engage with the Poster authors in more open and attractive surroundings.

We are especially pleased to have attracted Poster presentations for IBC2014 from so many different countries around the World. The fascinating topics of international research that will be presented by the authors, from both industry and academia, on the Friday and Saturday Showdays include: UHDTV-IP; HEVC in the Studio; Perception of 3D Audio; CGI and editing for 3D images;

Crowd-sourced Music; DTV Mobile Surveillance; Object-oriented Storytelling; Visual Quality Metrics; 4K Broadcast on a Budget; Adaptive Bitrates & Cloud-based Architectures for Multimedia; and more.

The Future Zone is open to all, thoughout the IBC Show; but we draw your attention in particular to the IET Reception, held in the Zone at 16:00 on Friday the 12th September. Here, you will be able to go round the exhibits, talk with the presenters, and network with the 'movers and shakers' in our industry, with some complimentary refreshments from the IET to help nourish your stimulated mind.

Professor David Crawford
IBC Conference Executive Producer

Selected content from IBC2014

ISSN 2041-5923

DVB-T2 LITE: exploiting broadcast HDTV networks for services to mobile receivers

G. Alberico A. Bertella S. Ripamonti M. Tabone

RAI - Research & Technology Innovation Centre, Italy

Abstract: In several countries DVB-T2 networks are being deployed for the delivery of domestic television over terrestrial channels. However, there is also an increasing demand for linear TV viewing on mobile and portable devices such as tablets. Whereas 4G mobile networks could provide services to these terminals using unicast delivery, the cheapest delivery mode to big audiences during peak hours or during major events is undoubtedly via broadcasting. For this reason DVB introduced a profile of its T2 standard called *T2-Lite*, orientated towards mobile reception.

To explore this further, RAI launched an experimental trial in the Aosta Valley where the same transmitting network was to be used to provide both HDTV services to domestic receivers and mobile TV services to portable devices.

This paper reports the positive results emerging from the trial which proved that services to mobile devices are feasible at marginal costs.

Introduction

Following analogue television switch-off in many countries, DVB-T and DVB-T2 terrestrial networks are being deployed or planned for the delivery of digital television services to domestic receivers. Most of these networks are designed to provide services to fixed receiving systems with roof-top antennas but there is also an increasing demand for TV viewing on mobile and portable devices, such as tablets and in-car screens.

Mobile TV services, such as those based on the DVB-H standard, were launched some years ago but had limited success, or even failed, for many reasons. Typical problems were: low transmission efficiency, high network costs, unclear business models and poor terminal performance.

Nowadays, major changes have occurred in the domain of receiving devices; with the success of tablets and smart-phones people have devices that are capable of displaying mobile or nomadic video in a form that is truly compelling. In addition, users are becoming accustomed to VoD and live content consumption on a great range of devices thanks to 3G/4G networks and Wi-Fi connections which are widely available at home or around.

Multiple cell handovers and network capacity, however, make live video streaming over a wireless network, a poor user experience. Furthermore, big audiences are seriously challenging those mobile operators' which attempt to serve customers with unicast delivery. Whenever 3G/4G networks reach capacity limits and are unable to deliver an acceptable quality of service, a possible solution is (apart from using more spectrum or increasing the number of transmitters) to off-load traffic onto other network technologies, for example Wi-Fi or broadcast.

From the technical and economical point of view broadcast networks are certainly the best delivery mechanism for reaching big audiences during peak hours or during major events. For this reason DVB introduced, in 2011, a new profile of its T2 standard called *T2-Lite*, aiming to better support portable and mobile reception.

Using the so-called FEF (Future Extension Frame) feature of the DVB-T2 system, *T2-Lite* services can be added to T2 broadcasts at the small expense of some capacity reduction.

The *T2-Lite* services can operate on-top of the existing '*T2-Base*' HD services in previously deployed networks and can thus obviate the need to build new dedicated mobile networks.

After experimenting in their laboratory with the first prototypes of DVB *T2-Lite*, RAI launched an experimental trial in Aosta Valley in 2013. Here, HDTV services for fixed reception and *T2-Lite* mobile TV services would coexist on the same frequency and would optimally exploit UHF spectrum while guaranteeing adequate signal robustness for many potential additional services having very different requirements.

The trial would validate the technical characteristics of the system: an evaluation of mobile coverage, a check on interoperability related to the implementation of a Single Frequency Network (SFN) using equipment from different vendors and, last but not least, a test of the behaviour of both static and mobile reception in the field.

The AOSTA VALLEY trial

The Aosta Valley region has very often been the scene of experiments by the RAI Research Centre, especially since the advent of digital technology in the '90s. For example, in 1995-96, at the dawn of Digital Radio, an intensive measurement campaign was carried out on DAB. When digital terrestrial television was in its infancy, in the late '90s, the possibility of transferring the DVB-T signal through an analogue radio link was explored. And again, in the early 2000s the Aosta Valley was the scene of an important pre-operational test of DVB-T.

The reason for using the Valley lies in its peculiar topography which is particularly complex from the viewpoint of signal propagation and reception. The numerous side valleys and the large number of transmitter sites needed to cover the area, also make the Aosta Valley perfect for testing single frequency networks. Furthermore, the possibility of travelling along a dense secondary road network, next to the main valley (which often climbs up steep mountain slopes frequently hidden from the transmitters), provides a demanding test case for the mobile reception of signals such as DAB digital radio or mobile TV.

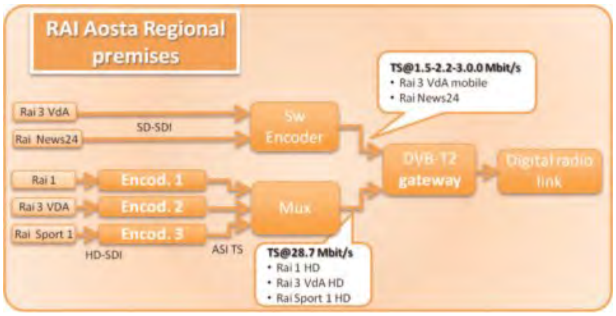


Figure 1 *The head-end*

System architecture

The 'bouquet' of services for the trial originated in the head-end located at RAI's Regional Centre in Aosta and consisted of three HD programmes for the *T2-Base* and of two services orientated to mobile reception for the *T2-Lite*. The system diagram is shown in Figure 1.

For the *T2-Base*, a Transport Stream (TS) at about 28.7 Mbit/s with three AVC-encoded live HD programmes was created using statistical multiplexing.

For the *T2-Lite*, a TS containing two AVC-encoded SD programmes was generated with a bit-rate between 1.5 and 3.0 Mbit/s, depending on the modulation scheme used during the tests.

The two TS streams were sent to a flexible T2 Gateway that allowed the modulation scheme and the network parameters for both the *T2-Base* and *T2-Lite* (Table 1) to be chosen.

For the *T2-Base*, it was decided to use the 256QAM constellation with FEC 3/4, allowing high transmission capacity while requiring a C/N ratio of about 20 dB, a value similar to that required to receive current DVB-T services. Since the *T2-Base* part is devoted to fixed

Table 1 Modulation scheme for *T2-Base* and *T2-Lite*

Parameter	DVB-T2 Base	DVB-T2 Lite
Constellation	256QAM, rotated	QPSK, rotated
FEC	3/4	1/3 – 1/2 – 2/3
FFT	32k	8k
Tg	1/128	1/16
Guard interval	28 ms	56 ms
Pilot pattern	PP7	PP4
Bit rate (Mbit/s)	28.7	1.5–2.0–3.0

Figure 2 *Geographical context*

reception with directional antennas, the PP7 profile of carriers was selected. The 32k FFT mode was chosen in order to maximize the guard interval duration which is different from that chosen for *T2-Lite*.

For the *T2-lite* part, a QPSK constellation was selected and it was decided to perform the tests (in the first phase, at least) with three values of FEC: 1/3, 1/2 and 2/3. For mobile reception, the optimal carrier profile found in the laboratory tests was PP4, which provides excellent strength even at high reception speeds. Finally, the 8k mode was used, being a good compromise between robustness to Doppler shift and guard interval length.

Finally, the gateway generates the T2-MI stream containing all the information necessary for the set-up of the modulators placed at the transmitting sites. With the aid of the time references 1 pps and a 10 MHz frequency obtained by GPS, all the synchronization information was available for the implementation of a single frequency network.

Network set-up

Four transmitters (Gerdaz, Salirod, Arpy, Courtil) were installed in the Aosta Valley region and their locations are shown in Figure 2. Each transmitter was equipped with a 50 W power amplifier except for Salirod which has a 100 W capability. Horizontal polarization, even though not ideal for vehicle reception, was used to take advantage of radiations from existing antennas.

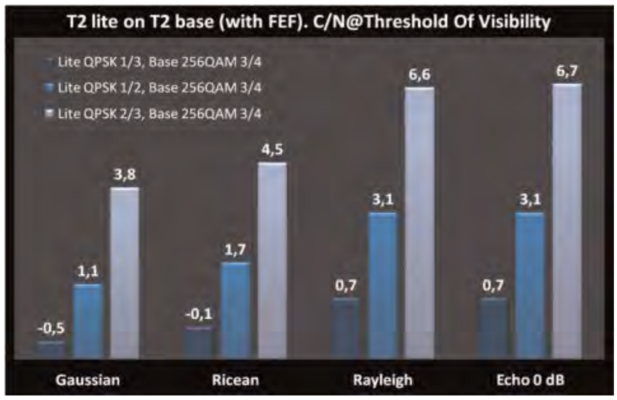


Figure 4 *T2-Lite performances*

The coverage simulations of the four transmitters allowed us to identify the areas of overlap of the various signals. A static delay of 44 ms was applied to the transmitter in Gerdaz of the *T2-Lite* stream in order to avoid inter-symbol interference. In this way, the zone of potential interference is moved into areas where signals are never superimposed. In potentially 'at risk' areas, the interferences are within the chosen guard interval of 56 ms.

Laboratory tests

Very extensive lab sessions allowed us to gain firsthand experience of the enormous potential of the system and to understand how mature the technology is for prompt exploitation.

Tests were mainly conducted on a prototype board equipped with one of the first available commercial chipsets capable of receiving both *T2-Base* and *T2-Lite*.

The tests employed all means for characterising the 'goodness' of the technology and its implementation on silicon: the sensitivity of the receiver, the performance with white Gaussian noise (AWGN), the performance in Rayleigh and Rice channel profiles to simulate respectively portable and fixed reception, the performance in 0 dB echo to evaluate the resilience in reflections and SFNs, the

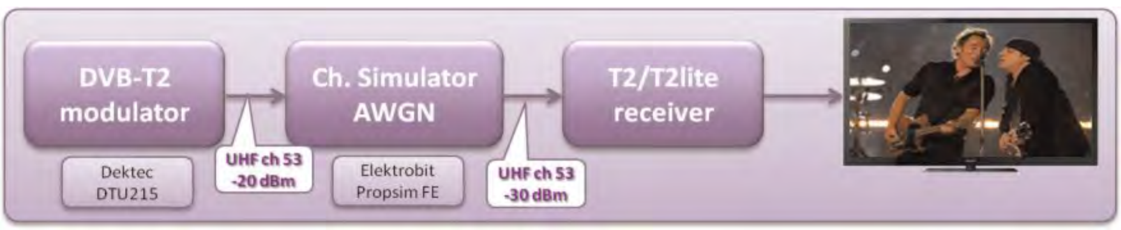


Figure 3 *Laboratory set-up*

effectiveness of the rotated constellations and the performance with different mobile channels at different speeds (profiles COST206 Typical Urban, Hilly Terrain and Rural Area). A full report of the results is available in (2) only in the Italian language.

The laboratory set-up is depicted in Figure 3.

In this paper only the data of interest for subsequent testing in the service area are reported.

The performance of *T2-Lite* in terms of C/N at the threshold of visibility (limited to the most robust modulations suitable for mobile usage), are shown in Figure 4. The three FECs shown in the graph (1/3, 1/2 and 2/3) are also those used during the field tests. The values obtained, are perfectly in line with what is expected from the theoretical point of view with very small implementation margins.

During the laboratory tests it was also observed that the time-to-re-sync of the receiver is very short. This is a particularly appreciable feature in mobile reception, where natural and man-made obstacles (bridges and tunnels, for example) can lead to a sudden loss of signal.

Other laboratory tests that simulated the mobile channel aimed to verify the maximum speed that the system was able to withstand.

The graph shown in Figure 5 reports the required C/N versus speed in km/h for UHF channel 53 (f = 730 MHz) which is the frequency used for testing in the Aosta Valley.

From this graph it can be seen that the maximum allowed speed with FEC 1/2 and 1/3 is around 240-280 km/h on a channel such as 53 which is located in the upper part of the UHF band. At lower frequencies the maximum speed could be even higher, compatible with a use not only in-car but also on high-speed trains.

To obtain this result, the pp4 profile of carriers was used which, compared to PP7 (commonly used for fixed reception), allows us to quadruple the maximum speed at the price of 10% reduction of available bit-rate. These results are obtained with the 8k FFT mode. If further increases of maximum speed are required, the 2k or 4k FFT mode could also be considered.

The lab results confirmed the effectiveness of the new *T2-Lite* profile, its efficacy on the mobile channel and the very good performance of the first hardware implementations.

Field tests

The tests in the service area were carried out by a car specially equipped for mobile measurements and by a van equipped with a telescopic mast up to 15 m above ground level for measurements at fixed locations.

The purpose of the mobile measurements was to evaluate the coverage area and to plot the map of field strength at the roof of the car. This data would be vital for future network planning.

The purpose of the fixed-point measurements was to determine the available field strength across the service area in order to verify consistency with the predictions obtained through computer simulations. The purpose was also to

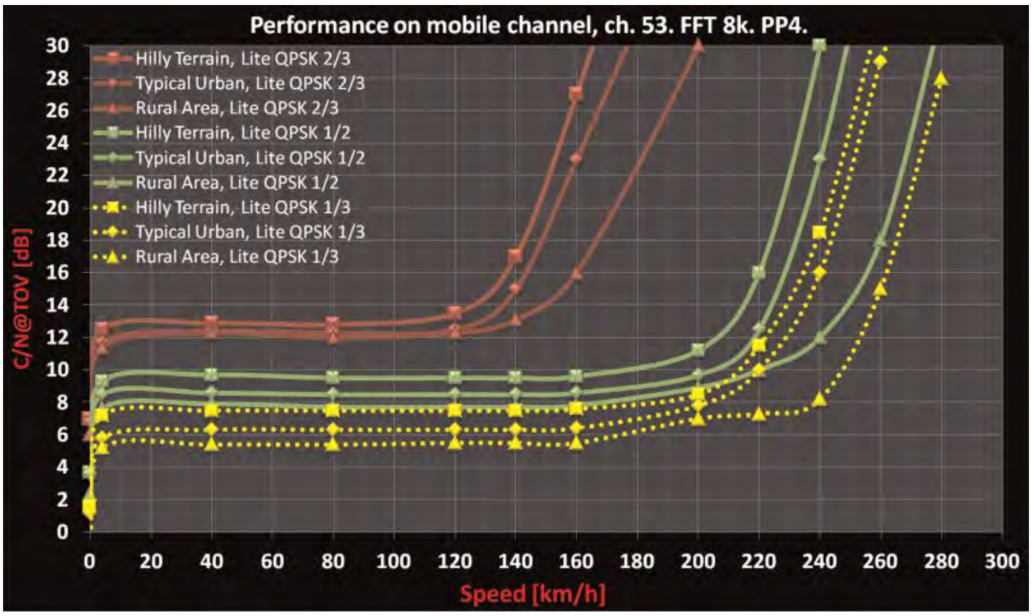


Figure 5 *T2-Lite performances on mobile channel*

check whether the transmission is as expected, with no problems in the radiating equipment. These measurements have also allowed us to assess the margin of *T2-Base* signal reception. In addition, it was also possible to check all the SFN synchronization issues of the network, verifying the parameters very accurately.

Mobile measurements

In the first phase of the field tests, the coverage of each single *T2-Lite* transmitter was verified using three different values of FEC: 1/3, 1/2 and 2/3. For each FEC, service availability over more than 400 km of road was recorded, thanks to the reception parameters logged by the *T2-Lite* evaluation board. With these data, recorded approximately once per second and correlated with the GPS position of the car, it was possible to get very detailed maps.

Figure 6 shows the mobile reception availability with FEC 1/2 for the Gerdaz transmitter. Similar maps are available for all the other transmitters.

The tests confirmed the excellent performance of the *T2-Lite* system that we had already seen in the laboratory. The differences in service availability due to FEC change are rather small between 1/3 and 1/2, and are slightly more accentuated with FEC 2/3.

Overall, despite the relatively low transmitting power (50 W), very good coverage was found on all types of road (motorway, main and secondary roads). Also in urban areas (some, densely populated) and in some tunnels, results were excellent, with complete absence of receiver unlocks or video artifacts. On the highway, no problems were encountered up to the maximum permitted speed of 130 km/h. In this regard, a test switching from PP4 to PP7 carrier profile was also performed and, confirming the laboratory results, the receiver lost the signal when traveling at speeds higher than 70-80 km/h in a radial direction with respect to the transmitter. Supported by these findings it can be reasonably assumed that *T2-Lite* should tolerate speeds up to 250 km/h, thus opening interesting prospects for use on high-speed trains.

Figure 6 *Mobile reception availability*
FEC = 1/2, transmitter: Gerdaz

During the data processing, a statistical figure was calculated which represents the value of EMF beyond which 95% of measurements fall, on segments of 300 m. These values were plotted on the map with colours that recall the mobile reception availability, thereby showing a graphical relationship between the electromagnetic field value and the service availability. This information is crucial during the network planning phase in order to properly design the transmitter power and the radiating system for the desired coverage. To achieve this result, the input power of a professional test receiver was measured and, given the antenna gain, the available field strength at the roof of the car could be calculated. Also in this case the data acquired have been associated with the location and speed of the car as obtained via GPS and the odometer.

Finally, the four-transmitter SFN was activated and, after the necessary set-up tests, the evaluation of mobile service availability began. Given the minimal differences in coverage obtained with FEC 1/3 and 1/2 values, it was decided to use the latter one only, being a better compromise between robustness and transmission capacity. The same routes used in previous tests were repeated, thus giving us a map of mobile service availability for the entire network (Figure 7).

Fixed-point measurements

The fixed measurements carried out with the van equipped as described above, allowed us to detect the level of the electromagnetic field available within the service area. These data allowed us to: check the reliability of the predictions obtained with the computer simulation, to check the evaluation of the reception margin of the *T2-Base* and to verify SFN synchronization.

As was done with mobile measurements, tests in the first phase were carried out by activating one transmitter at a time.

For each selected measurement point, the electromagnetic field at different altitudes was measured in order to find the maximum value. In addition, so that we could have precise information about the received signal and any anomalies, various reception parameters (such as MER, constellation, impulse response, etc.) were also measured and recorded for all points.

The measured field strength values are in very good agreement with the computer simulations thus confirming proper installation and operation of the transmitting equipment. In general, the electromagnetic field values are very high providing a very wide reception margin across the whole service area.

Successively, the four transmitters were switched on. First of all the synchronization of the network was verified using the impulse response measured at several locations.

Figure 7 *Mobile reception availability for the full SFN (4 transmitters). FEC = 1/2*

Software developed by RaiWay was used to determine the relative delay between the two signals. This software operates from a knowledge of the coordinates of the measuring point and any static delays present in the

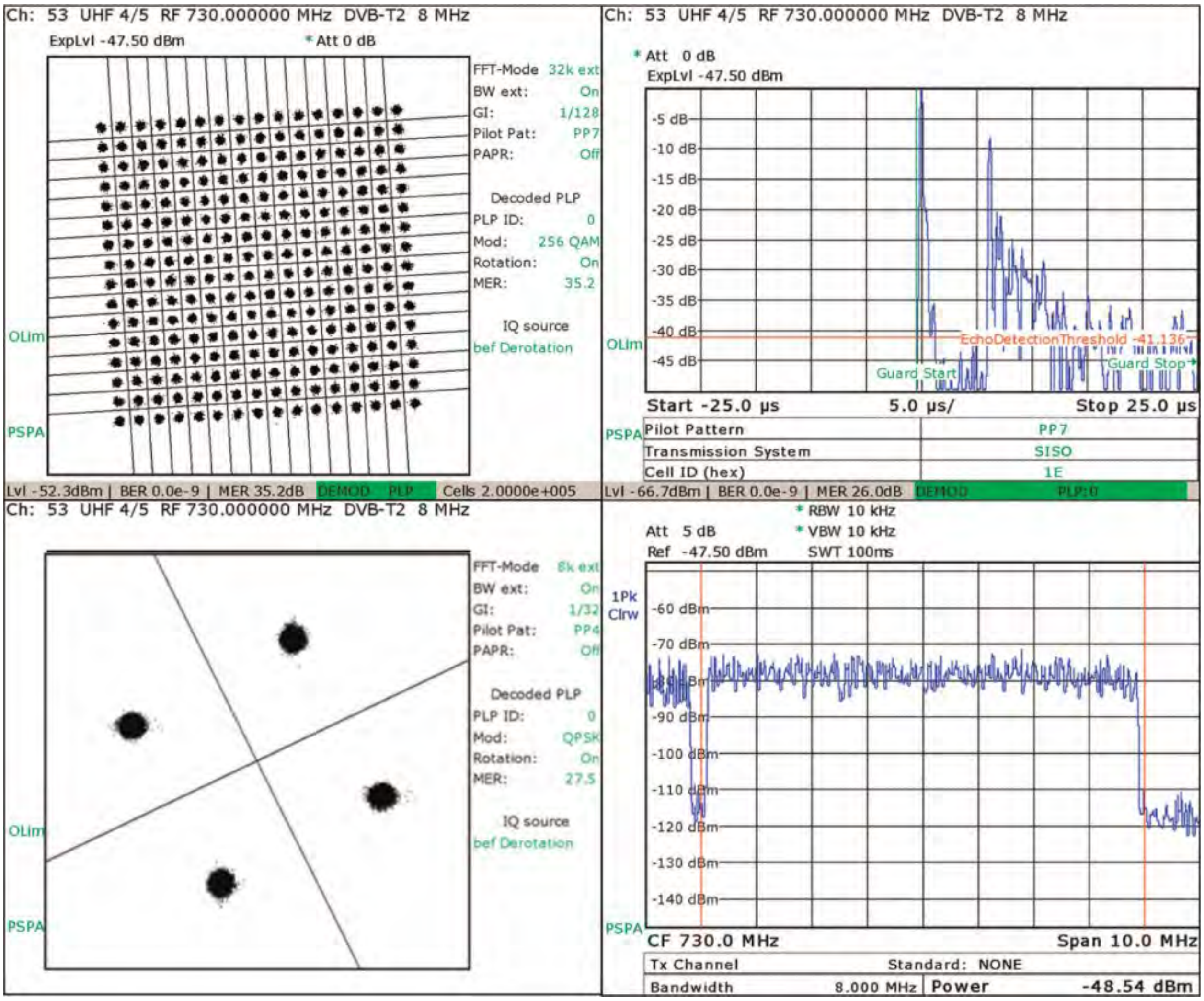


Figure 8 *Example of fixed-point measurements*

transmitters. The comparison between the calculations and the value obtained by the impulse response allowed us to validate the network.

Then, rotating the antenna in multiple positions for each of the measuring points, the spectrum of the signal, the impulse response, an overview of the reception parameters (MER, modulation scheme, C/N, input power etc.) and the constellation were recorded, for both *T2-Base* and *T2-Lite*.

An example of the data collected during the measurements campaign is shown in Figure 8.

Conclusions

The mixed *T2-Base*/*T2-Lite* field trials in Aosta Valley, conducted by the RAI Research Centre with the support and collaboration of RaiWay, led to the construction of a high-impact demonstrator for mobile TV.

The field trials allowed the robustness of the *T2-Lite* system to be evaluated in a region characterized by complex terrain, where natural and man-made barriers often obstruct the view of the transmitters. The same considerations also applied to the tests carried out in densely populated urban areas like Aosta and Saint Vincent and in the several tunnels distributed along the routes.

Using the so-called FEF (Future Extension Frame) feature of the DVB-T2 standard, *T2-Lite* services and "*T2-Base*" HD services can coexist with previously deployed networks, obviating the need for construction of new dedicated mobile networks. Potential infrastructure costs are therefore significantly reduced.

At present, the tests have been performed with a prototype receiver and a car roof-top antenna, so when targeting receiving devices such as tablets with integrated antennas higher field strengths or a few additional sites might be required to obtain a similar coverage.

Summarizing, *T2-Lite* showed, during tests in the laboratory and especially in the field, its enormous strength: an awesome ability to receive signals under the most severe conditions.

During the trial, a huge amount of data on electromagnetic field strengths at car roof-top level vs mobile service availability was collected, representing an important database for future network planning of mobile services.

Finally, a serendipitous result which rewarded all the effort put in the initiative, is that the test-bed is now being used by many TV and chipset manufacturers to test their own *T2-Base* and *T2-Lite* receivers!

Acknowledgments

The authors would like to thank their colleagues Arturo Gallo, Vittoria Mignone, Bruno Sacco for the support before and during the field tests and the colleagues of RaiWay for the support in the main transmitters installation and for the field measurements. Furthermore the authors would also like to thank Screen Service Broadcast Technology S.p.A. and Syes for the support during the installation of the transmitters.

References

[1] ETSI EN 302 755. V1.3.1. Digital Video Broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2). European Telecommunications Standards Institute. 2012

[2] A. Bertella, A. Gallo, S. Ripamonti, M. Tabone, 2013. DVB-T2 e DVB-T2 Lite: la sperimentazione in Valle d'Aosta. Elettronica e Telecomunicazioni. Dicembre 2013. pp. 12–34

ISSN 2041-5923

Interview – Gino Alberico

1. Tell us a bit about yourself and what you do

I graduated in Electronic Engineering from Turin Polytechnic and in 1988 I joined the R&D Centre of RAI Radiotelevisione Italiana, the Italian Public broadcaster, where I am now responsible for the "Study and Research" department.

I have been involved together with my research team, in the launch of digital TV and Radio services in Italy, both on terrestrial and satellite platforms, in the deployment of multimedia services and interactive applications for connected TVs and second screen devices, also combining TV and social networks.

I am a member of the EBU Technical Committee and Chairman of the EBU Strategic Programme on Broadcast Internet Services. In my career I have also contributed to the work of international standardisation bodies and industrial forums such as ETSI and DVB, as well as national organisations like HD Forum Italy and Tivùsat, the Italian free-sat platform.

2. What is your paper about?

The paper is about an experimental trial we have carried out in Aosta Valley, showing that DVB-T2 HDTV services for fixed reception and T2-Lite mobile TV services for portable devices can coexist on the same frequency.

Mobile services, based on the T2-lite profile defined by DVB in 2011, are added on-top of existing "T2-Base" HD services, using the so-called FEF (Future Extension Frame) feature of DVB-T2. At the expense of some capacity reduction, this approach allows to exploit already deployed transmitting networks, avoiding to build a new infrastructure dedicated exclusively to mobile services. The positive results of the trial have been encouraging and have confirmed that providing broadcast services to mobile devices is feasible at marginal costs.

3. What do you find exciting about this area of work?

The most exciting aspect of this kind of activity is certainly taking the first steps in the development of a new technology and finding the right technical solutions to problems never addressed before. Another amazing aspect of the trial is that our test-bed has been (and will be) used by different manufacturers for testing their T2-Base and T2-Lite receivers. Finally we are very satisfied of the results obtained with T2-Lite technology: the ability to receive the signal in the most severe conditions is really awesome.

4. What will you be hoping to see at IBC this year?

Of course I look forward to find tablets or portable devices equipped with T2-lite front-end or at least with external USB adapters. I am also very interested to see the latest developments about HEVC video coding as well as cloud-based video encoding and streaming systems. Another interesting topic is about new immersive audio technologies, spatial audio objects and their implications for the home user.

5. Do you believe that the future of television will be mobile? If so, will we soon see T2-Lite products in the shops?

I am convinced that digital terrestrial broadcasting is the best distribution platform for delivering high quality linear services to large audiences and ensuring universal and free-to-air access for users.

It's probably true that TV consumption will encompass many different situations, from the big screen in the living room to the nomadic viewing, from live to on-demand, so mobile TV will be just one of the possibilities complementary to one another. Consequently there will be certainly a potential market for portable devices with a T2-lite capable frontend.

6. Your paper describes both lab work and field trials. Tell us which you found more challenging.

Differently from the laboratory environment, where everything can be kept under control, field trial activity is definitely most challenging because of many additional complications and challenges. Just to mention a few of them: synchronising the SFN network using equipment from different vendors, checking synchronisation in the service area without, at least in the first phase of the trial, suitable instruments, adjusting transmitters relative delay in order to avoid interference in overlapping areas, and... walking in high mountains in the snow to reach the transmitting site for the installation of the equipment, to fix a problem or to carry out a SW upgrade not feasible remotely. However we were lucky to work in a beautiful region and this has allowed us to make lighter the obstacles that we have found during the trial.

7. Given the high reception speed of T2-Lite, will your future work involve collaboration with an Italian supercar company?

We would be delighted to broaden our trial campaign including tests on a supercar, but extending signal coverage to the Monza autodrome would be necessary to avoid fines for exceeding speed limits on normal roads. More seriously, watching TV on a car running at 300 km/h may not be recommended, unless you're the passenger, however similar reception conditions could be relevant for passengers on high speed trains.

8. Your trials were conducted in a beautiful valley. With an unlimited budget, where else in the world would you like to travel to do further trials?

Undoubtedly it would be fascinating to exploit radio waves propagation and "canyoning" effects in Arizona, perhaps by testing mobile TV reception on a small plane gliding between the Grand Canyon's walls.

Selected content from IBC2014

ISSN 2041-5923

Computational imaging for stop-motion animated video productions

Frederik Zilly Matthias Ziegler Joachim Keinert
Michael Schöberl Siegfried Foessel
Fraunhofer Institute for Integrated Circuits IIS, Department Moving Pictures Technologies, Germany

Abstract: Creating movies using stop-motion technology remains a fascinating method of storytelling, even in the era of digital cinema. The huge number of professional and semi-professional clips produced as fan-art, which can be watched on the internet, serve as testimony to the great interest in unleashing the creative potential of this art-form. However, producing content using stop-motion remains a cumbersome task, with high production costs especially for full-length movies. Consequently, a trend can be observed that even many successful television series which were originally produced using stop-motion, have been moved to computer animation.

Against this background, we propose in this paper a new production scheme for stop-motion-animated movies which has the potential to lower production costs while increasing the quality of the resulting content. By using a static multi-camera array and algorithms from the field of computational imaging, our technology permits the creation of artistic effects in post-production which are difficult to realise using conventional stop-motion methods. Our approach allows: changes in the depth-of-field; smooth camera moves along virtual camera paths; and up-sampling of the frame-rate of the stop-motion-video, all in high quality. All effects are computed and applied in post-production while all intrinsic and extrinsic parameters of the cameras remain fixed during the whole production. To demonstrate the practicability, we shall show within this paper, results from a stop-motion video which has been produced using the proposed methods.

Introduction

Henry Selick's famous stop-motion film *Coraline* published in 2009 is one of the most successful films shot with this technique. It had a budget of about $60 million but was able make about $125 million at the box office. It is of course not the only case of a successful stop-motion film. Other famous examples are productions of *Wallace and Gromit* and *Nightmare before Christmas*. Although shooting in stop-motion seems a little old-fashioned, it is regularly used for feature films as well as commercials and its slight stuttering appearance creates a unique impression on the audience.

Animation software suites constantly undergo improvements, however. They are hard competitors when it comes to the decision of whether a story shall be produced with classical stop-motion techniques or using computer-generated stop-motion.

In our opinion, there are good arguments for both approaches. Computer-generated stop- motion allows complete control to be exercised over the camera and scene at all times. Parameters can be changed and the final output can be adjusted to suit the display platform (e.g. cinema, TV, mobile platforms). On the other hand real stop-motion allows enthusiastic artists to use their full creative potential and the final result can score highly with its unique, incomparable look.

Classical stop-motion, however, also has some drawbacks compared with computer-generated animations. The effort required to set up a scene and to shoot a take is very high. Moreover, camera parameters must be fixed and creating dolly shots, camera pans or changing focus are tremendous tasks. Therefore, stop-motion tends to be more costly and takes more production time compared with full animation.

Against this background, we present a solution that can circumvent some of the drawbacks mentioned, while preserving the unique look of classical stop-motion films. In particular, we propose using a camera array to capture the light field of a scene and to postpone decisions concerning the camera set-up to post-production. Such decisions can include changes of focus, aperture or dolly shots.

The remainder of this paper is organised as follows: first, we describe the legacy stop-motion production process, its challenges and limitations. Then, we present our approach which is based on a camera array and aims to overcome some of the limitations of the legacy workflow. The visual effects which we make possible are explained and shown with example images. The fourth section presents details of our stop-motion test shoot made with our proposed workflow, followed by an outlook and concluding remarks.

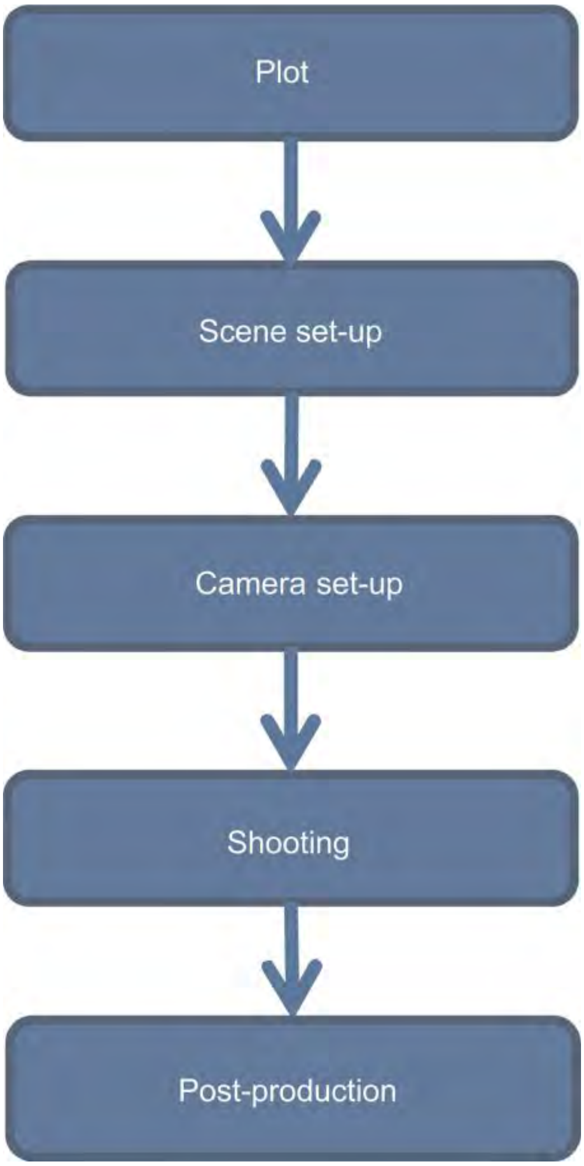


Figure 1 *Typical workflow in a stop-motion production*

Legacy stop-motion production workflow

Shooting a high quality stop-motion film is a tremendous task that requires a lot of material, time, technicians and – most important - enthusiastic artists. As with every feature film or TV show, producing a stop-motion film comprises different steps which are depicted in Figure 1: After the plot is completely defined, the scene set-up and camera settings are selected. The film is shot and finished in post-production.

The script is one of the most important factors in the making of a successful film. In the case of a stop-motion film, every movement of the actors and every camera pan needs detailed planning. In contrast to common movie-making it is generally not possible to shoot one scene several times to gain several different camera positions or slight variations from the plot. Because of the large effort and slow progress involved in the capture process, re-shooting a scene to correct any production errors will be costly and time-intensive.

From a technical point of view, many requirements need to be met in order to avoid visible artefacts in the final film. First of all, shooting a stop-motion video requires an indoor studio to maintain constant lighting conditions, because small variations of luminance will cause visible flickering in the film.

Typically, a scene set-up for a stop-motion film is spatially-limited as the actors in the film are puppets having sizes ranging from a few centimetres to about 30 centimetres for high quality characters having sharp detail with natural looks and movements. But this spatial limitation also limits the number of cameras that can be squeezed into the scene.

Beyond the requirements for scene set-up and puppets, the cameras and supporting tools also have a strong influence on

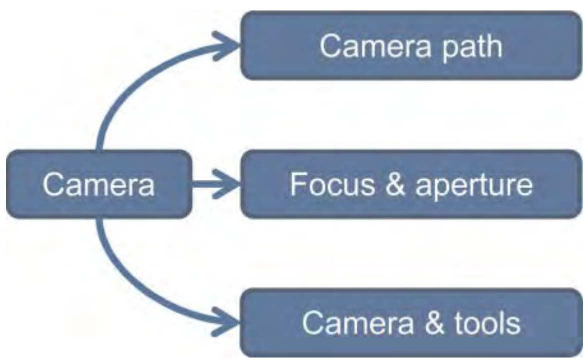


Figure 2 *The camera set-up in a stop-motion environment requires definition of camera paths, definition of focus and aperture and additional tools like gantries or tripods need to be included. These set-up decisions must be madein advance and cannot be changed*

the resulting image quality. Some of these aspects are mentioned in Figure 2 and will be outlined in the following. As with live action feature films, high quality images require high quality cameras. Unfortunately, such cameras are usually physically large. DSLR cameras are typically a good choice for stop-motion as they can deliver the required image quality. Additionally, each camera must be mounted on a robust tripod or similar device in order to avoid distortions by slight and random camera moves. Also, due to the size of the camera bodies, the inter-axial distance between the cameras' optical centres is too large for shooting stereoscopic 3D.

With a camera mounted on a tripod it is possible to generate a camera pan but it is not possible to move the camera through the scene. Generating precise dolly shots or shooting the film in 3D requires additional mechanical devices like precise sliders. Depending on their functionality, these devices can increase the time necessary to take a single frame as different camera positions need to be captured and the camera needs to be precisely moved to different positions. This can be done automatically by electronic drives or manually.

Another important aspect of the camera concerns the point of focus or depth of field (DOF) which is a widely used cinematic tool used to influence the visual impression; changes of focus and DOF within a scene are often desired. In the case of stop-motion, special care must be taken to generate a continuous-looking change of focus. Again, this can be accomplished by special tools which may be added to the camera to adjust its focus with an external drive. Similar considerations apply to the camera's aperture if its optics allow a seamless change. However, the combination of all the visual effects that are available in a normal film are hard to combine in a stop-motion set-up. Achieving a natural look when using camera movements and focus changes requires a lot of hard work and a lot of time.

All the technical aspects of lighting, scene and camera set-up which might hinder the artist from focussing on the artistic content must be considered. In order to reduce these dependencies and simplify the shooting of stop-motion films, we propose a new workflow that can enhance the creative process.

Proposed workflow

Instead of using one or more DSLR cameras to shoot the scene, we propose using a two- dimensional camera array for image acquisition. By using a camera array such as that presented by Foessel [1] and Zilly [2], we can capture the light field of a scene and, in combination with computer vision algorithms, we can simulate virtual cameras within the dimensions of the array. Changes of the DOF can be postponed to post-production and the artist doesn't need to worry about the camera set-up when capturing the scene. Beside a camera array, other implementations of light field cameras e.g. as presented by Ng [5] provide similar functionality, but suffer from a lower baseline. Therefore, flexibility to change the viewpoint of the virtual camera is limited compared with an array.

An example of a two-dimensional camera array is shown in Figure 3. It comprises a set of small, identical cameras that are arranged on a grid with identical line-of-sight and constant separation distance. This distance needs, in general, to be selected according to the scene. In the case of a stop-motion scene with small objects, the camera distance is chosen to be in the range of 4–6 cm in order to achieve high image quality and minimize the probability of visible artefacts. When shooting the take, the artist need not be concerned about camera settings. After each small change in the scene every camera in the array takes an image typically with a high DOF and therefore with a wide range of objects in sharp focus. In our array every camera uses the same settings such as focal point and exposure time.

In order to understand the possibilities that arise from a camera array, one needs to take a short look at its working principle. Due to their spatial distance, every camera sees the scene from a slightly different point of view. These differences are used to employ the algorithms for depth-image-based rendering (DIBR). More details on light field processing can be found in [3] and [4].

The processing chain for the array is depicted in Figure 4. After image acquisition, image rectification is the first computer vision algorithm to be used. This corrects small misalignments of the cameras. Precise alignment is important as it is required for the subsequent disparity

Figure 3 *Camera array in a 4x4 set-up*

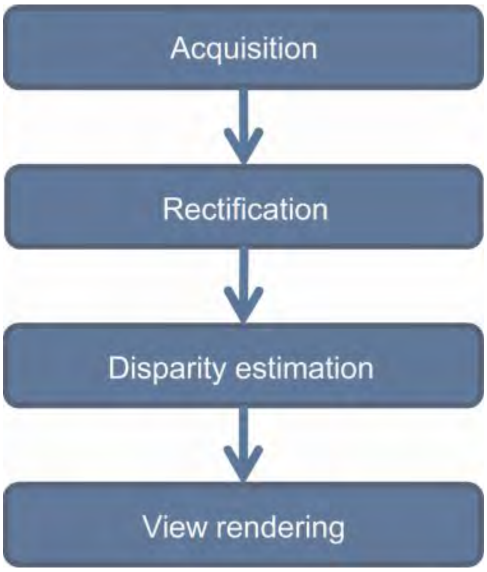


Figure 4 *Lightfield processing chain. Rectification and disparity estimation need to be executed before virtual camera positions can be rendered*

estimation. This estimation algorithm computes disparity from stereo image pairs in the vertical as well as the horizontal direction. These horizontal and vertical disparity maps are then merged to remove occlusions and other negative effects that can remain in the stereo disparity maps.

This step results in a set of dense disparity maps for every camera. They contain an implicit 3D model of the captured scene. These two steps are the most basic ones and need to be executed for every image in a take. All following processing steps rely on this data and therefore these steps require high precision. Errors here can cause visible artefacts in subsequent processes.

Based on disparity maps and rectified images, a rendering algorithm can compute parallax-compensated views from arbitrary positions in the array. The camera position can be defined independently from a specific point in time as it would be with a normal camera. This enables us to move the camera through the scene as a normal dolly would do. An example of a virtual dolly shot is depicted in Figure 5.

In a typical dolly shot, time moves forward linearly. With light field processing, however, we can slow down or even stop time and create a bullet-time effect that is well known from the movie *Matrix* and other feature films.

As mentioned earlier, the array captures the scene with high DOF. Using DIBR enables us to set the focus on a specific object and also to change it with time. This effect, of course, only works if DOF is reduced when compared with the original image. This is done by adding a synthetic aperture to the virtual camera. At the beginning of a scene we can set the focus on an object and smoothly change it to another object as time continues. Such a change of focus is depicted in Figure 6. At the beginning, the background is in focus (left image), changes to medium distance (centre image) and is finally in the front in the right image.

The dolly-zoom, also known as Vertigo-effect is another visual effect that is well-known from Hitchcock's *Vertigo*. In this effect, the camera moves backward while the focal length is changed such that the object in focus remains the same size. This is a challenging effect for a feature film and is even more challenging in a stop-motion film. By using our proposed workflow it is possible to move our virtual camera forward and also backward. Additionally, we can select the focal length of the virtual camera as required to achieve the desired effect in post-production. An example

Figure 5 *Example of a virtual dolly shot. The camera moves from a position in the left part of the scene to a viewpoint in the right part. The shift is done in a parallax-compensated way. Notice that the helicopter is seen from different perspectives*

Figure 6 *Example of a synthetic refocus: Starting with focus on the background in the left image, the focus moves to medium distance in the central image and ends on the foreground in the right image*

Figure 7 *Example images showing a Vertigo-effect. The virtual camera moves backward while zooming in. The policeman remains sharp and at same size, while the background gets smaller*

Figure 8 *Making-of picture from the stop-motion production showing the scene in the background and the backside of the 4x4 camera array in the foreground*

image sequence with a Vertigo-effect is depicted in Figure 7. It is especially good to see that the background seems to move away from the left to the right image while the policeman in the foreground stays at the same size.

Experimental stop-motion film

In order to test our proposed workflow and also our algorithms in a real-world set-up we decided to make a short stop-motion film. Our target was to create a video with a plot that covers about 20 seconds. The story is about a policeman drinking from a coffee cup in front of his police station. In the background a bad guy escapes from his prison, sneaks to a helicopter and tries to get away. At a frame-rate of 12fps this required at least 240 frames excluding additional frames before and after each scene.

The camera array employed has been seen in Figure 3. It consists of 16 identical cameras mounted in a 4x4 grid. Each camera has a resolution of 2048 x 1536 pixels that allows us to create the final film in full HD resolution. The distance between cameras is about 6 cm. A photo from the complete scene set-up is shown in Figure 8.

The camera is at a distance of about 80 cm from the scene. For the scene set-up, we decided to use a mixture of LEGO and natural materials while the background consisted of a printed poster.

Shooting relevant raw data for the film took about 3 days excluding scene set-up and ended up with about 400 frames. As each frame consisted of 16 single images this gives 6400 raw images. As soon as the first take was available, post processing could start using the proposed workflow and the image processing algorithms we have described. In post-production we created a film that contains all the visual effects discussed and combines them with a simple narrative.

During the development of all the necessary algorithms we learned that disparity estimation is the most critical point in the overall light field processing chain. This algorithm and supporting filters need to work as precisely as possible. Disparity maps need to be dense with no missing disparity

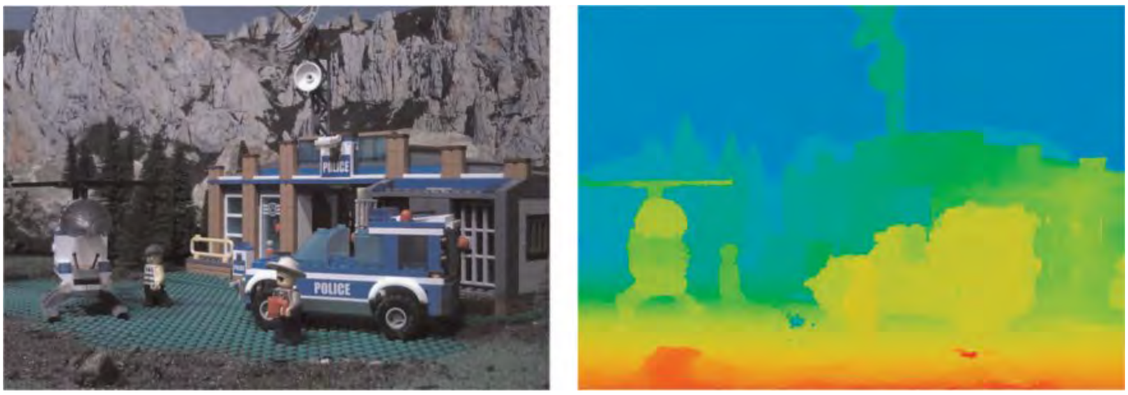


Figure 9 *Example image from a camera in the array and the corresponding disparity map. Blue areas denote low disparity; green denotes medium disparity and red areas are close to the camera*

values but the values also need to be correct with little difference from the ideal. Any such differences can lead to visible and disturbing artefacts in the final result. Figure 9 shows an example image as it is seen by one of the cameras in the array (left image) and the corresponding disparity map (right image). As can be seen, the disparity map shows small artefacts (e.g. on the ground and around the helicopter).

Nevertheless, we were able successfully to test our proposed workflow, to design visual effects in post-production and to modify them as necessary (Figures 5–7). Despite some deficiencies in the disparity maps which can be reduced manually if required, the resulting image quality was convincing.

Conclusions

Starting with a look at typical stop-motion production, we have described how shooting in stop-motion has some drawbacks compared with computer-generated animations, however, the former remains popular due to its unique character. Many of the difficulties discussed concern limitations in the stop-motion camera system.

By using a camera array we can overcome some of these limitations and can even provide more creative opportunities for artists, both on the set and in post-production. This is achieved by introducing techniques from computational imaging which simplify the camera set-up and allow the artist to focus on the story. We presented visual effects like dolly shots, refocusing and synthetic aperture that can be applied in post-production. Additionally, an example of the dolly-zoom has been presented, further demonstrating the creative possibilities available.

However, image quality is the most vital consideration when it comes to professional production and it was for this reason that we tested our technology with a short experimental production. The results proved the potential of our workflow. In our future developments we shall further improve our algorithms and acquisition systems to simplify application of our tools and we shall also improve image quality. In the longer term we shall also adapt this technology for video acquisition to make it available to a wider audience.

References

[1] FOESSEL S., ZILLY F., SCHÄFER P., ZIEGLER M., KEINERT J.: 'Light-field Acquisition and Processing System for Film Productions'. SMPTE 2013 Annual Technical Conference & Exhibition, Los Angeles, USA, Oct. 2013

[2] ZILLY F., SCHÖBERL M., SCHÄFER P., ZIEGLER M., KEINERT J., FOESSEL S.: 'Lightfield Media Production System using Sparse Angular Sampling'. ACM SIGGRAPH, Annaheim, USA, 2013

[3] LEVOY M., HANRAHAN P.: 'Light Field Rendering'. Proc. ACM SIGGRAPH, 1996, pp. 31–42

[4] FEHN C.: 'A 3D-TV approach using depth-image-based rendering (DIBR)'. Proceedings on Visualization, Imaging, and Image Processing, 2003

[5] NG R., LEVOY M., BRÉDIF M., DUVAL G., HOROWITZ M., HANRAHAN P.: 'Light Field Photography with a Hand-held Plenoptic Camera'. Computer Science Technical Report Stanford, USA, 2005

ISSN 2041-5923

Interview – Frauhofer Institute for Integrated Circuits

1. Tell us a bit about your project team. Is this the first time you've all worked together?

The group computational imaging & algorithms was established 3 years ago. It is part of the moving pictures technologies department in the Fraunhofer institute for integrated circuits IIS in Erlangen, Germany. Since then, we perform research in the field of high dynamic range imaging, lightfield image processing and multi-camera image processing. As our focus lies on applied research, we are always looking for opportunities to create algorithms which enable new production workflows. Our goal is to increase the creativeness of people in the movie business.

2. What is your paper about?

We have created algorithms which enable a multitude of visual effects to be applied in post-production. Interesting examples are refocusing, virtual camera paths and the vertigo or dolly-zoom effect. Thereby, our algorithms process data captured by a multi-camera rig. In this context, we have captured several sequences of stop-motion using a camera array consisting of 16 cameras. The multitude of perspectives can be used to generate new virtual views. Our paper describes the different effects which are possible when applying algorithms from the field of computational imaging.

3. Why are you all so excited about stop-motion animation?

On the one hand, the technique of stop-motion animation is suitable to create a laboratory environment for multi-camera image processing developments. The lighting conditions, the size of the scene and the position of the cameras and scene objects can be controlled precisely. Moreover, the requirements regarding hardware and equipment are relatively low and good results can be achieved using medium priced DLSR cameras. On the other hand, stop-motion animation has this special look. Real objects with real textures can be used, e.g. natural stones are part of our scenery which could not be generated in the same quality using computer animation. While working with the stop-motion material we learned to love the unique combination between reality and fiction which can be created using stop-motion animation.

4. Your technology brings stop-motion up to date but don't you think that it will eventually die out as an art-form?

Before we started our project, we assumed that only a very small number of stop-motion animated videos would be produced today and that the remaining ones would be created by fans and enthusiasts without any economic pressure. We learned how vital this scene really is, even growing due to the advent of medium priced DLSR cameras and *prosumer* solutions for PC-based video editing. However, stop-motion is also economically relevant for the advertising industry and children's programs. Last but not least, the commercial success of the LEGO movie show that there is a demand for this special look – even if the movie only looks like stop-motion animation while being computer animation. We believe that with new technologies such as our multi-camera image processing chain, we can revitalize this art-form by adding new creative options.

5. Did making your experimental production inspire any of you to take up stop-motion as a hobby?

Many ideas came up in the recent past, regarding interesting scene settings etc. The hope is to be able to capture them using a multi-camera array as we are excited about the new possibilities which come with the multi-camera image processing algorithms. The inspirational part can be seen as a hobby, and maybe some of us will start producing their own stop-motion animations videos, but this has not happened yet.

6. Are there any other traditional stop-motion methods which might benefit from your technology?

Our expectation is that the ability to change the focus and camera position in post-production is helpful for many stop-motion techniques, be it puppetoon, brickfilm, or even time lapse movies. We are looking forward to exploring new fields of application and hopefully we will be able to apply our algorithms to material shot using additional stop-motion methods.

7. You mention several classic stop-motion movies. If you could work with any creative team in the world, which would it be and what would you hope to learn?

There are so many teams and individuals all over the world which spend a huge amount of personal effort and enthusiasm, and put all of their heart and soul into stop-motion animation projects, so it is difficult to pick a single team. However, as we are doing applied research, we would like to learn more about specific requirements and workflows, as we try to understand what the artists need. In this context, and because we also work in the field of stereoscopic 3D production, it would be interesting to work with the team which created the movie Coraline in 3D, i.e. Laika, LCC. It is our expectation that their approach to capturing multiple images of the same scene using a single camera and a translation robot could be easily combined with our multi-camera image processing algorithms.

Selected content from IBC 2014

ISSN 2041-5923

THE ART OF BETTER PIXELS

D.G. Brooks

Dolby Laboratories, Inc., USA

Abstract: This paper gives an overview of the key experiments used to determine the parameters required for baseband video which combines both HDR (High Dynamic Range) and WCG (Wide Colour Gamut). We call this combination EDR (Extended Dynamic Range) for short.

Furthermore, it provides an insight into the EDR content creation process and proposes how the creative intent can be propagated throughout the distribution chain all the way to the TV receiver, in order to provide the viewer with the best possible experience on a wide range of EDR-capable TV receivers.

It is hard to exaggerate the impact that the cathode ray tube (CRT) had, and still continues to have, on the design of television standards. Our CRT heritage affects our HDTV digital standards in a number of key ways: (1) The colour gamut (as defined by the rare earth phosphors used). (2) The brightness was limited to ~100 candela/m^2 (units often referred to as "nits") not only to control large area flicker, but also to prevent the electron beam from spreading and reducing spatial resolution.

INTRODUCTION

The quality of colour television has been dominated by the physics of the cathode ray tube: in determining the spatial resolution, the image dynamic range, the colour gamut and even the frame rate of the images displayed.

UltraHD (UHD) as specified in Recommendation ITU-R BT.2020, was the first standard to partially break free from these constraints, however it retained the 100 candela/m^2 CRT reference brightness limit and the corresponding Electro-Optical Transfer Function (EOTF) based on the gamma characteristic of the CRT.

In April 2012, the United States made a submission to ITU-R Working Party 6C which proposed a new EOTF in order to enable the carriage of high dynamic range content through the 10 and 12-bit baseband interfaces specified in the then draft version of ITU-R BT.2020.

The new perceptual EOTF proposed was based not upon the gamma function of a CRT, but directly on the contrast sensitivity ratio of the human eye as measured by Barten and referenced in Report ITU-R BT.2246-2. The submission further proposed that, based on viewer preference testing, an HDR system should be capable of handling signals with a brightness range from 0 to 10,000 candela/m^2.

LOOKING TO THE FUTURE

There are 3 fundamental ways to improve moving image quality

- Increase spatial resolution – more pixels
- Increase the frame rate – faster pixels
- Increase the capability of each pixel to represent more colours and increased dynamic range – better pixels.

The ITU-R UHDTV Recommendation BT.2020 (1) published in August 2012 offered both more and faster pixels with an increased colour gamut, but did not consider the need for pixels with an increased dynamic range and brightness.

This vestige of the CRT lives on despite the increasing interest in high dynamic range systems in still photography.

What dynamic range is required for entertainment content?

Dolby developed the experimental display shown in Figure 1 to answer this question. This display consists of a digital cinema projector pointed at a 23" monochrome LCD

Figure 1 *Experimental HDR display*

panel. The digital projector image and the LCD panel are dual modulated to create a display capable of a black level of 0.004 nits and a peak white level of 20,000 nits.

Tests were performed to find out the preferred viewer experience for

- Black level
- Diffuse White level
- Highlight level

Content was carefully chosen to ensure that the image contrast ratio did not change while testing with average luminance level and largely achromatic images were used to avoid viewer preferences for more 'contrasty' or colourful images.

The full details of this experiment can be found in the papers published by Daly et al. (2) Daly et al. (3) and Daly, Kunkel and Farrell (4)

From Figure 2 it is evident that the current TV standards are inadequate as regards both dynamic range and brightness, with consumer preferences being orders of magnitude greater than today's television systems can provide.

Since the above results were first published, there have been questions as to whether these diffuse white and highlight levels are equally applicable to larger displays. An additional experiment was performed, with the same stimuli, using a 4-metre screen and a digital cinema projector. In both experiments the viewer was positioned 3 picture heights from the screen. The results of the later experiment are compared with small screen results in Figure 3.

Table 1 summarises these experiments and shows that a dynamic range of 22 F-stops is needed to meet 90% of viewer preferences for both small and large screen applications. One can further conclude that a television system which is capable of delivering content with a dynamic range from 0.001 nits to 10,000 nits would satisfy the vast majority of viewers on a range of consumer devices from tablets to very large LCD displays.

COMBINING HIGH DYNAMIC RANGE WITH WIDER COLOUR GAMUT

Recommendation ITU-R BT.2020 (BT.2020) provides next-generation standard dynamic range TV systems with

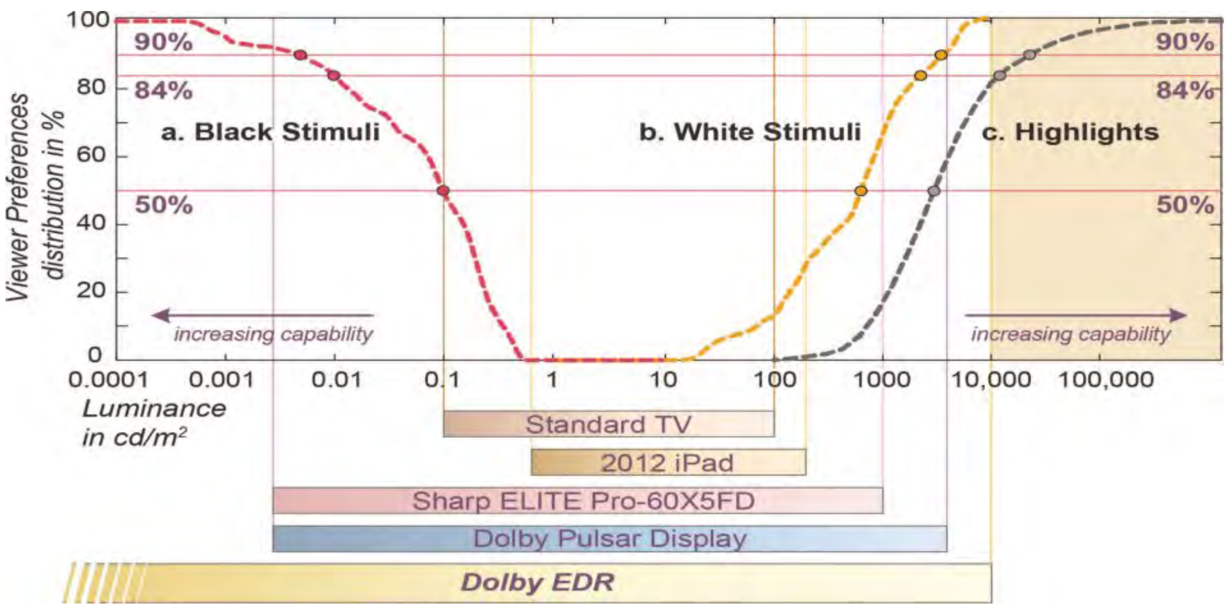


Figure 2 *Summary of dynamic range subjective tests*

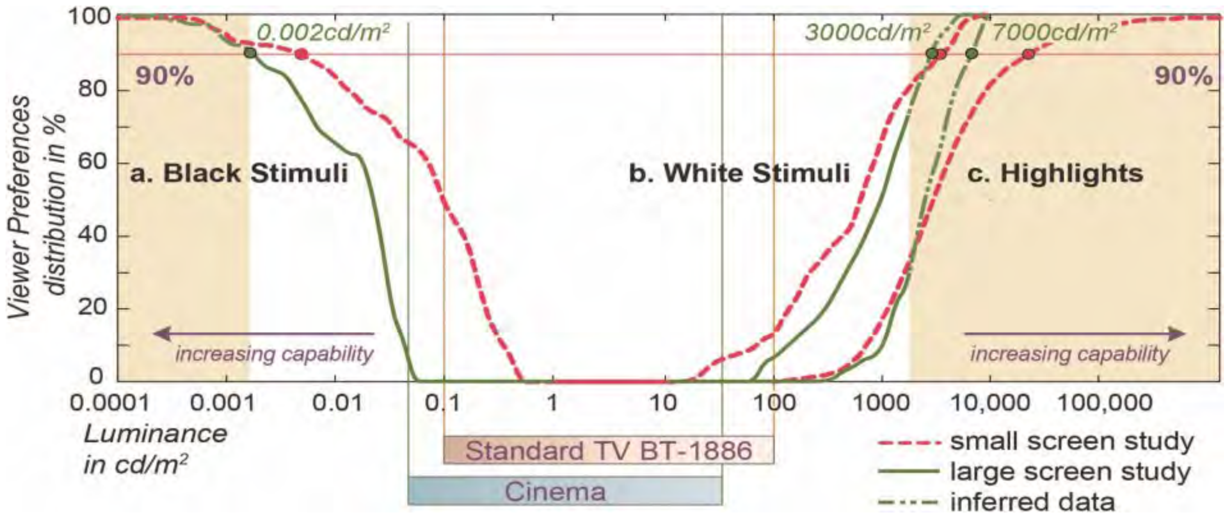


Figure 3 *Summary of large screen dynamic range subjective tests*

Table 1 Summary of Dynamic Range Experiments

	Luminance level to satisfy 90% of Viewers	
	Large Screen	Small Screen
Black	0.002 nits	0.005 nits
	21 F-stops	20 F-stops
Diffuse White	3,000 nits	4,000 nits
	1 F-stop	2 F-stops
Highlights	7,000 nits	20,000 nits

a much wider colour gamut than current Recommendation ITU-R BT.709 (BT.709) (5), consequently any new television system must be capable of delivering both wider colour gamut and high dynamic range content.

Having now defined the required colour volume needed to satisfy the vast majority of viewers, how can this be represented in a practical television system?

Colour Volume

Traditionally we have used the colour "horseshoe" diagram to represent the colour gamut of a television signal such as shown in Figure 4, however for each colour shown there is a corresponding maximum luminance. White is the brightest colour, as television uses an additive colour system.

An alternative representation, which allows both the colour gamut and the dynamic range to be represented, is shown in Figure 5. Figure 5 also shows both today's BT.709 colour space and the new EDR colour space, encompassing both the high dynamic range requirement from 0 to 10,000 nits and the colour gamut as defined in BT.2020.

This representation has the advantage of showing all possible colours at all available luminance levels.

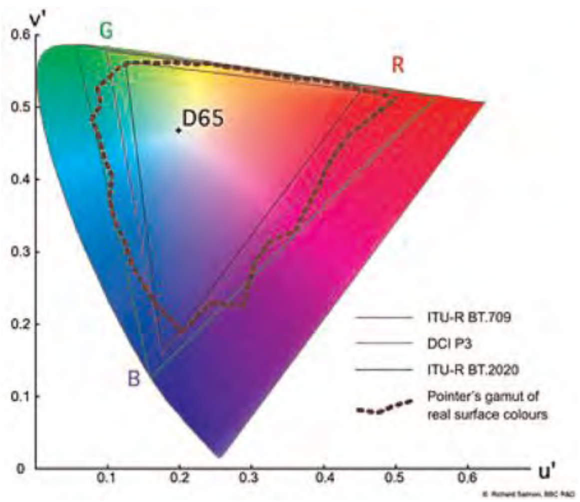


Figure 4 *The colour "Horseshoe"*

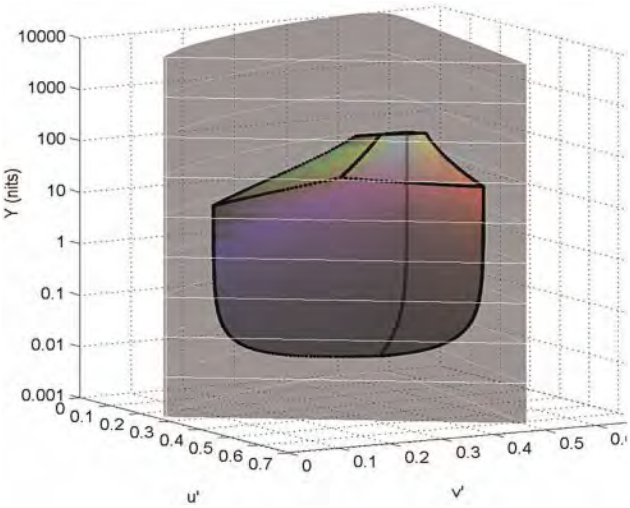


Figure 5 *Colour volume EDR v. BT709*

EDR Colour Volume Quantization

What bit-depth is required to accurately represent this EDR colour volume shown in Figure 5? To answer this question requires an investigation into both luminance and chrominance elements.

Luminance Quantization

Report ITU-R BT.2246-2 (6) makes use of Barten's model for the contrast sensitivity function of the human visual system to ensure that the contouring artefacts in the BT.2020 UHDTV standard are acceptable.

Using the existing Electro-Optical Transfer Function (EOTF) of Gamma 2.4, as defined in Recommendation ITU-R BT.1886 (7), would require a 15-bit representation to match the contrast sensitivity or contouring performance of the human visual system as shown in Figure 6.

However, for a luminance range of 10,000 nits, Gamma 2.4 does not accurately match the contrast sensitivity function of the human eye. Gamma wastes bits by coding bright areas which are well below the Barten threshold. Alternatively, using a 13-bit log representation wastes bits in the darker regions.

To resolve these issues a new Perceptual Quantizer (PQ) (8) was developed which follows the Barten curve to ensure the optimum use of the bits available. From Figure 6, it can be seen that 12-bit PQ is below the visible threshold for contouring artefacts whereas 10-bit PQ is above.

Colour Quantization

Whilst using PQ as the EOTF can provide the solution for monochrome images, an additional metric is required to ensure that colour images are also artefact free. The JND Cross test, Figure 7, was developed for this purpose.

Each square in Figure 7 represents a block of pixels which have been perturbed by 1 code word away from the grey background. All possible variations of this in RGB result in 26 different colour patches.

If the picture looks like a flat grey image, then there is no visible quantization. If some patches are visible, then some contouring or banding is likely.

This test is then performed for multiple different grey levels. The industry-accepted metric to judge the visibility of these patches is CIEDE2000 (9) colour difference formula.

Subjective tests by Nezamabadi et al. (10) indicate that for noise-free images such as animation, CGI and graphics a DE2000 threshold of ~2.5 is needed to avoid contouring

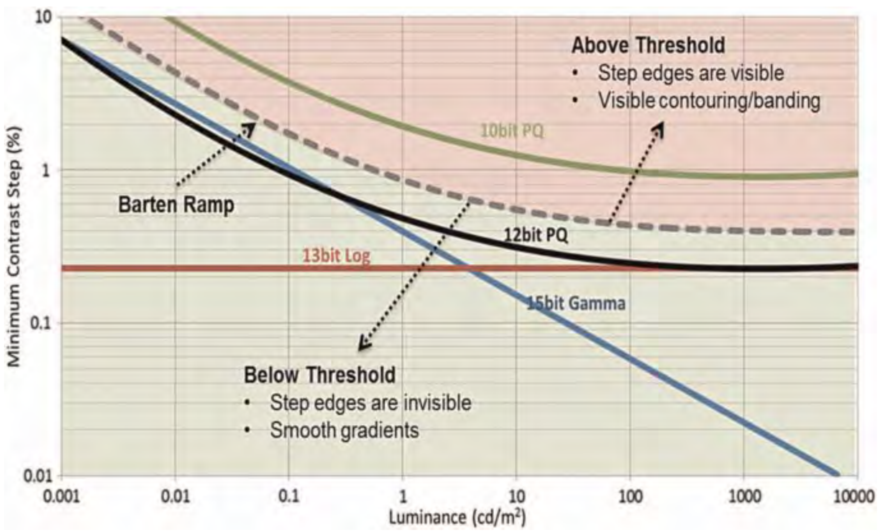


Figure 6 *Barten's model over 10,000 nit range*

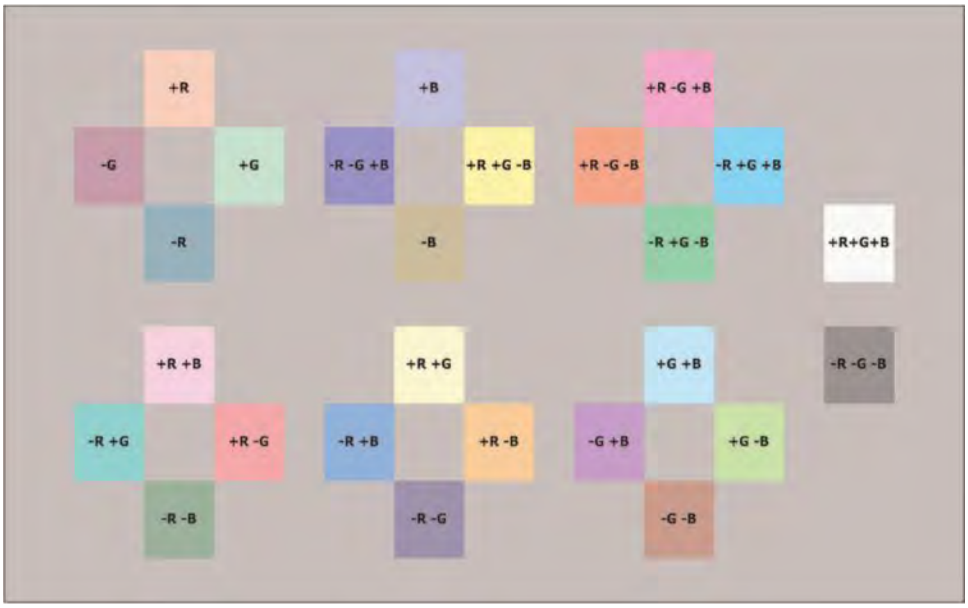


Figure 7 *The JND cross (colours exaggerated)*

artefacts. When noise is present, this threshold increases to ~5.

It should also be noted that the DE2000 metric starts to lose accuracy below about 3 nits; this corresponds to the point where the human visual system begins to switch between photopic and scotopic vision.

Combining the use of this metric with the JND colour patches allows the quantification of possible colour quantization artefacts. The results are shown in Figure 9.

EDR Colour Volume Conclusions

For noise-free EDR images such as animation, CGI or graphics, using 12-bit PQ ensures that both monochromatic and colour banding artefacts are below the visible threshold. This matches the MovieLabs specification for next-generation content (11).

For captured images that contain some noise, using a 10-bit PQ representation is adequate, but care needs to be taken especially with graphics. However, this is an area well understood by broadcasters who deal daily with these issues for today's 8-bit BT.709 transmissions.

Based on the above studies, a detailed submission was made in July 2012 by the United States to ITU-R Working Party 6C proposing the PQ EOTF for High Dynamic Range/Wider Colour Gamut images (12). A Rapporteur Group was subsequently set up to examine the options. Work in this group is on-going. In parallel with this effort, the PQ EOTF is in the process of being

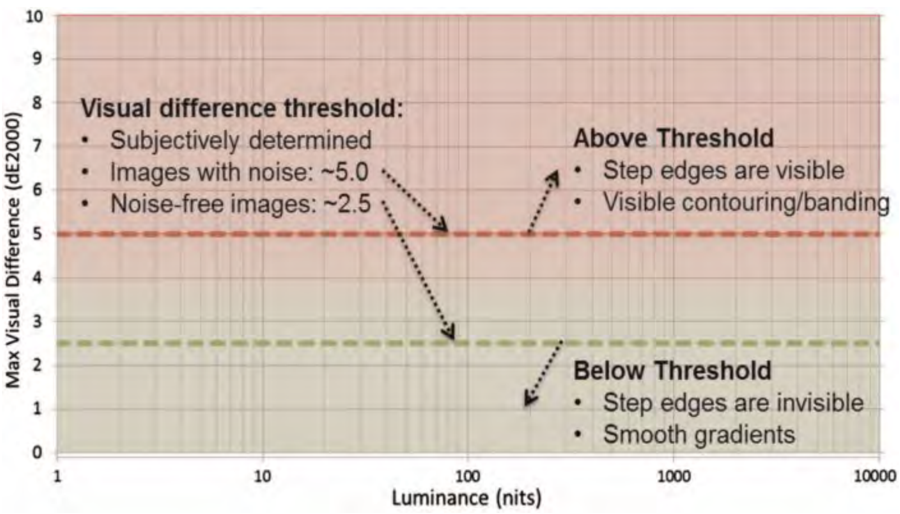


Figure 8 *CIEDE2000 overview*

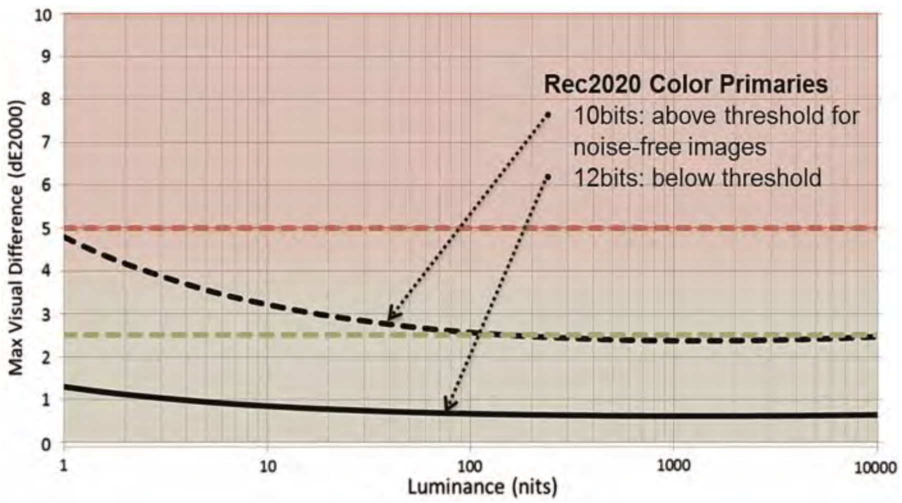


Figure 9 *EDR evaluation using DE2000*

standardized by the Society of Motion Picture and Television Engineers (SMPTE) (8).

GETTING EDR CONTENT TO THE HOME

Having established the definition for EDR content, how best can this content be transmitted to the home and presented to the viewer?

The following criteria were defined to meet this requirement:

• Single production workflow required for both SDR and EDR content

• Compatibility with existing off-line and real-time infrastructures

• The highest quality images, matching the creative intent, should always be available to the home viewer

• Image quality will increase, in both professional and consumer applications, as display technology evolves - the transmission system must not be the bottleneck

• The transmission system should allow backwards compatibility with today's BT.709 or BT.2020 systems

• Should be independent of any spatial resolution, colour gamut or frame rate

• The EDR transmission system should use industry standard codecs

• A bit-rate-efficient transmission method is required for EDR

• New EDR TVs must be capable of mapping the transmitted EDR images to the TV display's native colour volume (defined by the display's black level, peak white level, colour temperature and colour primaries) thereby remaining as true as possible to the original creative intent within the confines of the display's capabilities

An end-to-end system overview

Figure 10 provides an end-to-end architecture / workflow for off-line content creation and distribution which meets the requirements described above.

The ability to create EDR content has been limited, not by today's cameras, but by existing 100nit broadcast reference monitors. The development of the Pulsar monitor, which has a peak brightness of 4000nits with a DCI-P3 colour gamut, plus a set of EDR tools as plug-ins to commonly used off-line production suites, meet the needs of the creative.

In parallel with this process, image metadata is automatically generated and stored along with any creative input. This metadata is used downstream to guide the display management block in the TV receiver, based on both automatic and creative input captured during production. The definition of this metadata has begun and is now in the process of becoming a SMPTE standard (13).

To meet the practical requirements of backwards compatibility with today's BT.709 and future BT.2020 transmission systems, a dual-layer codec architecture has been developed which allows the existing standards to be transmitted as the base layer. An enhancement layer contains the information necessary to recreate the EDR signal. Figure 11 shows a block diagram of this architecture. It is worth noting that this dual-layer codec works outside of the coding loop (unlike MVC or SVC,

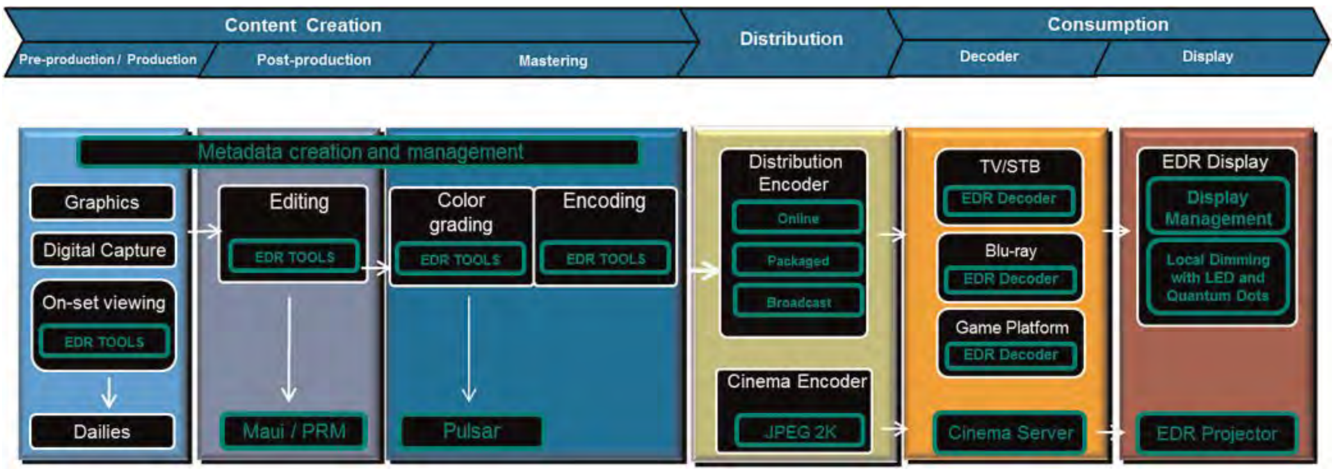


Figure 10 *Off-line EDR system architecture/workflow*

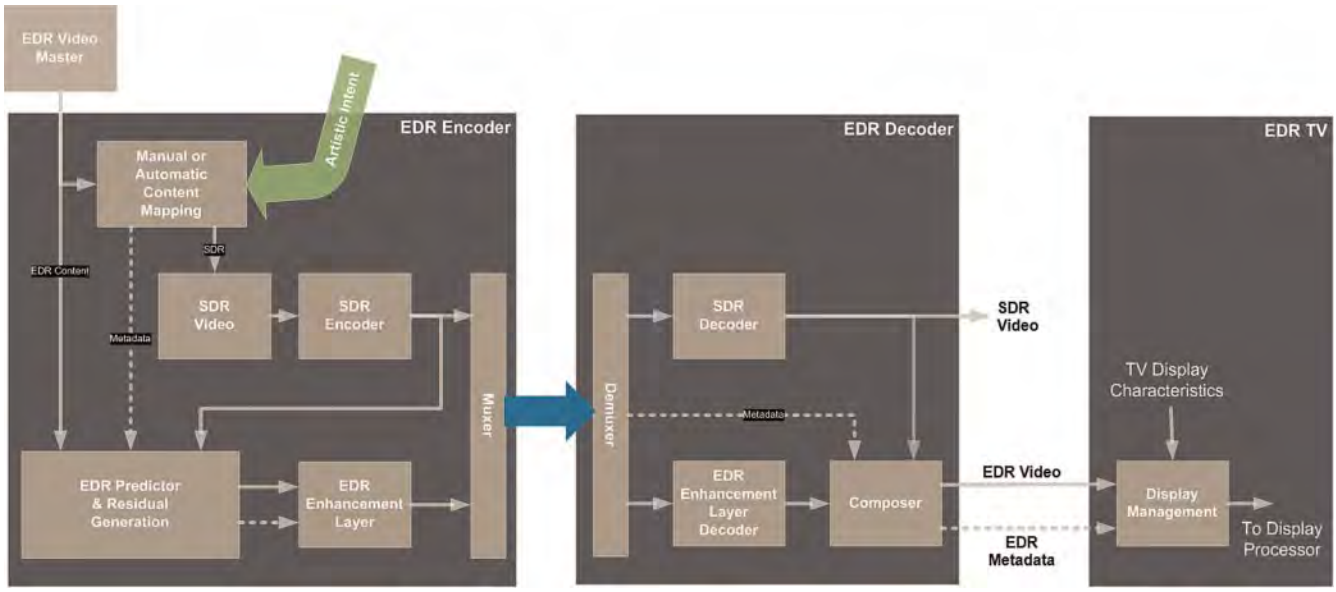


Figure 11 *The EDR dual-layer codec*

for example) so that no changes are required to either encoders or decoders.

Moreover, the dual layer codec as shown in Figure 11 allows for either an 8-bit AVC or an 8/10-bit HEVC base layer for compatibility with existing services. Similarly the EDR input signal can be either 10-bit PQ or 12-bit PQ and either will be faithfully reconstructed at the dual decoder output.

Using this dual-layer codec also allows existing TVs to receive the base layer standard dynamic range signal, whilst new EDR televisions can receive the EDR signal and adapt the content to match the capabilities of the display. The EDR enhancement layer only increases the bit-rate by ~25% compared with the backwards-compatible base layer.

The capabilities of the EDR TV will evolve over time as new technologies such as quantum dots allow wider colour primaries and brighter displays. By the use of a combination of television's inherent display characteristics plus the EDR metadata transmitted, the display management function inside the TV is able to adapt optimally the incoming EDR images to match each display's characteristics whilst maintaining the creative intent.

CONCLUSION

This paper outlines the experiments made to determine the image dynamic range required for television and cinematographic entertainment purposes.

Using this information, a series of tests and experiments were performed to determine the baseband bit-depth required to ensure that no contouring artefacts would be visible.

In making these tests, the traditional TV gamma 2.4 non-linear curve was proven to be inefficient for high dynamic range images and a new Perceptual Quantizer was designed to match the characteristics of the human visual system.

Both 10 and 12-bit PQ EDR signals can be encoded by the dual-layer codec proposed. The SDR base layer can be coded as either 8 or 10-bit depending on the backwards compatibility requirements.

Existing SDR TV receivers will ignore the enhancement layer and display the SDR content as today.

New EDR receivers will decode both base and enhancement layers and combine these together to faithfully reproduce EDR images at either 10 or 12-bit PQ depending on the source. The TV receiver will also incorporate a new display management block which maps the content to the evolving characteristics of each display by using the metadata present in the EDR stream.

ACKNOWLEDGEMENTS

The author would like to thank his colleagues for all their work in developing EDR (Dolby Vision) from Image Research & Development in Burbank, Sunnyvale, San Francisco and Yardley. I would also like to thank my wife, Naomi, for encouraging me to write this paper over a beautiful long spring weekend.

Figure 4 is courtesy of Richard Salmon at the BBC.

REFERENCES

[1] Recommendation ITU-R BT.2020 'Parameter values for ultra-high definition television systems for production and international programme exchange', August 2012

[2] DALY S., KUNKEL T., SUN X., FARRELL S., CRUM P.: 'Preference limits of the visual dynamic range for ultra high quality and aesthetic conveyance'. SPIE Electronic Imaging, HVEI, Burlingame, CA, Feb 2013, 2013

[3] SCOTT DALY, TIMO KUNKEL, XING SUN, SUZANNE FARRELL, POPPY CRUM: 'Viewer Preferences for Shadow, Diffuse, Specular, and Emissive Luminance Limits of High Dynamic Range Displays'. SID Display Week, paper 41.1, Vancouver, Canada, (2013)

[4] SCOTT DALY, TIMO KUNKEL, SUZANNE FARRELL: 'Using Image Dimension Dissection to Study Viewer Preferences of Luminance Range in Displayed Imagery'. Invited talk, IDW, International Workshop on Displays, SID, Sapporo, Japan, 6 Dec 2013, (2013)

[5] Recommendation ITU-R BT.709-5 'Parameter values for the HDTV standards for production and international programme exchange', April 2002

[6] Report ITU-R BT.2246-2 'The present state of ultra-high definition television', November 2012

[7] Recommendation ITU-R BT.1886 'Reference electro-optical transfer function for flat panel displays used in HDTV studio production', March 2011

[8] Draft SMPTE Standard 2084 'High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays'

[9] CIE Publication 142:2001 "Improvements to industrial colour difference evaluation", Commission Internationale de l'Eclairage, Vienna, Austria (2001)

[10] NEZAMABADI M., MILLER S., DALY S., ATKINS R.: 'Color signal encoding for high dynamic range and wide color gamut based on human perception'. Color Imaging at SPIE' s Electronic Imaging Conference, (2014)

[11] MovieLabs 'Specification for Next Generation Video' – Version 1.0, 2013

[12] Input document USA002 to ITU-R Working Party 6C, 'Parameter values for UHDTV systems for production and international programme exchange' April 2012

[13] Draft SMPTE Standard 2086 'Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images'

Selected content from IBC2014

ISSN 2041-5923

Source-timed SDN video switching of packetized video

Thomas G. Edwards
FOX Network Engineering & Operations, USA

Abstract: The broadcast industry is considering a move from a serial digital interface (SDI) infrastructure to a computer-networked infrastructure. Professional media networking will enhance flexibility and agility of the broadcast plant, reduce and simplify cabling, provide for format agnostic transport, and will allow media organizations to benefit from the economies of scale of COTS networking hardware.

One key element of video processing is synchronous (clean) switching between video sources. The accuracy of the SMPTE RP 168 recommended switch point is of the order of 10 ms. Unfortunately, the current speed of networking switch flow-rule changes using software defined networking (SDN) is of the order of 1 ms to 10 ms and the speed of rule changes is not highly deterministic.

To solve this problem, source-timed video switching separates the actions of updating SDN flow-rules with the timing of the actual video switch. This technique first adds new SDN flow-rules to the networking switch that match a packet header value that is not currently being transmitted. Then the video source changes its output's packet header value to the new value in the flow-rule at the precise RP 168 video switch point. The new packet header value matches the new flow-rules, thus the effect of these two steps achieves precisely timed synchronous video switching (which can be compared with the actions of "preview" and "take" on a video switcher).

FOX has developed a proof-of-concept of source-timed video switching using live HD-SDI sources packetized using SMPTE 2022-6 and synchronously switched on a COTS 10 Gbps Ethernet switch.

Introduction

Professional media networking is the use by professionals of a networked infrastructure to process media. Although the broad definition includes file-based workflows, the greatest challenge of professional media networking today is the carriage of real-time audio and video flows over IT networks.

Ethernet switches are readily available today at affordable prices that can switch terabits per second of traffic in a non-blocking fashion. For example, a 50-port 10 gigabit per second Ethernet (10 GbE) switch carrying six HD-SDI flows bi-directionally per port, can serve as a 300×300 uncompressed HD-SDI non-blocking video router.

One of the key elements of video processing is the synchronous ("clean") switching between video sources. If video is switched in the middle of a frame, there will be undesirable video artifacts, thus it is desirable for the switch to occur during the vertical blanking interval (VBI). There is often ancillary data transmitted during the VBI, so care needs to be taken not to damage that data.

SMPTE has defined [1] a switching point in Recommended Practice 168 (RP 168), such that the effects of any signal discontinuity in the processing chain due to switching is minimized. The line designated for the switching point is chosen to be after the vertical sync, but early in the VBI. This is to ensure that ancillary signals transmitted during the vertical interval remain with the video frame that they are associated with. Also ancillary data is excluded from the line following the switch line to allow for receivers to synchronize to EAV/SAV (End/Start of Active Video) before data becomes present to avoid its loss. This exclusion is reflected in several SMPTE ancillary data standards. Finally, the recommended switch area on the switching line is a region relatively within the middle of the active video to provide some tolerance for inaccurate timing and also to provide a reasonable "guard area" to protect the SAV and EAV data flags.

With packetized video, it is unclear how many of the restrictions in RP 168 remain truly relevant. As every bit should survive transport within a clearly defined timing matrix, there are no concerns about synchronization to EAV/SAV. And as video packets should fully tile the raster without bit-timing jitter, the switch could occur earlier in the VBI (if care is taken not to bisect ANC data at packet boundaries). However until our experience allows us to loosen the tolerances on the switching point, it may be wise to implement switching in the RP 168 area.

This paper will examine the different methods of performing synchronous switching of uncompressed video, and will concentrate on a novel method of switching that utilizes a relatively new networking technique known as software defined networking (SDN). In SDN, Ethernet switches expose an API such as OpenFlow [2] to allow for precise control of packet flow-rules, including matching packet headers on layer 2, layer 3, and TCP/UDP fields and then performing actions on the matched packets (such as sending them to a specific output port, dropping them, re-writing a packet header, etc.). SDN breaks the "old model" of networking where switches implemented their own specific algorithms for setting up packet-forwarding rules (such as spanning tree and IP routing protocols), and replaces them by having a separate "controller" that monitors the network as a whole, is directed by applications (such as broadcast-specific applications), and then sets up flow-rules on Ethernet switches in the data center to achieve the desired application-directed goals.

SMPTE 2022-6

SMPTE 2022-6 [3] carries the entire SDI signal including all ancillary data (such as embedded audio, Closed Captioning data, and Active Format Descriptor) in an uncompressed fashion. It uses an RTP payload to deliver the SDI data using UDP. The High Bit Rate Media (HBRM) Payload Header of 2022-6 includes video format identification, a frame count, a sample clock frequency, and a video timestamp. FOX has developed a Wireshark Lua packet dissector for 2022-6 available at: https://github.com/FOXNEOAdvancedTechnology/smpte2022-6-dissector. An example of the output of this dissector can be seen in Figure 1.

The payload for each RTP datagram in 2022-6 is 1376 octets of SDI data. Since SDI typically has 10-bit samples, often a sample is split between two RTP datagrams. The last datagram of the frame ends with zero padding if there is not enough SDI data left in the frame to fill all 1376 octets. Also the last RTP datagram in a frame has its marker bit set. Total data overhead for 2022-6 carried by Ethernet over the original SDI bit stream is about 5.4% to 5.9% depending on header options. Due to the single uniform size of the datagram for all video formats, there

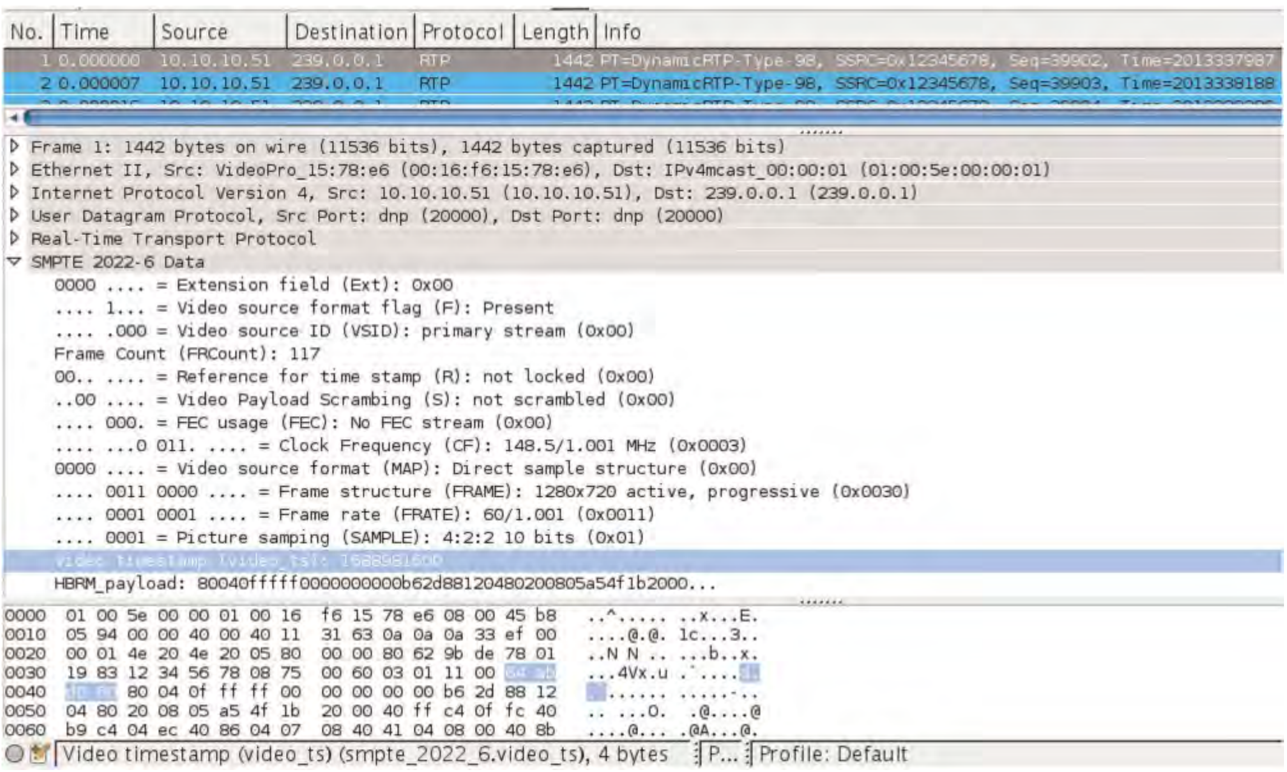


Figure 1 *SMPTE 2022-6 Wireshark dissection*

typically is not an integer number of RTP datagrams per line (see Figure 2, the diagonal white lines are RTP datagram boundaries of 2022-6).

Switching 2022-6 in the RP 168 area on the boundary of datagrams can be achieved by counting datagrams from the end-of-frame set RTP marker bit as in Laabs [4]. Table 1 lists valid RTP datagram boundaries in the RP 168 area for 2022-6 streams.

Through the efforts of the Video Services Forum, there has been extensive work on interoperability between multiple vendors for 2022-6 streams, as described in Kouadio [5]. Although 2022-6 was originally conceived of as an inter-facility data stream, some vendors are now exploring its use within the broadcast plant. IP cores for conversions between SDI and 2022-6 are commercially available for the major FPGA types.

Methods of switching packetized video

Three potential strategies for synchronous switching of real-time packetized video streams are shown in Figure 3.

Previous work has attempted to solve the problem of precisely timed synchronous video switching through precise timing of packet flow-rule changes at the Ethernet switch ("switch-timed"), or by using buffering to allow for flow-rules to be precisely changed at the destination ("destination-timed"). Unfortunately, both of these techniques have significant real world drawbacks.

Switch-timed video switching

The switch-timed video switching technique suffers from the fact that today, common-off-the-shelf (COTS) Ethernet switches generally cannot provide the temporal accuracy of flow changes required for synchronous video switching. To change flows within the RP 168 area for 720p/59.94 video, the required accuracy would need to be within the length of time represented by one 2022-6 datagram, or about 7.4 ms. Unfortunately, reported OpenFlow rule update rates of COTS network devices from Bilbery et al. [6] and Mogul et al. [7] range from 100 to 1000 per second, implying an upward limit of temporal precision between 1 ms and 10 ms. It is possible that packet-forwarding rule update rates on COTS network devices will become faster over time but they are not fast enough for reliable synchronous video switching today.

Destination-timed video switching

Destination-timed video switching requires the destination to receive the video flow to be switched into before ending the reception of the video flow to be switched out of ("make before break"). During that time, the destination buffers data from both video flows and the destination device itself can determine the precise video switch point. This solution can be readily implemented using IGMP, as in Laabs [4]. However during the period between joining the new multicast group and leaving the old multicast group, one or more data paths through the network will have to carry twice the bandwidth of the media flow. In particular, the "last mile" Ethernet switch port that the receiving device is directly attached to is likely to be limited to handling half of the media flows that it could otherwise carry in a

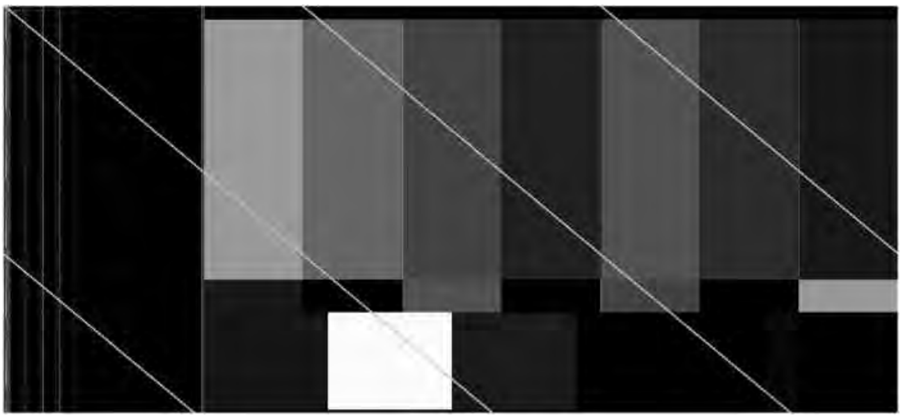


Figure 2 *SMPTE 2022-6 datagram boundaries*

Table 1 RP 168 switch points for 2022-6

	720p	1080i (F1)	1080i (F2)
RTP Datagram #'s in frame	20–21	26–27	2272–2273
Boundary, pixels after SAV end	738	831.4	629.8

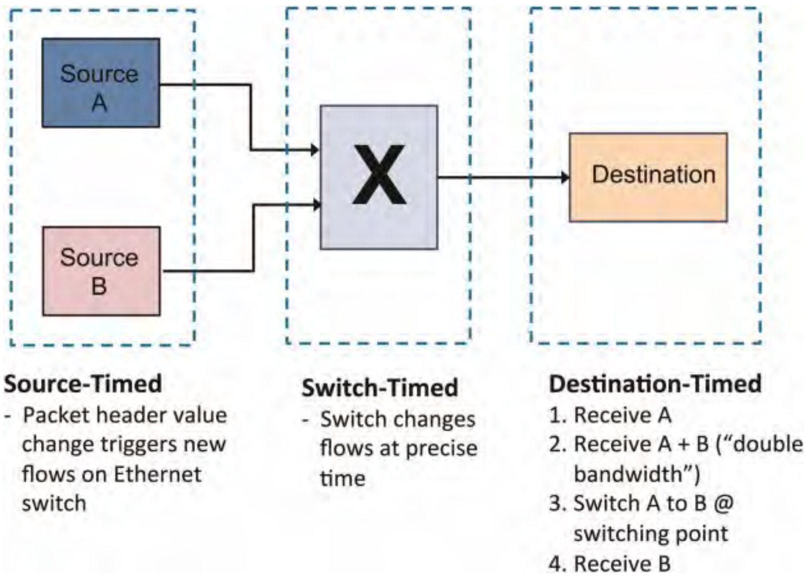


Figure 3 *Three ways to switch packetized video*

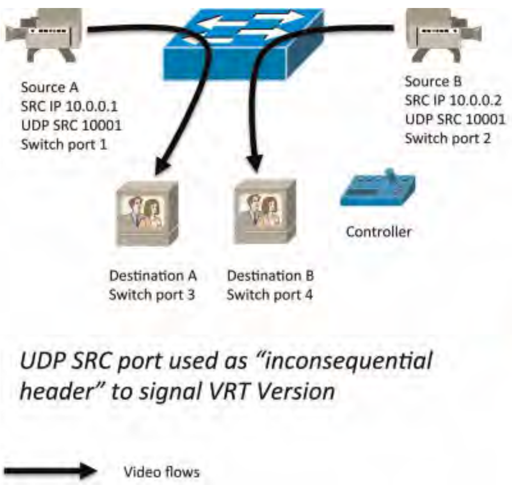
steady-state environment to avoid potential blocking, representing a significant waste of bandwidth capacity.

Source-timed video switching

The concept of source-timed video switching is to separate the temporally inaccurate process of updating Ethernet switch packet-forwarding rules from the actual precise timing of a synchronous video switch. To do this, a specific element of the packet header is selected as the "timing signal" match field whose value will be used to trigger a precise flow change by matching rules previously configured using SDN on the network devices. Preferably this header will have little or no impact on other stream processing functions. In our proof-of-concept, the UDP source port value was used.

Figure 4 shows the initial state of an example source-timed video switching network. There are two video flow sources and two video flow destinations. A notional "video routing table" (VRT) shows the relationship between flow sources and destinations. The VRT has a version number (in this case 10001), which will be used to match the timing signal match field of the packet flow-rules. Initially in this network, a flow from Source A is sent to Destination A, and a flow from Source B is sent to Destination B. The flows have the UDP source port 10001, which matches the flow-rules on the switch to carry out the routing as directed by the VRT. This example only shows one-to-one flows, but the flows could also be one-to-many.

A request to change video flows is received by the SDN controller in Figure 5. The notional VRT is first updated to reflect flows from Source A to be sent to Destination B, and flows from Source B to be sent to Destination A. This VRT has the new version number 10002. The controller responds by adding new flow table rules to the Ethernet switch to steer packets that have UDP source ports



Initial Video Routing Table (VRT):

VRT #10001	Source A	Source B
Destination A	X	
Destination B		X

Initial SDN switch flow table:

IP SRC	UDP SRC port	Action
10.0.0.1	10001	Forward Port 3
10.0.0.2	10001	Forward Port 4

Figure 4 *Source-timed switching initial state*

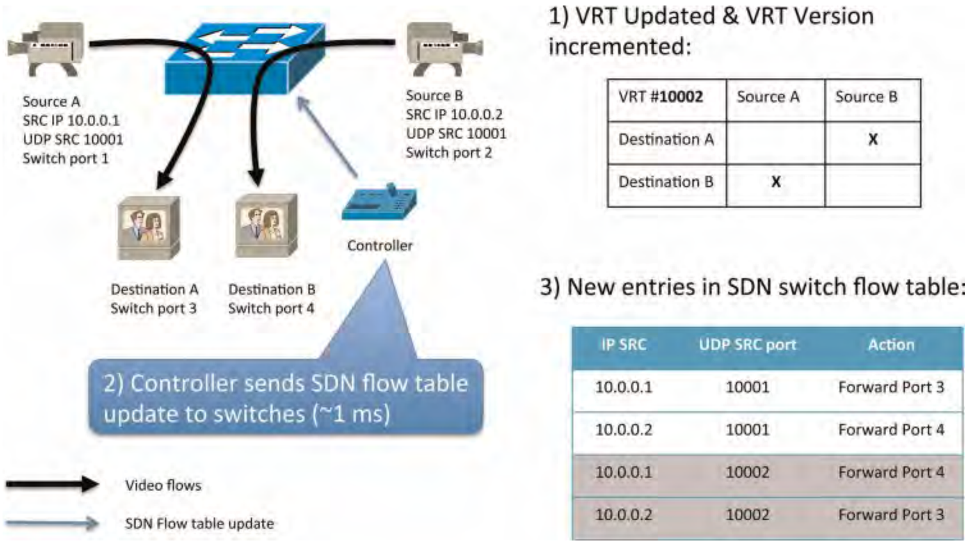


Figure 5 *Source-timed switching switch request*

matching the new VRT version number (a process that could take tens of milliseconds). During this time, the sources are still transmitting flows with UDP source port 10001, so the flows are still following the paths as per VRT version 10001.

After enough time has passed for the new flow-rules to have been reliably added to the Ethernet switch using SDN, the controller informs the sources of the new VRT version number, as illustrated in Figure 6. This prepares the sources to change the UDP source port of their emitted flows at the precise time of the next RP 168 video switching point.

In Figure 7, the sources have changed their UDP source port to 10002 at the precise video switching point, and since the flows now match the new flow-rules on the switch, they are steered as directed by FRT version 10002.

After this point in time, the old (now unused) flow-rules may be removed from the switch, either programmatically or using OpenFlow ***idle_timeout***.

Proof of concept

A proof-of-concept was developed to test and demonstrate source-timed SDN synchronous video switching.

Hardware

A schematic diagram of the proof-of-concept hardware is shown in Figure 8. Uncompressed HD-SDI video sources were from an SSD-based player and a live camera. Both

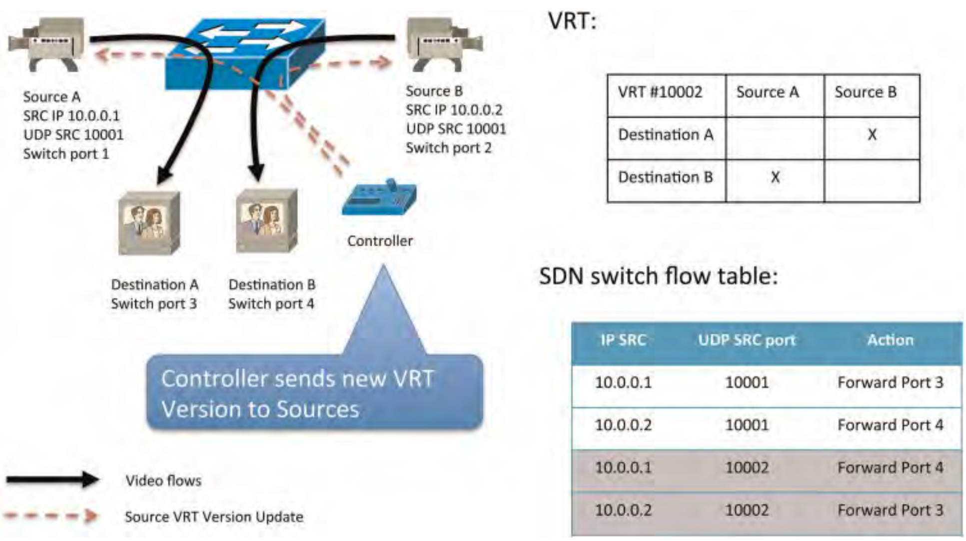
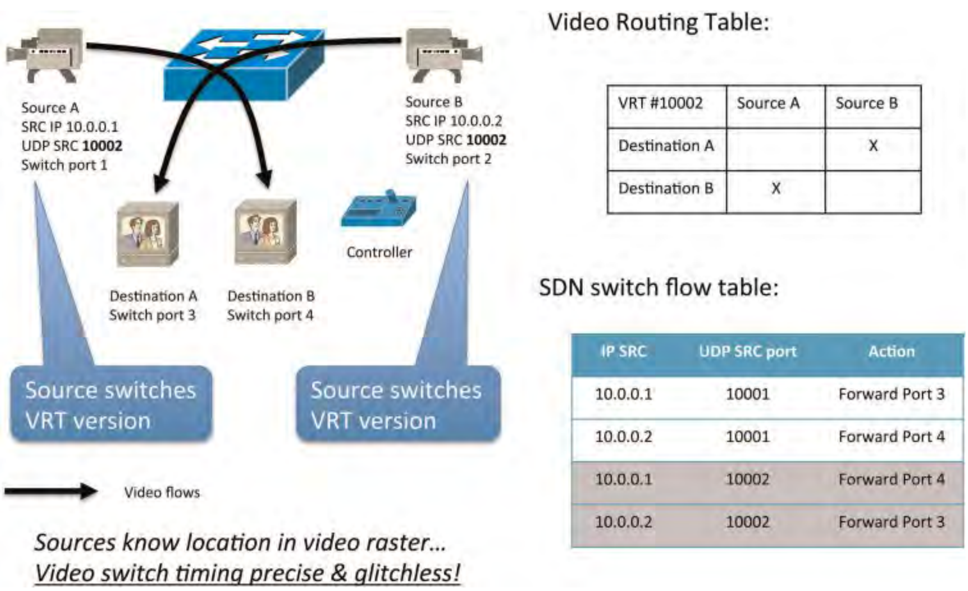


Figure 6 *Source-timed switching VRT version update*



VRT #10002	Source A	Source B
Destination A		X
Destination B	X	

IP SRC	UDP SRC port	Action
10.0.0.1	10001	Forward Port 3
10.0.0.2	10001	Forward Port 4
10.0.0.1	10002	Forward Port 4
10.0.0.2	10002	Forward Port 3

Figure 7 *Source-timed switching, sources change VRT version*

video sources emitted 720p/59.94 fps video and they were synchronized using a signal from a blackburst generator.

The uncompressed HD-SDI video from the sources was packetized using a 2022-6 gateway. The gateway device has four HD-SDI inputs, four HD-SDI outputs, and two 10 GbE SFP+ ports. One gateway was used for packetization of both HD-SDI inputs, which allowed RTP and 2022-6 HBRM metadata to be consistent between the two streams. This gateway also de-packetized one video flow output from a particular port of the Ethernet switch into HD-SDI. A second gateway was used only for a de-packetization of a second video flow output from another port of the Ethernet switch. HD-SDI outputs from the gateways were displayed on professional monitors.

A server was used to host software implementing the source-timed packet header changes as well as SDN orchestration. The server had one 3.3 GHz Intel Xeon E3-1230 V2 processor, 8GB of RAM, and ran Ubuntu Linux. Its Ethernet card was a dual Intel 82599EB-based 10 Gbps SFP+ NIC. The NIC was operated using “DNA (Direct NIC Access)” DMA-enabled drivers from Ntop.org. Attached to the server was a USB-connected button box that emulated pressing a space bar on a keyboard to command a switch of the video flows.

The Ethernet switch was an Arista 7050S-52, a 52-port 10 Gbps Ethernet switch with SFP+ interfaces and a 1GbE out-of-band management (control) port.

Server software

A controller was written in Python to orchestrate the source-timed flow change operation. When the button box was pushed, the controller contacted a DirectFlow agent on the

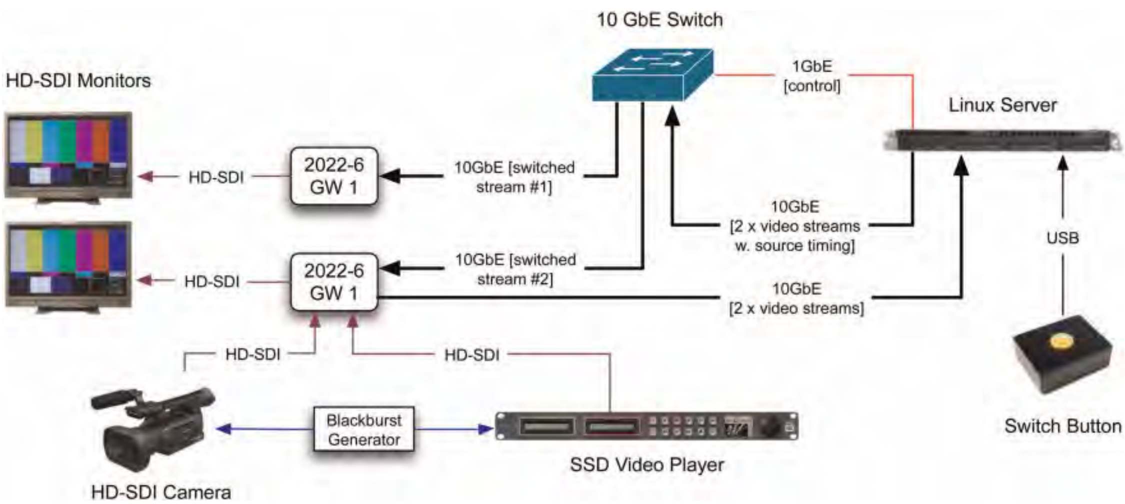


Figure 8 *Source-timed switching proof of concept schematic diagram*

switch to install new flow-rules. After a delay of 100 ms, the controller then commanded a source-timed packet header adjustor application (also running on the server) to change the UDP source ports of the video flows at the next RP 168 switch point. After another 100 ms delay, the controller then contacted the DirectFlow agent on the switch to remove old flow-rules.

The source-timed packet header adjustor received incoming 2022-6 video flows from the sources on one NIC of the server, and re-broadcasted them out of another NIC with modified UDP source ports. The adjustor was written in C using the Ntop.org PF_RING high-speed packet processing socket library.

Flow Steering with DirectFlow Agent

Arista's DirectFlow API allows custom flow-rules to be added by agent processes running on a switch using semantics similar to OpenFlow. In this case, DirectFlow-rules were used to match multicast UDP traffic from a specific IP source address and UDP source port and to direct that flow to a specific destination Ethernet port, while also changing the VLAN.

A Python agent script was scheduled to run when the switch is first powered. The script listens on the 1GbE out-of-band management port for UDP command packets. The payload of the UDP packet contains instructions from the orchestration system to the switch. Upon detecting a valid new command packet, the Python script adds or removes the appropriate DirectFlow-rules via the Arista EOS Python API.

Figure 9 shows an example of a flow-rule that directs UDP datagrams with UDP source port 20001 from IP source 10.10.10.52 to be output on port Ethernet 17 with VLAN ID 2.

Results

The proof-of-concept delivered a system capable of synchronous switching of real-time uncompressed HD video sources at the RP 168 switch point using a COTS Ethernet switch. PCAP captures of the video switching points were made at the output ports of the Ethernet switch to show a clean switch from one source to another at the proper point in the video raster. In casual testing among users in our lab and at a broadcast industry trade show, no "bad switches" were detected.

```
Flow DirectFlow_Video_52_to_port_Ethernet17:
   match:
     Ethernet type: IPv4
     source IPv4 address: 10.10.10.52/255.255.255.255
     IPv4 protocol: UDP
     source TCP/UDP port or ICMP type: 20001
   actions:
     set VLAN ID to: 2
     output interfaces: Ethernet17
```

Figure 9 *Example DirectFlow-rule*

In automated "torture testing" of switching as fast as possible for many hours, there were some occasional video glitches in the switching of the order of one out of 10,000 switches. This may be due to occasional de-synchronization of the 2022-6 flows, possibly leading to one too few or one too many datagrams for a frame at the destination. The 2022-6 standard contains no metadata for de-packetizers to know where in the video raster a particular datagram comes from. The only hint regarding a datagram's location in the raster is that the datagram that ends a frame has its RTP marker bit set. Additional metadata in 2022-6 that informs de-packetizers about the position of datagrams in the raster could allow them to better handle lost datagrams or those that duplicate data in a particular section of the raster.

Conclusions

Source-timed SDN video switching is a way to achieve high temporal accuracy of packetized flow changes despite the comparatively slow and temporally inaccurate nature of typical SDN APIs. Also it does not require the "double bandwidth penalty" of destination-timed switching. Philosophically, it shows how there can be a beneficial collaboration between COTS networking devices built for general enterprises that can process billions of packets per second according to slowly changing rules, and the video devices from our specialized industry that must be highly temporally accurate.

Acknowledgments

The author would like to thank Warren Belkin and Andreas Bechtolsheim from Arista Networks for support in implementing the PoC, Ntop.org for several key elements of software in the PoC, Steven Wallace from Indiana University InCNTRE for discussions about SDN and OpenFlow, Al Kovalick for his suggestions regarding packetized video switching theory, and Clyde Smith and Richard Friedel from FOX Network Engineering & Operations for providing the resources and guidance to carry out this work.

References

[1] SMPTE RP 168:2009. Definition of Vertical Interval Switching Pont for Synchronous Video Switching

[2] Open Networking Foundation, 2013. OpenFlow Switch Specification Version 1.3.3, https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.3.3.pdf

[3] SMPTE 2022-6:2012. Transport of High Bit Rate Media Signals over IP Networks (HBRMT)

[4] LAABS M.: 'SDI Over IP – Seamless Signal Switching in SMPTE 2022-6 and a Novel Multicast Routing Concept', *EBU Technical Review*, 2012, Q4

[5] KOUADIO A.: 'SDI over IP interoperability tests'. presented at the EBU Network Technology Seminar 2013, Geneva, CH, June 2013

[6] BILBERRY J., PALMER M., SULLIVAN R.: 'Network Service Security Through Software-Defined Networking', http://institute.lanl.gov/isti/summer-school/cluster-student-projects/network-service-security-through-software-defined-networking

[7] MOGUL J.C., TOURRILHES J., YALAGANDULA P., SHARMA P., CURTIS A.R., BANERJEE S.: 'Devoflow: Cost-Effective Flow Management for High Performance Enterprise Networks'. Proc. 9th ACM SIGCOMM Workshop on Hot Topics in Networks, Hotnets-IX, 2010, pp. 1:1–1:6

Selected content from IBC2014

ISSN 2041-5923

Live event experiences - interactive UHDTV on mobile devices

O.A. Niamut[1] *G.A. Thomas*[2] *E. Thomas*[1]
R. van Brandenburg[1] *L. D'Acunto*[1] *R. Gregory-Clarke*[2]
[1]*TNO, NL*
[2]*BBC R&D, UK*

Abstract: This paper reports on the latest developments in tiled streaming. As an extension of HTTP adaptive streaming, it retains all the benefits of this streaming technology, while adding the potential of interaction when UHDTV is consumed on mobile devices. In particular, we discuss the underlying principles and aspects, such as multi-layer resolution scaling, spatial segmentation and adaptive streaming. Then we present insights from a number of technology validation tests and demonstrations, such as a live dance performance in Manchester in May 2013; a training tool for professional skiers employing tiled streaming in Schladming, host of the Alpine Skiing World Championship 2013; and tests incorporating 'augmented reality'-style overlays in an athletics stadium in preparation for a trial at the 2014 Commonwealth Games. Finally, we report on the status of on-going standardization efforts in the MPEG-DASH ad-hoc group, where tiled streaming is considered as a new feature, referred to as Spatial Relationship Description.

Introduction

The momentum behind UHDTV is quickly gathering and UHDTV displays are entering the market at an increasing pace. And although the production of UHD content is lagging behind, online content providers such as Netflix, Amazon and YouTube have announced their plans for releasing TV series and films in UHD formats such as 4K. With UHD format recommendations and requirements emerging from ITU [1] and DVB (https://www.dvb.org/news/uhdtv-new-evidence-and-new-questions-for-dvb), with the specification of a 4K Blu-ray format (http://www.hollywoodreporter.com/behind-screen/ces-as-ultra-hd-train-669587) and with several live 4K trials over existing broadcast (http://www.broadbandtvnews.com/2014/05/27/4k-smash-for-french-open/) and CDN infrastructures (http://www.iptv-news.com/2014/05/vienna-state-opera-streams-in-4k-with-elemental/), the future for UHD (initially with 4K pixels) looks bright. For mobile devices, the expectations and benefits of UHD formats are less clear. That is, 4K UHD tablets and smartphones featuring limited screen sizes are less-than-ideal candidates for displaying 4K UHD video. And with an 8K UHDTV standard emerging, the discrepancy between native content resolution and screen rendering resolution will widen. While it is possible to enable regular UHDTV experiences on tablets and smartphones, this experience can be enriched by: (i) allowing end-users freely to extract a region-of-interest and navigate around the ultra-high resolution video, and (ii) adding scalable 'augmented reality'-style overlays to the video. Such an approach requires efficient delivery and media-aware network-based processing in order to support mobile terminals and bandwidth limitations in the access networks.

An emerging technology, referred to as 'tiled streaming', enables interaction with streaming video in such a way that end-users can enjoy the full UHD resolution, even if their device is not capable of rendering and displaying the video in its entirety. As an extension of HTTP adaptive streaming, it retains all the benefits of this streaming technology, while adding the potential of interaction when consuming UHDTV on mobile devices. This paper reports on the latest developments in tiled streaming. In particular, we: (i) discuss the underlying principles and aspects, such as multi-layer resolution scaling, spatial segmentation and adaptive streaming, and (ii) present insights from a number of technology validation tests and demonstrations, such as a live dance performance in Manchester; a training tool for professional skiers employing tiled streaming in Schladming (host of the Alpine Skiing World Championship in 2013); and tests incorporating 'augmented reality'-style overlays in an athletics stadium in preparation for a trial at the 2014 Commonwealth Games. We further report on the status of on-going standardization efforts in the MPEG-DASH ad-hoc group, where tiled streaming is considered as a new feature, referred to as Spatial Relationship Description.

Related work

With recent capturing systems for panoramic and omnidirectional UHD video, new types of media experiences are possible where end-users have the ability freely to choose their viewing direction and zooming level. Many different examples of such interactive video delivery have been demonstrated or deployed. In the entertainment sector, companies like Immersive Media (http://immersivemedia.com/) and Mativision (http://www.mativision.com/) offer web streaming and mobile app solutions to cover events with a 360-degree video camera. However, such solutions rely on streaming or downloading a complete spherical panorama to end-user devices, where the final rendering of the interactive viewport takes place. In an alternative approach, KDDI (http://www.talkandroid.com/22426-kddi-brings-zoom-enhance-for-hd-movie-streaming-on-smartphones/) has shown a solution where all rendering takes place on the server side. Here, a low-powered and low-resolution mobile phone sends a spatial request to the server, requiring the server to reframe and rescale the content accordingly before compression and streaming to the end-user device. Tiled streaming has emerged as a scalable and bandwidth-efficient approach to interactive UHD. Initially proposed by Mavlankar [2] and further developed in [3–5], interactive UHD is enabled by a multi-layer tiling approach where the video is split into multiple independently-encoded tiles which are re-stitched in the end-user device.

In addition to interactive video delivery, panoramic and UHD video also provide a good opportunity for adding hotspots and overlays. That is, due to the static nature of the background, additional and interactive overlay graphics can be positioned in the world reference frame rather than having to account for camera motion. Such hotspots were initially incorporated into static panoramic systems, such as Quicktime VR, introduced in 1994. More recently, systems using video panoramas have been developed for use in sports broadcasting, where real-time data relating to player tracking can be overlaid on the scene prior to selecting a window for broadcast (http://www.stats.com/pdfs/SportVU_SonyDAV.pdf). Approaches for adding interactive overlay graphics to conventional web video at the client side are starting to appear (https://popcorn.webmaker.org/), but these are not yet generally being applied to interactive delivery of panoramic and UHD video.

Tiled streaming for zoomable video

Zoomable video allows users to selectively zoom and pan into regions of interest (ROI) within the video. Such interaction typically requires dynamic cropping of ROIs in the source video as well as unicast streaming of the cropped ROIs. The concept of tiled streaming addresses the limitations of today's networks when streaming high resolution content, as well as allowing new applications in video streaming such as interactive panning and zooming. Tiled video provides a better user experience than with predefined or dynamically cropped regions of interest. Moreover, it provides better image quality for the selected region than by simply enlarging pixel dimensions. Finally, the adaptive version of tiled streaming offers a new adaptation dimension after bandwidth, resolution and quality, i.e. the ability at a given bandwidth to choose between full-frame video in low quality and a spatial area in higher quality.

Basic concepts of tiling

A tiled video can be obtained from a single video file or stream by partitioning each individual video frame into independently-encoded videos. Tiles are thus defined as a spatial segmentation of the video content into a regular grid of independent videos that can each be encoded and decoded separately. We denote the tiling scheme by MxN where M is the number of columns and N is the number of rows of a regular grid of tiles. See Figure 1 for two examples of regular tiling grids. The tiling and subsequent separate encoding of tiled videos leads to a reduced compression performance, due to reduced exploitation of spatial correlation in the original video frame being limited to tile boundaries. This compression performance loss can be reduced by using multiple resolution layers. Each additional layer originates from a lower resolution version of the original video frame, tiled into a grid with fewer tiles. If the tiling is small enough (such as thumbnails), the bit-rate overhead of using another resolution layer is affordable. This multi-resolution tiling increases the quality of user-defined zooming factors on tiles. Once a user

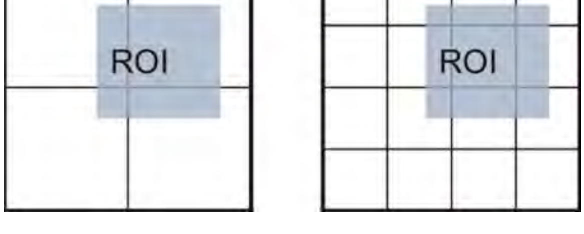


Figure 1 *Example of 2x2 (left) and 4x4 (right) tiling grids. Depending on the tile size, a single ROI overlaps multiple tiles, and thus requires multiple tiles for reconstruction*

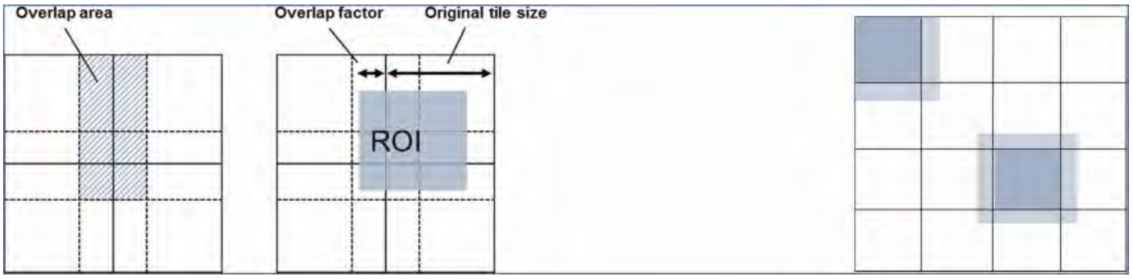


Figure 2 *(left) Tiling scheme with overlapping tiles (2x2 grid) and an overlap factor of one third. (right) Overlapped tiling scheme with an overlap factor of one fourth*

zooms into a region of the content, the system will provide the highest resolution tiles that are included in the requested region.

Optimizing performance on low-powered devices with overlapping tiling

Figure 2 provides an example of an overlapping tiled grid, as used in [6]. Such a tiling is beneficial for mobile devices that are equipped with a single hardware decoder only. In this case, the total number of tiles required to reconstruct the requested ROI can be reduced to one, while the ROI size and total number of tiles remain approximately the same. The effectiveness of overlapping tiles in reducing the number of tiles required to reconstruct a given ROI is determined by the overlapping factor, which gives the relative overlap (per planar direction) of a particular tile in relation to its size. Choosing a larger overlapping factor results in larger overlapping areas, and thus in fewer tiles being required to reconstruct a given ROI. The downside of this overlap is that these redundant pixels result in a larger amount of data to be stored on the server side. Also, overlapping tiles result in heterogeneous tile sizes.

Tiled adaptive streaming

As shown in Figure 3, spatial segmentation can be complemented with the temporal segmentation of HTTP Adaptive Streaming (HAS). The scalability properties of HAS enable zoomable video to be available to a large number of users thanks to efficient bandwidth utilisation, cacheability and simpler inter-tile synchronization. Tiled streaming can be integrated within HAS by having each video tile individually encoded and then temporally segmented according to any of the common HAS solutions (e.g. MPEG DASH [7] or Apple HLS (HTTP Live Streaming, see http://tools.ietf.org/html/draft-pantos-http-live-streaming-13). This leads to a form of tiled adaptive streaming, where all tiles are temporally aligned such that segments from different tiles can be recombined to create the reassembled picture. An advantage of using HAS for the delivery of spatial tiles is that the inherent time-segmentation makes it relatively easy to resynchronise different spatial tiles when recombining tiles into a single picture or frame. As long as the time segmentation process makes sure that time segments between different spatial tiles have exactly the same length, the relative position of a frame within a time segment can be used as a measure for

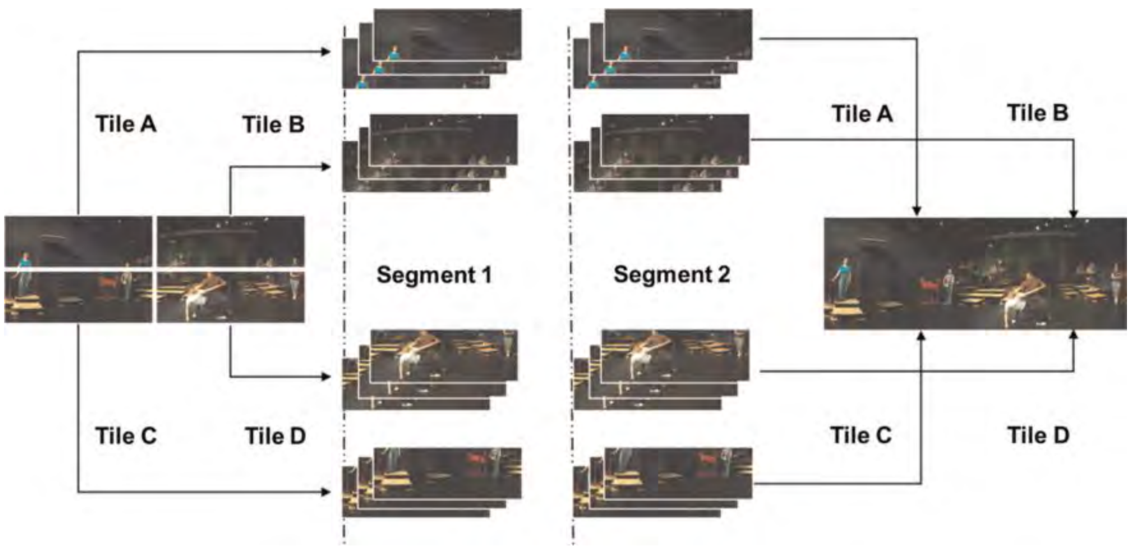


Figure 3 *With tiled HAS, a video is tiled in a certain grid. Each tile is encoded and segmented using HAS segments. In this example, the grid is 2 by 2*

the position of that frame within the overall timeline. For example, frame number n within time segment s of tile A is synchronised with frame number n within time segment s of tile B. On the client side, timestamps provided by the segment container can be used to ensure perfect synchronisation between the segments that make up the final viewport to be rendered on the screen of the end user.

Applications of tiled streaming technology

In this section we present insights from a number of recent and planned technology validation tests and demonstrations, such as: (i) the FascinatE system, used during a live dance performance in Manchester; (ii) the iCaCoT system, used as a training tool for professional skiers in Schladming, February to April 2014; and (iii) preparations for tests including 'augmented reality'-style overlays in an athletics stadium for the 2014 Commonwealth Games.

Fascinate system

Tiled streaming was employed for the distribution of panoramic video sequences in the context of the European-funded project FascinatE (Format-Agnostic SCript-based INterAcTive Experience, see http://www.fascinate-project.eu/). A live system was demonstrated during a dance performance in Manchester in May 2013, see Figure 4 and 5. Here, a multi-camera audio-visual scene representation, including both panoramic 6K and regular 1080p video content, was spatially tiled and temporally segmented. For tiling, a dyadic tiling approach was used, ranging from 1x1 to 8x8 tiling grids, resulting in a multi-resolution set of panoramic video layers, with the original resolution of 6976 by 1920 pixels as a base layer.

iCaCoT System

Tiled streaming was further incorporated into the iCaCoT (Interactive CAmera-based COaching and Training, see http://www.experimedia.eu/2014/02/20/icacot/) training application, in the context of the European-funded project EXPERIMEDIA (Experiments in live social and networked media experiences, see http://www.experimedia.eu/). Here, the goal was to provide ski coaches in Schladming with a tablet application through which they could provide their trainees with real-time feedback. This was achieved using a combination of a set of static high-resolution cameras along the ski-slope, tiled streaming and advanced trick-play and drawing features, see Figure 6. In collaboration with Schladming2030, the Austrian venue partner, we performed several experiments with a live 4K-tiled streaming system. A significant challenge was to cope with the particular conditions of the experiment

Figure 4 *Panoramic 6K image of the FascinatE live demonstration, based around the performance of 'Deeper than all roses', a composition from Stephen Davismoon, featuring rock band Bears?Bears! and live performance artists Joseph Lau and Shona Roberts*

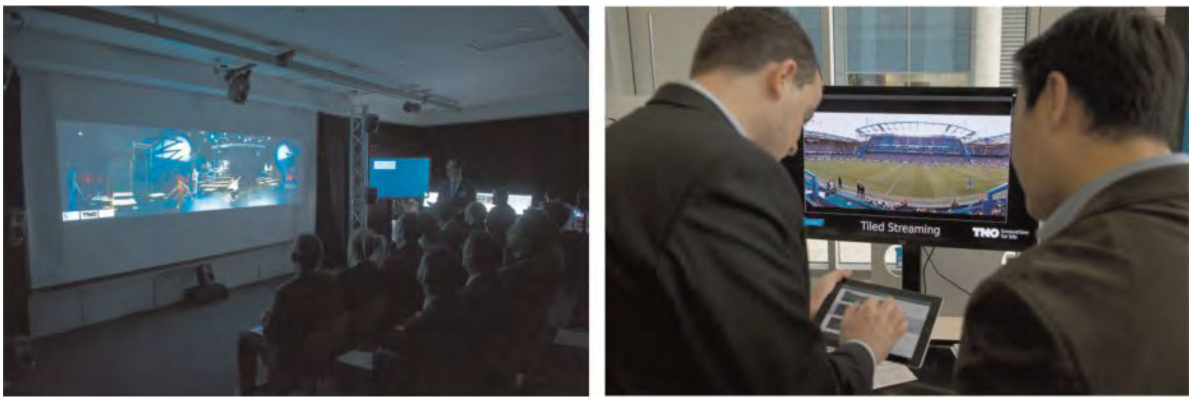


Figure 5 *The live performance was shown in a separate room, delivered via tiled streaming. The presenter could navigate through the video panorama using a tablet. Additional footage from earlier recordings was also shown on tablets*

Figure 6 *Interactive camera-based coaching and training using tiled streaming for video navigation. Several application screenshots, with the third screenshot showing user-drawn markers helping to provide the skier useful feedback*

location, as the material had to be protected from extreme weather conditions in terms of temperature, humidity, and so on. As the on-site connectivity to the open Internet ruled out any off-site video processing such as tiling and encoding, the overall system had to be stand-alone and compact.

For tiling of 4K video at 24 fps in real-time, a software-based approach was not a viable solution. Therefore, a hardware-accelerated pipeline based on the Intel Media SDK (Intel Media Software Development Kit, see http://software.intel.com/en-us/vcsource/tools/media-sdk-clients) was used. This pipeline, running on a single machine, is able to decode the original panorama, to produce the tiles and to independently encode them in H.264 at 24 fps. In practice, the pipeline can handle in parallel more than 10 tiles at roughly 1920x1080 resolution.

Commonwealth games 2014 system

These experiments aim to verify that navigation around a high resolution video using tiled streaming will contribute to a higher sense of interaction and engagement amongst users, particularly in the case of large-scale, event-based programming. This form of content often contains several distinct regions of interest, and a number of different things that a user may want to look at in more detail. Hence, we hypothesise that the user experience could be further enhanced through the use of overlaid, interactive graphics which provide extra information about the scene.

An example of such an application is an athletics event, which typically features several different track and field events occurring simultaneously. As well as being able to pan and zoom around the scene as they wish, the user can also be presented with data which offers more detail about what they are looking at. This could include the locations and times of the sports taking place on the day in question, the names of the athletes that are visible, the current height of the high jump bar, and so forth. The intention is to provide the type of rich data that a user would typically be interested in anyway, but that they would ordinarily have found either from burnt-in graphics provided by a broadcaster, or else a self-initiated search. Presenting this information as optional overlays in this way allows the user to access the level of detail they want, when they want it, without having to leave the application. We plan to test interactive overlays on panoramic video as part of a closed trial of a prototype system, known as the Venue Explorer, at the Commonwealth Games in Glasgow in July-August 2014. The plan is to capture a wide-angle view of the athletics stadium using a 4k camera, encode this as a set of overlapping tiles, and stream these to an HTML5-based client using MPEG-DASH. Data relating to sports events (including live updates) will be used to render interactive overlays in the client, according to options selected by the viewer. Initial work on the system used video captured at the London Anniversary Games in summer 2013 and is shown in Figure 7.

Standardisation of tiled streaming

MPEG-DASH (ISO/IEC 23009) is the adaptive streaming technology standardised by MPEG. After having published a first edition of the standard [7], MPEG experts are now

Figure 7 *A possible user interface for the Venue Explorer system*

aiming to extend the original scope of adaptive streaming to new 'use-cases'. Tiled streaming use-cases and their relevancy were presented to the MPEG-DASH working group during the 104th MPEG meeting in April 2013. There, MPEG-DASH experts acknowledged the usefulness of such use-cases and agreed to start a so-called Core Experiment. The goal of this Core Experiment was to steer the group effort towards a technical solution which would permit these tiled streaming use-cases. The discussions within the Core Experiment reached a consensus among the group at the 107th MPEG meeting last January. Consequently, MPEG has initiated the publication process of this new feature which should be completed during 2015.

Conceptually this new feature, called Spatial Relationship Description (SRD), allows an author of the MPEG-DASH media presentation description (MPD) to describe how the various tiles are spatially related with each other. This description handles both intra and inter layer relationships. Thus far, the standard only allows for an *AdaptationSet* to define perceptually-equivalent content. Therefore, describing different tiles under the same *AdaptationSet* would violate this rule. With the SRD feature, the concept of a tile, as defined in this paper, is mapped onto the *AdaptationSet* element of the MPD. It is also important to note that this new mapping decouples the tile concept from a particular video. To offer backwards compatibility, it remains possible to benefit from the regular properties of *AdaptationSets*, namely the availability of several *Representations* in different bit-rates, codecs, resolutions, and so on. In practice, the new MPEG-DASH SRD feature will specify a set of parameters in order to describe tiles with respect to a common reference space. These parameters are: (x,y), respectively, horizontal and vertical positions of the tile; (w,h), respectively, width and height of the tile; and (W,H), respectively, width and height of the reference space. All these values are expressed in an arbitrary unit as chosen by the MPD author.

Future work

In subsequent developments, we aim to improve the live tiling process, such that it can handle a range of UHD formats, including higher resolutions and frame rates. Incorporation of multi-sensor systems will also be considered. In the near future, we aim to produce live 4K footage and to deliver this via tiled streaming to end-users in a large-scale user trial. This allows us to determine the impact of live 4K-tiled streaming on a regular production environment and to measure the effects of tiled streaming on bandwidth and latency in a regular content delivery setting. Furthermore, the ways in which users interact with panoramic UHD video and overlays will be studied in a series of user trials, which will inform future developments. It is also planned to evaluate the use of interactive panoramic video with overlays in other application scenarios.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreements no. 248138 and no. 287966.

References

[1] ITU-R Recommendation BT.2020, "Parameter values for ultra-high definition television systems for production and international programme exchange". August 2012

[2] MAVLANKAR A., AGRAWAL P., PANG D., HALAWA S., CHEUNG N-M., GIROD B.: 'An Interactive Region-Of-Interest Video Streaming System For Online Lecture Viewing'. Special Session on Advanced Interactive Multimedia Streaming, Proc. of 18th International Packet Video Workshop (PV), Hong Kong, December 2010

[3] KHIEM N., RAVINDRA G., CARLIER A., OOI W.T.: 'Supporting Zoomable Video Streams via Dynamic Region-of-Interest Cropping", in'. Proc. of 1st ACM Multimedia Systems Conf., Scottsdale, Arizona, 22–23 February 2010, pp. 259–270

[4] BRANDENBURG R., VAN NIAMUT O., PRINS M., STOKKING H.: 'Spatial Segmentation For Immersive Media Delivery'. In Proc. of 15th Int. Conf. on Intelligence in Next Generation Networks (ICIN), Berlin, Germany, 4–7 October 2011, pp. 151–156

[5] QUAX P., ISSARIS P., VANMONTFORT W., LAMOTTE W.: 'Evaluation of Distribution of Panoramic Video Sequences in the eXplorative Television Project'. NOSSDAV'12, Toronto, Canada, 7–8 June 2012

[6] PANG D., HALAWA S., CHEUNG N-M., GIROD B.: 'ClassX Mobile: Region-of-Interest Video Streaming to Mobile Devices with Multi-Touch Interaction'. In Proc. of MM'11, Scottsdale, Arizona, USA, November 28–December 1 2011

[7] MPEG, "ISO/IEC DIS 23009-1 Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats", January 5, 2012.

Selected content from IBC2014

ISSN 2041-5923

High frame-rate television: sampling theory, the human visual system and why the Nyquist-Shannon theorem doesn't apply

K.C. Noland

BBC Research and Development, UK

Abstract: In this paper an analysis of video motion portrayal in terms of classical sampling theory is presented, that uses measurements of the human visual system to determine the highest temporal frequency that the frame-rate should be able to support. A discussion of the limitations of a traditional sampling theory approach is also presented. Additional analysis is given of the effects of camera shuttering and display processing, which are essential parts of any real system. This is the first time that such analysis has been carried out in the context of ultra-high definition television.

Introduction

Ultra-high definition (UHD) television standards are in the process of being formalised, and there is currently much discussion about suitable parameter choices. Higher frame-rates have received particular attention, with subjective tests carried out by the European Broadcasting Union [1] having demonstrated that simply increasing spatial resolution produces only a small quality improvement, and that using a higher frame-rate can significantly improve the perceived quality.

Digital video capture is a three-dimensional sampling process, to which it is possible (although not necessarily desirable) to apply traditional sampling theory principles. The frame-rate relates to the time dimension. After a review of video motion artefacts and related work, this article presents an explanation of how sampling theory can be used to find a suitable frame-rate, assuming that no strobing can be tolerated. This requires a limit for the highest temporal frequency in the signal. A model of the human spatio-temporal contrast sensitivity function (CSF) is used to find the highest temporal frequency needed.

Then follows a discussion on the effects of eye movements, which mean that standard sampling theory is insufficient to model motion artefacts. Measurements of eye tracking speeds are used to model increased blur perception during tracking, and then applied to the frame-rate calculations. Another consequence of eye tracking is that the aliasing artefacts that can cause strobing are masked. The perception of what would traditionally be called aliasing is still poorly understood, and requires extensive further research. One way to balance blur and strobing is to control the camera shuttering and display hold times, and this article also presents a theoretical analysis of these effects.

Motion artefacts in video

There are three important artefacts that are affected by the frame-rate: large area flicker, which is the result of the whole screen being refreshed at the frame-rate, visible on displays that have some blanking between frames; motion blur, caused by objects moving while the camera shutter is open; and strobing, caused by having too few frames per second for the motion to appear smooth. Motion blur and strobing can be traded against each other by varying the time that the camera shutter remains open for each frame capture, but the only way of improving them both together is to increase the frame-rate.

The perception of motion is also affected by eye movements. When the eye follows a moving object during eye tracking, the relative speed between the eye and the object is reduced, and hence the image moves more slowly across the retina than when the eye is still. This increases sensitivity to motion blur but decreases sensitivity to strobing. For displays with a long on-time additional motion blur is created in the eye as it moves while the displayed image remains still for a frame period. This is sometimes called "retinal slip".

Previous research

The ITU-R Report BT.2246 [2], *The present state of ultra high definition television*, contains an overview of subjective experiments carried out by Japanese Broadcaster NHK on frame-rates for television. The results show that around 80 frames per second (fps) were needed to eliminate flicker with 100 degrees field of view, much lower than the frame-rates required to reduce motion blur and strobing to acceptable levels. For motion blur it was found that the camera shutter length should be less than about 3 ms. Further experiments showed great improvements in the perceived strobing at 120 fps over 60 fps, with additional perceptible improvements at 240 fps. Strobing was also shown to decrease with a longer camera shutter. Overall subjective ratings confirm that 120 fps offers a big improvement in subjective quality over 60 fps, and 240 fps is better still.

Kuroki et al. [3] show that in 24 frames per second progressive and 30 frames per second interlaced formats visual acuity is much reduced in comparison to direct viewing as soon as any movement occurs. This is clear evidence that the frame-rates tested are insufficient for a transparent system. Subjects also rated blur and jerkiness for material at frame-rates from 62 to 500 fps, and a dramatic reduction in both artefacts was reported at 125 fps over 62 fps, then flattening off above 250 fps. Jerkiness was also rated worse with short shutter times. These experiments independently verify the results presented in ITU-R BT.2246 [2].

In much earlier experiments with high definition (HD) television, Stone [4] demonstrated significant perceived improvements when increasing the frame-rate from 50 to 80 fps. He reports that although reducing the camera shutter time to 50% also improved the quality of motion, this approach caused severe problems when converting to other frame-rates due to the strong temporal aliases in the signal. More recently, Armstrong et al. [5] give descriptions of the arguments for higher frame-rates, and report observations of significant quality improvements in material displayed at 100 fps over conventional rates.

Tonge [6] and Watson [7] both took a more analytical approach. Tonge suggested that a frame-rate of up to 2000 fps would be needed for television pictures to appear exactly equivalent to continuous motion, based on measured limits of dynamic resolution. He also noted that with eye tracking the rate can be reduced to the critical flicker frequency. Watson analysed visibility of artefacts according to whether they appear in the "window of visibility", which he defined as a threshold on the CSF measured by Robson [8].

Applying sampling theory principles to video frame-rate

Classical sampling theory teaches that the sampling frequency must be at least twice the highest frequency in the signal. This is a fundamental principle of digital signal processing, and is usually attributed to Nyquist [9] and Shannon [10]. When applied to video capture, this translates to a frame-rate of at least double the highest temporal frequency—the rate at which an individual pixel changes its brightness, most often as a result of an object moving past that pixel position. The rate at which the brightness changes depends on the amount of spatial detail in the object and its speed. This relationship can be formalised as:

$$t = r\nu \qquad (1)$$

where t is the temporal frequency, r is the spatial frequency and ν the velocity [6]. A flashing object can also create a temporal frequency, in which case the flash frequency should also be incorporated into equation 1, but this occurs relatively rarely in real scenes.

The frame-rate must therefore be at least $2t$, and can be calculated for any r and ν. The pixel format enforces an upper limit on the spatial detail that can occur, but it is much more difficult to determine the highest velocity that we want to represent. To find a suitable upper limit on velocity, the next section proposes using measurements of the human visual system to perceptually match the spatial and temporal resolution of a format. First it is assumed that the viewer has a fixed gaze, then a model of eye tracking is incorporated.

Human contrast sensitivity function

The human CSF models sensitivity to different spatial frequencies, moving past at different velocities. Our ability to resolve spatial detail drops both with higher spatial frequency and with higher velocity. Laird et al. [11] fit a model of the CSF originally proposed by Kelly [12] to their own experimental data, as shown in figure 1. Subjects maintained a fixed gaze during the experiments, so the data does not take eye tracking into account.

Using the contrast sensitivity function to find a limit on velocity

Since it is not possible to put an upper limit on the velocities that may occur in a scene, it is proposed instead to restrict the representable velocities to those that would cause loss of spatial detail due to motion blur in the signal to be no worse than loss of spatial detail caused by limiting the spatial resolution. Our ability to resolve the highest spatial

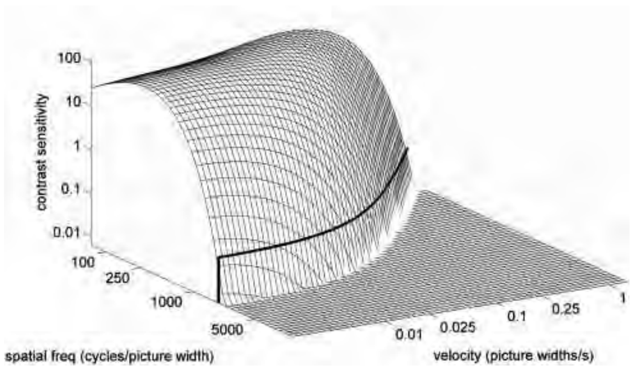


Figure 1 *Contrast sensitivity function model, for a viewing distance of 1.5 H and a fixed gaze at the centre of the screen. The thick line shows the threshold of spatio-temporal frequencies to be represented to perceptually match the spatial resolution of UHD-1*

frequency of a particular format is represented by the height of the CSF at that spatial frequency and at a velocity of zero, indicated by the thick vertical line in figure 1, which is at 1920 cycles per picture width. In order to be able to represent all moving objects that we can resolve just as well as a static object of the highest spatial frequency, it is necessary to follow one contour line in the CSF, as illustrated by the thick curved line in figure 1. Each point along the line has an associated spatial frequency and velocity, from which a frame-rate can be calculated using equation 1. The highest of these frame-rates is proposed as the overall required rate.

The CSF data is specified in units of cycles per degree subtended at the retina, and degrees per second at the retina. To use the more convenient units of cycles per picture width and picture widths per second it is necessary to take account of the viewing distance. Full details of the conversion are provided by Noland [13]. Figure 1 is plotted using the new units, for a distance of 1.5 times the screen height (1.5 H).

Figure 2 shows the calculated frame-rates for three viewing distances. With 720 horizontal pixels viewed at 3H the required frame-rate is a little below 50 fps. This matches the standard definition format in the UK, which has an interlaced field-rate of 50 fields per second. The calculated frame-rate then increases approximately linearly with the number of horizontal pixels up to around 140 fps for 1920 horizontal pixels. Beyond this point there are no more contrast sensitivity measurements, so it is not valid to extrapolate the lines shown. At viewing distances of 1.5 H and 0.75 H, that are recommended for UHD with 3840 and 7680 horizontal pixels respectively, the results suggest that without eye tracking 140 fps would also be suitable. This is because the recommended viewing distances are all designed to put the pixel structure at the limit of human spatial acuity, so the same frame-rate is needed to match the spatial resolution. Viewing from further away brings no benefit in spatial resolution beyond the limit of acuity, with

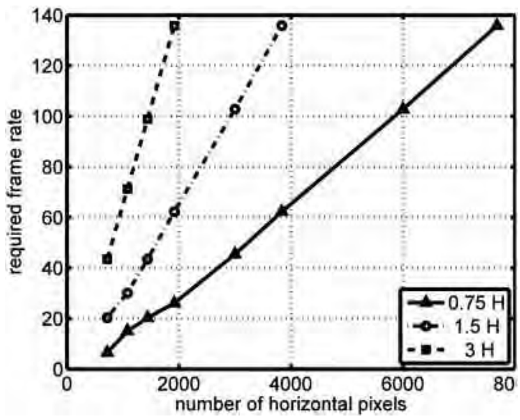


Figure 2 *Frame-rate required to perceptually match various common horizontal resolution values, with viewing distances of 0.75 H, 1.5 H and 3 H and a fixed gaze, assuming no strobing can be tolerated*

ideal Nyquist-Shannon reconstruction at the display. Further discussion of this is given in [13].

Eye tracking model

The CSF model is based on data collected with a fixed gaze, but under normal viewing conditions the eye is free to follow moving objects, which has the effect of improving our ability to resolve spatial detail. This implies that objects can move much faster before we are unable to resolve their spatial detail than when the eye is fixed, and hence a higher frame-rate would be required to keep motion blur to an acceptable level.

A model of human eye tracking, based on experimental data, was incorporated into the CSF model. Daly [14, pp. 185–187] describes how on average humans track motion at about 82% of the object speed, and finds a maximum tracking speed of about 80 degrees per second. The model is shown in figure 3.

To adapt the frame-rate calculations for eye tracking, all velocities in the CSF were scaled up by $1/(1-0.82) \approx 5.6$, to give the speed of the tracked object

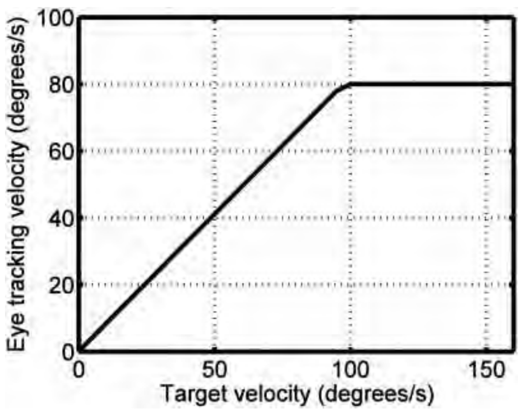


Figure 3 *Eye tracking model*

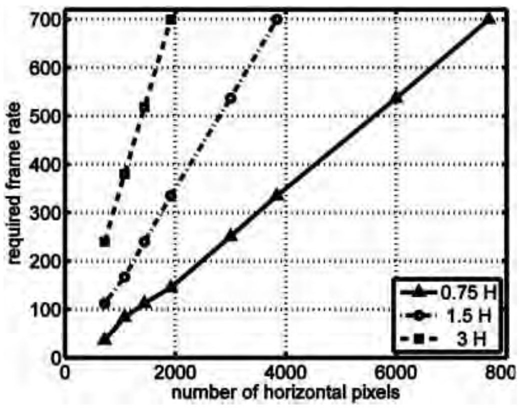


Figure 4 *Frame-rate required to perceptually match various common horizontal resolution values, with a viewing distance of 0.75H, 1.5H and 3H, during eye tracking, assuming no strobing can be tolerated*

that would create an image moving at the original, unscaled speed on the retina during tracking. An upper limit of 80 degrees per second was also imposed on the tracking velocity. The corresponding maximum trackable object velocity is (80/0.82) = 97 degrees per second.

Figure 4 shows the calculated frame-rates according to the contrast sensitivity model with incorporated eye tracking. The figures demonstrate just how important an effect eye tracking is in the discussion of motion portrayal and frame-rates. Frame-rates of up to 700 fps are now coming into consideration.

Why the Nyquist-Shannon theorem doesn't apply

Before dwelling on the clearly impracticable figure of 700 fps, it is important to consider the assumptions made to reach it. The Nyquist-Shannon theorem assumes that no aliasing can be tolerated. However, this is not necessarily the case. In fact, television systems have always been temporally undersampled [15]. The reason this does not look more objectionable is that the moving eye is an important part of video signal reconstruction, and is not taken into account by traditional sampling theory, which requires no aliasing only from the

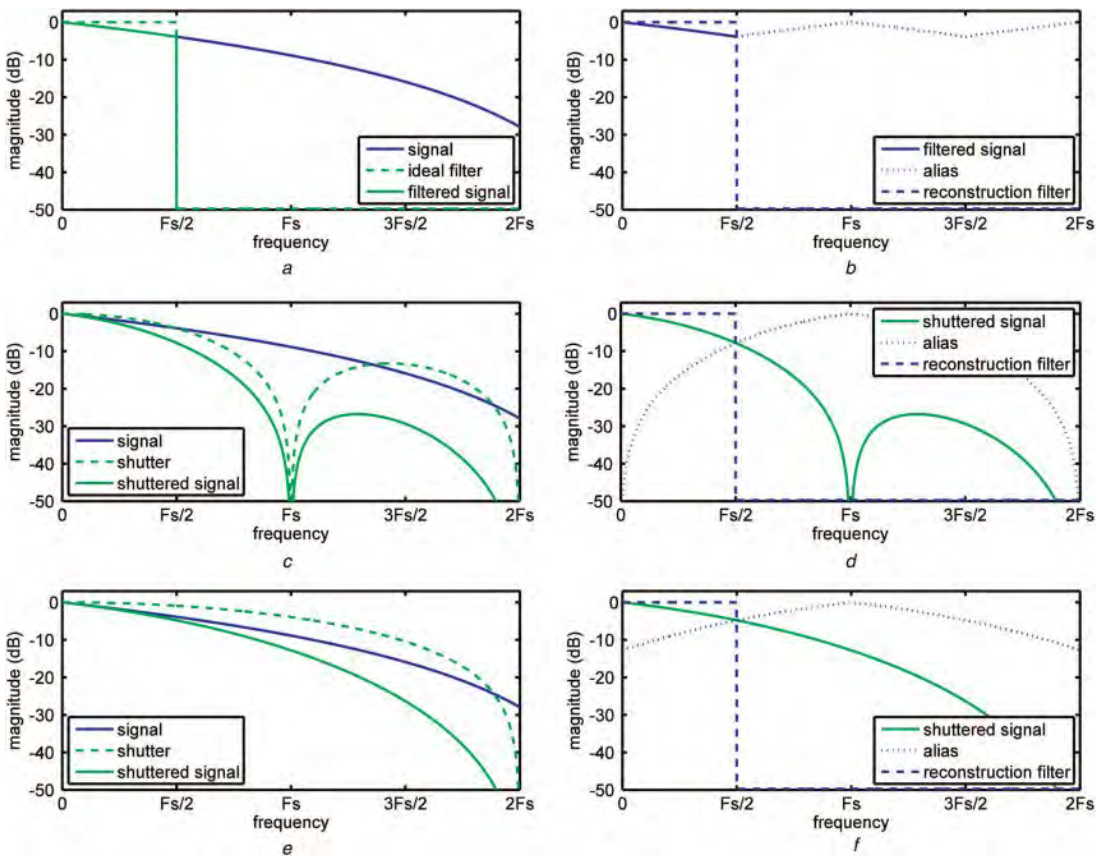


Figure 5 *Illustration of camera shutter effects*
a Nyquist-Shannon sampling
b Nyquist-Shannon sampling with alias and ideal reconstruction filter
c 100% camera shutter
d 100% camera shutter with alias and ideal reconstruction filter
e 50% camera shutter
f 50% camera shutter with alias and ideal reconstruction filter

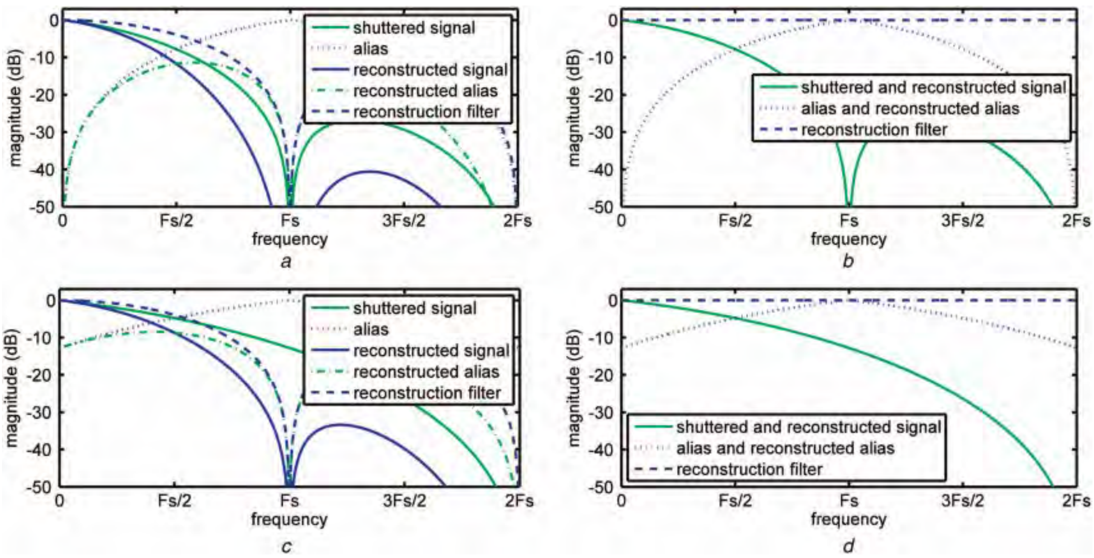


Figure 6 *Illustration of display effects*
a 100% camera shutter followed by sample-and-hold display
b 100% camera shutter followed by impulsive display
c 50% camera shutter followed by sample-and-hold display
d 50% camera shutter followed by impulsive display

perspective of a fixed position on the screen. Eye tracking can be interpreted as a kind of motion-compensated interpolation, where parts of the signal that would traditionally be called aliases from the perspective of a fixed position on the screen are correctly unwrapped by the eye to be perceived at the right speed. This means that, for tracked motion to look smooth, it may only be necessary for the frame-rate to be above the critical flicker frequency.

However, the eye can only track one velocity at a time, and any untracked motion will still be subject to strobing effects. Little is currently known about how the visibility of strobing artefacts varies with different kinds of motion, frame-rates, camera shutter times and display apertures, but that knowledge is essential to make a complete judgement about suitable frame-rates. If signal aliases are to be eliminated, the calculations suggest that a frame-rate of 700 fps is needed to maintain resolution in tracked motion. However, if it is found that some strobing can be tolerated, and it is assumed that aliases are the sole cause of strobing artefacts, the strict criterion of having no aliasing in the signal could be relaxed, and hence the required frame-rate could be reduced. Better understanding of strobing perception would also allow an informed judgement to be made regarding the best balance between blur and strobing for a more realistic frame-rate such as 100 fps.

The effects of camera Shuttering and display sample-and-hold

One way to balance blur and strobing effects is to vary the amount of time for which the camera shutter is open. The longer the shutter time, the more blurred the image. Conversely, the shorter the shutter time, the further an object will have moved between one frame and the next, and so the more strobing will be present. Camera shuttering can be interpreted as a filtering operation: the light entering the camera is integrated for the time that the shutter is open, which is equivalent to convolution with a rectangle function prior to sampling. Viewed in the frequency domain, this corresponds to a multiplication of the signal spectrum with a sinc function, as illustrated in figure 5.

Figure 5a shows the unrealistic case that was assumed by the Nyquist-Shannon analysis, where a so-called "brick wall" anti-alias filter passes all content below half the sampling frequency, and removes everything at higher frequencies. Figure 5b shows the alias in the sampled signal, which does not overlap at all with the baseband. When instead a 100% camera shutter is used as the anti-alias filter, as in figure 5c, the null of the frequency-domain sinc function appears at the sampling frequency. Upper passband frequencies are slightly attenuated, and stopband frequencies, whilst attenuated, are not completely removed. However, due to the position of the null, the alias (figure 5d) has a very low level around zero frequency, and hence strobing is not severe. With a 50% shutter (figure 5e), the frequency-domain sinc function has its null at twice the sampling frequency. This causes much less signal attenuation than a 100% shutter, and hence much less blur, but also means that the notch in the alias (figure 5f) is no longer conveniently positioned at D.C. and so strobing becomes highly visible in untracked motion.

The reconstruction filter has been shown as an idealised "brick wall" filter in figure 5, but reconstruction filtering is in reality provided by the display in conjunction with the human visual system. It is therefore instructive to additionally consider the effects of different kinds of temporal display response on the signal.

A conventional liquid crystal display (LCD) will have close to a 100% on-time. Viewed as a filtering operation, this multiplies the sampled spectrum by a sinc function, as illustrated in figure 6 for a 100% camera shutter (figure 6a) and a 50% camera shutter (figure 6c). In both cases the display sample-and-hold causes additional attenuation, resulting in more motion blur. Aliases in the signal due to the capture process are also attenuated, although not completely removed. Strobing visibility is further masked in sample-and-hold displays when the eye is moving, by the addition of motion blur as a result of "retinal slip".

If instead an impulsive display response is used, where the display is on for a very short proportion of the frame period, the display essentially leaves the signal spectrum unchanged. Motion blur will not be made worse by the display in this case, but it cannot be removed. For a long camera shutter (figure 6b), blur will still be present in the signal, and for a short camera shutter (figure 6d) the strobing will be highly visible in non-tracked motion. If blur and strobing are well-balanced during capture, it would be preferable to use a display with a short hold time, since this would alter the signal least. However, at 50 or 60 fps flicker is still visible on large impulsive displays, so additional processing such as motion-compensated frame insertion is needed, which adds its own motion artefacts. For a frame-rate above the critical flicker frequency impulsive displays are an attractive option.

Conclusion

An analysis of video sampling in terms of classical Nyquist-Shannon theory has been presented, and used together with models of the human contrast sensitivity function and of eye tracking to suggest a frame-rate that perceptually matches a given spatial resolution, in the sense that spatial detail is not lost disproportionately when objects move. The frame-rates proposed should be regarded as approximate, due to their reliance on experimental averages. More significantly, the analysis does not take into account any tolerance we may have to strobing effects. It is likely that lower frame-rates will be acceptable if a degree of aliasing in the signal can be permitted, but to confirm this further research into the visibility of strobing artefacts is needed.

One method of controlling the balance between blur and strobing is to vary the camera shutter and display aperture times. An analysis of these effects illustrated that there will always be some attenuation of wanted signal with a finite aperture time, so some temporal oversampling may be desirable. A shorter camera shutter time can reduce the attenuation, but strongly boosts the level of any aliases, which cause strobing especially in non-tracked motion. A better understanding of the severity of strobing artefacts is also essential to a decision on suitable camera and display processing for a given frame-rate.

References

[1] EBU. 2013. Subjective testing confirms importance of frame-rate for UHDTV. https://tech.ebu.ch/news/subjective-testing-confirms-importance-o-07aug13

[2] ITU-R. 2012. The present state of ultra high definition television. ITU-R BT.2246–1

[3] KUROKI Y., NISHI T., KOBAYASHI S., OYAIZU H., YOSHIMURA S.: 'A psychophysical study of improvements in motion-image quality by using high frame-rates', *J. Soc. Information Display*, 2007, **15**, (1), pp. 61–68

[4] STONE M.A.: 'A variable-standards camera and its use to investigate the parameters of an HDTV studio standard'. BBC Research Dept. Report RD 1986/14 Report, 1986

[5] ARMSTRONG M., FLYNN D., HAMMOND M., JOLLY S., SALMON R.: 'High frame-rate television'. BBC Research White Paper WHP, 2008, p. 169

[6] TONGE G.J.: 'Time-sampled motion portrayal'. Proc. 2nd Intl. Conf. Image Processing and its Applications, 1986

[7] WATSON A.B.: 'High frame-rates and human vision: A view through the window of visibility', *SMPTE Motion Imaging J*, 2013, **122**, (2), pp. 18–32

[8] ROBSON J.G.: 'Spatial and temporal contrast-sensitivity functions of the visual system', *J. Opt. Soc. Am.*, 1966, **56**, (8), pp. 1141–1142

[9] NYQUIST H.: 'Certain topics in telegraph transmission theory', *Trans. American Institute of Electrical Engineers*, 1928, **47**, (2), pp. 617–644

[10] SHANNON C.E.: 'A Mathematical Theory of Communication', *Bell System Technical J.*, 1948, **27**, (3,4), pp. 379 to 423, 623–656

[11] LAIRD J., ROSEN M., PELZ J., MONTAG E., DALY S.: 'Spatio-velocity CSF as a function of retinal velocity using unstabilized stimuli', *Proc. SPIE-IS&T Electronic Imaging*, 2006, **6057**

[12] KELLY D.H.: 'Motion and vision. II. Stabilized spatio-temporal threshold surface', *J. Opt. Soc. Am*, 1979, **69**, (10)

[13] NOLAND K.C.: 'The Application of Sampling Theory to Television Frame-rate Requirements'. BBC R&D White Paper WHP, 2014, Vol. 282

[14] DALY S.: 'Engineering observations from spatiovelocity and spatiotemporal visual models' in Christian J. van den Branden Lambrecht(editor), *Vision Models and Applications to Image and Video Processing*, 2001, Kluwer Academic Publishers

[15] DREWERY J.O.: 'TV-the grand illusion', *IEE Colloq. Motion Reproduction in Television*, 1995 pp. 1/1–1/7

Selected content from IBC2014

ISSN 2041-5923

Development of 8K-UHDTV system with wide-colour gamut and 120-Hz frame frequency

T. Soeno T. Yasue K. Kitamura K. Masaoka Y. Kusakabe H. Shimamoto T. Yamashita Y. Nishida M. Sugawara

NHK (Japan Broadcasting Corporation), Japan

Abstract: NHK has been researching Super Hi-Vision (SHV) as a next-generation broadcasting system. SHV aims at an ultra-high definition television (UHDTV) system with a 7,680 $\times$ 4,320 resolution (8K resolution), 120-Hz frame frequency, and wide-colour gamut, which we refer as 'full-specification SHV', although its R&D was initiated with limited parameter values owing to the available technologies.

In this paper, we discuss the equipment developed for a full-specification SHV system. Utilising CMOS image sensors for capturing 8K images at 120 Hz and a prism designed for the wide gamut, we produced the world's first full-specification SHV camera system. A new signal interface for studio equipment was developed to transmit the full-specification SHV signal with a data-rate of approximately 144 Gbit/s using a single optical cable. In addition, we developed an 8K/120Hz liquid crystal display and 8K/120Hz wide-colour gamut laser projector. We used these pieces of equipment to produce full-specification SHV video content, and confirmed the improvements in the image quality of 8K resolution, 120-Hz frame frequency, and wide-gamut colorimetry.

Introduction

We have been researching and developing an extremely high-resolution video system called Super Hi-Vision (SHV) which is the ultimate two-dimensional television. The video parameters of UHDTV systems have been standardised in Recommendation ITU-R BT.2020 [1], SMPTE ST 2036-1 [2], and ARIB STD-B56 [3]. Among these parameters, SHV aims at a system with a 7,680 $\times$ 4,320 resolution (8K resolution), 120-Hz frame frequency, 12-bit coding format, and wide-colour gamut. We refer to such a system as 'full-specification SHV' [4] although its R&D was initiated with limited parameter values owing to the available technologies. The realisation of full-specification SHV requires the evaluation of moving images and the development of equipment such as a camera, an interface, and a display. This paper reports on the development of such a system.

First, we describe the world's first full-specification SHV camera system. For constructing this camera, we developed a CMOS image sensor (which is the key device of a camera system) for capturing 8K images at 120 Hz and a prism designed for the wide-colour gamut. Next, we describe a new interface, which is based on multilink 10-Gbit/s streams, which can transmit a full-specification SHV signal with a data-rate of approximately 144 Gbit/s using a single optical cable. Then, we explain monitoring equipment for the UHDTV signals. We developed an 8K wide-colour gamut laser projector, a colour-gamut monitor, and a liquid crystal display. We produced SHV content at full-specification using the aforementioned equipment and confirmed the improvements in the image quality of 8K resolution, 120-Hz frame frequency, and wide-colour gamut colorimetry.

Camera system

Camera design

With the aim of producing a full-specification SHV camera, we set the following requirements.

- The camera shall be of a size which promotes mobility and operability for indoor and outdoor shooting.

Table 1 Image sensor specifications.

Image size	25 mm (H) × 14 mm (V)
Optical image format	1.7 in.
Active pixels	7,680 (H) × 4,320 (V)
Pixel size	3.2 mm (H) × 3.2 mm (V)
Frame frequency	120 Hz

- The camera shall capture wide-colour gamut images.

- The camera shall capture 8K resolution (7,680 × 4,320 pixels) images with a frame frequency of 120 Hz.

- The camera system shall maintain the same sensitivity and signal-to-noise ratio (SNR) as conventional SHV camera system [5].

- The system shall perform signal processing in real time.

First, we developed a CMOS image sensor which could capture 8K resolution images at a frame frequency of 120 Hz to achieve the above requirements. Table 1 lists the specifications of the developed image sensor.

Next, we designed the colorimetry in the camera. The colour gamut of the SHV, as specified in Rec. ITU-R BT.2020 is wider than that of HDTV, as specified in Rec. ITU-R BT.709 [6]. The values of the primaries and white points specified in Rec. BT.2020 and Rec. BT.709 are listed in Table 2. Figure 1(a) shows the ideal camera spectral sensitivity curves for rec. BT.2020 primaries. The curves are not physically realisable because of their negative sensitivities. Therefore, we designed practical spectral sensitivity curves composed of only the positive main lobes of the ideal curves by eliminating the second positive lobe and negative regions [7], as shown in Figure 1(b). The camera system applies a 3 × 3 linear matrix (LM) to approximate the effect of the negative regions.

The block diagram of the camera system is shown in Figure 2. The camera head consists of three CMOS image sensors on headboards for each colour (red, green, and blue), the wide-gamut prism described above, and an optical transmission unit. Dense wavelength division multiplexing (DWDM) is used for signal transmission between the camera head and the camera control unit (CCU) because we would prefer to use the same cables for conventional HDTV cameras to transmit the SHV signal The image data are transformed into sixteen 10-Gbit/s streams and then output to the CCU. The CCU performs signal processing including fixed-pattern-noise (FPN) cancellation, colour correction by using a linear matrix, signal level control, gamma correction, and detail enhancement in real time. After that, the output signal is

Table 2 Comparison colour primaries and white point between Rec. ITU-R BT. 2020 and Rec. ITU-R BT.709.

	Rec. BT.2020		Rec.BT.709	
	CIE x	CIE y	CIE x	CIE y
Red primary (R)	0.708	0.292	0.640	0.330
Green primary (G)	0.170	0.797	0.300	0.600
Blue primary (B)	0.131	0.046	0.150	0.060
Reference white (D65)	0.3127	0.3290	0.3127	0.3290

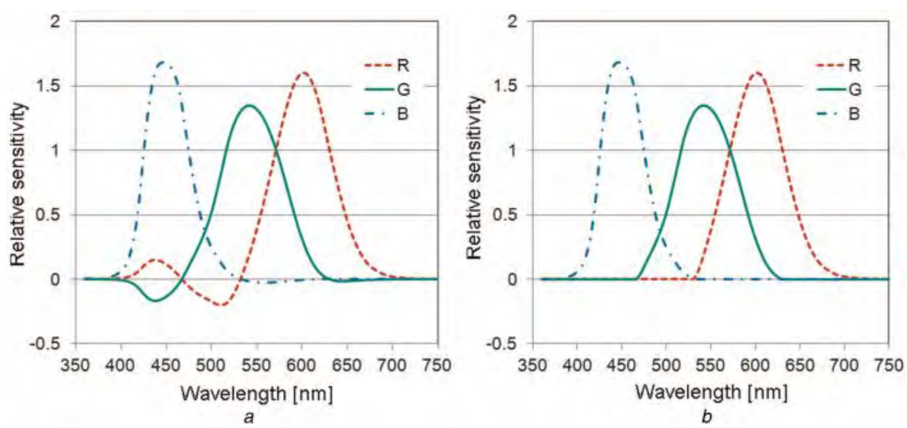


Figure 1 *Spectral sensitivity for the SHV wide-colour gamut: (a) ideal curves and (b) curves composed of the positive main lobes of the curves with ideal spectral sensitivity*

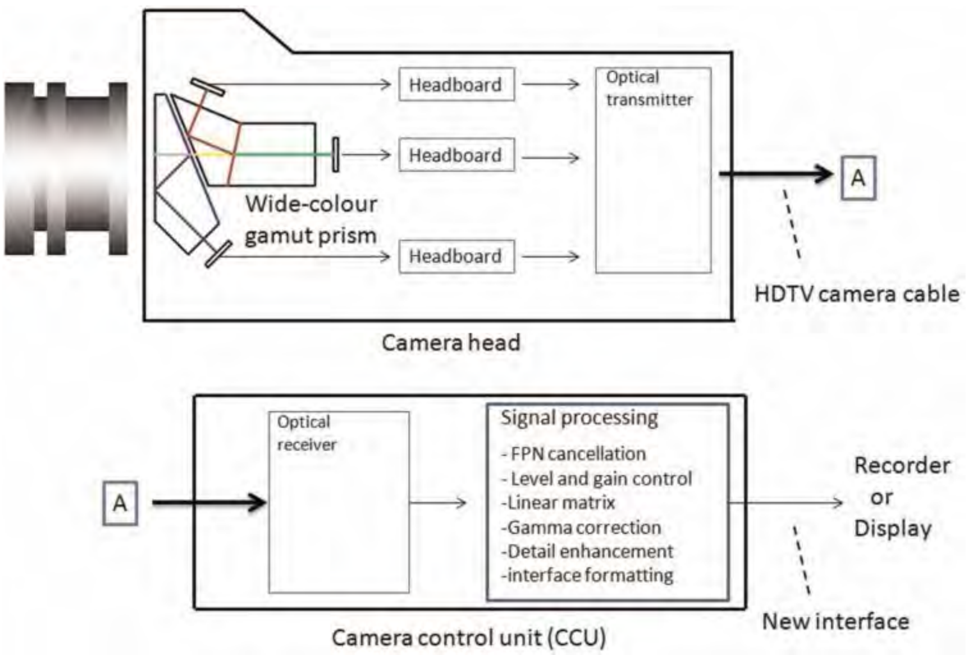


Figure 2 *Block diagram of the camera system*

Figure 3 *Appearance of the camera head*

formatted into a new interface described in the following section. Figure 3 shows the appearance of the camera head.

Spatial resolution and SNR

An example of a reproduced image is shown in Figure 4. We constructed a chart corresponding to one-fourth of the entire screen. In the image, a modulation response at 2,000 TV lines, which corresponds to 4,000 TV lines by conversion, was observed. We also measured the modulation transfer function (MTF) of the luminance component (Y) by using the 'slanted edge' method with the ISO-12233 resolution chart [8]. Figure 5 shows the characteristics of the measured MTF in the vertical direction. The calculated points denote the theoretical MTF calculated by multiplying the respective component MTFs of the system,

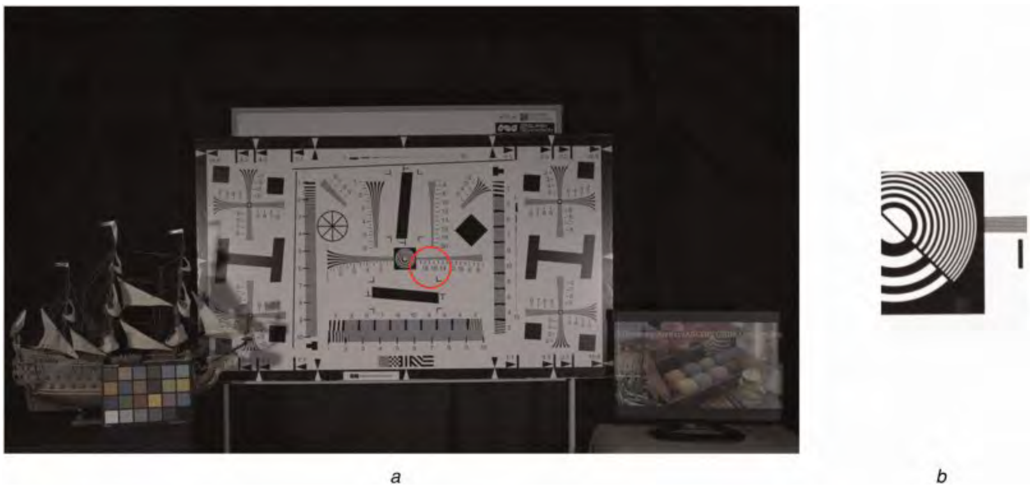


Figure 4 *(a) Example of reproduced image (7,680 × 4,320 px) (b) Enlarged view of the image corresponding to the red circle in (a), to measure around 2,000 TV lines*

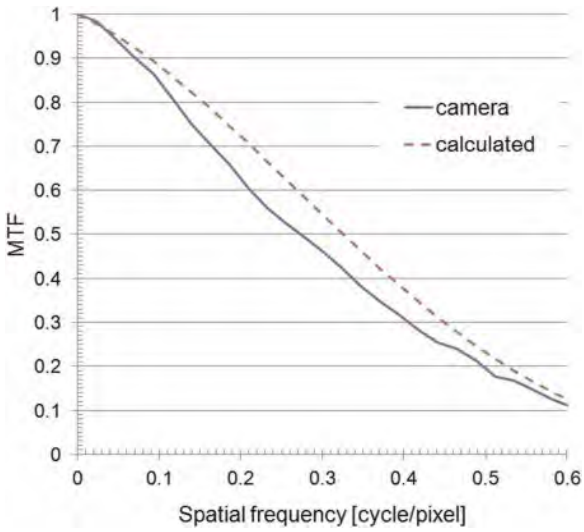


Figure 5 *MTF characteristics of the camera*

the ideal aberration-free-lens MTF, and the image sensor MTF [9] when the f-number of the lens is set to 4.0. We can observe that the MTF at the Nyquist frequency is approximately 20%. From the resolution measurement presented above, this camera has sufficient ability to capture 8K resolution images.

Measurement of the camera's SNR revealed a value of 50 dB (sensitivity setting: luminance level is 2,000 lx; lens aperture sets at F# 5.6), which is better than that of a conventional SHV camera system [5], despite having a frame frequency which is twice as high.

Capturing moving objects

We captured both a moving object and a one-frame image while the camera was operated at frame frequencies of 120 Hz and 60 Hz (Figure 6). The motion blur of the image with a frame frequency of 120 Hz was less than that with a 60-Hz frame frequency. From the results, we confirmed that the frame frequency of 120 Hz improves the image quality, and the camera system can capture images at 120 Hz.

Colour reproducibility

The colour-reproduction characteristics of the camera system were measured using a 24-patch transmission

Figure 7 *Integrating sphere and 24-patch transmission colour checker chart displaying high-saturation colour by doubling two charts*

colour checker chart [10]. We displayed high-saturation colour by doubling two transparent charts [11]. An integrating sphere was used to create a light source with a uniform luminance distribution, and the colour temperature of the light source was 6810 K. Figure 7 shows the appearance of the integrating sphere and colour chart. The colour samples from No. 1 through No. 24 and from No. 25 through No. 43 were obtained using one transmission chart and two doubling charts, respectively.

The tristimulus values of each colour patch were measured using a spectroradiometer (Topcon's SR-UL1) and were set as the reference values. The colour differences (DE) between the reference values and the camera system values were calculated using a CIE LAB colour-difference formula. An LM was derived via a calculation which minimises the average of the colour differences of these colour patches. The following LM is used for colour correction of the capturing system:

$$\begin{bmatrix} 1.16 & -0.21 & 0.05 \\ -0.02 & 1.12 & -0.10 \\ -0.01 & -0.03 & 1.03 \end{bmatrix}$$

. The average value and maximum of DE were 1.0 and 2.9 (No. 38), respectively. Figure 8 shows the xy chromaticities of colour patches measured by the spectroradiometer

Figure 6 *Example images of a moving object at frame frequencies of (a) 120 Hz and (b) 60 Hz*

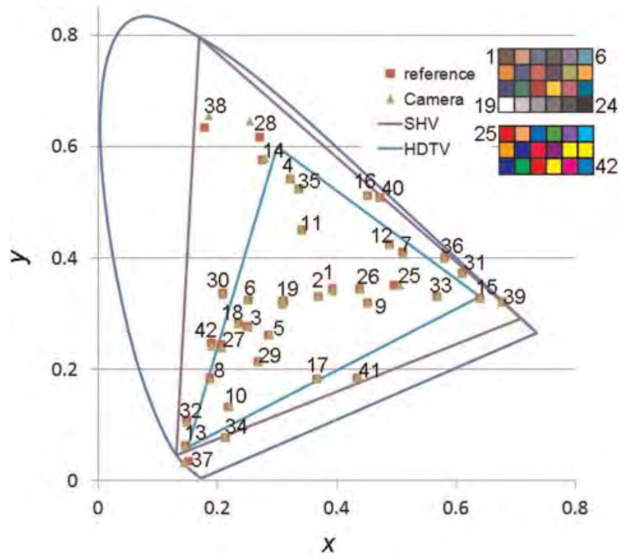


Figure 8 *xy chromaticities of colour patches*

(reference) and captured by the camera in the CIE 1931 chromaticity diagram and standard RGB primaries. The numbers beside the points are the patch numbers of the colour checker.

Table 3 Connector specifications.

Number of fibres	24 (12 × 2)
Fibre type	Multimode fibre
Connection loss	Less than 0.75 dB
Insertions/withdrawals	More than 5,000 times
Equilibrium tensile loading of connectors	250 N
Lock mechanism	Bayonet-type lock

Studio equipment interface

Interface requirements

The requirements of the interface used to connect to UHDTV devices are as follows:

- The interface should support all UHDTV formats specified in Rec. ITU-R BT.2020 in the unified method for interface dissemination;
- The interface must be compact so that it can be incorporated into UHDTV devices, have low-power operation, achieve signal transmission with a single cable, and be easily connected; and
- The interface should transmit signals at a transmission distance of approximately 50 m because the interface connects devices in a studio or outside broadcast van.

Interface design

The payload data-rate for the transmission of a full-specification SHV signal is approximately 144 Gbit/s. It is difficult to transmit a full-specification SHV signal via a single coaxial cable with current technologies; therefore, we adopt an optical device.

We adopted an optical multilink transmission system which is composed of a multicore, multimode fibre-optic cable and parallel fibre-optic transceiver because the aim of the interface is to transmit a UHDTV signal over a short length with low cost. Furthermore, we employed 10-Gbit/s optical transceivers which are practically available with

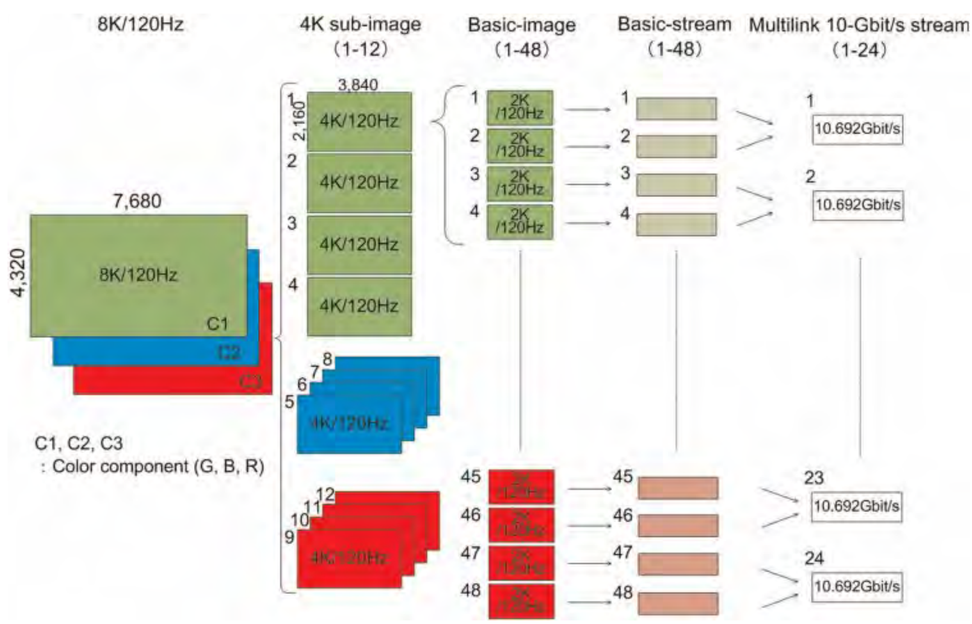


Figure 9 *Overview of the signal mapping of a full-specification SHV image into proposed 10-Gbit/s multilink interface*

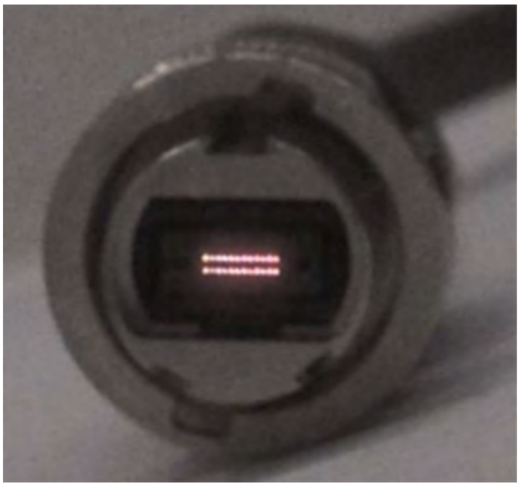


Figure 10 *Connector appearance*

current technologies. The target bit error rate is set to no less than 10^{-12}, which is defined by the existing standards using a 10-Gbit/s signal [12].

Signal mapping

Figure 9 shows an overview of a full-specification SHV image mapped into multilink 10-Gbit/s streams. First, each colour component signal (R, G, B) is divided into four 4K sub-images with a sample structure of 3,840 × 2,160 pixels. Various UHDTV signals (7,680 × 4,320 or 3,840 × 2,160 pixels; sampling structures of 4:4:4, 4:2:2, and 4:2:0) can be similarly handled using this method. Next, a 4K sub-image is divided into basic-images. Then, the basic-stream is obtained by streaming a basic-image. The basic-streams are multiplexed, and 8b/10b encoding is performed. 10-Gbit/s data are obtained by serialising the 8b/10b encoding data. The mapping method is capable of switching frames because the frame frequency of the original signal is maintained; namely, the method supports a 120-Hz signal.

On the basis of the aforementioned interface system, a prototype interface which connects UHDTV devices by using parallel fibre-optic transceivers and a multimode fibre-optic cable was developed. Moreover, the practicality and feasibity of the system were verified [13].

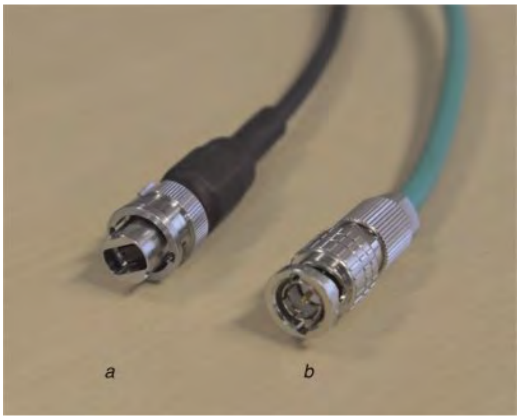


Figure 11 *Appearance of the new interface optical cable (a), and a coaxial cable (b)*

Table 4 Specifications of the wide-colour gamut laser projector.

Device	1.3 in Liquid Crystal on Silicon (LCOS)
Spatial resolution pixels	Equivalent to 7,680 × 4,320
Frame frequency	120 Hz
Light source	3,000 ANSI lumen
Contrast	5,000 : 1
Bit depth	12 bits
Light source	RGB semiconductor laser (R: 639 nm, G: 532 nm, B: 462 nm)

Standardisation

This system was proposed to the Association of Radio Industries and Businesses (ARIB), which is a Japanese standardisation organisation, and was standardised as ARIB STD-B58 in March 2014 [14].

Because the intensity of the multiple-fibre push-on (MPO) ordinarily adopted in optical multilink transmission is insufficient in terms of the strength for broadcasting use, a new connector standard was determined. We aimed to ensure an operability and a shape similar to a BNC connector. Table 3 lists the specifications of the

Figure 12 *– Wide-colour gamut projector*

Table 5 – Primaries of the wide-colour gamut laser projector.

	CIE x	CIE y
Red primary	0.7157	0.2837
Green primary	0.1781	0.7846
Blue primary	0.1342	0.0455

Figure 13 *Waveform and colour gamut monitor*

standardised connector and its appearance in shown in Figure 10. We developed the full-specification SHV camera system described in the previous section, a video routing switcher, and a display which incorporates the interface. Further, we also developed an optical cable for connecting equipment. Figure 11 shows the appearance of the optical cable compared to a coaxial cable, where the optical cable is similar in size to the coaxial cable.

Monitoring equipment

We have developed a wide-colour gamut SHV projector operating at 120 Hz by modifying a prefabricated projector (15). The light sources of the SHV projector were replaced with red, green, and blue lasers corresponding to the wide-gamut colorimetry. The wide-colour gamut projector specifications are listed in Table 4, and the projector appearance is shown in Figure 12. The wavelengths of the lasers used as the light sources are slightly different from the wavelengths corresponding to the three primary colours defined by the standardised format; however, they were adopted in consideration of the cost and feasibility of the devices using present technology. The chromaticity points of the three primary colours displayed by the projector were measured using a spectroradiometer. The results are listed in Table 5, and it is confirmed that the projector achieved approximately equal primaries to Rec. ITU-R BT.2020. We also developed a waveform monitor which displays the xy chromaticity distribution of the wide-gamut RGB signals (Figure 13).

Additionally, we also developed an 85-inch liquid crystal display (LCD) with 8K resolution, operating at 120 Hz and incorporating the new interface described in the previous section. The colour gamut of the LCD is considerably wider than that of HDTV, but needs to be

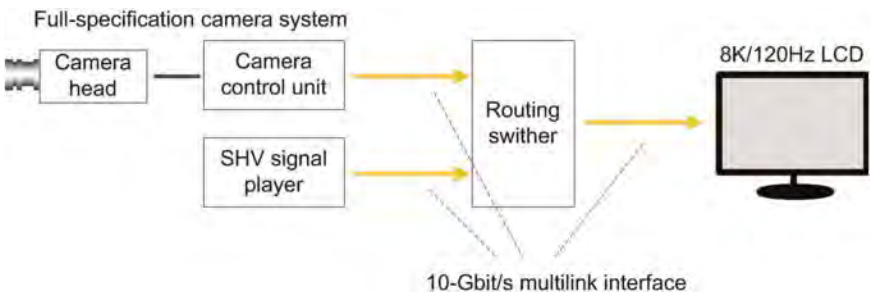


Figure 14 *System diagram of full-specification SHV demonstration*

Figure 15 *Full-specification SHV demonstration at the NHK STRL Open House 2014*

further widened to achieve that of Rec. BT. 2020. We plan to improve the colour filters and the backlights of the LCD.

Demonstration of full-specification SHV

We used the aforementioned equipment to produce full-specification SHV content to evaluate the video quality. The effectiveness of a full-specification SHV was confirmed in a demonstration at the NHK Science and Technology Research Laboratories (NHK STRL) Open House held in May 2014 [16]. Figure 14 shows the demonstration system. The new interface is incorporated in all of the equipment (camera, signal player, routing switcher, and LCD) which all operates at a frame frequency of 120 Hz. The improvements in the image quality of an 8K resolution, 120-Hz frame frequency and wide-colour gamut system were verified. Figure 15 shows the appearance of the demonstration.

Conclusions

In this paper, we have described the equipment developed to create a full-specification SHV system: the world's first full-specification camera system, an interface and a display.

The camera system can capture 8K resolution images at a 120-Hz frame frequency. We developed an interface which can transmit a full-specification SHV signal with a data-rate of approximately 144 Gbit/s via a single optical cable. We plan to devise a monitoring method for the optical signals, and discuss the possibility of audio data insertion into an ancillary signal domain. We also developed monitoring devices; an 8K/120Hz LCD, 8K/120Hz wide-colour gamut laser projector and a waveform monitor which supports wide-gamut signals. We confirmed the capability of the full-specification SHV system and the effectiveness of the 8K resolution, 120-Hz frame frequency, and wide-colour gamut colorimetry.

References

[1] Recommendation ITU-R BT.2020, 2012. Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange

[2] SMPTE ST 2036-1, 2013. Ultra High Definition Television – Image Parameter Values for Program Production

[3] ARIB STD-B56 (1.1), 2014. UHDTV System Parameters for Programme Production. March 2014

[4] NISHIDA Y., MASAOKA K., SUGAWARA M., OHMURA M., EMOTO M., NAKASU E.: 'Super Hi-Vision System Offering Enhanced Sense of Presence and New Visual Experience', *The Best of IET and IBC*, 2011, **3**, (2011), pp. 5–10

[5] YAMASHITA T., FUNATSU R., YANAGI T., MITANI K., NOJIRI Y., YOSHIDA T.: 'A Camera System Using Three 33-Megapixel CMOS Image Sensors for UHDTV2', *SMPTE Motion Imaging Journal*, November/December 2011, pp. 24–30

[6] Recommendation ITU-R BT.709-5, 2002. Parameter Values for the HDTV Standards for Production and International Programme Exchange

[7] MASAOKA K., SOENO T., KUSAKABE Y., YAMASHITA T., NISHIDA Y., SUGAWARA M.: 'UHDTV System Colorimetry and Wide-Gamut Color Reproduction', *Asia-Pacific Broadcasting Union TECHNICAL REVIEW*, January-March 2014, pp. 2–7

[8] ISO-12233, Photography - Electronic Still-Picture Cameras Resolution Measurements

[9] BOREMAN G.D.: 'Modulation Transfer Function in Optical andElectro-Optical Systems', *SPIE PRESS*, 2000, pp. 31–57

[10] TE188 Color Rendition Chart, Image Engineering, Frechen, Germany

[11] MASAOKA K., NISHIDA Y., SUGAWARA M., KATO S., SAITA A.: 'Spectral Sensitivity and Color Reproduction for SHV Wide-Gamut Camera', *ITE Technical Report.* 2011, Vol. 35, No. 45, pp. 83–86

[12] SMPTE ST 435-2, 2012. 10 Gb/s Serial Digital Interface – Part2: 10.692 Gb/s Stream – Basic Stream Data Mapping

[13] SOENO T., NISHIDA Y., YAMASHITA T., KUSAKABE Y., FUNATSU R., NAKAMURA T.: 'Development of a Multilink 10 Gbit/sec Mapping Method and Interface Device for 120 Frames/sec Ultrahigh-Definition Television Signals', *SMPTE Motion Imaging Journal*, May/June 2014, pp. 29–38

[14] ARIB STD-B58 (1.0), 2014. Interface for UHDTV Production Systems. March. 2014

[15] SHIMAMOTO H., KITAMURA K., WATABE T., OHTAKE H., EGAMI N., KUSAKABE Y., NISHIDA Y., KAWAHITO S., KOSUGI T., WATANABE T., YANAGI T., YOSHIDA T., KIKUCHI H. 2012 '120 Hz Frame-Rate Super Hi-Vision Capture and Display Devices', *SMPTE Motion Imaging Journal*, March, 2013, pp. 55–61

[16] NHK Science and Technology Research Laboratories Open House 2014. Full-Specification 8K Super Hi-Vision Production System

Selected content from IBC2014

ISSN 2041-5923

HEVC subjective video quality test results

T.K. Tan[1] *M. Mrak*[2] *R. Weerakkody*[2] *N. Ramzan*[3]
V. Baroncini[4] *G.J. Sullivan*[5] *J.-R. Ohm*[6] *K.D. McCann*[7]

[1]*NTT DOCOMO, Inc., Japan*
[2]*British Broadcasting Corporation, UK*
[3]*University of West of Scotland, UK*
[4]*Fondazione Ugo Bordoni, Italy*
[5]*Microsoft Corporation, USA*
[6]*RWTH Aachen University, Germany*
[7]*Zetacast, UK*

Abstract: The primary goal for the development of the new High Efficiency Video Coding (HEVC) standard was to achieve a substantial improvement in compression efficiency relative to its predecessor, H.264/AVC. However, most comparisons to date had used simple objective measurements, rather than a formal *subjective* evaluation. Though objective tests are relatively easy to perform, their results can only be indicative of the actual quality perceived by viewers.

This paper describes the verification tests that were conducted by the Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 (VCEG) and ISO/IEC JTC 1/SC 29/WG 11 (MPEG). Both objective and subjective evaluations were conducted, comparing the HEVC Main profile to the H.264/AVC High profile. The tests used sequences with resolutions ranging from 480p to Ultra High Definition (UHD), encoded at various quality levels. The results showed that the subjective performance of HEVC is actually better than would be predicted from the objective test results. In 86% of the test cases, the subjective test results showed that HEVC required half or less of the bit-rate of the H.264/AVC reference to achieve comparable quality. The average bit-rate savings for test sequences at UHD, 1080p, 720p and 480p resolutions were found to be approximately 64%, 62%, 56% and 52%, respectively.

Introduction

We are currently witnessing something that has become a once-in-a-decade event in the world of video compression: the emergence of a major new family of video compression standards.

The mid-1990s saw the introduction of MPEG-2, the first compression standard to be widely adopted in broadcasting applications. H.264/AVC appeared in the mid-2000s, offering the same subjective quality at approximately half the bit-rate. Now a new standard, High Efficiency Video Coding (HEVC), has been developed that promises a further factor of two improvement in compression efficiency for the mid-2010s. As a consequence, HEVC can likely be used to broadcast Ultra High Definition (UHD) video while using transmission bit-rates that a decade ago were required for High Definition (HD), and two decades ago for Standard Definition (SD) resolution.

The HEVC standard has been jointly developed by the same two standardization organizations whose previous collaboration resulted in both MPEG-2 and H.264/AVC: the ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG).

The initial edition of the HEVC standard was completed in January 2013 and it is published by ISO/IEC as ISO/IEC 23008-2 (MPEG-H Part 2) and by ITU-T as Recommendation H.265 [1]. This first version supports applications that require single-layer 4:2:0 sampled video with 8 or 10-bit precision. A second edition, which extends the standard to support contribution applications with tools that enable 4:2:2 and 4:4:4 sampled video as well as 10, 12 or 16-bit precision, was completed in April 2014. Further amendments are currently being developed to extend the standard by efficient scalability and stereo/multiview compression, and to allow representation of 3D (video plus depth) content.

This paper presents results obtained during the recent JCT-VC verification testing of HEVC. The purpose of this test was to verify that the key objective of HEVC had been achieved: providing a substantial improvement in compression efficiency (e.g. by a factor of two) relative to its predecessor, ITU-T Rec. H.264 | ISO/IEC 14496-10 Advanced Video Coding (AVC) [2]. The test sequences used for the verification testing were deliberately chosen to be different from those that had been used during the development of the standard, to avoid any possible bias that the standard could have towards those sequences.

This paper is organized as follows. The test methods, material and laboratory set-up are described in the section of verification test methodology. The subjective test results and analysis are described in the section on verification test results. Finally, the conclusions are drawn in the last section.

Verification test methodology

Test methods

The subjective test method adopted for this evaluation is Degradation Category Rating (DCR) [3]. This test method is commonly adopted when the material to be evaluated shows a range of visual quality that distributes well across all quality scales. This method was used under the schema of evaluation of the quality (and not of the impairment); for this reason a quality rating scale made of 11 levels was adopted, ranging from "0" (lowest quality) to "10" (highest quality).

The structure of the Basic Test Cell (BTC) of DCR is shown in Figure 1. The structure comprises two consecutive presentations of the video clip under test. First the original version of the video clip is displayed followed by the coded version of the video clip. Then a message asking the viewers to vote is displayed for 5 seconds. A mid-grey screen displaying for one second precedes the presentation of both video clips. All the video material used for these tests consists of video clips of 10 seconds duration.

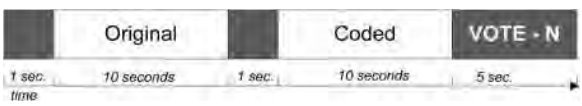


Figure 1 *DCR BTC*

Test material

The test material for the verification test was encoded using HM12.1 (the reference software of the HEVC codec) and JM18.5 (the reference software of the AVC codec).

Four picture resolutions were tested: UHD-1 (2160p), 1080p, 720p and 480p. Each resolution was represented by 5 test sequences, giving a total of 20 test sequences.

For each test sequence 4 test points with different bit-rates were chosen to be encoded using the HM12.1 and JM18.5 software. The four JM18.5 bit-rates were chosen such that the quality levels spanned the entire range of the mean opinion score (MOS) scale. The HM12.1 test points with indices $i = 0, 1, 2, 3$ were then selected such that the bit-rates $R_{\mathrm{HEVC}}(i)$ were approximately half the bit-rates of the corresponding JM18.5 reference points $R_{\mathrm{AVC}}(i)$, i.e. $R_{\mathrm{HEVC}}(i) \approx \frac{1}{2}.\ R_{\mathrm{AVC}}(i)$.

Further detail, including the coding conditions used, the quantization parameter (QP) selection and encoded bit-rates can be found in Annex B of [3].

Laboratory Setup

For subjective assessment, the laboratories were set up according to [3], except for the selection of the display and the video play-out server.

The TVLogic LUM-560W 56" professional LCD monitor was used for UHD sequences, the Pioneer 50" PDP-5000EX monitor was used for 720p. Both monitors were set to their native resolutions of 3840×2160 and 1920×1080 pixels, respectively.

For the 1080p and 480p video clips, the monitor used was the HP 30" ZR30W set at its native resolution of 1920×1080 pixels. Two simultaneous displays were used with one viewer per screen.

Play-out of the 720p and 480p video clips was at the native resolution of the monitors (1920×1080) using the central area of the screen; the remaining part of the screen is set to a mid-grey level (128 in 0–255 range).

The viewing distance varied according to the physical dimensions and the native resolutions of the monitors; this led to the viewing distance varying from 1.5H to 3H, where H is equal to the height of the screen.

Verification test results

Subjective test results

In the analysis of the results, the mean opinion score (MOS) values and the associated confidence interval (CI) were computed for each test point. From the raw data, the

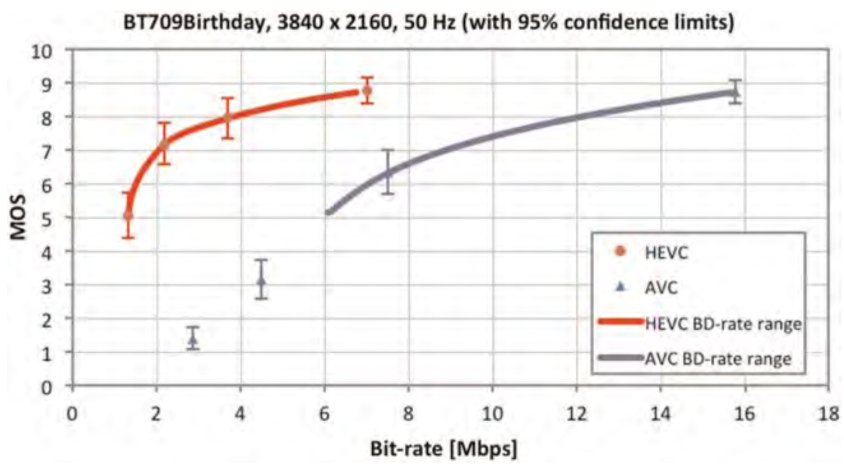


Figure 2 *Example of UHD sequence*

reliability of each viewer was calculated by computing the correlation index between each score provided by a viewer to the general MOS value computed for that test point. A correlation index greater than or equal to 0.75 was considered as valid for the acceptance of the viewer.

The MOS values together with the CI were then plotted against the bit-rate for each sequence tested. Figures 2 through 5 show examples of the plots for each of the resolutions tested. The complete data set of the twenty sequences can be found in [3].

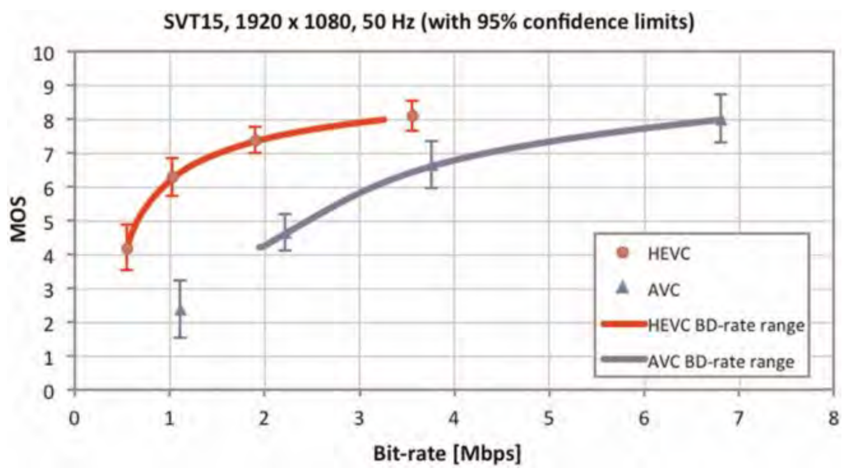


Figure 3 *Example of 1080p sequence*

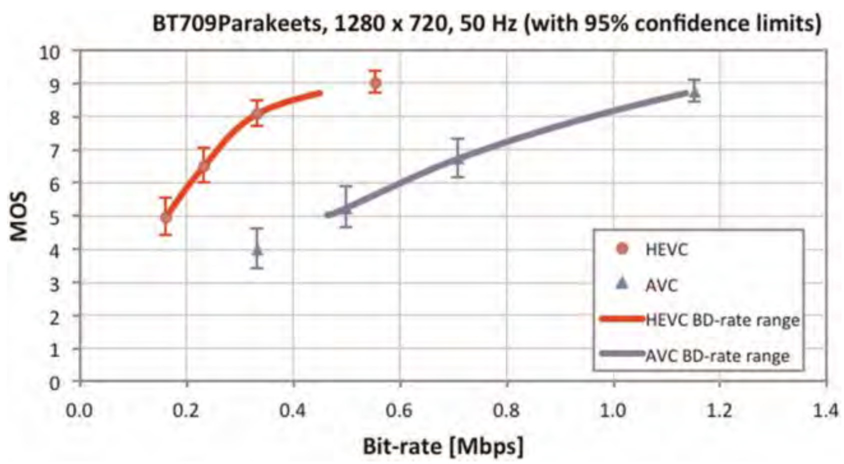


Figure 4 *Example of 720p sequence*

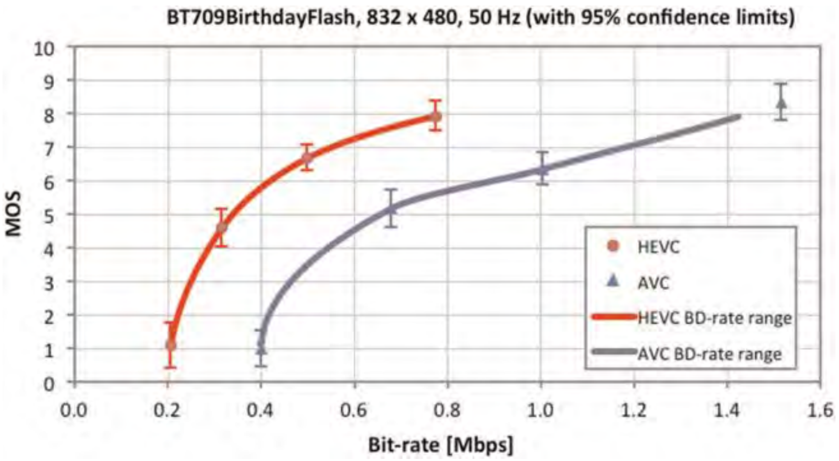


Figure 5 *Example of 480p sequence*

Analysis

All the plots in Figures 2 through 5 show that the rate-distortion (RD) curves of the HEVC test points are located substantially to the left of the RD-curves of the AVC reference points. This shows that HEVC is achieving a substantial bit-rate reduction relative to AVC.

An example of the analysis of the test points is shown in Figure 6. All the test points were classified into the following cases where the HEVC test points $P_{\mathrm{HEVC}}(i)$ have the same subjective quality (overlapping MOS confidence intervals (CI)) as the AVC reference points $P_{\mathrm{AVC}}(i)$ where the bit-rates of AVC are:

(a) greater than double the bit-rates of HEVC ($\mathrm{CI}_{\mathrm{AVC}}(m)$ overlaps with $\mathrm{CI}_{\mathrm{HEVC}}(n)$ and $m > n$)
(b) equal to double the bit-rates of HEVC ($\mathrm{CI}_{\mathrm{AVC}}(m)$ overlaps with $\mathrm{CI}_{\mathrm{HEVC}}(n)$ and $m = n$),
(c) less than double the bit-rates of HEVC ($\mathrm{CI}_{\mathrm{AVC}}(m)$ overlaps with $\mathrm{CI}_{\mathrm{HEVC}}(n)$ and $m < n$), and $\mathrm{CI}(i)$ corresponds to the confidence interval of $R(i)$ and both are associated to test point $P(i)$.

45% of the HEVC test points achieved approximately 50% better coding efficiency than AVC. Another 41% of the HEVC test points achieved more than 50% better coding efficiency than AVC. In contrast, only 14% of the HEVC test points achieved a coding efficiency of less than 50% better than AVC.

Even though the inspection of the RD-curves and the overlapping MOS CI gave high confidence that HEVC was achieving a bit-rate reduction over AVC that was very likely above 50%, the test data points were not sufficiently dense to quantify precisely the amount of the bit-rate reduction. In order to estimate the coding efficiency improvement achieved the approach described in the next section was employed.

MOS BD-rate

In this section, the average bit-rate savings of HEVC compared to AVC for each sequence are computed from the MOS scores in the same manner as in [5][6], to further quantify the bit-rate savings achieved. The Bjøntegaard Delta (BD) rate measure described in [7][8]

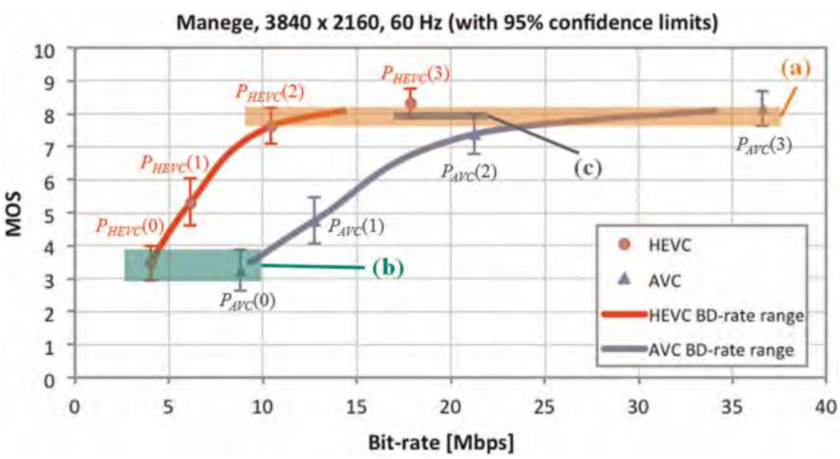


Figure 6 *Examples of test points with overlapping CI*

Table 1 Average bit-rate savings

Resolution	Sequence	MOS BD-rate
UHD-1 (2160p)	BT709Birthday	−75%
	Book	−66%
	HomelessSleeping	*
	Manege	−56%
	Traffic	−58%
1080p	JohnnyLobby (LD)	−70%
	Calendar	−52%
	SVT15	−69%
	sedofCropped	−53%
	UnderBoat1	−68%
720p	ThreePeople (LD)	−48%
	BT709Parakeets	−66%
	QuarterBackSneak	−58%
	SVT01a	−73%
	SVT04a	−36%
480p	Cubicle (LD)	−45%
	Anemone	−42%
	BT709BirthdayFlash	−49%
	Ducks	−72%
	WheelAndCalender	*
Average		−58.7%

was used with the MOS scores taking the place of the Peak Signal-to-Noise Ratio (PSNR) and a piecewise cubic interpolation instead of cubic spline interpolation.

Average bit-rate savings were calculated where the same MOS scores for both HEVC and AVC could be interpolated from subjective test results, as shown in

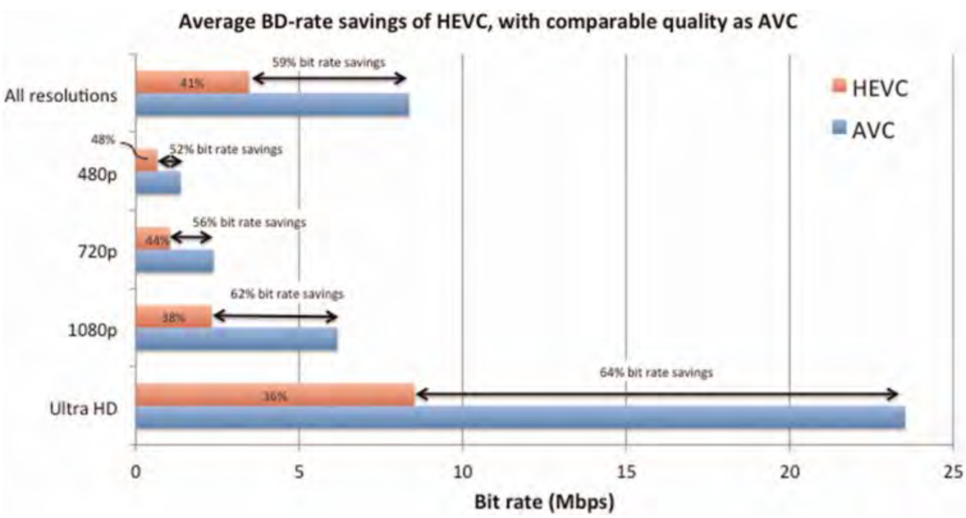


Figure 7 *Average bit-rate savings (measured by BD-Rate) of HEVC compared to AVC*

Figures 2 through 5. The interval over which the BD-rate is averaged is shown as solid lines in each of these plots. Table 1 shows the average bit-rate savings calculated by this measure for the 20 sequences.

Bit-rate savings were not calculated for the sequences HomelessSleeping and WheelAndCalender, since they did not satisfy the criteria for an accurate calculation: MOS values exhibiting a smooth curve, with an averaging interval interpolated from at least three MOS scores with monotonically increasing bit-rate.

The average bit-rate saving calculated from these results confirmed that the HEVC Main profile achieved the same subjective quality as AVC High profile while requiring approximately 59% fewer bits. It was noticeable that higher resolution sequences tended to achieve slightly greater savings than lower resolution sequences.

Figure 7 shows the average bit-rate savings for test sequences with UHD-1, 1080p, 720p and 480p resolutions, which are estimated at approximately 64%, 62%, 56% and 52%, respectively. The bit-rates shown in this illustration correspond to the average of the highest bit-rate points over all sequences at each resolution.

Table 3 of [3] shows that the BD-Rate savings remain consistently high when only the results of test points with quality suitable for broadcasting (MOS greater than or equal to 7, corresponding to "good" or "excellent") are considered.

Conclusions

Analysis of the subjective test results show that HEVC at half or less than half the bit-rate of the AVC reference achieved comparable quality in 86% of the test cases.

By applying a MOS BD-rate measurement on the results of the subjective test, it was found that the HEVC Main profile achieves the same subjective quality as AVC High profile while requiring on average approximately 59% fewer bits.

It can therefore be concluded that the HEVC standard is able to deliver the same subjective quality as AVC, while on average only requiring half or even less than half of the bit-rate used by AVC. The initial objective of the HEVC development of substantial improvement in compression compared to previous state of the art has therefore been successfully achieved.

Acknowledgment

The authors thank the contributors of the video test sequence material for permitting its use for the tests and this publication, and their collaborators in the JCT-VC for their assistance in developing the test plan and analyzing its results.

References

[1] ITU-T and ISO/IEC JTC1, "High Efficiency Video Coding, ITU-T Rec. H.265 and ISO/IEC 23008-2", April 2013

[2] ITU-T and ISO/IEC JTC1, "Advanced Video Coding for Generic Audio-Visual Services, ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC)", March 2005

[3] TAN T.K., *ET AL.*: 'HEVC verification test report', Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCTVC Q1011 (ISO/IEC JTC1/SC29/WG11 N14420), 17th Meeting, Valencia, ES, 27 March – 4 April 2014

[4] International Telecommunication Union Standardization Sector; Recommendation ITU T P.910 "Subjective video quality assessment methods for multimedia applications"

[5] TAN T.K., FUJIBAYASHI A., SUZUKI Y., TAKIUE J.: 'Objective and Subjective Evaluation of HM5.0', document JCTVC-H0116, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, San Jose, CA, February 2012

[6] OHM J.-R., *ET AL.*: 'Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)', invited paper to IEEE CSVT, Special Issue on Emerging Research and Standards in Next Generation Video Coding (HEVC), December 2012

[7] BJØNTEGAARD G.: 'Calculation of Average PSNR Differences Between RD Curves', ITU T SG16/Q6, 13th VCEG Meeting, Austin, Texas, USA, April 2001, Doc. VCEG-M33

[8] BJØNTEGAARD G.: 'Improvements of the BD-PSNR model', ITU-T SG16/Q6, 35th VCEG Meeting, Berlin, Germany, July 2008, Doc.VCEG-AI11

ISSN 2041-5923

Introduction to *Electronics Letters*

In 2014 *Electronics Letters* is celebrating its 50th year of publication. Launched in 1965, just two years before the first IBC, over the last five decades *Electronics Letters* has published over 43,000 papers and seen its scope evolve to reflect the amazing changes and advances in electronics since the 1960s.

Electronics Letters[1] is a uniquely multidisciplinary rapid publication journal with a short paper format that allows researchers to quickly disseminate their work to a wide international audience. *Electronics Letters* broad scope involves virtually all aspects of electrical and electronic technology from the materials used to create circuits, through devices and systems, to the software used in a wide range of applications. The fields of research covered are relevant to many aspects of multimedia broadcasting including fundamental telecommunication technologies and video and image processing.

Each year the *Electronics Letters* editorial team and the executive committee of the IET Multimedia Communications Community come together to select a small number of papers from the relevant content of Electronics Letters to appear in this publication.

This year all three of the chosen papers were also highlighted in the free magazine style news section included as part of each issue of Electronics Letters. The news section includes articles based on some of the best papers in each issue, providing more background and insight into the work reported in the papers. The news articles associated with all three papers are also included in this 2014 volume of The Best of IET and IBC.

Respectively addressing terahertz communications, issues encountered in evolving current optical networks for new services and developing components for increasingly multi-standard wireless devices, the papers selected give an idea of how research published in *Electronics Letters* will impact both the immediate and long term future of the broadcast media industry.

We hope you will enjoy reading these papers and features as examples of our content, and if you like what you read, all our feature articles are available for free via our web pages[1].

The *Electronics Letters* editorial team

[1]www.theiet.org/eletters

Selected content from *Electronic Letters*

ISSN 2041-5923

Coherent THz communication at 200 GHz using a frequency comb, UTC-PD and electronic detection

G. Ducournau[1] Y. Yoshimizu[2] S. Hisatake[2] F. Pavanello[1] E. Peytavit[1] M. Zaknoune[1] T. Nagatsuma[2] J.-F. Lampin[1]

[1]Institut d'Electronique, de Microélectronique et de Nanotechnologie (IEMN), UMR CNRS 8520, Université Lille 1, Avenue Poincaré, B.P. 60069, 59652 Villeneuve d'Ascq, France

[2]Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka 560-8531, Japan

E-mail: nagatuma@ee.es.osaka-u.ac.jp

Abstract: A coherent terahertz (THz) link at 200 GHz, with a variable data rate up to 11 Gbit/s, featuring a very high sensitivity at the receiver, is investigated. The system uses a quasi-optic unitravelling carrier photodiode (UTC-PD) emitter and an electronic receiver. The coherent link relies on an optical frequency comb generator at the emission to produce an optical beat note with 200 GHz separation, phase-locked with the receiver. Bit error ratio testing has been carried out using an indoor link configuration, and error-free operation is obtained up to 10 Gbit/s with a received power $<2\ \mu W$.

Introduction

To serve new services such as video on mobile terminals, the demand for bandwidth is growing every year, and new ways for wireless data transmission are being investigated. Terahertz (THz) communications are thus very promising [1] as their open a huge space for new high data-rate services. THz communications have been intensively studied over the past few years and systems are now emerging based on photonic and/or electronic technologies. Among these different systems, coherent THz links pave the way for high performances in terms of data rate and sensitivity at receiver circuits. For example, using photonics at the emission, Nagatsuma *et al.* [2] have achieved real-time coherent 100 GHz up to 11 Gbit/s and Li *et al.* [3] reported 200 Gbit/s in the W-band using off-line demodulation and signal processing. Above the W-band, Koenig *et al.* [4] reported very recently up to 100 Gbit/s for a 20 m link at 237.5 GHz, also using the off-line detection. At the electronic side (both at emission and reception), Antes *et al.* [5] achieved 30 Gbit/s at 240 GHz, using off-line processing. Future real-time THz coherent communication systems will require a locking between the transmitter (Tx) and the receiver (Rx) to supress any residual frequency drift between Tx and Rx, downconverted into random amplitude modulation at the receiving mixer. For this reason, off-line detection can be used to overcome the locking problem, in order to highlight the intrinsic characteristics of the THz link. At the photonic side, a very convenient way to create the dual-tone optical signal is to use optical frequency combs, where the combs remain locked on a microwave reference. One of the key issues to realise coherent communication by photonics-based THz generation is the phase locking. In the proposed system, the phases of two optical sidebands extracted with an arrayed waveguide grating (AWG) have been locked to the optical frequency comb.

Experimental setup

The data link is described in Fig. 1.

Transmitter description

The system is first composed of the optical frequency comb generator (OFCG), described in Fig. 2. The phase of a

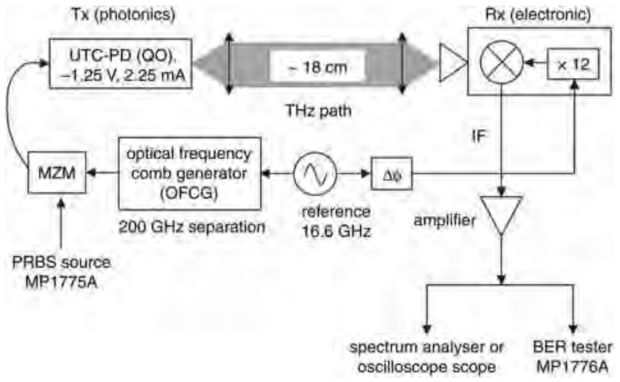


Figure 1 *Experimental setup of 200 GHz coherent data link*

CW: continuous wave; D f: fixed electrical phase shifter

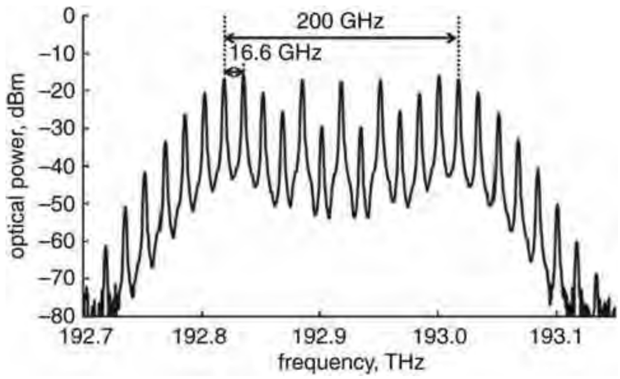


Figure 3 *Optical spectra obtained at OFCG output*

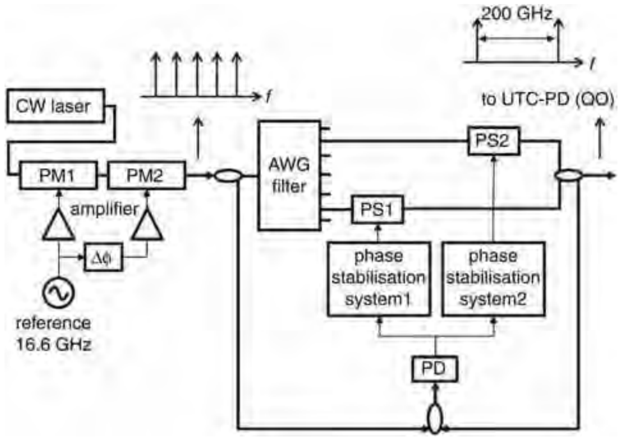


Figure 2 *OFCG details*

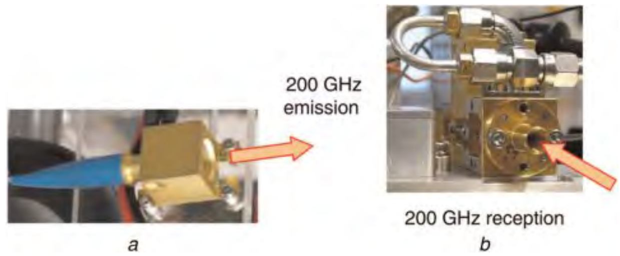


Figure 4 *Views of quasi-optic UTC-PD module and electronic receiver, working at 200 GHz*

a Quasi-optic UTC-PD module
b Electronic receiver, working at 200 GHz

single frequency continuous-wave (CW) laser is modulated with cascaded optical phase modulators to generate the OFC. The modulation frequency was 16.6 GHz. Dual-optical carriers with 200 GHz frequency separation are extracted with an AWG filter. The phases of the two optical carriers fluctuate independently due to temperature fluctuations and acoustic noise in the optical fibre cables between the AWG filter and the optical coupler. The coherence of the THz wave to be generated as a carrier for the wireless link is dependent on the stability of the phase differences between the optical carriers. In our system, the phases of the two carriers are locked to that of the OFC to generate a coherent THz wave, the spectrum of which is given in Fig. 3. Details of the phase stabilisation system are described in [6, 7].

The dual-optical signal is then modulated by a Mach-Zehnder amplitude modulator with a variable data rate up to 11 Gbit/s amplified with an erbium-doped fibre amplifier and feeds a quasi-optic unitravelling carrier photodiode (UTC-PD) module, as shown in Fig. 4. This UTC-PD has a 0.2 A/W responsivity, and is integrated with a log-periodic antenna. First, the quasi-optic UTC-PD module was power calibrated in free space using an Erickson PM4. Using a 1 mA photocurrent, the modulated THz signal emitted power was found to be around 3 mW (−25 dBm) at 200 GHz and 1.5 mW with a 0.7 mA photocurrent. The THz path was composed of two polymer lenses. This THz path was preliminarily tested using a quasi-optical vector network analyser (VNA), leading to THz losses of ∼3 dB at 200 GHz frequency. However, it is worth mentioning that this value just estimates the losses as the first lens was fed by a waveguide horn antenna during the VNA experiment and was fed by the integrated silicon lens during the data transmission experiments.

Receiver description

The receiver is first composed of a WR5.1 waveguide corrugated horn (Fig. 4). The detected signal feeds a sub-harmonic mixer (SHM), pumped with a local oscillator at 100 GHz. This SHM exhibits 8 dB conversion losses at 200 GHz. The intermediate frequency (IF) signal is then amplified with a wideband amplifier (33 dB gain). When the optical beat note is controlled with the phase stabilisation system, the 200 GHz carrier is phase-locked to the reference synthesiser. At the receiver side, the mixing produced is centred at DC as long as the carrier frequency remains in phase with the local oscillator. Under these conditions, the baseband data are recovered, and a limiting amplifier is used before eye diagrams display or bit error ratio (BER) measurements.

Results

In the experiment, the coherent operation was obtained using the same reference to feed the comb generator and the receiver circuits. This enables one to address easily the real-time capability of the THz link, without off-line data processing. Indeed, we succeed to realise true BER testing using the combination of the MP1775A tester (Anritsu) for psuedorandom binary sequence (PRBS) generation and the MP1776A (Anritsu) for BER measurement with $2^{15}-1$ bit sequences, up to 11 Gbit/s. Up to 10 Gbit/s data rate, the slope of the BER curves is similar for the different tested data rates, as shown in Fig. 5. Moreover, an ~3 dB penalty was obtained when increasing the data rate from 5 to 10 Gbit/s. Indeed, an increase of the data rate by a factor of 2 while keeping the signal constant should decrease the signal-to-noise ratio by ~3 dB. The error-free operation (defined here for BER $<10^{-11}$) is obtained at 5 Gbit/s for 0.7 mA in the UTC-PD, corresponding to a 1.5 mW power at the UTC-PD output. This should correspond to ~0.75 mW <−30 dBm) received power. For 10 Gbit/s, the error-free operation was obtained with 1 mA photocurrent in the UTC-PD, corresponding to ~−28 dBm received power.

The BER curve obtained for the 11 Gbit/s data rate presented a reduced slope, and the achieved BER performance was limited to few 10^{-6}. We assume that this comes from the limited bandwidth available at the heterodyne receiver. To verify this assumption, a frequency response of the receiving mixer was carried out. In this experiment, the mixer was tested under waveguide conditions using a continuous- wave source coming from a VNA, and the frequency was tuned between 180 and 220 GHz. Fig. 6 presents the results. From this curve, a 20 GHz bandwidth +10 GHz around the carrier

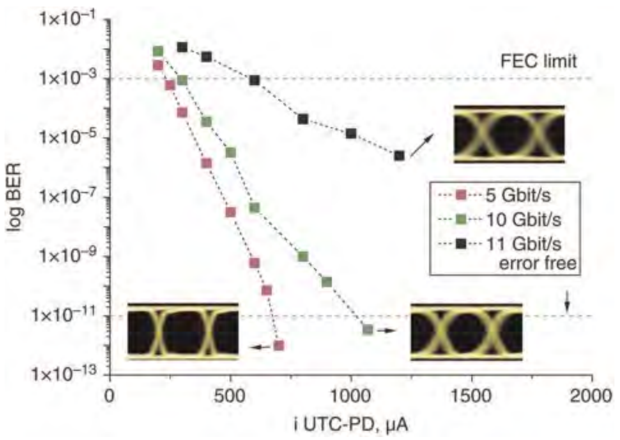


Figure 5 *BER characteristics, using 11 Gbit/s limiting amplifier*

Inset: Eye diagrams for highest photocurrents tested

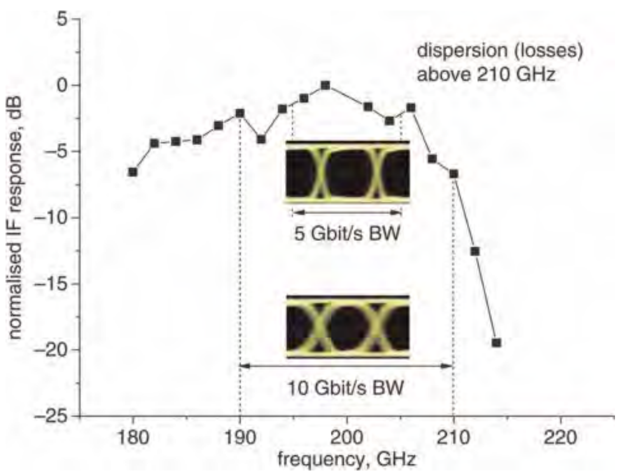


Figure 6 *Frequency response of heterodyne receiver including IF wideband amplifier*

frequency) enables an up to 10 Gbit/s error-free operation. For higher data rates, a very large amount of amplitude distortion affects the receiver, which is highlighted in Fig. 6 by an increase of the conversion losses (decreasing IF response). For that reason, the eye diagrams beyond 10 Gbit/s were degraded and the BER performance was limited. In our last experiment, an uncompressed high-definition television data stream (1.5 Gbit/s) was used to test the system with real applicative signals, and the stable operation (error-free video transmission) was obtained up to 0.6 mA photocurrent, corresponding to ~500 nW received power.

Conclusion

A coherent transmission system was demonstrated at 200 GHz, using a photonics-based emission, and featuring a very low power requirement at the reception. Error-free operation was obtained up to 10 Gbit/s data rates, with a limitation due to the receiver bandwidth, not by the comb or the UTC-PD transmitter. Next steps will concern high distance transmission using powerful photomixers at the emission. Moreover, in a future outdoor THz coherent scheme application, locking on the global positioning system (GPS) will lead to a common reference between Tx and Rx, avoiding the need for the common reference used here.

Acknowledgments

This work has been partially supported by the Agence Nationale de la Recherche and Japanese Science and Technology Agency within the framework of the 'WITH' ANR/JST Franco-Japanese project, and was partly supported by the Ministry of Higher Education and Research, Nord-Pas de Calais Regional Council and FEDER through the Contrat de Projets Etat Région 'Campus Intelligente Ambiante' (CPER-CIA) 2007–2013.

References

[1] FEDERICI J., MOELLER L.: 'Review of terahertz and subterahertz wireless communications', *J. Appl. Phys.*, 2010, **107**, p. 111101

[2] NAGATSUMA T., HORIGUCHI S., MINAMIKATA Y., YOSHIMIZU Y., HISATAKE S., KUWANO S., YOSHIMOTO N., TERADA J., TAKAHASHI H.: 'Terahertz wireless communications based on photonics technologies', *Opt. Express*, 2013, **21**, (20), p. 23736

[3] LI X., YU J., ZHANG J., DONG Z., LI F., CHI N.: 'A 400G optical wireless integration delivery system', *Opt. Express,* 2013, **21**, (16), p. 187894

[4] KOENIG S., LOPEZ-DIAZ D., ANTES J., BOES F., HENNEBERGER R., LEUTHER A., TESSMANN A., SCHMOGROW R., HILLERKUSS D., PALMER R., ZWICK T., KOOS C., FREUDE W., AMBACHER O., LEUTHOLD J., KALLFASS I.: 'Wireless sub-THz communication system with high data rate', *Nat. Photonics*, doi: 10.1038/NPHOTON.2013.275

[5] ANTES J., KOENIG S., LOPEZ-DIAZ D., BOES F., TESSMANN A., HENNEBERGER R., AMBACHER O., ZWICK T., KALLFASS I.: 'Transmission of an 8-PSK modulated 30 Gbit/s signal using an MMIC-based 240 GHz wireless link'. IMS Conf., Seattle, WA, USA, June 2013

[6] YOSHIMIZU Y., HISATAKE S., KUWANO S., TERADA J., YOSHIMOTO N., NAGATSUMA T.: 'Wireless transmission using coherent terahertz wave with phase stabilization', *IEICE Electron. Express*, 2013, **10**, (18), p. 20130578

[7] YOSHIMIZU Y., HISATAKE S., KUWANO S., TERADA J., YOSHIMOTO N., NAGATSUMA T.: 'Generation of coherent sub-terahertz carrier with phase stabilization for wireless communications', *J. Commun. Netw.*, 2013, **15**, (6), pp. 569–575

Selected content from *Electronic Letters*

ISSN 2041-5923

Minimising nonlinear Raman crosstalk in future network overlays on legacy passive optical networks

D. Piehler

NeoPhotonics, 2911 Zanker Road, San Jose, CA 95134, USA
E-mail: david.piehler@neophotonics.com

Abstract: There is a desire to overlay future optical access networks onto legacy passive optical networks (PONs) to provide increasingly advanced services by filling empty wavelength bands of legacy gigabit rate PONs. Nonlinear Raman crosstalk from new wavelengths onto legacy RF video services (1550–1560 nm band) *and* onto the legacy digital downstream at 1490 nm, however, can limit the number and the launch power of new wavelengths. As an example, straightforward physical-layer adjustments at the optical line terminal that increase the number of new, 10 Gbit/s channels launched in the 1575–1580 nm band by a factor of 16 without increasing the Raman penalty on the video signal are illustrated. A physical-layer (RF) filter modifies the on–off-keyed signal feeding each 10 Gbit/s transmitter, suppressing the RF Raman crosstalk on the video signal by 9 dB while incurring a power penalty on each 10 Gbit/s link of $<$0.5 (2.0) dB with (without) forward error correction. The previous Raman mitigation work used non-standard line-coding to shape the 10 Gbit/s electrical spectrum. In addition, polarisation-interleaving of new wavelengths lowers the worst-case RF DC crosstalk by $\sim$3 dB in fibres and it limits and stabilises DC crosstalk in low polarisation mode dispersion fibre links.

Introduction

Passive optical networks (PONs) with broadband optical power splitters assign discrete wavelength bands to various downstream and upstream services. In principle, unallocated wavelength bands can be utilised for overlaying future (higher-bandwidth) services on the legacy fibre plant without disrupting the legacy services. It is clear that the terminal equipment of both the legacy and future networks must provide a certain level of optical isolation between signals to limit any service degradation due to 'linear' optical crosstalk. This Letter outlines two steps that will limit the impact of 'nonlinear' optical crosstalk due to the nonlinear Raman effect when launching multiple wavelengths carrying new services into a legacy PON.

As an example, this Letter considers the nonlinear Raman limitations associated with launching multiple 10 Gbit/s non-return-to-zero (NRZ) modulated signals in the 1575–1580 nm band into a legacy gigabit rate PON co-propagating with a 1550–1560 nm RF video overlay and a 1.25 or 2.5 Gbit/s 1490 nm digital (NRZ) downstream signal.

The nonlinear Raman effect mediates the transfer of optical power from shorter to longer wavelengths (DC crosstalk), while the RF crosstalk flows in both directions. The effect is robust, requiring no phase matching and it can facilitate network-failing interference between wavelengths separated by up to 220 nm [1]. The most well-known impact of RF crosstalk in PONs is degradation of the carrier-to-noise ratio of a broadcast $\sim$1555 nm RF video signal ($\sim$17 dBm launch power) from a relatively weak ($\sim$3 dBm launch power) co-propagating 1490 nm digital signal [2]. The less well-known DC Raman crosstalk can cost the downstream 1490 nm signal one or more dB in optical power by the time it reaches a subscriber [3]. Since PON standards assume equal loss budgets for upstream and downstream, ignoring wavelength-dependent fibre loss, the 1490 nm downstream signal has an ~0.10 dB/km 'effective' fibre gain compared to the 1310 nm upstream. This more than compensates for any actual DC Raman loss.

The number and the total optical power of 10 Gbit/s signals are limited by *both* the DC crosstalk (Raman loss) which attenuates the legacy 1490 nm signal, and by the RF crosstalk onto legacy RF video channels. This Letter outlines two tools which when used together allow 16 10 Gbit/s wavelengths to be launched compared to one without using the tools, with a minimal impact on the 10 Gbit/s links.

The first tool, the PHY-filter, modifies the RF power spectral density of each 10 Gbit/s NRZ signals in the 50–300 MHz region where the RF Raman crosstalk is most severe. Earlier work by Colella *et al.* shows that spectral shaping of the interfering electrical spectrum by modification of its bit-patterns reduces RF Raman crosstalk [4]. In this Letter, the spectrum is shaped at the physical-layer without changing the bit-pattern, meaning that the existing bit-encoding and decoding standards and silicon can be used at the optical line terminal (OLT) and at the subscriber terminals. The second tool involves minimising polarisation-dependent Raman crosstalk by polarisation-interleaving the new wavelengths.

RF PHY-filter

The RF crosstalk on the video channel is proportional to a frequency-dependent walk-off parameter which depends on the fibre interaction length, attenuation, chromatic dispersion and separation ($D\lambda$) between the video wavelength and the interfering wavelength(s) [1, 2]. Since the optical power drops precipitously after the optical splitter, the interaction length is that of the optical trunk line (OTL) between the OLT and the splitter.

As illustrated in Fig. 1, the walk-off parameter is generally highest at lower frequencies. Thus, nonlinear Raman

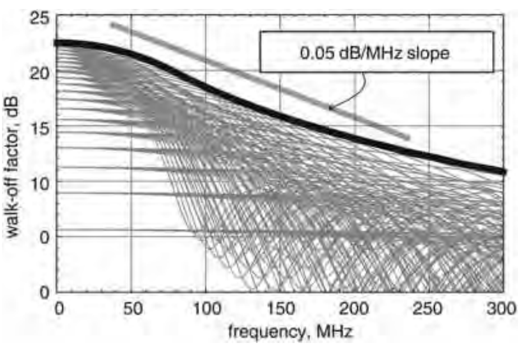


Figure 1 *Calculated walk-off factor [1, 2] against RF frequency for OTL distances from 1 to 20 km, and $D\lambda$ = 15 to 30 nm*

Solid black line is maximum value for all scenarios. (Assumptions: fibre dispersion = 18 ps/(nm km), attenuation = 0.2 dB/km)

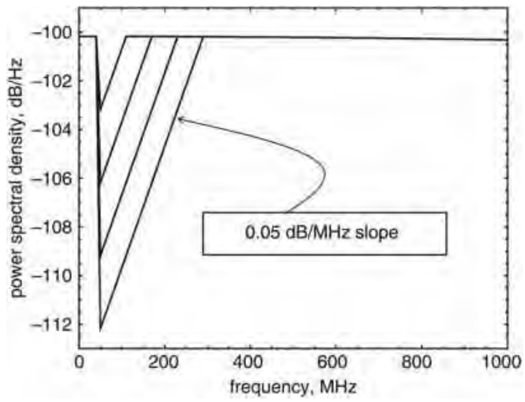


Figure 2 *Power spectral density of PHY-filtered 10-Gbit/s NRZ signal for filters of depth 3, 6, 9 and 12 dB and positive slope of 0.05 dB/MHz*

crosstalk impacts the low-frequency RF video channels which begin at ~50 MHz. The intention of the PHY-filter is to lower the power spectral density of the interferer's electrical spectrum in this region. In principle, this filter can be an RF filter placed between the transmitter's electronic driver and its laser or optical modulator. From Fig. 1, a triangular filter with a maximum attenuation at ~50 MHz followed by a +0.05 dB/MHz ramp will accommodate a wide variety of deployment scenarios.

Fig. 2 illustrates how various PHY-filter shapes modify the sinc-shaped RF spectrum of a 10 Gbit/s NRZ signal. As long as each of the new 10 Gbit/s transmitters are fed by uncorrelated bit streams, Raman crosstalk adds incoherently, meaning that a 9 dB drop in crosstalk from each transmitter allows their number to increase by a factor of 8 while keeping the total RF power transferred to the video signal constant.

Impact of PHY-filter on 10 Gbit/s link

For baseband systems in general, flat passbands and well-shaped high-frequency cutoffs are design objectives and irregularities in the frequency response are rarely studied [5, 6]. The impact from the triangular RF filters of Fig. 2 on a 10 Gbit/s, $2^{18}-1$ pseudorandom bit stream (PRBS) was numerically simulated by an application in the frequency-domain onto the Fourier transform of the NRZ PRBS. An inverse Fourier transform produced an output time-domain waveform. The PHY-filter has two main effects on the bit stream. The primary impact is that a filtered random bit stream contains a (Gaussian) noise waveform $S_{\text{filter}}(t)$ with a power spectral density in the same frequency band as the PHY-filter.

As a second-order effect, the post-filter waveform, having lost energy, exhibits overall amplitude compression. This

Table 1 Calculated optical power penalties at 10^{-3} and 10^{-12} BER for $2^{18}-1$ PRBS with various triangular filter depths

Filter depth (dB)	RF pass (%)	*s*	*d*	Penalty 10^{-3} BER (dB)	Penalty 10^{-12} BER (dB)
3	99.6	0.0130	0.9979	0.02	0.08
6	98.7	0.0247	0.9930	0.11	0.51
9	97.6	0.0397	0.9861	0.25	1.49
12	96.3	0.0546	0.9777	0.48	3.78

Letter defines the amplitude of the post-filtered signal as $d = \langle P_{one}\rangle - \langle P_{zero}\rangle$. ($P_{one} - P_{zero} \equiv 1$ for the pre-filter bit stream.) The measured values of d and the standard deviation, s, of the noise amplitude for the waveforms generated by the PHY-filters on the NRZ PRBS are listed in Table 1.

With this information, standard techniques are used to calculate the optical power penalty due to the PHY-filter. Assumptions on the optical transmitter and the receiver are: laser transmitter RIN, −140 dB/Hz; extinction ratio, 10 dB; receiver sensitivity, 0.9 A/W; thermal noise, 28 pA/$Hz^{\frac{1}{2}}$; avalanche photodiode (APD) multiplication factor, 15; APD ionisation ratio, 0.7; and electrical bandwidth, 6.7 GHz.

If forwarded error correction (FEC) is (is not) used to improve a 10^{-3} bit error ratio (BER) to 10^{-12}, 9 dB of nonlinear Raman crosstalk cancelation will induce a 0.25 (1.49) dB optical power penalty. Thus, 9 dB of (RF) Raman crosstalk reduction is possible with a modest power penalty on the interfering 10 Gbit/s signals especially if FEC is used.

Low-frequency filtering is known to impact the non-random bit streams [6]. The example PHY-filters avoided the <50 MHz region to minimise the impact of long consecutive-bit strings. The simulated 9 dB triangle filter generated the Gaussian noise waveforms when up to 30-digit sequences of all ones or zeros were randomly added to the PRBS.

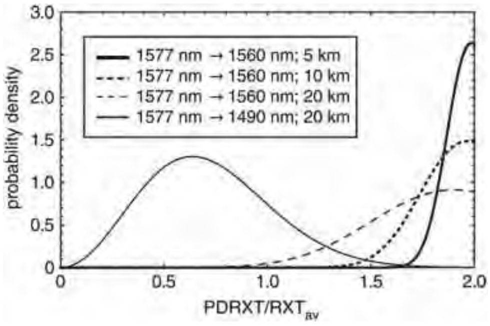


Figure 3 *Probability density for polarisation-dependent Raman crosstalk when ~1577 nm signals are launched with aligned polarisation states*

Although the PHY-filter may be realised as an actual RF filter, digital signal processing and an appropriate digital-to-analogue converter (DAC) can convert the PRBS into the post-filter waveform. A more compact and cost-effective realisation computes the PHY-filter's noise waveform, S_{filter} (t), in real-time from the PRBS, creates this analogue waveform with a 100 to 300 MHz DAC, and applies it to the unfiltered 10 Gbit/s NRZ bit stream as a (low-frequency) bias.

Polarisation-dependent RF Raman crosstalk

Nonlinear Raman crosstalk is highly polarisation-dependent, involving only parallel components of interacting lightwaves. Polarisation mode dispersion (PMD) scrambles relative polarisation states of signals at optical frequencies, v_1 and v_2 as they co-propagate along a fibre of length z and PMD D_p according to the autocorrelation $\langle \hat{s}(z, v_1)\, \hat{s}(z, v_2)\rangle = \exp\left[-D_p^2 z\,(\mathrm{D}\,v)^2/3\right]$ [7]. In a low-PMD fibre link ($D_p = 0.02$ ps/$km^{\frac{1}{2}}$), all signals in the 1575–1580 nm band will remain 95% polarisation-correlated for the first 27 km. If the 10 Gbit/s signals are launched with all polarisations aligned, the polarisation-dependent Raman crosstalk (PDRXT $= RXT_{max} - RXT_{min}$ [in dB], over all polarisations) can be calculated according to [8]. Fig. 3 illustrates the calculated statistical distributions of PDRXT for RF crosstalk from the ~1577 nm band onto a video signal at the (worst-case) wavelength of 1560 nm for various OTL lengths using a Raman gain coefficient, g_R, of 0.22/(W·km) and a total launch power (in the ~1577 nm band) of +20 dBm.

At all OTL lengths, the expected values of PDRXT/RXT_{av} are nearly 2, indicating a 'worst-case' polarisation RF Raman crosstalk that is twice the average crosstalk. ($RXT_{av} = [RXT_{max} + RXT_{min}]/2$ [in dB]). If, however, the signals in the ~1577 nm band are polarisation-interleaved and launched with alternating polarisation states, the expected PDRXT/RXT_{av} would be nearly zero, and the worst-case RF crosstalk would be the average Raman crosstalk. Thus, polarisation-interleaving enables the number of new 10 Gbit/s in the ~1577 nm band wavelengths to double (to 16), without changing the 'worst-case' RF crosstalk.

Polarisation-dependent DC Raman crosstalk

A simple calculation indicates that +20 dBm net power launched in the ~1577 nm band will cause a (polarisation-averaged) Raman loss of 1.8 dB in a co-propagating 1490 nm signal at the end of a 20 km OTL. ($g_R = 0.61/(\mathrm{W{\cdot}km})$). Fig. 3 shows that with polarisation-aligned ~1577 nm band signals, the 'expected' polarisation-dependent loss is 0.7, equivalent to 2.4 dB loss at the 'worst-case' polarisation. There is a small probability (4%) that Raman loss can be >3 dB and fluctuate over a 2.3 dB range as polarisation states evolve over time. If the ~1577 nm band signals are polarisation-interleaved, Raman loss is constant at 1.8 dB, which can be compensated by the 0.1 dB/km 'effective' fibre gain mentioned in the Introduction Section. The DC Raman crosstalk limits the net ~1577 nm band launch power to about +20 dBm. For low-PMD links, polarisation-interleaving may be required to minimise and stabilise the DC Raman crosstalk.

Conclusion

The described PHY-layer techniques applied at the PON OLT can mitigate RF and DC Raman crosstalk when new wavelengths overlay legacy PONs. Polarisation-interleaving may be necessary to manage nonlinear Raman crosstalk in low-PMD fibre plants.

References

[1] PIEHLER D., MEYRUEIX P., RISHI G., CHEN L.P., COPPINGER F.: 'Nonlinear Raman crosstalk in a 125-Mb/s CWDM overlay on a 1310-nm video access network'. Optical Fiber Communications Conf., Los Angeles, CA, USA, February 2004, paper FE8

[2] COPPINGER F., CHEN L.P., PIEHLER D.: 'Nonlinear Raman cross-talk in a video overlay passive optical network' Optical Fiber Communications Conf., Atlanta, CA, USA, March 2003, paper TuR5, doi: 10.1109/OFC.2003.1247660

[3] AGATA A., MURAKAMI A., IMAI K., HORIUCHI Y., EDAGAWA N.: 'Influence of stimulated Raman scattering in video overlay EPON systems on downstream data signal'. 31st European Conf. Optical Communications, Amsterdam, The Netherlands, September 2005, paper We.4.5.7, doi: 10.1049/cp:200504165

[4] COLELLA B., EFFENBERGER F.J., SHIMER C., TIAN F.: 'Raman crosstalk control in passive optical networks'. Optical Fiber Communications Conf., Anaheim, CA, USA, March 2006, paper NWD6, 10.1109/OFC.2006.215670

[5] YONEYAMA M., AKAZAWA Y.: 'Bit error rate degradation due to a dip in the frequency response of an equalizing amplifier', *J. Lightwave Technol.*, 1999, **17**, (10), pp. 1766–1773, doi: 10.1109/50.793748

[6] JORGESEN D., MARKI C.F., ESENER S.: 'Improved high pass filtering for passive optical networks', *IEEE Photonics Technol. Lett.*, 2010, **22**, (15), pp. 1144–1146, doi: 10.1109/LPT.2010.205114436

[7] KARLSSON M., BRENTEL J., ANDREKSON P.A.: 'Long-term measurement of PMD and polarization drift in installed fibers', *J. Lightwave Technol.*, 2000, **18**, (7), pp. 941–951, doi: 10.1109/JLT.2004.825330

[8] LIN Q., AGRAWAL G.P.: 'Statistics of polarization dependent gain in fiber-based Raman amplifiers', *Opt. Lett.*, 2003, **28**, (4), pp. 227–229, doi: 10.1364/OL.28.000227

Selected content from *Electronic Letters*

ISSN 2041-5923

Compact inductorless CMOS low-noise amplifier for reconfigurable radio

Abdelhalim Slimane[1] *Fayrouz Haddad*[2] *Sylvain Bourdel*[3]
Sid Ahmed Tedjini-Baïliche[1] *Mohand Tahar Belaroussi*[1]
Trabelsi Mohamed[4] *Hervé Barthélemy*[2]

[1]*Centre de Développement des Technologies Avancées, Cité du 20 Août 1956, Baba Hassen, Algiers, Algeria*
[2]*IM2NP Aix-Marseille University, Technopôle Château-Gombert, Marseille, France*
[3]*IMEP-LAHC, Université de Grenoble, Minatec, 38000 Grenoble, France*
[4]*Ecole Nationale Polytechnique d'Alger, Avenue Hacen Badi – EL Harrach, Algiers, Algeria*
E-mail: aslimane@cdta.dz

Abstract: A compact reconfigurable CMOS low-noise amplifier (LNA) is presented for applications in DCS1800, UMTS, WLAN-b/g and Bluetooth standards. The proposed LNA features first a current reuse shunt-feedback amplifier for wideband input matching, low-noise figure and small area. Secondly, a cascode amplifier with a tunable active LC resonator is added for high gain and continuous tuning of bands. Fabricated in a 0.13 mm CMOS process, the measured results show $>$20 dB power gain, $<$3.5 dB noise figure in the frequency range of 1.8–2.4 GHz, return losses S_{11} and S_{22} lower than -12 and -14 dB, respectively, with a moderate IIP3 of -11.8 dBm at 2.4 GHz. It consumes 9.6 mW from a 1.2 V supply voltage, while occupying an active silicon area of only 0.052 mm^2.

Introduction

Spectacular recent developments in wireless technology are pushing the boundaries of communication standards and placing higher requirements on next generation wireless systems. Nowadays, RF front ends based on parallel narrowband paths, the single wideband path and the single multi-band path are the only ways to implement multi-standard architectures [1–3]. However, design specifications for such architectures increase the integration issues of radio sections, leading to high costs and high power consumption. Therefore, the design of reconfigurable radios, based on maximum sharing of RF building blocks and a minimum of passive inductors, is expected to deal with multi-band/multi-standard transceivers regarding cost and energy saving. In this context, recent works are still focusing on low-noise amplifier (LNA) reconfigurability to meet multi-standard requirements. In [2–4], the designed LNAs employed either discrete or continuous tuning to cover different standards, whereas passive inductors caused a large silicon area. In contrast, inductorless LNAs are also investigated for efficient silicon area purposes. However, their drawbacks are either poor performance in multi-band LNAs [5] or the selectivity issue in wideband LNAs.

In this Letter, a compact reconfigurable CMOS LNA is proposed for a single-path multi-band front-end using continuous tuning for low sensitivity to process variations. The following Section describes the LNA circuit. Measurement results and comparisons are then provided.

Reconfigurable LNA design

Fig. 1 illustrates the proposed reconfigurable LNA that satisfies all the above-specified standards. On the basis of the two cascaded amplifiers for high gain purposes, the first stage offers a good trade-off between wideband matching and low-noise figure, whereas the second is more selective by including a tunable active inductor load.

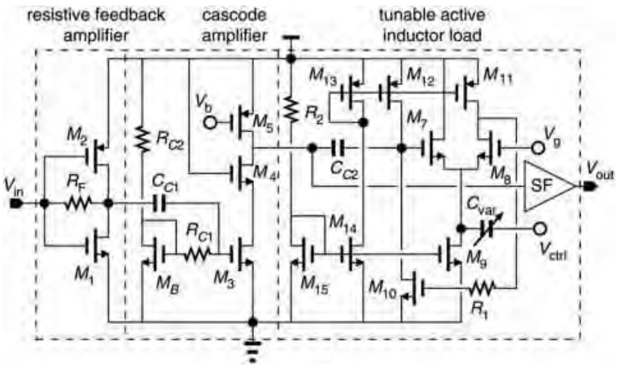


Figure 1 *Schematic of proposed reconfigurable multi-standard LNA*

Shunt-feedback current reuse amplifier

The shunt-feedback amplifier is a topology that offers wideband input matching with the advantages of design simplicity, relatively low-noise figure and small die area compared to those based on passive inductors. The amplifier analysis in terms of input impedance (Z_{in}), gain (A_v) and noise factor (F) is given as follows:

$$Z_{in} = \frac{R_F + R_L}{1 + g_m R_L} \tag{1}$$

$$A_v = \frac{R_L(1 - g_m R_F)}{R_F + R_L} \tag{2}$$

$$F \simeq 1 + \frac{1}{R_s R_F}\left(\frac{1}{g_m} + R_s\right)^2 + \cdots$$
$$\cdots + \frac{1}{R_s}\left(1 + \frac{R_s}{R_F}\right)^2\left(\frac{1}{g_m^2 R_L} + \frac{g}{a g_m} + \frac{da}{5 g_m}\right) \tag{3}$$

where parameters g and d depend on the process technology and a is a transistor ratio of transconductance to conductance.

From (2) and (3), maximising the transconductance g_m and the load resistance R_L is necessary to achieve high gain and a low-noise figure for the first stage according to the Friis formula. Therefore, a current reuse topology is used instead of a conventional shunt-feedback to increase the equivalent g_m to $g_{m_{M1}} + g_{m_{M2}}$ and the equivalent R_L to $R_{ds_{M1}} || R_{ds_{M2}}$ for a required DC current. As a result, the noise factor becomes completely dependent on g_m, and R_F. Furthermore, the design constraints are relaxed by solving (1) for the input return loss $S_{11} < -10$ dB, which leads to choosing R_F for a good trade-off between high g_m and a low-noise figure.

Cascode amplifier with active inductor load

As shown in Fig. 1, to provide high gain and good output/input isolation, the second amplifier uses a cascode topology formed by two stacked transistors M_3 and M_4 and supplied through transistor M_5. The cascode amplifier is loaded by the tunable active inductor based on a gyrator-C network. It consists of two connected back-to-back transconductors g_{mA} implemented by the differential pair (M_7, M_8) and g_{mB} is implemented by the common source amplifier M_{10}. To increase the quality factor, a resistance R_1 is inserted between the two transconductors. Thus, the equivalent inductance and series resistance of the active inductor are approximated as follows [6]:

$$L \simeq \frac{2}{g_{mA} g_{mB}}\left(C_{gs_{10}}\left(1 + \frac{R_1}{r_{oA}}\right) + C_{oA}\right) \tag{4}$$

$$R_s \simeq \frac{2}{g_{mA} g_{mB}}\left(\frac{1}{r_{oA}} - v^2 C_{oA} C_{gs_{10}} R_1\right) \tag{5}$$

where C_{gs10} is the gate–source capacitance of the transistor M_{10}, r_{oA} and C_{oA} are, respectively, the output resistance and capacitance of the differential pair transconductor. The transistor M_{15} and the resistor R_2 are used to bias the active inductor, whereas M_{12} and M_9 mirror a ratio of the reference current in M_{10} and the differential pair (M_7, M_8), respectively. A controlled varactor C_{var} is also added at the source of the differential pair (M_7, M_8) to tune the quality factor against frequency.

Experimental results and discussion

In addition to the previous simulated results in [7], the LNA prototype is fabricated using 0.13 mm CMOS technology. The micrograph of the die is shown in Fig. 2, where the total die area including pads and decoupling capacitors is 0.165 mm^2. However, the active area including decoupling capacitors is 0.052 mm^2 (250 × 210 mm). The fabricated LNA includes a source follower as an output buffer to drive the 50 V input impedance of the network analyser.

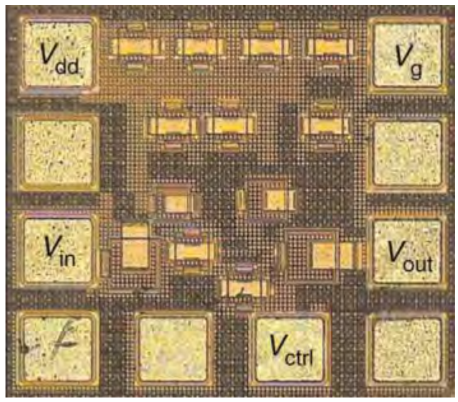


Figure 2 *Micrograph of fabricated multi-standard LNA*

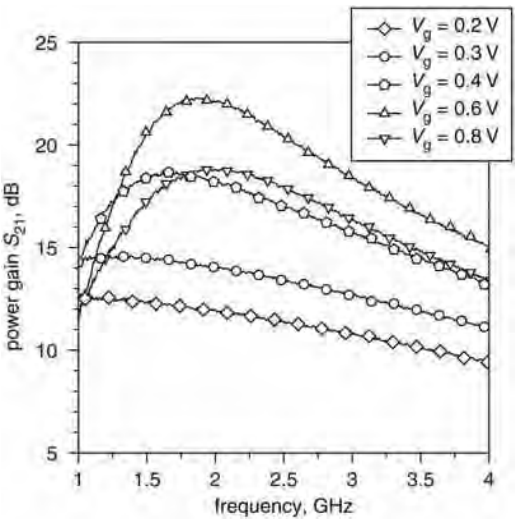


Figure 3 *Measured power gain for different values of V_g*

Fig. 3 depicts the applied V_g effect on gyrator efficiency. For $V_g = 0.2$ V, a flat power gain is observed indicating that the gyrator is still off. However, by increasing gradually V_g, the power gain increases to reach its maximum at $V_g = 0.6$ V, meaning that the gyrator achieves its high quality factor at this voltage. Then, for a fixed $V_g = 0.6$ V as shown in Fig. 4*a*, the power gain is tuned from 1.8 to 2.4 GHz by a continuous tuning of the varactor's voltage (V_{ctrl}), where more than 22 dB is performed at 1.8 GHz for $V_{ctrl} = 0$ V. However, when V_{ctrl} increases, the inductor's quality factor decreases, leading to a lower power gain of 20.6 dB at 2.4 GHz for $V_{ctrl} = 1.2$ V. Owing to the impact of the first stage on noise contributions, the measured NF is not affected by the V_{ctrl} variations where it slightly varies between 3.2 and 3.5 dB in the frequency range of 1.8–2.4 GHz as shown in Fig. 4*b*.

In Fig. 4*c*, the current reuse stage, in charge of input matching, achieves a good S_{11} lower than -12 dB for all bands of interest. A good output matching is also ensured by a low $S_{22} < -14$ dB. As a result, the continuous tuning by V_{ctrl} increases the robustness of the LNA against process variations and allows an easy frequency shift to the target standards. By including the buffer nonlinearities, the IIP3 is also measured across the band, showing moderate values from -16 to -11.8 dBm as given in Fig. 4*d*. From a supply voltage of 1.2 V, the LNA power consumption is about 13.8 mW including the buffer and 9.6 mW without it.

Operating as narrowband LNAs at one frequency f_0 at a time, the LNA figure-of-merit (FoM) is given as follows [8]:

$$\mathrm{FoM[GHz]} = \frac{\mathrm{Power\ Gain[abs]} \cdot \mathrm{IIP3[mW]} \cdot f_0\mathrm{[GHz]}}{(F-1)\mathrm{[abs]} \cdot P_{DC}\mathrm{[mW]}} \tag{6}$$

In Table 1, a performance summary of the recent published

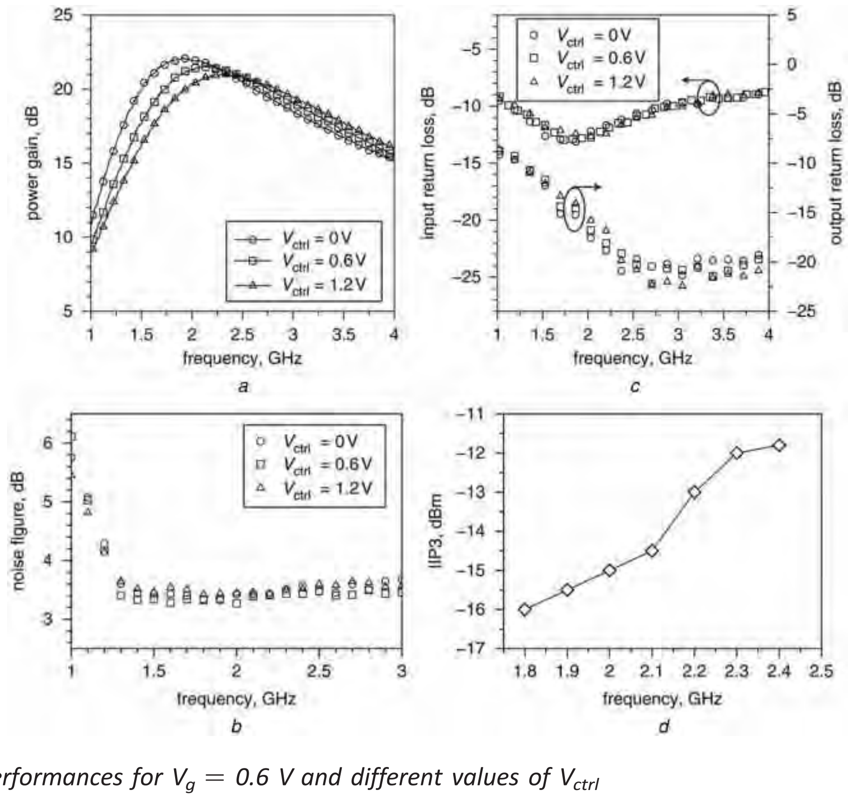


Figure 4 *Measured performances for $V_g = 0.6$ V and different values of V_{ctrl}*

a Power gain
b Noise figure
c S_{11} and S_{22}
d IIP3

Table 1 Performance summary and comparison with state-of-the-art reconfigurable CMOS LNAs

Reference	[1]	[3]	[4]	[5]	This work
Technology (mm)	0.13	0.13	0.13	0.18	0.13
Tuning	Discrete	Discrete	Cont.	Cont.	Cont.
Passive inductors	Yes	Yes	Yes	No	No
Frequency (GHz)	1.8/2.1/2.4	2.4–5.4	1.9–2.4	0.75–3	1.8–2.4
S_{21} (dB)	23–29[a]	22–24	10–14[b]	24–28	20.6–22.1
NF (dB)	5.2–5.8	2.2–3.1	3.2–3.7	4.8–10	3.2–3.5
III3 (dBm)	−7.5/0/−4.8	−16– −21	−6.7	−24	−16–11.8[c]
V_{dd} (V)	1.2	1	1.2	1.8	1.2
P_{DC} (mW)	24	4.6	17	42	9.6
Area (mm^2)	0.6	0.49	0.083	0.09	0.052
FoM (GHz)[d]	2.58	1.41	0.22	0.005	1.59
FoM/area	4.3	2.88	2.7	0.05	30.6

[a]: LNA + mixer gain; [b]: Gain voltage; [c]: Including buffer nonlinearities; [d]: FoM at $f_0 = 2.4$ GHz

multi-standard CMOS LNAs is given for comparison. The FoM of the proposed LNA is as good as the other reported FoMs. Besides, the relevant parameter of performance and cost-saving design is the ratio of FoM to chip area [9]. Occupying an active silicon area of only 0.052 mm^2, the proposed LNA demonstrates a better ratio of FoM to chip area with a good robustness against the process variations by means of continuous tuning.

Conclusion

In this Letter, a reconfigurable CMOS LNA is implemented in an active silicon area of only 0.052 mm^2. On the basis of cascaded amplifiers and an active LC load, the fabricated LNA performs more than 20 dB of gain and <3.5 dB of noise figure, while consuming 9.6 mW from a 1.2 V supply voltage in the frequency range of 1.8–2.4 GHz. One can conclude that the proposed LNA could be a good candidate to address DCS1800, UMTS, WLAN-b/g and Bluetooth standards for low power and low cost.

References

[1] LISCIDINI A., BRANDOLINI M., SANZOGNI D., *ET AL.*: 'A 0.13 mm CMOS front-end for DCS1800/UMTS/802.11b–g with multiband positive feedback low-noise amplifier', *IEEE J. Solid-State Circuits*, 2006, **41**, pp. 981–989

[2] PARK J., KIM S.-N., ROH Y.-S., *ET AL.*: 'A direct-conversion CMOS RF receiver reconfigurable from 2 to 6 GHz', *IEEE Trans. Microw. Theory Tech.*, 2010, **58**, pp. 2326–2333

[3] FU C.-T., KO C.-L., KUO C.-N., *ET AL.*: 'A 2.4–5.4 GHz wide tuning-range CMOS reconfigurable low-noise amplifier', *IEEE Trans. Microw. Theory Tech.*, 2008, **56**, pp. 2754–2763

[4] EL-NOZAHI M., SANCHEZ-SINENCIO E., *ET AL.*: 'A CMOS low-noise amplifier with reconfigurable input matching network', *IEEE Trans. Microw. Theory Tech.*, 2009, **57**, pp. 1054–1062

[5] ANDERSSON S., SVENSSON C.: 'A 750 MHz to 3 GHz tunable narrowband low-noise amplifier'. 23rd NORCHIP Conf., 2005, pp. 8–11

[6] ABDALLA M.A.Y., PHANG K., ELEFTHERIADES G.V.: 'Printed and integrated CMOS positive/negative refractive-index phase shifters using tunable active inductors', *IEEE Trans. Microw. Theory Tech.*, 2007, **55**, pp. 1611–1623

[7] SLIMANE A., BELAROUSSI M.T., HADDAD F., *ET AL.*: 'A reconfigurable inductor-less CMOS low noise amplifier for multi-standard applications'. IEEE 10th Int. New Circuits and Systems Conf. (NEWCAS), 2012, pp. 57–60

[8] BREDERLOW R., WEBER W., SAUERER J., *ET AL.*: 'A mixed-signal design roadmap', *IEEE Des. Test Comput.*, 2001, **18**, pp. 34–46

[9] CHEN H.-K., CHANG D.-C., JUANG Y.-Z., *ET AL.*: 'A compact wideband CMOS low-noise amplifier using shunt resistive-feedback and series inductive-peaking techniques', *IEEE Microw. Wirel. Compon. Lett.*, 2007, **17**, pp. 616–618

Communities
The Institution of Engineering and Technology

About the Network

The **Multimedia Communications Network** is a free to join global community bringing together broadcasting, communication and technology professionals interested in electronics and creating dynamic media for different devices in multiple locations.

The network enables and supports transformational behaviour through technology evolution, and revolution across government and public services; business and commerce; education and entertainment.

If you work or have an interest in broadcasting or telecommunications then come and join our Community.

Supporting sector:

Information & Communications

www.theiet.org/multimedia