Vol 4 | ISSN 2041-5923
2012

IET Journals
The Institution of Engineering and Technology

The Best of IET and IBC

INSIDE Papers and articles on electronic media technology from IBC2012 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

VOLUME 4 2012

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*, 2012. This is the fourth volume[1] 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 and the IET's flagship journal, *Electronics Letters*.

In this volume, with IBC celebrating 20 years since moving to the RAI in Amsterdam, we start by taking a look back at the first IBC held there in 1992 in our editorial on page 3. What has happened to some of the exciting technologies grabbing people's imagination in 1992? How have they fared and what might the next two decades hold for today's hot topics?

Next up we have the seven papers chosen as the best contributions to IBC2012 by the IBC technical papers committee in conjunction with the executive team of the IET Multimedia Communications Network executive team. This includes overall winner of best in conference for IBC2012, 'Fully automatic stereo-to-multiview conversion in autostereoscopic displays'. The papers are given a proper introduction on page 6.

We include an interview with Kazuya Kitamura, winner of IBC2012 best young professional contribution as lead author of 'Development of 33 megapixel 120 Hz CMOS image sensor and experimental colour camera system', which is also published in this volume.

From *Electronics Letters* this year we have three papers chosen from those published since IBC2011; this time each is accompanied by a supplementary article, a feature introduced in *Electronics Letters* to provide readers with more background and behind-the-scenes information on some of the research published in each issue.

It is encouraging that, despite the world's current economic difficulties, ingenuity, innovation and investment in media engineering continue undiminished, and we remain delighted to be able to publish papers of a high standard which manage to combine technical brilliance, real-world relevance and crystal clarity in a way which also makes them entertaining to read.

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 and extend our best wishes for a successful and stimulating IBC2012.

Dr Nicolas Lodge
Chairman IBC Technical Papers Committee
&
The IET Multimedia Communications Network Executive Team

[1]For previous volumes see www.theiet.org/ibc

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

1992: broadcasting at the analogue-digital crossroads

If you were researching the history of broadcasting technology over the last half-century, your first port of call would almost certainly be the Proceedings of IBC's conferences. Since 1967, when it was founded, IBC has been the focus for discussion and debate of the latest ideas and technologies in broadcast media, and the Proceedings provide a complete record of this progress. Not only is it possible to trace the development of many technologies and services that have now become commonplace, but it is also fascinating to cast a retrospective eye over those ideas which simply failed to make it in the real world. Succession technologies, market competition, diverse strategies and international politics have all played out on the conference platforms over the years and have all shaped the ideas and opinions of the session audiences.

IBC's growth, assisted by the growth of its exhibition, was so rapid that it outstripped several venues before moving to take up its current home at Amsterdam's RAI Centre 20 years ago. IBC was then a biennial event and, unusually, the conference that year was held in July because the usual September slot was unavailable.

So what were the major broadcast technology issues in 1992?

A look back at the Proceedings reveals that this was arguably IBC's most significant year. Much of the world had been desperately trying to enhance its tired and traditional analogue television systems through a variety of extensions, in such a way that the new signals were compatible with the existing receivers. They all aimed at wide 16:9 pictures (an important potential new market for set manufacturers) and they all promised sharper resolution by using digital signal processing (DSP) to filter and fold multidimensional image spectra. But television transmission remained a fundamentally analogue affair.

Not everyone was convinced that enhanced-analogue was the way to go. Experimental all-digital systems such as SPECTRE, STERNE, HD-DIVINE and DigiCipher were presented and demonstrated at IBC1992 and some nations, notably the US, believed that imminent digital technology would obviate the need for any intermediate enhanced-analogue market. But digital terrestrial television still had huge unknowns and had not even been properly trialled. It is interesting to note that in 1992 it took a large R&D team nearly 2 years just to design and build a prototype video codec and modem using banks of DSP chips.

The conference keynote addresses reflected the decisive crossroads that broadcasting faced: '*Enhanced Widescreen in Europe – Why?*', '*16:9 - Format of the Future*', '*PALplus – Terrestrial Widescreen on Track*' and '*Digital Television Broadcasting – Dreams, Decisions and Destinies*'.

Looking immediately after IBC1992 we see the world scramble for all-digital TV really taking-off and the focus was on terrestrial transmission with its challenging multipath propagation characteristics (digital cable and satellite transmission being much better understood). In the US the communications regulator had already organised a competition between four candidate all-digital systems and its experts were busy assessing the results. A year later the so-called Grand Alliance was formed to 'cherry-pick' the best ideas from these candidates. This led to the standardisation of the US advanced television system in 1995.

In Europe, the European Commission supported a major research collaboration which brought together several individual research projects to create a flexible terrestrial system based on multi-carrier COFDM technology. This drew upon working relationships established through earlier collaborations in HDTV, digital audio and PAL enhancement, and was the technical seed from which the DVB forum emerged in 1993. DVB completed technical development of its digital satellite and cable standards within a year, and the more complex DVB-T terrestrial specification followed in 1995.

It was at IBC1992 that we find the first mention of the Japanese ISDB initiative, in a historical paper that is still widely cited today. At the time ISDB defined a digital satellite broadcasting system but the acronym was later to encompass a range of delivery media including ISDB-T, the Japanese COFDM-based terrestrial standard, ratified in 1998.

But what became of enhanced TV in Europe?

From the middle of the 1980s European researchers had been developing an enhancement to the D2-MAC analogue satellite standard which would compatibly carry HDTV. The project employed hundreds of engineers from every corner of Europe and was called HD-MAC. Its technology involved sophisticated multi-dimensional signal processing and motion-compensated interpolation and was Europe's answer to Japan's analogue HDTV satellite technology, MUSE.

By 1992 HD-MAC was mature, technically successful and enjoying its biggest year. Like 2012, it was an Olympics year, and HD-MAC broadcasts were made from both the winter games in Albertville and the summer games in Barcelona. For the summer games, 700 HD-MAC receivers were positioned in public venues across Europe and it is estimated that 350,000 people saw the pictures. Demonstrations were also made at IBC'92 and seven conference papers addressed HD-MAC technology.

Above *Before RAI, the IBC beach circa 1982*

HD-MAC's biggest year was also to be its dying gasp, however. The European standard D2-MAC, on which it was based, failed to thrive commercially because satellite operators chose to broadcast with cheaper PAL in bands where D2-MAC was not mandated. HD-MAC was abandoned in 1993.

Japan's MUSE system fared better. Just eight months before IBC1992 it was used for the world's first regular HDTV service. Remarkably, MUSE broadcasts continued until 2007.

Perhaps Europe should have learned its lesson about compatible enhanced-analogue television, but even before HD-MAC was completed, it had embarked on another major programme, this time based on PAL. Again: a huge collaborative effort, a big investment, much signal processing and some impressive pictures. IBC1992 featured an entire conference session devoted to PALplus and it was the occasion when PALplus prototype hardware was shown for the first time. The first regular broadcasts were on German pay TV in December 1994 and the system was officially launched the following year.

Although broadcasters in nine countries adopted PALplus, few programmes were actually transmitted in it and receiver take-up was disappointing. In the late 90s, widescreen pictures were often transmitted (without PALplus) and were watched in 'letterbox' format on 4:3 receivers. Gradually, viewers came to accept this and many bought widescreen sets which simply expanded the letterbox to fill the screen. So a modest market in widescreen receivers did emerge while Europe waited for digital television, but it was not driven by analogue enhancements.

What other new technologies appeared at IBC1992?

In 1992 broadcasting researchers were especially busy developing video compression technologies. It had only recently been practical to carry out an 8×8 discrete cosine transform (DCT) in hardware at video rates, and implementing motion-compensated prediction still proved quite a challenge. The ISO/IEC MPEG group had been formed some years earlier but it would not complete the specification of MPEG-2 until 1995. Given this, it is remarkable that several ambitious HD codecs using hybrid DCT algorithms were reported and demonstrated at IBC1992.

High definition standards conversion and slow motion processing were other active areas of image research. Motion-compensated interpolation was in its infancy and handling interlaced pictures was always tricky. IBC1992 offered a session on HD standards conversion where researchers presented their latest algorithms.

Digital Audio Broadcasting, as we know it today, began as a European collaboration in 1987. In 1990 the audio compression, modulation and error correction were chosen and the first trials were conducted. By 1992 the specification was almost complete and public trials were being planned. In IBC's propagation session, progress on service planning for the COFDM-based services was reported. These studies were to attain increased significance as COFDM would also be chosen for DVB-T transmission.

Above *Up to date, two decades at RAI*

Stereoscopic television is familiar today but it is exciting to discover that at IBC1992, Japan's NHK discussed its 3D *Hi-Vision* system. Their paper is remarkable for the psychological studies it reported on such aspects as preferred viewing distance, screen size, aspect ratio and perceived naturalness. Curiously, it has strong echoes of NHK's award-winning paper on psychological aspects of *Super Hi-Vision* at IBC in 2011.

Audio description for people with visual impairment is also a familiar service now, but it was at IBC1992 that its first European initiative, AUDETEL, was unveiled. In 1994 the consortium would carry out the first national trial of described television to 140 special receivers using a digital speech commentary conveyed through teletext data. AUDETEL also explored the art of audio description and established practices which are still used today.

For those of us lucky enough to have been young technologists in 1992 and to have attended the first IBC in Amsterdam, it was a very exciting time at the beginning of a pioneering digital decade. But in 20 year's time, will today's young technologists regard 2012 as an exciting crossroads? Very likely. A new 'internet' generation has since grown up with social media and targeted advertising, and these ideas now threaten to overturn traditional broadcasting models. Surely, a revolution as significant as the coming of digital TV ever was!

Nicolas Lodge and Martin Russ

ISSN 2041-5923

Introducing the winners

The winner of this year's best overall paper, by Christian Reichert and his colleagues, is entitled 'Fully automatic stereo-to-multiview conversion in autostereoscopic displays'. Dealing with the commercially important area of glasses-free 3D displays, it discusses in detail the multiple-stage image processing required to produce the complex interwoven driving signals they require. The real-time prototype is described and its performance is both demonstrated and evaluated.

A novel perspective and forward-thinking ideas are discussed in 'Project Fresco – tileable technology for television's big future in the home'. Fitting very large screens into normal homes is a problem that has not faced most of us yet, but here is presented a potential solution, which involves tiling the wall space available with display tiles. Furthermore, the images shown on them can be multiple and of differing sizes – and they don't even need to be rectangular.

The judges were also especially impressed by the paper 'Asset values increased – new technologies for the automatic transfer and digital preservation of analogue video-cassettes'. In a novel approach, the raw analogue RF signals stored on old magnetic videotapes are digitised and stored, allowing sophisticated software algorithms to recover from them the best pictures possible.

A fascinating analysis, which hints at the future influence of social media on broadcasting, is given in 'Likes, tweets and diggs: the impact of social media on viewing behaviour'. This is a key strategic paper which surveys and analyses the social behaviour of television viewers at different times and across different programme genres. A thought-provoking read.

How well do we really know the performance of video delivery through the internet? 'Impact of delivery ecosystem variability and diversity on internet video quality' looks at the variability of a host of internet system components such as: encoder formats and profiles, CDNs, ISPs, streaming protocols and video players. It reports on how this variability affects video quality and makes recommendations for maximising it.

'ISDB-Tmm mobile multimedia broadcasting in Japan' gives a big picture look with a thorough overview of the design and implementation of this new service, which carries both real-time and file-based media. Launched in April this year, one day after analogue shutdown had vacated the spectrum, the transmission technologies are discussed together with their performance in field trials.

Another example of ambitious and exciting engineering is 'New equipment for 22.2 multichannel sound production and reproduction'. This paper describes several facets of the 3D audio technologies which have been designed and built to accompany next-generation ultra-high definition television.

Difficult choices

The process of selecting the 'Best of' always leaves some excellent submissions that, owing to the need for a finite list, do not make it into the final selection. So we pause here to mention three such papers from this year's conference. They are highly recommended reads for anyone wanting to learn more about some of the multimedia topics represented at IBC2012:

(IET)'iTapeless – file-based acquisition the smart phone way'
S. R. Knee
Pressing equipment intended for consumer use into professional use often leads to fascinating applications. This paper shows that not only is BYOD (bring your own device) thinking not restricted to the office environment, it may provide a glimpse of one possible future for outside broadcast.

'Depth-based content creation targeting steroscopic and auto-stereoscopic displays'
F. Zilly, N. M. Gutberlet, C. Riechert, R. Tanger and P. Kauff
Covering three camera and multi-camera approaches, with clear illustrations and explanations' this work provides an excellent introduction to an alternative approach for 3D visuals. If you thought that two cameras were all you needed for 3D...

'Immersive 3D-stereo surround cinema'
S. Peikert, J. Gerhardt and M. Schiewe
3D on flat or slightly curved screens is not the only way to do it. Projecting onto the inside of domes can produce even more immersive 3D surround cinema. This work provides a wide-ranging view of the practical challenges involved in realising a dome cinema system.

Dr Nicolas Lodge
Chairman IBC Technical Papers Committee
&
Martin Russ
Chairman of The IET Multimedia Communications Network Executive Board

Selected content from IBC2012. Winner of the Best Paper award

ISSN 2041-5923

Fully automatic stereo-to-multiview conversion in autostereoscopic displays

C. Riechert F. Zilly P. Kauff J. Güther R. Schäfer

Image Processing Department, Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Einsteinufer 37, Berlin 10587, Germany

E-mail: christian.riechert@hhi.fraunhofer.de

Abstract: Watching 3D without glasses will be the future of 3D home entertainment. However, current stereoscopic 3D content is not suited to drive autostereoscopic displays and a 3D format conversion is required. In this paper a system is presented which is capable of high-quality stereo-to-multiview conversion in real time. This allows playback of 3D Blu-ray or any other stereoscopic 3D video content without prior adaptation on most existing autostereoscopic displays in real time. The algorithms described include image rectification, fast and reliable disparity estimation, adaptive post-processing and up-sampling of disparity maps, rendering of virtual views and interweaving these views into a format suitable for autostereoscopic displays.

1 Introduction

Currently most 3D content is only available as stereo productions and so far storage and transport media have only been standardised for stereo video. On the other hand, autostereoscopic displays which appear on the market require more than two views, in general five to ten or even more. Therefore fully automatic stereo-to-multiview converters at the consumer side are required to display available stereo content on autostereoscopic displays.

In this paper, a system is described which is capable of high-quality stereo-to-multiview conversion. It has been developed at Fraunhofer Heinrich Hertz Institute and is running in real time as a software solution on a high-performance computer. The conversion algorithm itself consists of five main parts:

- a preprocessing part which analyses and rectifies the incoming stereo pair in order to ensure that only horizontal disparities exist between left and right image.
- a very fast disparity estimator, which generates temporally and spatially consistent disparity maps
- an adaptive multi-step filter unit for post-processing and up-sampling of the disparity maps to full resolution
- depth image based rendering (DIBR) unit for rendering an arbitrary number of virtual camera views as required by a specific autostereoscopic device
- a post-processing and interweaving unit for filling disocclusions in the rendered views and merging all views to an output image suitable for the output device

The algorithms used in this solution were specifically designed to be suitable to be ported onto hardware platforms such as FPGAs or SoCs.

2 Related work

Fully automatic stereo-to-multiview conversion is an extremely challenging field, since rectification as well as disparity estimation is highly ambiguous, and image-based rendering is very sensible to errors in the underlying disparity maps. Therefore very few solutions exist so far. In Berretty *et al.* [1] a video plus depth (V + D) approach is described, which relies on the ability of the target autostereoscopic screen to do the actual view rendering on its own and thus only provides a disparity map. In Lang

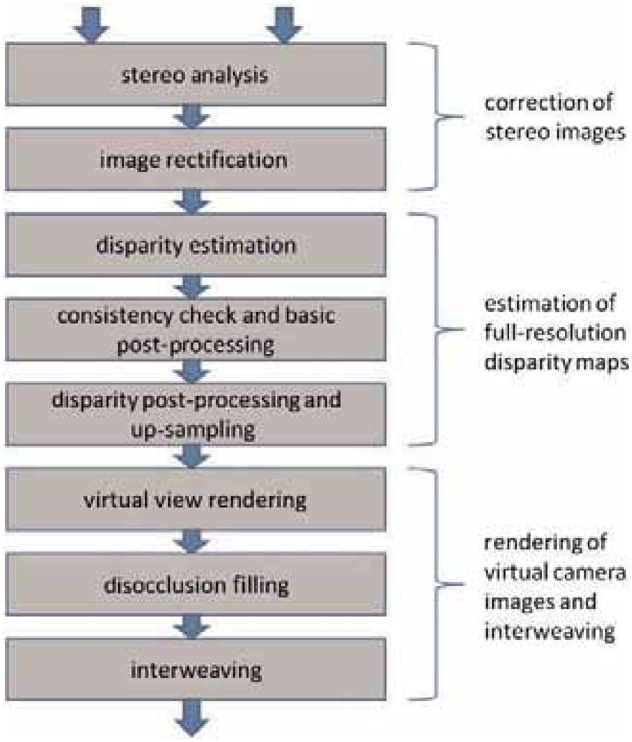


Figure 1 *Diagram of complete stereo-to-multiview conversion chain*

et al. [2] a different approach is presented, which is based on image-based saliency estimates and stereoscopic warping, instead of dense disparity map estimation. Both approaches have the common problem that the depth impression on an autostereoscopic display is rather limited, because V + D as well as saliency maps only allow little extrapolation without image degradation.

In this paper a complete processing chain for stereo-to-multiview conversion including rectification, dense disparity estimation and multiview depth (MVD) rendering as illustrated in Fig. 1 is presented. The concept of image rectification in order to align epipolar lines of stereo image pairs is well known in the literature (Hartley *et al.* [3]). The estimation of rectification parameters in real time and without prior knowledge of the relevant camera parameters, on the other hand, is addressed properly only recently by systems like the stereoscopic analyser (STAN) (Zilly *et al.* [4, 5]).

With rectified images, a disparity between the input image pair can be estimated. There already exist a wide variety of stereo matching algorithms. A comprehensive overview is, for example, provided by Scharstein *et al.* [6]. Most of these algorithms have in common that they are either too slow to be used in a real time environment in HD resolution, or the quality is not sufficient to use the resulting disparity maps for DIBR. In this paper we therefore use the recently developed line-wise hybrid recursive matcher (L-HRM) coupled with a bilateral-filter-based post-processing and up-sampling step, which is real time capable and provides high-quality disparity maps.

The disparity maps are used to render virtual camera views via depth image based rendering (DIBR) (Mark [7], McMillan [8]). Initially only one image and the corresponding disparity map (V + D) were used. Later approaches improved the render results by providing additional occlusion information. Layered depth video (LDV) (Zitnick *et al.* [9]) added additional occlusion layers to the disparity map and multiview video (MVD) (Müller *et al.* [10], Smolic *et al.* [11]) added additional image and disparity pairs at other camera positions to fill disoccluded regions in the rendered image. For application in a stereo-to-multiview conversion the MVD approach is most suitable, because there are already two images at different positions available and disparity maps can easily be estimated for both images.

3 Stereo analysis and correction

The disparity estimation, and by that the whole processing chain, is extremely dependent on a well rectified input image pair. While for 3D Blu-ray content, the left and right view can generally be assumed to be fairly well rectified, this is definitely not true for live camera images. To guarantee rectified input images, the stereoscopic analyser (STAN) [4, 5] is used to analyse the input images and rectify them if necessary.

3.1 Feature point based stereo analysis

The stereo analysis by the STAN is done by first detecting a set of reliable image feature correspondences between left and right input image. As a feature detector the STAN uses the semantic kernels binarised (SKB) detector (Zilly *et al.* [12]). The SKB detector is a very fast blob detector and is optimised to find reliable feature point matches between stereo image pairs. These matches are used in a RANSAC-based optimisation step to determine differences between important physical camera parameters of the left and right cameras. These parameters include the difference in vertical tilt, rotation, keystone, zoom, and differences in vertical position. They can be used to calculate a pair of rectification matrices, which, if applied to the input camera images, produce a rectified output image pair.

4 Disparity estimation

After rectification, the disparity maps are estimated using the line-wise hybrid recursive matcher (L-HRM). The L-HRM is based on the hybrid recursive matcher (HRM) (Atzpadin *et al.* [13], de Haan *et al.* [14]). Though the HRM is real-time capable for small image sizes, it cannot take advantage of multicore environments owing to its recursive structure, which prevents parallelisation by design. The L-HRM breaks up part of the recursion in the HRM which allows for a parallel execution of lines and columns.

The L-HRM consists of two essential parts: the line- and column-wise recursion and the pixel recursion. The pixel recursion step introduces new disparity candidates to the estimator and is essential for the initialisation of the disparity estimation and scenes with fast changing content. The line- and column-wise recursion propagates disparities of surrounding pixels and previous frames to the current pixel. This ensures spatial and temporal stability of the resulting disparity maps.

4.1 Pixel recursion

The pixel recursion of the L-HRM is highly similar to the pixel recursion of the HRM. In order to find a reliable disparity estimate the neighbourhood of the current pixel is examined. Following the principle of optical flow, a disparity update value is calculated on the basis of spatial gradients and gradients between left and right input image as shown in (1):

$$d_{n+1}(x, y) = d_n - \Delta I(x, y, d_n) \times \frac{\vec{\nabla} I_L(x, y)}{\|\vec{\nabla} I_L(x, y)\|^2} \qquad (1)$$

The gradient between left and right image is approximated by the intensity difference of corresponding pixels as calculated by the current disparity estimate, as shown in (2):

$$\Delta I(x, y, d_n) = I_L(x, y) - I_R(x + d_n, y) \qquad (2)$$

Since vertical disparities are of no concern, the spatial gradient reduces to a simple derivation in the x-direction. As with optical flow the disparity is updated iteratively choosing a different position in the neighbourhood in each iteration. The disparity update value with the smallest absolute intensity difference between corresponding pixels in left and right image is chosen as the new candidate for the line-wise respectively column-wise recursion.

4.2 Line-wise and column-wise recursion

In the line-wise respectively column-wise recursion the disparity candidate from the pixel recursion is compared to further candidates from previously calculated pixels in the neighbourhood and the previous frame. This is done using block comparisons of the current pixel's neighbourhood with the corresponding region in the second stereo image. As a quality measure the normalised cross-correlation is used.

The final disparity is calculated in a two-step approach. First a line-wise recursion is performed. In this step three disparity candidates are considered for each pixel: the horizontally preceding pixel's disparity, the disparity of the same pixel in the previous frame and the resulting disparity of the pixel recursion. The initial disparity estimate for the pixel recursion is the best candidate of the first two candidates. The direction of the line-wise recursion is switched from left-to-right to right-to-left for each consecutive line.

The second step is the column-wise recursion. Here, only two candidate disparities are considered: the resulting disparity of the preceding line-wise recursion and the vertically preceding pixel's disparity. As in the line-wise recursion the direction is switched from top-to-bottom to bottom-to-top on a column-wise basis.

4.3 Consistency check and basic post-processing

The consideration of the previous disparity value as candidate at a given pixel ensures a high degree of temporal stability. The resulting disparity maps may still have false disparities, especially in occluded regions. Thus, a left-right consistency check and a small median filter to remove isolated false disparities are applied as shown in Fig. 2.

5 Post-processing and up-sampling

In a refinement and up-sampling step the disparity maps estimated by the L-HRM are post-processed to reduce artefacts and to resample the disparity maps to full resolution if necessary. For this, different variations of cross-bilateral filters are applied iteratively, similar to the methods introduced in Kopf *et al.* [15] and Riemens *et al.* [16]. First the raw disparity map is filtered by a cross-bilateral median filter of rather large size, similar to the one

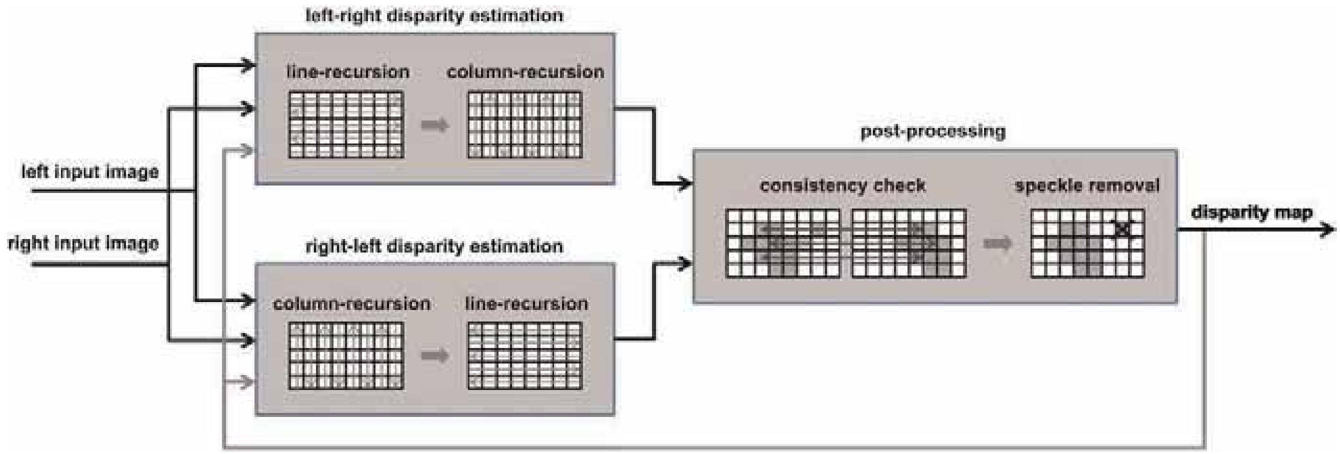


Figure 2 *Disparity estimation scheme of L-HRM including left-right consistency check*

proposed in Müller *et al.* [17]. Afterwards a cross-bilateral up-sampling filter with a small filter window is applied. Each application of the up-sampling filter increases the horizontal and vertical resolution of the disparity map by a factor of 2 until full resolution of the input images is reached. To ensure temporal stability during the refinement, in the last iteration of the up-sampling filter the disparity of the previous frame is also taken into account. Unlike the L-HRM, the post-processing uses RGB images as input instead of luminance images.

5.1 Cross-bilateral median filtering

The cross-bilateral median filter essentially works like the common cross-bilateral filter, but instead of a linear combination of all disparity values in the filter window, the weighted median of all disparity values is calculated with weights as shown in (3):

$$w(x_i, y_j) = e^{-\frac{(x_i-x_0)^2+(y_j-y_0)^2}{2\times\sigma_R^2}} \times e^{-\frac{(I(x_i,y_j)-I(x_0,y_0))^2}{2\times\sigma_C^2}} \quad (3)$$

To increase performance the first factor in the equation is set to 1 in this application, thus instead of applying a Gaussian range filter a uniform one is used. The weighted median filter ensures that only existing disparity values are possible output values. This prevents the introduction of averaged non-existing disparity values at disparity edges with similar colours in the input image. For this application a typical filter window size of 25×17 pixels is used.

5.2 Cross-bilateral up-sampling

As the up-sampling filter a common cross-bilateral up-sampling filter is employed. The filter window sizes of the up-sampling filters are significantly smaller than the pre-up-sampling filter size, typically 5×5. Instead of the rather complex filter window topology used in [16], a standard rectangular window is used, since the rather small reduction in execution time owing to non-rectangular windows could not offset the loss in quality. In the last iteration of the cross-bilateral up-sampling the disparity values of the previous frame are also taken into account. It showed that considering only both the current and the most recent frame is enough to ensure temporal stability.

6 Virtual view rendering

The full-resolution dense disparity maps which are the output of the previous processing step can now be used to render an arbitrary number of virtual views. Since the input images have been rectified in a previous step, the view rendering simplifies to a horizontal pixel shift depending on the pixel's disparity value. For the view interpolation–that is, virtual views between left and right original view–an MVD approach is used; for view extrapolation a V + D approach with adaptive disparity map pre-filtering is applied.

In both cases, the rendering is done in two distinct steps. First, the disparity map is rendered to its new position. For this forward mapping is used because backward mapping is not possible. Forwards mapping has the disadvantage that in general not every target position is hit by a forward mapped pixel and there might be pixels in the target image which were not hit at all. In order to have these pixels filled and to prevent depth errors resulting from rounding artefacts of the rendered disparity map, a continuous disparity mapping instead of a per-pixel mapping is used. For this, two consecutive disparities are warped to their exact target positions and the disparities at all integer pixel positions in between are interpolated linearly.

This works fine as long as there is no disparity gap between consecutive disparities. In this case there are two possible outcomes. In the first case, the two pixel positions change order during warping. This happens when one object occludes another. In the second case a disoccluded region, a region where no image information is available, is created. Therefore simple gap detection was implemented. If the

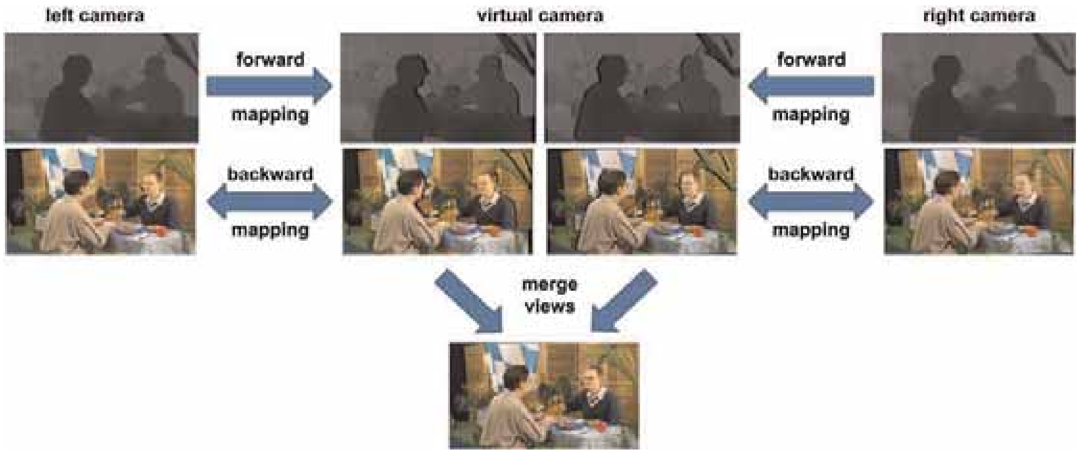


Figure 3 *View interpolation using MVD2 rendering (image source: 3D4YOU [18])*

absolute disparity difference of two adjacent pixels excesses a threshold, a gap is assumed and no disparities in between are interpolated. Since it is possible that multiple source positions are warped to the same target position, a depth test is applied for each warped pixel. Only if the new depth is nearer to the camera than the current depth is the target pixel's disparity overwritten.

When the disparities have been warped, the actual image can be rendered via backward mapping in a second render step. The target pixel's colour value can be determined by any kind of standard interpolation filter in the source image.

6.1 View interpolation using MVD2

The MVD2 approach (illustrated in Fig. 3) is used for all views rendered between left and right original view. Both the left and the right view are warped to the virtual camera position independently as described above. In a second step they are combined to a single virtual view. Both original views are used, because there can be disocclusion in the virtual view from one camera that can be filled by the virtual view of the other camera. The mixing is done using the distances of the target view to the source views as mixing coefficients. To avoid visible colour differences in image areas where only one view is available, the borders of these areas are smoothly blended with the rest of the image.

6.2 View extrapolation using video-plus-depth

If the target view is not between left and right original view, only the nearest view is used to render the virtual view. In this case the disparity map is pre-filtered to avoid large disparity gaps, which would result in disocclusions in the target view. The filtering is done using an adaptive Gaussian blur, whose variance σ depends on the disparity distance of adjacent pixels. Specific care is taken that foreground objects are not excessively stretched by the filtering, because in general a stretching of the background introduces less visible artefacts. The V + D render scheme is illustrated in Fig. 4.

7 Disocclusion filling and interweaving

Even though in both view interpolation and view extrapolation various methods to reduce disoccluded areas are used, not all disocclusion could be taken care of. Thus, in the last step these rather small disocclusions have to be filled in a visually pleasing way. Since the whole algorithm chain is supposed to run in real time and the disocclusions

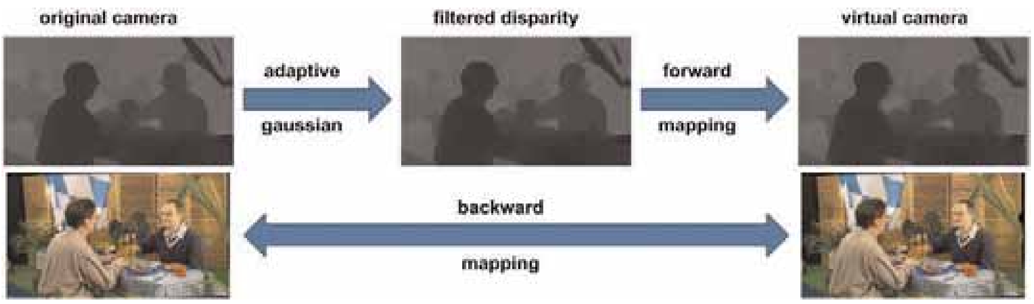


Figure 4 *View interpolation using V + D rendering (image source: 3D4YOU [18])*

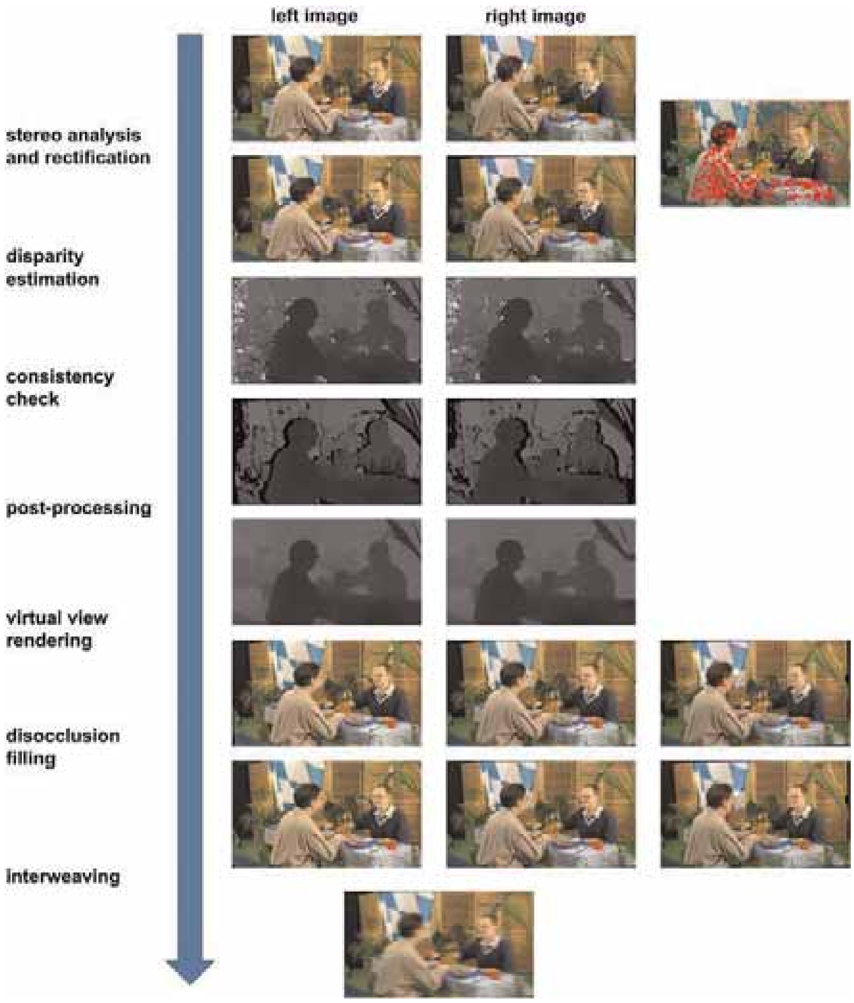


Figure 5 *Results of each step in stereo-to-multiview processing chain (image source: 3D4YOU [18])*

at this point are rather small, a sophisticated inpainting is not feasible and a simple pixel repetition is used. For this, the disparities of the rendered image are analysed and disoccluded areas are filled with the colour of the nearest background pixel.

A typical effect which occurs in rendered virtual views is that foreground objects look as if they were stamped into the background. This is caused by sharp object edges introduced by the DIBR. In natural images the change between foreground and background is rather smooth. Thus, an adaptive smoothing of the image at disparity edges is applied. This also reduces visible artefacts introduced by disparity flickering at object edges.

7.1 *Interweaving*

In a final step, all rendered views have to be interwoven into a single image of full display resolution in a display specific pattern. Since an Alioscopy autostereoscopic display is used, eight views are rendered and the Alioscopy pattern is applied to merge them to a single full HD resolution image, which can be shown directly on the screen.

8 Evaluation

The whole processing chain was implemented on a dual-processor Intel Xeon X5690 system with two Nvidia Geforce GTX 590 graphics cards. The pre-processing including the STAN-based stereo analysis and rectification, as well as the disparity estimation, was executed on the CPU. For this, the initial disparities were only estimated on a 4×4 times subsampled image pair. The processing time saved by the sub-sampling was instead used to make several disparity estimation iterations per frame. This ensures a much better adaptation of the disparity maps to fast changing scenes. The disparity map post-processing and up-sampling, as well as the rendering and disocclusion filling, was implemented as CUDA code and executed on the GPU.

In Fig. 5 the results of each consecutive step of the processing chain are visualised. The complete stereo-to-multiview conversion solution runs at a minimum of 24 fps using as input a pair of Full HD images. If not otherwise specified, all processing – in particular the view rendering– is done at full resolution as well. This enables us to support future autostereoscopic displays with higher resolution than HD with only minor adaptations to the processing chain.

9 Conclusion

In this paper a processing chain was presented which enables high-quality live playback of conventional stereo video content on an autostereoscopic display. The current implementation of the chain runs in real time as software on any recent high-performance PC. The involved algorithms were designed to be suitable for implementation on hardware platforms such as FPGAs or SoCs.

10 References

[1] BERRETTY R.-P.M., *ET AL.*: 'Real-time embedded system for stereo video processing for multiview displays'. *Proc. SPIE-IS&T Electronic Imaging*, 2007, Vol. 6490, p. 649014

[2] LANG M., *ET AL.*: 'Nonlinear disparity mapping for stereoscopic 3D'. SIGGRAPH '10, Los Angeles, CA, USA, July 2010, vol. 29, p. 75

[3] HARTLEY R.I.: 'Theory and practice of projective rectification', *Int. J. Comput. Vis.*, 1999, **35**, (2), pp. 115–127

[4] ZILLY F., KAUFF P., EISERT P.: 'Joint estimation of epipolar geometry and rectification parameters using point correspondences for stereoscopic TV sequences'. 3DPVT, Paris, France, May 2010

[5] ZILLY F., KAUFF P., EISERT P.: 'An image-based assistance tool for stereo shooting and 3D production'. ICIP 2010, Hong Kong, September 2010

[6] SCHARSTEIN D., SZELISKI R.: 'A taxonomy and evaluation of dense two-frame stereo correspondence algorithms', *Int. J. Comput. Vis.*, 2002, **47**, (1–3), pp. 7–42

[7] MARK W.R.: 'Post-rendering 3D image warping: visibility, reconstruction and performance for depth-image warping'. PhD thesis, University of North Carolina at Chapel Hill, 1999

[8] MCMILLAN L.: 'An image-based approach on three-dimensional computer graphics'. PhD thesis, University of North Carolina at Chapel Hill, 1997

[9] ZITNICK C., *ET AL.*: 'High-quality video view interpolation using a layered representation'. SIGGRAPH'04, Los Angeles, CA, USA, August 2004, vol. 23, no. 3, pp. 600–608

[10] MÜLLER K., *ET AL.*: 'View synthesis for advanced 3D video systems', *EURASIP J. Image and Video Process.*, 2008, doi:10.1155/2008/438148

[11] SMOLIC A., *ET AL.*: 'Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems'. Proc. ICIP 2008, San Diego, CA, USA, October 2008

[12] ZILLY F., *ET AL.*: 'Semantic kernels binarized – a feature descriptor for fast and robust matching'. CVMP'11, London, UK, 2011

[13] ATZPADIN N., KAUFF P., SCHREER O.: 'Stereo analysis by hybrid recursive matching for real-time immersive video conferencing', *IEEE Trans. Circuits Syst. Video Technol.*, 2004, **14**, (4), pp. 321–334

[14] DE HAAN G., *ET AL.*: 'True-motion estimation with 3-D recursive search block matching', *IEEE Trans. Circuits Syst. Video Technol.*, 1993, **3**, (5), pp. 368–379

[15] KOPF J., *ET AL.*: 'Joint bilateral upsampling'. ACM Trans. Graphics (SIGGRAPH'07), July 2007, vol. 26, no. 3

[16] RIEMENS A.K., *ET AL.*: 'Multi-step joint bilateral depth upsampling'. *Proc. Visual Communications and Image Processing (VCIP)*, 2009, vol. 7257

[17] MÜLLER M., ZILLY F., KAUFF P.: 'Adaptive cross-trilateral depth map filtering'. 3DTV-Conf., Tampere, Finland, June 2010, pp. 1–4

[18] 3D4YOU, FP7 project funded by European Commission. FP7-ICT-2007-215075

Selected content from IBC2012

ISSN 2041-5923

Project Fresco – tileable technology for television's big future in the home

A. Ashley[1] *M. Costello*[1] *K. Murray*[2] *S. Parnall*[1] *J. Walker*[2]
[1]*NDS, Staines, TW18 4EX, UK*
[2]*NDS, Southampton, SO53 3YU, UK*
E-mail: kamurray@nds.com

Abstract: The continual evolution of display technologies offers the potential for new, immersive and realistic experiences of media 'lifestyle' displays constructed from modular tiles. This generates two questions, 'what could we do with this?', and 'how could we build it?'

In Project Fresco we are developing, evolving and describing a vision addressing the first of these questions, a vision that is not constrained by today's displays or conventions. This vision is illustrated by, and has been informed through, experiences with our first prototype.

To address the second question we have developed an architecture to realise and explore our vision. This scalable architecture supports an arbitrary number of displays, synchronises across multiple tiles, integrates devices in the home as equal partners and can be implemented today. Further, it adapts content to whatever displays are available and to the involvement of the user.

1 Introduction

The choice of type and size of television for the home is so often a compromise between the extremes of an exciting viewing experience when the device is switched on and the wall or corner space occupied by a dark and dull object when the device is switched off. And, when it is on, the size of the picture may well be inappropriate for the type of content and engagement of the occupants of the room.

Science fiction, as we discussed in Walker et al [1], provides a vision of the future including television based on wall-sized displays. At the same time technology is bringing this closer to a reality through advances in graphics and display technology such as thin-bezel, tileable televisions, and cheap, powerful graphics capabilities.

In this paper, believing in the inevitability of these technological trends and the opportunities they create, we set out our vision and its realisation. We then describe the technologies and components required to implement this vision with a sophisticated and intuitive user experience that matches content and mood, supplemented with additional content and so-called domotic feeds (that is material concerning the home). In doing this we build on lessons learned from our first prototype implementation and describe key aspects of our second prototype.

2 Vision and first prototype

Our vision of the future is of a viewing environment with large displays constructed from modular tiles. These displays are frameless, unobtrusive, ultra high-definition and ambient. They can be adapted to fill or partially fill one or more walls of a room, and will co-operate to provide an integrated experience. The opportunity is to open up possibilities way beyond the limits of today's devices through:

- content comprising multiple visual elements that can be adapted spatially and temporally, freeing the user from the need to choose a single element, or the system from having to impose overlays

- shared, co-operative usage of the displays, with connected companion devices becoming personal extensions

- supporting connected applications and services operating in a more streamlined, integrated manner, reflecting and effecting changes in viewer engagement in TV content

- dynamic adaption to, and control over, the environment of the displays, and adapting to the wallpaper and lighting

- introducing domotic content in a sympathetic manner

Based on this vision, a prototype was constructed and demonstrated with great success at both IBC 2011 (Clover [2]) and CES 2012. This prototype has a single display occupying most of one wall as shown in the photograph in Fig. 1. This single display was constructed from six tiles, driven from one device. A companion device used to control and interact with the system can also be seen.

From this prototype (described in Walker et al [1]) we have learned much, not least of which is the popularity and potential of this approach to television. In developing the second prototype we are building on our experiences and exploring two aspects: the practical construction of a system and the impact of multiple displays.

2.1 Lessons learned from the first prototype: immersion

Television programmes have a natural flow and pace and will have points at which viewers are likely to be extremely immersed and engaged in the content. In our system we have introduced the concept of 'immersion' that is key to the way that the presentation of content is adapted to reflect this flow and pace. Put simply, the more immersed in the content a viewer is, the greater emphasis that is placed on the core video, and the less immersed they are the more emphasis comes to be placed on related content which may then be introduced. Fig. 2 shows an example of immersion adapting the presentation.

2.2 Lessons learned from the first prototype: the importance of companion devices

Living with the first prototype has shown the importance of companion devices as an integral part of the experience. There are several reasons for this: their use as a vehicle for adapting and manipulating the content, including controlling the immersion level, for interactions with the content that naturally lead back to the main display, and to provide a personal interaction within a larger shared experience. All of these require companion devices to be an integral part of the system.

Figure 1 *First prototype, IBC2011*

3 An architecture: building the vision

In realising our vision we have developed an overall architecture that builds on the experiences of the first prototype. This architecture describes both the physical components of the system and the logical components that can exist anywhere on any device.

3.1 Architectural requirements

To set the scene for the architecture we start with the main (though not exhaustive) requirements:

- the system should be highly scalable, in terms of the number and shape of displays in the system and the number and topology of tiles used in each display (displays no longer need to be rectangular)

- the architecture should be flexible in orientation, positioning and shape of the displays

- the system should support synchronised presentation across all displays and tiles

- the system should be cost efficient and we assume that the system will be constructed and extended using small, cheap building blocks

3.2 Structure

Fig. 3 shows the main components of our architecture, of which the synchronisation server (sync server) is covered in more detail later in the paper. Although this diagram shows a single client for each display, it is equally valid to have multiple clients each driving a subset of the tiles in a single display. The main goals of scalability and flexibility are addressed through a modular approach, where the main modules are 'client devices', combined with flexibly located software components such as the layout engine and synchronisation server.

3.3 Client devices

A client device is connected to one or more tiles, and is responsible for rendering content for the tiles connected to it. The client device selects only the media components that are to be shown on its connected tiles – this is a well-known technique in computer graphics referred to as culling or clipping. This approach means that each client only needs sufficient hardware to perform media decoding

Figure 2 *Low immersion (left) and high immersion (right) example*

and graphics rendering for the tiles it supports (video decoding currently requires full decoding of any stream, even if it is partially displayed, but future coding technologies, or profiles of existing technologies, may allow for this to be limited in the case of arbitrary resolution video), and so a system can scale by adding client devices (which could even be part of the tiles).

3.4 Layout engine

The layout engine is a software component that can be flexibly located, e.g. in a client device or a home gateway or even the cloud. The functionality of this is largely as developed for the first prototype (as described in Walker et al [1]) which is to take the immersion level and details of the content to be presented, including domotic and personal content, compare these with the capabilities of the displays (number, size, shape, location and resolution) and generate a layout. This layout is a list of media items that are to be displayed and their size and position on the displays. However, our experiences have shown that the layout engine is best treated as a separate component that communicates the resultant layout to the client devices. This approach assists the cleanness of the architecture, the scalability of the clients (no client needs a 'global' view) and aspects of co-ordination and synchronisation between client devices.

3.5 Companion devices

The companion devices provide a means of control and interaction with the system, and are treated as first class citizens with full access to all parts of the system. They communicate with the layout engine which acts on their commands and updates the layout as needed. The companion devices also monitor the output of the layout manager and produce a 'mirror' of what is shown on the main displays, providing a natural interaction point. To address scalability issues of bandwidth and resolution within the companion device 'mirror', they typically make use of alternate, low resolution static graphics.

3.6 Layout metadata

Layout metadata is required for every piece of content displayed by the system. This metadata can either be bespoke authored, or be one of a set of generic layouts. This information is used by the layout engine to determine what should be displayed, where these items should be displayed, and their relative priorities. Generally, though

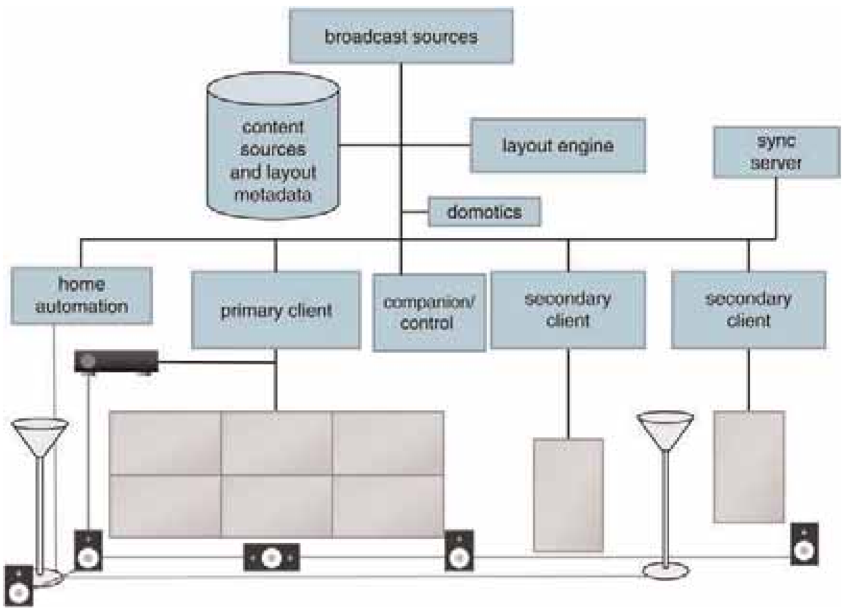


Figure 3 *Architectural components of system*

not necessarily, this information is time synchronised to the content.

4 Synchronisation of client devices

The architecture we have described above is based on many independent client devices that are rendering content to provide a single unified impression of a large display rather than a group of unrelated displays. This means that both the tiles and the separate clients need to be synchronised.

4.1 Tile synchronisation

Although two tiles presented with the same signal at the same time may not display the pictures synchronously, we believe this is not an issue. First, we performed simple tests using a graphics card with a digital splitter driving identical conventional displays and observed differences in output time of a maximum of 5 ms. Secondly, video game demands have driven down response time (the time from the signal arriving to being visible) to values of 2 ms and below (Carter [3]). Finally, newer technologies such as OLED have the potential to reduce these values even further.

4.2 Client rendering synchronisation

As the tiles (or groups of tiles) are potentially driven by different client devices, the clients must be synchronised. If each client was connected to an ideal tile with a response time (or latency) of 0 ms, then a timing error between the clients could result in the case shown illustratively in Fig. 4. Here, the client driving tile 2 is running two frame periods behind the client driving tile 1.

4.3 Synchronisation methods

The mechanisms for achieving synchronisation fall into two broad camps:

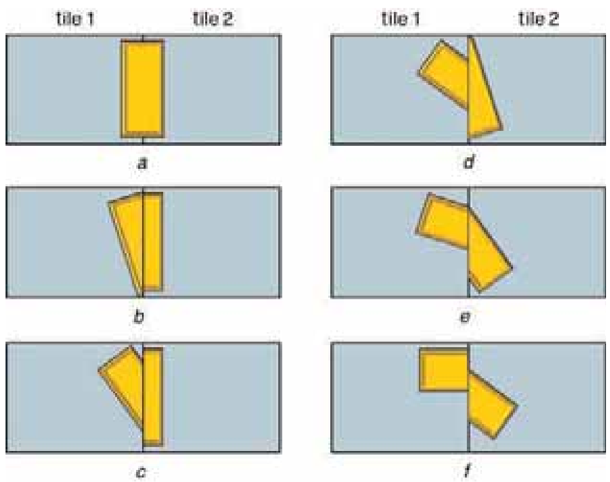


Figure 4 *Example of unsynchronised clients*

- synchronised clocks, where all client devices have the same clock and agree to generate output (e.g. produce the next frame) at the same time

- barrier methods, where all client devices wait for each other to reach a given point (e.g. a new frame is prepared) and progress (e.g. display the frame) only when they have all reached that point

The best known, and most widely used, example of synchronised clocks is the internet NTP protocol (Mills et al [4]), which under ideal circumstances can achieve from approximately 1 ms to 10 ms accuracy without specialised modifications to the hardware or software of the client. A similar, but more accurate, effect can be achieved through the MPEG-2 system clock (ISO [5]), in a broadcast signal. Unfortunately, in our architecture, such a signal would need to be carried over a home network, which reduces the ease with which this clock can be used for this style of synchronisation owing to the significantly increased jitter, which in turn increases clock lock times.

Barrier synchronisation mechanisms have been used in multiple client systems, such as in the Scalable Adaptive Graphics Environment (SAGE) described by Jeong et al [6]. This approach has its own challenges as its nature requires an exchange of messages on each frame which results in a period of time when clients are idle awaiting confirmation to proceed (even the simplest network message typically takes around a millisecond, which is a significant fraction of a 60 Hz frame, especially where there are multiple clients resulting in multiple messages), and this can reduce frame rates and consequentially the smooth behaviour of the system.

4.4 Implemented solution

The solution we are using is a hybrid of clock synchronisation and barrier techniques. We use standard mechanisms to replicate the time from a master video stream in one client device to all other client devices that require synchronisation with the master. This is done via the synchronisation server to minimise serialisation between clients. The slaved clients perform clock drift compensation, and implement a local time-based barrier to control output. As all slave clients play back the same stream each client effectively blocks (i.e. makes use of a barrier) until the desired time is reached – much as in traditional decoding. The difference from traditional decoding is that the block can involve an interaction with the synchronisation server to check and update the current master media time.

Whilst this mechanism could be applied to blocking on each video frame, we apply it to groups of audio frames. This fits simply into a number of architectures where audio is often the clock master. By using multiple audio frames the frequency of synchronisations can be reduced (we are currently running with inter-synchronisation intervals of approximately 200 ms, but indications are that this can be increased with no adverse effect). Furthermore, as our current architecture only has one audio output on the master, we are able to adapt the audio on the slave devices without impacting the playback experience, and thereby use the audio output as the locally corrected clock. This can be done either by altering the number of audio samples presented to the underlying hardware or using standard audio playback rate alteration techniques.

Our on-going analysis of this architecture currently shows instant, stable synchronisation of playback suitable for A/V synchronisation between client devices over a reasonably loaded network. This analysis has also shown visually acceptable playback of a video split across two client devices. Audio has far tighter requirements (e.g. avoiding phase shift) and we have had some positive results in splitting the audio across devices. However, the fact that our second prototype only has one audio output means that we are not currently pursuing this, though for certain cases such as multi-room, audio synchronisation will be important.

4.5 Extending synchronisation to graphics

Graphics, unlike video, typically requires state as input to the process of generating an image. Examples of the state might be the position or rotation of an object that is to be rendered. Without agreement on the state, the result of otherwise synchronised tiles could be like that shown in Fig. 4. Thus, for graphics, in addition to agreeing when to display a frame, synchronisation must include what to display.

This can be achieved through simple extensions of the video solution, by assuming the presence of a (possibly hidden) video stream and placing certain constraints on the graphics. First, all graphical operations are triggered at a start time related to a video stream, and secondly are based on an animation timeline again tied to the video stream. The agreed start time ensures starting synchronisation. The animation timeline, combined with the video stream, allows each client to correctly determine the state it needs. This is done by using a predicted frame rate to calculate the point in the timeline that the next frame will be displayed, and so the state required to generate that next frame. As different clients may have different graphical loads, the frame rate can vary between them. To overcome this messages are included with the synchronisation mechanism so that all clients will quickly converge to the same frame rate (that of the slowest client).

5 Non-rectangular video

The majority of today's broadcast video is landscape rectangular, but our explorations have shown how compelling other shapes can be, especially when combined with the possibility of blending the video into the background; an example of this is shown in Fig. 5*a*. Although non-rectangular video is not new, having been present in MPEG-4 (as described in Ohm [7]) and the DVD standard, broadcast does not support this to our knowledge.

There is a range of ways that non-rectangular video can be supported, most of which are combined with a conventional rectangular video (as this provides the simplest deployment through re-use of existing video decoders). The desired shape and background blending for the video is typically conveyed by one of three means:

- using specific values, or ranges of values, in the content: this is exemplified by chroma or luma keying, but has the limit of allowing only 'on' or 'off' blending

Figure 5 *Alpha mask and compression*

a Example of use of alpha mask
b Alpha mask
c Error on compressed alpha mask

- using an additional plane or picture that conveys the alpha values: this approach can convey blending information to improve the visual quality

- by specifying a shape, e.g. using vector graphics

We chose to work with alpha planes, the second case listed above, as most graphics hardware supports this. Furthermore, we had concerns that compression of a signal using luma or chroma keying would be susceptible to blending errors as well as placing additional constraints on the source material to avoid the key values.

Fig. 5*b* shows a small section of the alpha plane used with the example shown in 5*a* (which was taken from our first prototype). The result of compression errors typical of a lossy broadcast transmission is then shown in Fig. 5*c*. Akin to the early challenges with chroma-key systems, these errors can result in an image that stands out from the background, rather than one that is blended into the background. In our first prototype, this was largely avoided by both using slightly higher bitrates and quantised alpha planes equivalent to lower bit depths (to minimise the errors that compression introduced). Our main challenge is the absence of direct integration of alpha channels into the usual YUV broadcast formats, and we feel this is an area that the industry should explore, possibly through the use of the existing auxiliary picture mechanisms.

6 Conclusions

Our exploration started by looking at how television might evolve in response to anticipated developments in display technology, and our first prototype explored the possibilities for a new way of looking at television that these developments will open up. In this paper we have described our vision and aspects of the work we are undertaking in the development of the second prototype: developing a practical, flexible, cost effective in-home architecture capable of supporting several modular tiled displays.

This paper has purposefully emphasised the aspects of work related to scalable synchronised clients running with modular tileable displays. There are many other aspects of this work which space does not allow us to cover here, including the evolution of content material suitable for such systems, the metadata authoring, management, carriage and synchronisation, and technical means of integration of a range of companion devices.

Project Fresco is allowing us to explore and show a future in which television is not marginalised but naturally reflects the viewers' engagement. It is a future in which television is simultaneously shared and personalised, and seamlessly includes many media forms. It is a future in which, we believe, television will maintain a dominant and important role in our lives.

7 Acknowledgments

The authors wish to thank others within NDS who have contributed to this project in many ways.

8 References

[1] WALKER J.G., MURRAY K.A., PARNALL S.J.: 'Right pictures, wrong size? Evolving the TV experience for tomorrow's home displays'. Proceedings of the 2011 International Broadcasting Convention, Amsterdam, The Netherlands, 2011

[2] Julian Clover, 2011, 'NDS puts TV on all Surfaces', Broadband TV News, http://www.broadbandtvnews.com/2011/09/12/nds-puts-tv-on-all-surfaces/

[3] Jamie Carter, March 2011, 'The Best TV for Gaming: what you need to know', Tech Radar, http://www.techradar.com/news/television/best-tv-for-gaming-what-you-need-to-know-935647

[4] D. Mills (U. Delaware), J. Martin (ISC), J. Burbank, W. Kasch (JHU/APL), March 2010, Network Time Protocol, RFC 2905, IETF, http://www.ietf.org/rfc/rfc5905.txt

[5] International Standards Organisation, 2007, 'Information technology — Generic coding of moving pictures and associated audio information: Systems', ISO/IEC 13818-1

[6] JEONG B., RENAMBOT L., JAGODIC R., SINGH R., AGUILERA J., JOHNSON A., LEIGH J.: 'High-performance dynamic graphics streaming for scalable adaptive graphics environment'. SuperComputing 2006, Tampa, FL, USA, November 2006, http://www.evl.uic.edu/files/pdf/SAGE-SC06-final.pdf

[7] Jens-Rainer Ohm, July 2005, Introduction to MPEG-4 Video (arbitrary shape), http://mpeg.chiariglione.org/technologies/mpeg-4/mp04-vid(s)/index.htm

Selected content from IBC2012

ISSN 2041-5923

Asset values increased – new technologies for the automatic transfer and digital preservation of analogue video-cassettes

J. Burghardt J. Houpert T. Mayer

Cube-Tec International GmbH, Anne-Conway-Straße, D-28359 Bremen, Germany
E-mail: juergen.burghardt@t-online.de

Abstract: Millions of video-cassettes are still stored in shelves carrying valuable content waiting to be transferred into files before decomposing. Time is running short: not only tape stock decays but also video tape recorder (VTR) availability gets critical and VTR service experts are becoming rare. As it is the final transfer, a technology for highly automated large-scale and lossless transfer which recovers and retains the original signal on tape is required.

Direct tape transfer is a new technology which directly converts the off-tape signal from analogue video tapes into digital files without passing through conventional playback circuits, and without its quality degrading multiple A/D- and D/A-conversions of standard video playback equipment.

By digitising the RF tape signal directly, all subsequent signal processes can be done by software with a much higher resolution than any conventional VTR allows. Intelligent signal regeneration by smart dropout prevention and digital concealment algorithms allow signal recovery far beyond the possibilities of traditional dropout compensation circuitry.

The technical details of the direct tape transfer technology, which was developed in a joint research project by the Fraunhofer Institute for Integrated Circuits IIS and Cube-Tec International, are explained.

1 Overview of signal processing path

Over several decades, VTR's have been the most successful recording devices for TV and professional video production. With Sony's BetacamSP and Panasonic's M-II Formats, analogue video cassette became the workhorse format which could record video signals in full specification regarding bandwidth, SNR and pulse-response characteristics. Half-inch cassettes were used in all areas of TV production, from electronic news gathering (ENG) to high-end studio and live recordings. It is obvious that still many millions of cassettes are stored in shelf archives. Many of those contain very valuable and high-quality content, which represents an important asset and needs to be preserved for future use. Playback of tapes and ingest into today's file-based infrastructures still rely on the nearly 20 year old signal processing capabilities of analogue VTRs containing several simple 8-bit A/D and D/A conversions, as well as old style analogue dropout compensation methods.

A new approach to rescue analogue content on tapes by means of modern signal processing algorithms was developed as a joint cooperation between the German Fraunhofer IIS research institute in Erlangen and Cube-Tec International.

The idea was to replace all real-time conventional VTR playback processing with digital signal processing, using sophisticated high-resolution file-based processing in real-time computer systems, which totally avoid any video A/D and D/A conversion artefacts. It allows selectable signal depth at more than 10-bit, and makes use of sophisticated dropout prevention and concealment and creates comprehensive sets of metadata to be used in subsequent quality control and restoration processes.

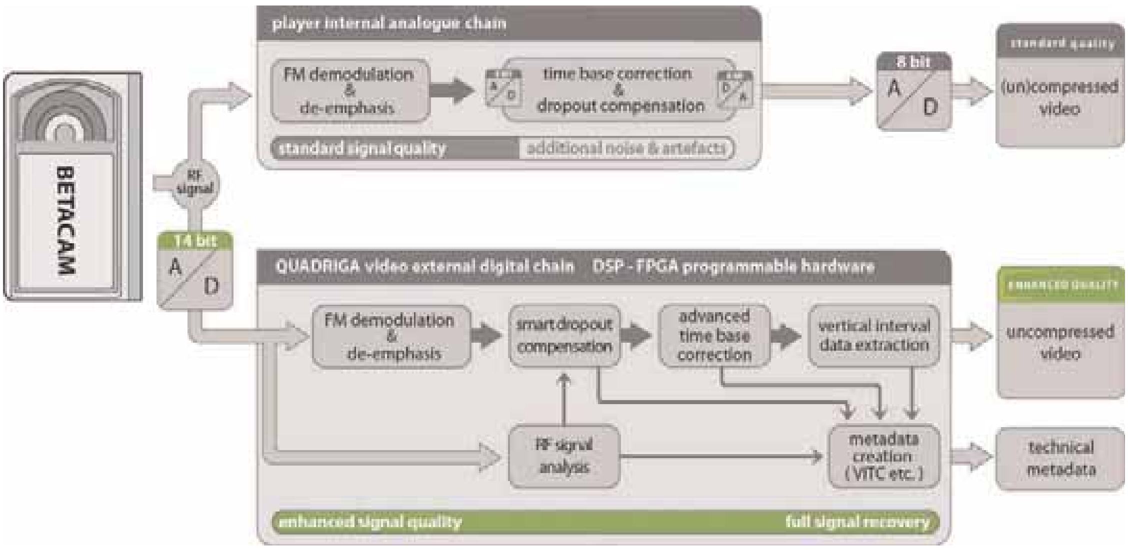


Figure 1 *Comparison of conventional and SW-based BetacamSP playback processes*

Fig. 1 shows a comparison between conventional VTR playback processing and the new software-based demodulation of analogue VTR signals. The figure refers to Betacam cassette format, because in Europe Betacam and BetacamSP were by far the most used video cassettes in broadcast and professional applications.

The upper path shows the playback processes of a conventional VTR. The off-tape RF signal passes an analogue FM-demodulation and de-emphasis circuit. The time base corrector (TBC) corrects for any jitter introduced by the mechanical tape drive and synchronises both luminance and chrominance channels as part of the component recording Betacam format. Dropout detection is done by detecting RF-levels below a fixed level threshold. Signal defects caused by dropouts are compensated by replacing the affected portions of the video signal with video information from previous lines. In Betacam VTRs the dropout compensation is part of the TBC. The digital signal processing inside the TBC allows easy interim storage of video as complete lines. This video storage can be used to replace the defective parts of the signal with correctly demodulated video from previous lines.

One drawback of this conventional signal path is obvious. Video signals are demodulated back to base band before being transferred into a file. Several A/D and D/A conversions are necessary, when the video output is fed to an ingest station for transfer into server-based infrastructures. Even worse, VTRs are equipped with simple 8-bit converters which were commonly used in the 20 years old circuit design.

The lower path in Fig. 1 shows the new approach. The off-tape RF signal is directly digitised by a state-of-the-art high-performance and high-density A/D converter. The digital data stream output is a 100% digital representation of the original analogue RF signal on tape including all dropouts and all signal timing and jitter characteristics from tape. Playback processes like demodulation and de-emphasis are fully implemented by software using up to 16-bit signal processing. It is obvious that no video A/D – D/A concatenation takes place, because of the total digital approach. Tape playback is done in real time, but by nature, all following software-based processes are non-real time. High-performance processing power is needed to secure real-time capturing. Even multipath processing for improved results (e.g. RF signal analysis) becomes possible.

One of the important aspects is the possibility to prevent dropouts by adaptively lowering the dropout detection level to the lowest possible level, before invalid video signals are demodulated. Even if dropouts cannot be recovered, smart concealment algorithms can be implemented as in use in today's digital VTR's to further improve dropout compensation. This system is called 'smart dropout prevention and concealment'.

Each processing step can be monitored and logged, from which comprehensive sets of metadata are generated to be used for further descriptions and quality control of the entire transfer process. In an automatic scalable approach of mass tape digitisation it is essential that video, audio and time code are captured simultaneously.

2 RF signal digitisation

In analogue VTRs like BetacamSP the video signal is frequency modulated before being recorded on tape. The three channel video component input signal (Y, R-Y, B-Y) is multiplexed into two channels, one for luminance (Y) and the second for a time-compressed and multiplexed signal carrying both chrominance signals (R-Y/B-Y). Before the frequency modulation (FM) both signals are fed through a pre-emphasis circuit which effectively boosts high-frequency signal components. The inverse de-emphasis on the playback side lowers high-frequency signal components including noise. The combination of pre- and de-emphasis effectively improves the signal-to-noise performance of the overall analogue recording and playback.

However, maximum frequencies of the FM signal in BetacamSP can reach up to 12 MHz owing to the video level overshoot at high level signal edges caused by the high pass filter of the pre-emphasis during recording. This determines the minimum specifications of the RF–A/D conversion. The system uses a two channel A/D converter, one channel for luminance RF, the other for chrominance RF. The sampling rate per channel is 25 MHz. The A/D level resolution can be up to 16 bit, but 14 bit is actually used, which is more than sufficient compared to 8-bit resolution in standard VTRs. This generates a raw data stream of up to 800 Mbit/s. Digitising a 90 minute tape would require approximately 450 MB of raw RF data to be stored plus audio, TC and metadata.

Playback is done in real time by a conventional tape deck because a tape transport and its servo control are needed. A simple modification of the playback VTR is necessary in order to decouple the RF signal from the playback amplifiers. This RF signal is level adjusted to the input level range of the external A/D converter and connected to a spare BNC connector at the backplane of the VTR. Fig. 2 shows a picture of the RF amplifier designed to fit into a spare PCB slot in a Sony BVW-Series VTR.

This modification does not affect any standard function of the VTR. Full conventional playback is maintained and can be used for video monitoring purposes. Any other Betacam playback device can be modified for RF signal decoupling; however, BVW-Series editing VTRs are known to have the most robust and reliable tape transport system.

3 PB signal path

As explained earlier, all PB signal processing is done in software fully simulating all of the conventional playback steps. Signal processing depth is up to 16 bit, rather than 8 bit in conventional VTR's and furthermore 8 bit A/D – D/A concatenation artefacts can be fully avoided. Therefore signal quality output can be maintained at the highest possible level. The following signal processes are implemented:

- FM demodulation
- noise cancelling (de-emphasis)
- Y-C synchronisation and time base correction
- dropout detection and smart dropout correction and concealment

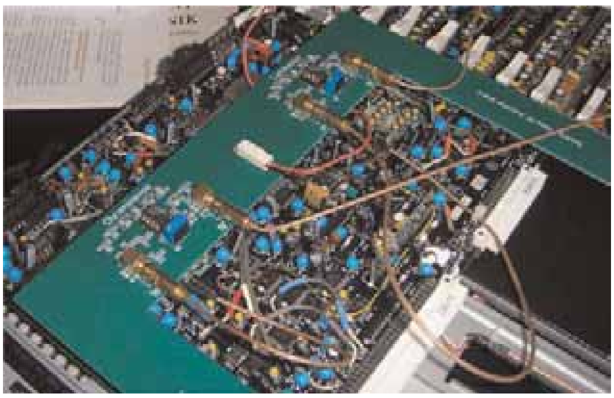


Figure 2 *RF decoupling amplifier for BVW-Series VTRs*

4 FM demodulation

FM modulation during recording converts video levels into frequencies. The following table shows the frequency allocations for the FM modulation of a BetacamSP recording system.

Fig. 3 shows the result of a first implementation of a software-based, straightforward FM demodulation. No further signal processing is applied at this stage. The signal is the 'raw' signal as recorded on tape. One can clearly see the signal overshoots at high level steps owing to the pre-emphasis circuit on the recording side. The amount of noise is still significant because de-emphasis is not applied at this stage. Betacam specific reference pulses for synchronisation of luminance and chrominance and for time base correction can be detected.

5 De-emphasis

It is essential for good signal quality, that de-emphasis circuits fully match the format specifications. Extensive tests were made in order to optimise the de-emphasis characteristics and improve signal quality.

Fig. 4 shows the result of an optimised de-emphasis. Optimisation is done by adjusting the frequency response at different signal levels. Different levels of multi-burst or sweep signals are used, because the de-emphasis characteristic is nonlinear and dependent on frequency and signal level. Attenuation of high frequencies at lower signal levels is higher compared with high frequency signal components at higher levels [1].

Table 1 Betacam SP FM frequency allocation [2]

Y-FM carrier frequency (100% white)	8.8 MHz
Y-FM carrier frequency (Sync)	6.8 MHz
Y-FM deviation (without sync)	1.43 MHz
C-FM carrier frequency max. video level	5.6 MHz
C-FM carrier frequency min. video level	6.6 MHz
C-FM deviation	1.0 MHz

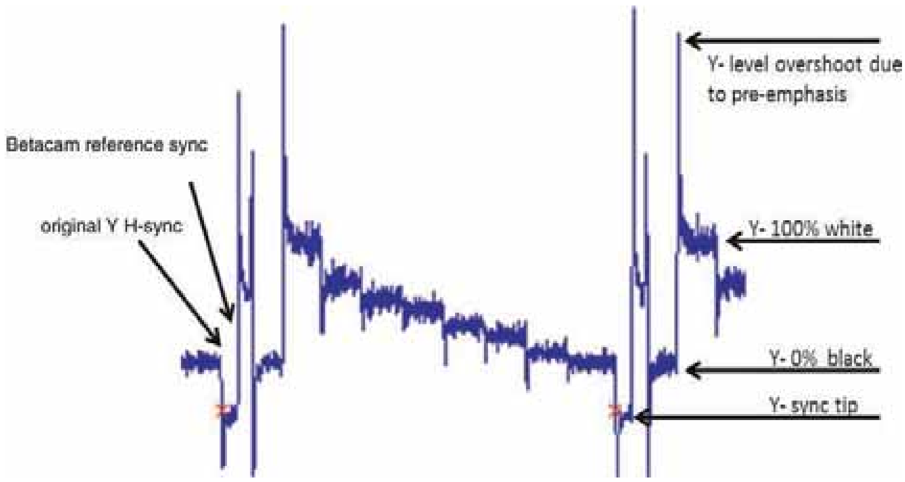


Figure 3 *Demodulated Y signal with pre-emphasis (one horizontal line)*

A multi-burst signal with exactly known frequencies is used to finely adjust the de-emphasis filter characteristics as shown in Fig. 4.

One can clearly see the level rise at high frequencies around 3–4 MHz owing to the recording pre-emphasis and the adjustment to 'flat response' over the entire frequency range after de-emphasis optimisation. The overall theoretically achievable amount of noise reduction of up to 18 dB becomes obvious.

6 Detection of sync pulses and Y-C synchronisation

Luminance and chrominance analogue component signals are recorded in two individual channels on Betacam tape using two separate video recording and playback heads.

Mechanical tape transports introduce jitter in the playback signals because of the complex mechanical head to tape transport design. Wow, flutter, mechanical friction dependent on the tape surface, and tape transport conditions generate timing errors over a wide range of frequencies. Time base correction of video signals in relation to the studio video reference is therefore an essential part of a broadcast VTR. In two channel VTRs like Betacam, additional synchronisation between both channels is of highest importance. Time differences between Y and C are specified to be shorter than 20 ns [2]. The sync detection and synchronisation of both video signals are done in several steps.

1. coarse detection of H- and V-sync position

2. fine H-sync detection by oversampling and interpolation

3. Betacam sync-pulse detection

4. line resampling to 720 pixels per active line for the luminance signal

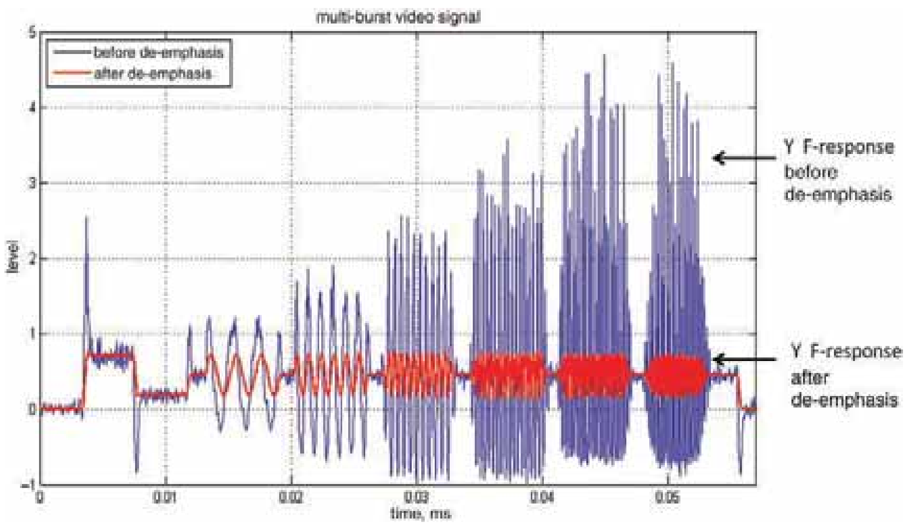


Figure 4 *Y-frequency response after optimised de-emphasis*

5. line resampling to 360 pixels per active line for both colour difference signals

6. Y-C synchronisation

7. chrominance time stretching

For A/D conversion of the PB RF signal, a free running sampling rate of 25 MHz is used. The exact sync timing reference, which is the 50% level point of the H-sync pulse, cannot be detected accurately by the free running sampling process. A sync detection tolerance of maximal 40 ns owing to the 25 MHz sampling rate, however, cannot be accepted because Y/C timing specifications in Betacam are less than 20 ns [2]. The method to exactly find the 50% level reference uses interpolation between adjacent samples around the H-Sync timing reference point, detected in the first step of the coarse sync detection. Fine sync detection by interpolation is shown in Fig. 5.

Time base correction is an essential function of the playback process because video signals need to be synchronised to a reference studio sync. Conventional VTRs normally detect PB timing errors and synchronise write clock of TBC memories to these errors by means of a phased locked loop (PLL). Read out of the memory is done by read clocks which are locked to the stable studio reference sync.

Software-based processes use resampling methods. The actual number of samples per line is detected. By resampling to the exact number of 720 samples per active line, correction of the time base is achieved automatically. This method is even more accurate than TBCs in conventional VTRs can achieve, because timing errors within one line are averaged out. In conventional VTRs, only timing differences of adjacent H-sync pulses can be detected and jitter within one video line is not corrected.

For time compressed chrominance signals in component VTRs time stretching and de-multiplexing into both R-Y and B-Y components is necessary. Time stretching is automatically done by the resampling process when 360 samples per line are created for each of the colour components.

7 Smart dropout prevention and concealment

Dropouts occur during VTR playback more often than people expect. Dropouts are level drops in RF-signal due to dirt on the tape or inside the tape transport or debris on the head drum. Worn out heads create more dropouts than clean or new heads. Mechanical tape damages or decomposing of the tape itself are also reasons for dropouts. Generally dropouts occur whenever the tape loses direct mechanical contact to the playback heads. Fig. 6 illustrates the RF-level drop and describes the dropout characteristics of length and level drop. Fig. 7 shows a histogram of momentary and accumulated dropouts at real playback conditions.

Most of the dropouts are shorter than 10 μs. But there are also significant numbers of dropouts up to 1 ms, which means that several lines of video can be affected. These dropouts are visible and can clearly be recognised in the video picture. Dropouts shorter than one video line can be nearly fully compensated and will not be visible.

Dropouts in conventional VTRs are detected by drops of the RF level below a fixed threshold, independently

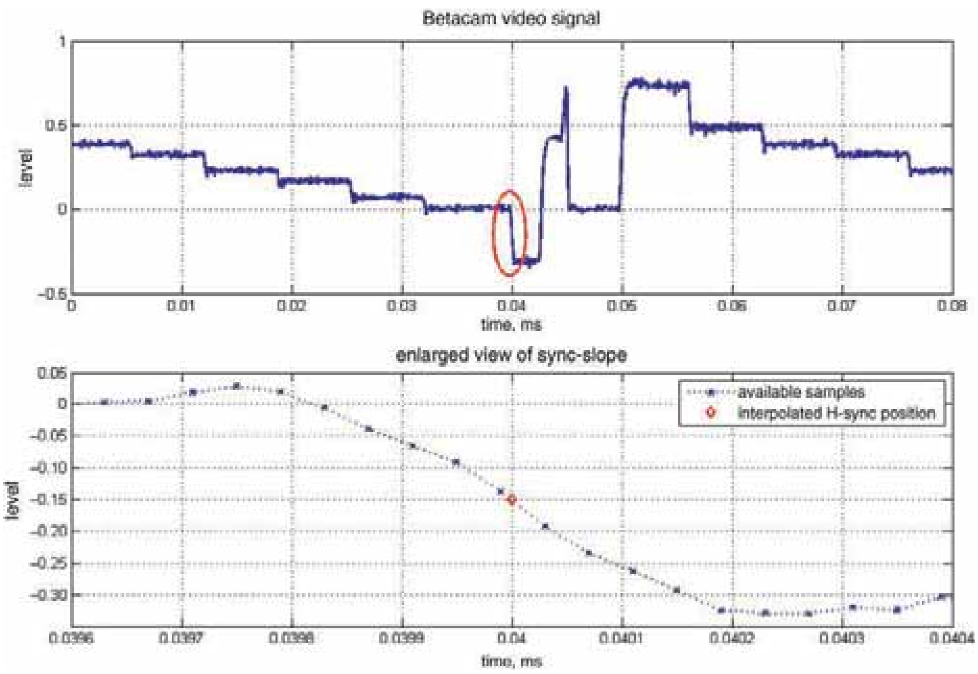


Figure 5 *H-sync detection by linear interpolation*

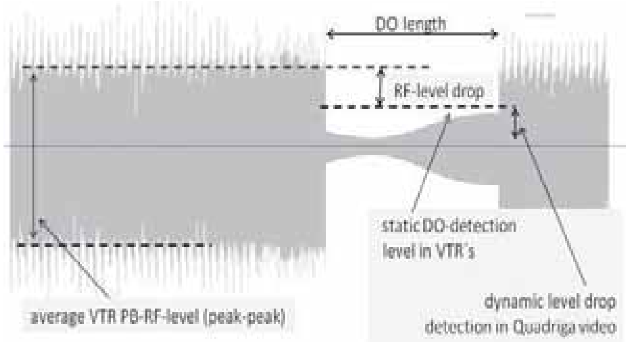


Figure 6 *RF level drop and dropout detection*

whether or not FM demodulation would still be possible. In the software-based RF level analysis, this level drop can be analysed and dropout detection level can be lowered until FM demodulation becomes impossible. By this method signals can be recovered from low RF, in cases where conventional dropout detection would detect dropouts. This software-based RF level analysis is a dropout prevention method, rather than dropout compensation. It effectively lowers the dropout rate and thus improves signal recovery. Fig. 8 shows a snapshot of a real dropout caused by an RF level drop. RF levels below a certain threshold can generate illegal video signals as shown in the lower graph.

In addition to the dropout detection improvement, newer and better compensation methods are implemented. Conventional dropout compensators simply replace the signal containing dropouts by video information from the previous line. In particular dropouts as long as several video lines create visible errors and unwanted picture disturbances.

The new approach uses newer dropout concealment algorithms as known from modern digital VTRs. Three-dimensional pixel error concealment is used to recalculate defective pixels from adjacent pixels in horizontal, vertical and diagonal directions as well as temporal interpolation using pixels from previous and following fields.

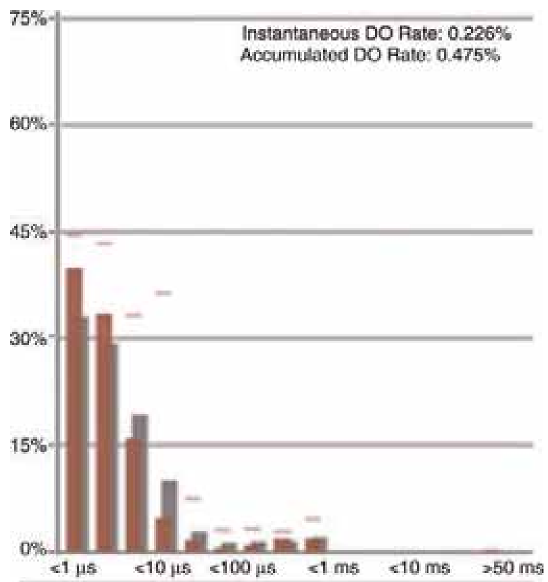


Figure 7 *Dropout histogram Betacam playback*

8 Audio and time code digitisation

Audio and time code is digitised in a conventional way by using 24-bit audio A/D converters.

Keeping video and audio in perfect synchronisation is essential for correct transfer of content from tape to file. In an offline process, where video, audio and TC is kept separately, the risk of losing lip synchronisation is high. It is therefore important that successive file processing of video and audio uses one single source for time reference. Longitudinal time code (LTC) is digitised in parallel with audio and video and is stored as reference time-stamp metadata to keep lip sync and to be transferred into the final target file format, which in most cases will be MXF.

9 Implementation and system aspects

Direct RF signal digitisation, as described in this paper, is marketed as 'QUADRIGA Video' by Cube-Tec International GmbH. The system consists of the modification kit for VTRs, an external A/D converter system for the RF signals, audio and time code and the PC system which runs the playback software and metadata creation. The output file format is DPX, which allows any other coding system to create the desired target coding and wrapping format.

Several ingest channels including coding to the desired target file format need to be controlled to ensure efficiency of the entire archiving factory. Fig. 9 illustrates this approach.

QUADRIGA Video is used for highly automated and multichannel transfer of legacy tape archives. Comprehensive sets of technical metadata that are used for description and quality control of the transfer process are generated. The following list shows the current implementation of metadata. Depending on the user requirements many other metadata can be created.

- exact dropout position within a video line
- logging of all prevented and concealed dropouts
- H- and V-sync timing errors logged before time base correction
- Y- and C- timing errors before correction
- video, blanking and sync levels before gain control
- detection of vertical interval subcarrier (VISC) reference pulse (frequency and phase)

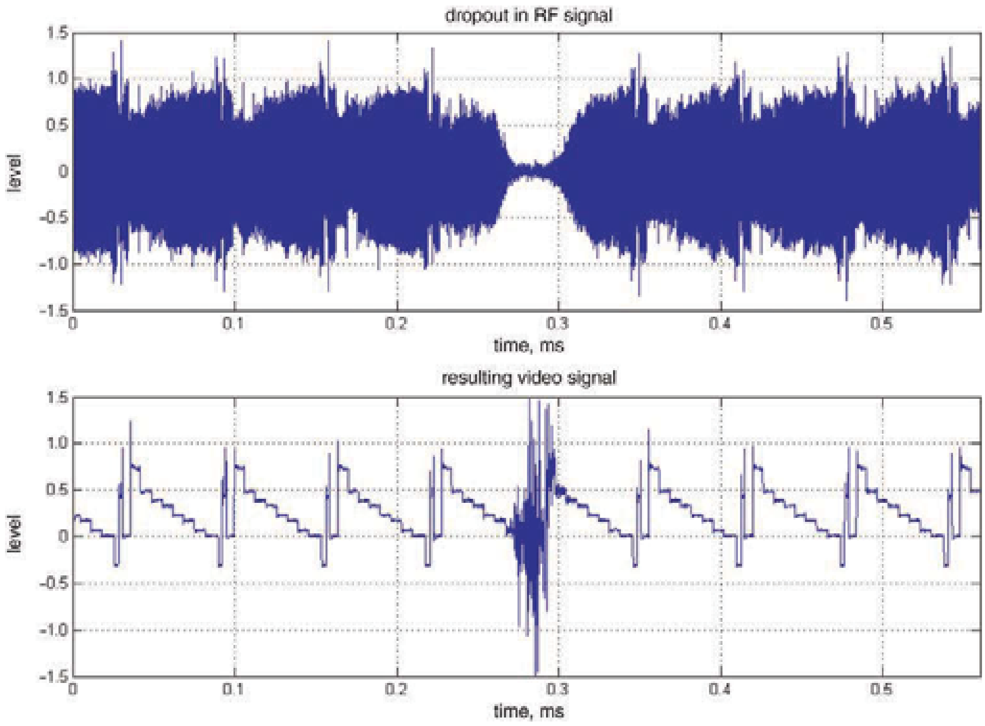


Figure 8 *Real dropout of RF playback signal and demodulated video*

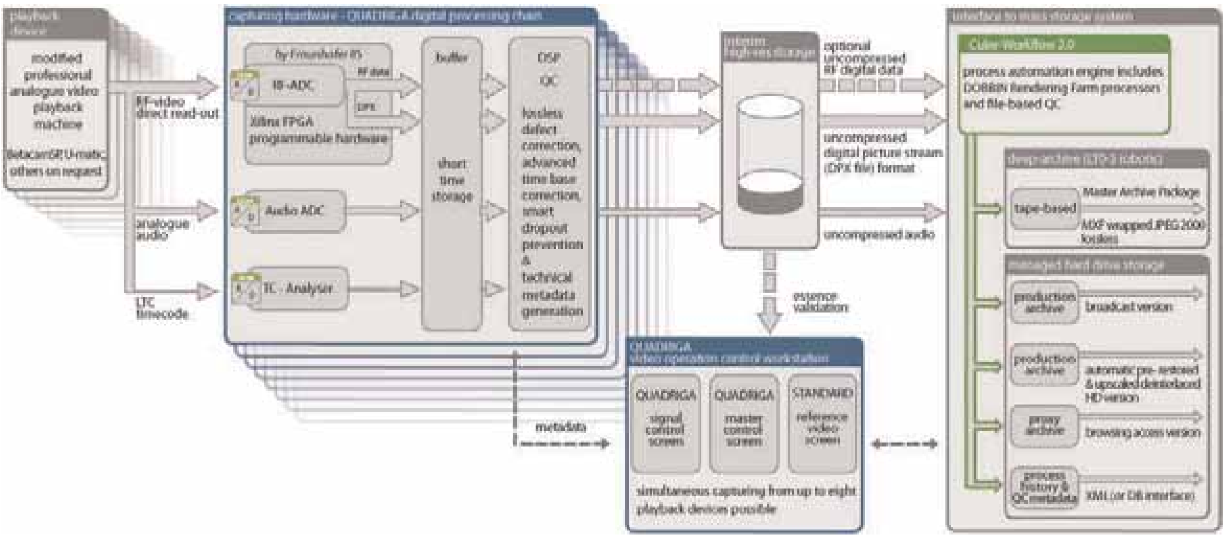


Figure 9 *Scalable video tape ingest system*

- detection of 8V-ID pulses
- logging of vertical interval time code (VITC)
- plausibility of 8V-ID, LTC, VITC and VISC

10 Conclusion

Direct digitisation of RF playback signals in analogue VTRs and successive software-based signal processing provide many advantages over conventional VTR playback.

- signal quality can be improved, because high bit-depth processing is possible and multiple AD/DA conversions can be avoided
- dropouts can be prevented by smart dropout level detection
- better dropout concealment can be used to mask non-correctable dropouts
- higher confidence in the transfer process, because RF signal analysis allows detection of PB errors, tape transport miss-alignments for both playback and recording

• metadata creation for better quality control and documentation

• flexibility to transfer into any target coding and wrapping format

In summary, this new approach for the transfer of analogue tapes into digital files is the most comprehensive method to secure valuable content which is still stored in shelves and waiting to be transferred into the file-based future. Historically and culturally important material can be preserved with the highest quality possible.

11 Acknowledgments

The author would like to thank his colleagues T. Mayer and J. Houpert from Cube-Tec International and the team from the Fraunhofer Institute for Integrated Circuits for their contributions to this work.

12 References

[1] BURGHARDT JUERGEN: 'Handbuch der professionellen Video Recorder' (Vereinigtes Filmcontor zu Essen e. V., 1994), ISBN: 3-930 524-01-5

[2] SONY BVW-75P Service Manual

[3] SONY BVW-75P Operational Manual

Selected content from IBC2012

ISSN 2041-5923

Likes, tweets and diggs: the impact of social media on viewing behaviour

C. Connor[1] *N. North*[1] *T. Weiss*[2]

[1]*GfK*
[2]*Red Bee Media*
E-mail: nick.north@gfk.com

Abstract: The advent of social media has enabled viewers to interact with their favourite television programmes in new and exciting ways. However, little research has been undertaken to understand in detail how consumers engage with social media around television programming, and what value this provides to broadcasters. The results of new research by GfK and Red Bee Media into the behaviours of users on social media are described in this paper: the different ways that viewers interact with programming; attitudes towards social media and its influence on viewing choices; and what the audience values from a social TV experience. To highlight the challenges of predicting mainstream future behaviour, it compares and contrasts the views of the average consumer with the views of those working in the industry and how closely (or not) the two align. We conclude with an analysis of how content owners can successfully engage with audiences beyond broadcast.

1 Introduction

Television has proved an enduring topic of conversation for over 50 years. Social media provides a platform for this conversation, enabling viewers to share opinion across a virtual sofa with friends before, during and after a broadcast, and to interact with the programme makers themselves. Consequently this has led to far reaching changes in the television industry ranging from content commissioning, through to new user interfaces and a fresh breed of companion applications.

It is anticipated that social media will play a key role in shaping the future of television. Understanding the future impact of social media on television viewing will inform a broadcaster's strategy, in its own *proactive* social media campaign planning; and also in building *reactive* mechanisms to harness consumer opinion to shape programming, scheduling and content discovery.

At the time of writing, there has been little careful evaluation of the value of social media to a broadcaster's performance. What is the impact of social media on viewing choice, on enhancing or enriching the viewing experience? How can social interaction drive appointment to view and programme loyalty? Understanding these will influence a programme brand's commitment to a social media presence, to ensure that its social media budget is being invested most efficiently and effectively.

For the broadcaster, what meaning or value can be placed upon a Facebook like? How representative or influential is social media? Do tweets have disproportionate influence on a broadcaster's decision-making process?

1.1 Background

Red Bee Media is one of the world's leading media management companies providing multiplatform technology and creative solutions to broadcasters, content rights holders, platform operators and brand owners.

GfK is one of the world's leading market research companies, with more than 11,000 experts working to discover new insights into the way people live, think, shop, interact and consume media in over 100 countries, every day.

For our Media & Entertainment clients, our focus is to understand the way that media fits into consumers' lives, offline and online across devices, platforms and services: how consumers engage with content across TV, print, online, mobile, on social media and more. We are leaders in audience measurement, capturing media choices and audience profiles. We track the sales of physical and digital entertainment products – VoD, DVDs, music, books and games – and we identify emerging trends. To complement the hard facts, we measure the drivers of consumers' media choices, their loyalties and their changing patterns of consumption and spend.

The media and entertainment industry drives and is driven by changing technologies and changing consumer expectations and behaviours. To ensure we can continue to provide our clients with relevant information, we pay particular attention to emerging trends, to identify the mainstream of the future.

1.2 Predicting the future – ignorance and bias

When assessing future risk or opportunity, we listen to our clients and we listen to consumers. In this paper, we wish to highlight the opportunities for media owners in forging deeper bonds with consumers through social interaction. Surveying specialists in broadcast and social media, we gain an expert view of key trends, informed by awareness of technological developments not generally known by consumers.

At the same time, we also wish to reflect on some of the threats to the accuracy of any predictions about future behaviours which are created by a combination of availability bias and confirmation bias. What comes most readily to mind – typically determined by significant events in our own experience – shapes and can distort our perception of the probability of the occurrence of an event. And if we have invested significant time or money in something that offers us a perceived benefit, we tend to believe that this is worthwhile, and that others will share our opinion. How could anyone live without an ipad, a PVR, Facebook?

In preparation of this paper for IBC, to gain the expert view and also to highlight differences between the tech-savvy, tablet-wielding media industry executives and the TV viewing population, we conducted research among a sample of industry executives, drawn from a database supplied by Red Bee Media, together with consumer research among a representative sample of the online population.

To address the core concerns of this paper, relating to the use of social media in conjunction with television viewing, we have used this primary research in conjunction with numerous additional sources from Red Bee Media and GfK:

- we have drawn on Red Bee Media's continuous monitoring of tweets and Facebook likes relating to TV programme content running in the UK and the Netherlands. The Facebook likes are drawn from a sample of over 200,000 users, who have installed TV guides using Red Bee Media's Content Discovery Platform (CDP)

- with kind permission of ITV in the UK, we have analysed the role of social media in driving viewing choice from a representative panel of 8,000 people in the UK administered by GfK on behalf of the UK's largest commercial broadcaster

- from our Media Efficiency Panel in Germany, in which we measure television viewing, online and mobile behaviour and consumer purchases, we have been able to analyse simultaneous use of social media and television viewing

2 Influence of social media

The world has become increasing social. 82.4% [1] of the world's online population are socially connected, with Facebook the number 3 site globally. From a media perspective, this has created new and exciting opportunities for consumers to share their opinions about programmes; for programme brands to get closer to consumers; and for media brands to create platforms for viewers to fully interact and engage with their favourite programmes.

A consumer's Facebook profile may carry a rich array of brands and programmes they like. But what does it all mean? A quote from *Social Media Futures: The future of advertising and agencies in a networked society* aptly sums up the situation today: 'Knowing lots of people is not the same as influencing lots of people' [2].

If we analyse a sample of Facebook users who have downloaded Red Bee Media's CDP powered Facebook TV guide application, we find very significant differences between top liked shows and TV audience size (see Table 1). None of the top ten viewed shows (according to BARB) feature in the most 'liked' programmes.

Table 1 Top 10 programmes by Facebook likes

Most 'liked' programmes	Audience size [3]
The Simpsons	1,669,900
Family Guy	869,200
Top Gear	1,407,900
The Inbetweeners	1,495,800
The Big Bang Theory	2,043,500
Blackadder	143,700
Scrubs	162,100
Star Wars	Not broadcast
South Park	282,500
The Young Ones	Not broadcast

Evidently, liking a programme is not time-dependent, it remains as a profile characteristic long after the programme has aired. This then raises a question of the relevance or value of fan status of a brand no longer available. With the 'like' symbol now a familiar feature on many websites; understanding how consumers perceive and use such features in relation to TV brands is increasingly important in any prediction of future performance. More relevant would be to analyse the content of Facebook posts by fans of programmes, which are not readily available.

If we look instead at Twitter, we have more time-sensitive activity that we can readily analyse. Of course, Twitter does not have the market penetration of Facebook, with only 10% reach in the UK, but if we consider it a measure of the zeitgeist it may be expected to reflect the current conversation about television, in terms of programmes discussed and the volume of tweets proportionate to audience size.

Looking at Table 2, showing the audience size of the top-tweeted shows, it is certainly the case here that the Twitter activity reflects the current broadcast schedule more (but not entirely). Britain's Got Talent (BGT) attracts higher Twitter participation than its BBC competitor The Voice. But Made in Chelsea attracts disproportionately high volumes of Twitter activity relative to audience size.

If we analyse the pattern of activity throughout each episode (see Fig. 1), it is also interesting to see how there are more interactions for Made in Chelsea while the programme is being broadcast, whereas for both The Voice and Britain's Got Talent, viewers are waiting for natural breaks in the programme to comment.

If we compare the top Facebook liked programmes versus the top tweeted about programmes, in the UK and also in the Netherlands, what is striking is the very international composition of the most liked programmes versus the local market broadcast focus of the tweeted programmes: in the Netherlands, 9 of the top 10 tweeted shows are Dutch.

Table 2 BARB audience size of most tweeted shows

Most tweeted programmes	Audience size
Britain's Got Talent	10,631,500
Made in Chelsea	916,800
The Apprentice	7,459,500
Crimewatch	Not broadcast
Homeland	3,490,900
Question Time	2,933,900
The Voice UK	10,271,000
One Tree Hill	537,600
EastEnders	8,072,800
90210	540,800

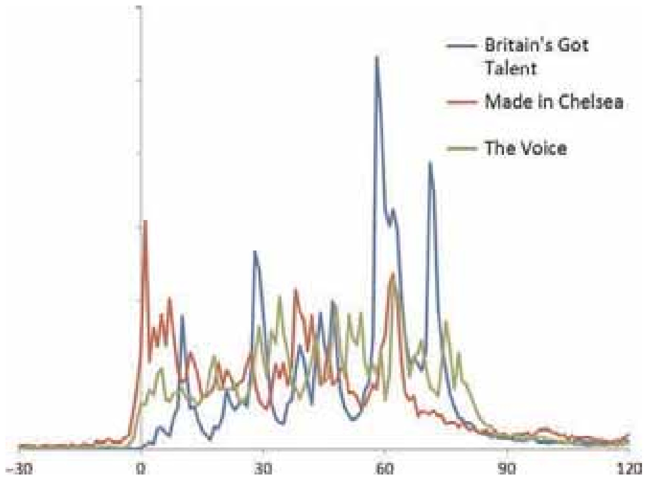


Figure 1 *Comparative analysis of twitter activity by programme*

An analysis of volume of tweets and programme audience size shows that Twitter does indeed reflect the current programme schedule, but there is little correlation between the size of audience and volume of tweets.

But does Twitter reflect the opinion of the audience?

We analysed the key themes relating to a recent talent show broadcast in various European markets. When we studied the opinions about the show from a nationally representative sample, we saw some key themes emerge. When we conducted a similar analysis of social media comments relating to the same show, we found that the conversation was much more oriented towards the private lives of the celebrities featured, much less about the content of the programme itself.

3 Drive to viewing?

Clearly liking a programme is an important factor in future viewing choice. But it is also interesting to note that social media likes or recommendation on their own are yet to account for significant role in viewing decisions. Table 3 details the ten programmes that show the highest influence of online sources (including social networks and blogs) in driving viewing behaviour. This list has been assembled from data of a UK representative panel managed by GfK on behalf of the UK's largest commercial broadcaster, ITV. On a daily basis, respondents are asked questions about what they watched on TV yesterday, how they came to view, and the quality of their engagement with the programmes viewed. Even among these programmes with the strongest social media drivers to view, previous viewing behaviour, listings (either EPG or magazines) and channel hopping still dominate viewing decisions. Previous viewing

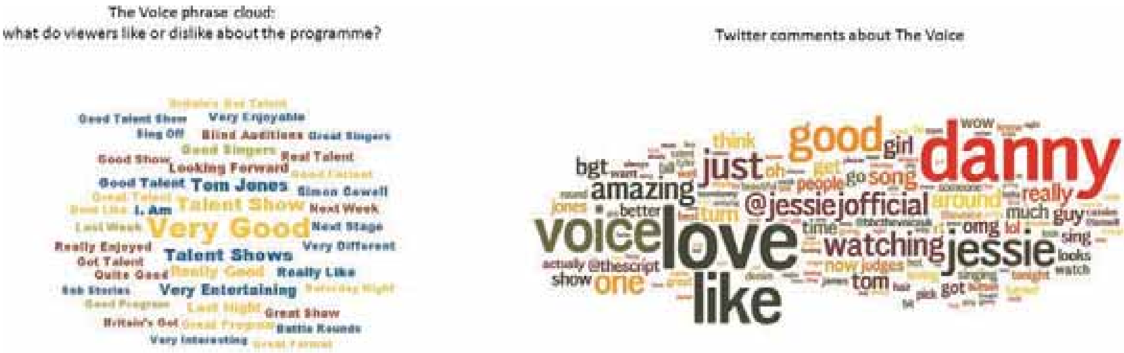


Figure 2 *ITV panel comments about The Voice vs Tweets about The Voice*

Table 3 Top ten programmes by online recommendation in quarter 1, 2012

	Watch regularly/ watched before	Came across via EPG	Saw in TV Listings	Channel hopping	Saw a promotion on TV	Heard it talked about on the radio	Read something about it in newspapers/ mags	Saw a poster	Word of mouth (physical)	Online
	%	%	%	%	%	%	%	%	%	%
Stargate SG-1	94	52	38	26	47	42	47	42	20	16
American Idol: Performance	86	2	–	6	13	6	5	1	3	10
Obese: A Year to Save My Life	64	20	10	13	23	15	20	29	5	10
American Idol: Auditions	88	4	15	4	7	3	–	–	9	8
Crufts 2012: Best in Show	75	9	14	2	15	–	–	–	5	6
Grand Designs	85	12	11	15	7	6	9	6	6	6
Make Bradford British	33	5	15	1	46	2	5	–	8	5
Agatha Christies Poirot	96	9	6	–	–	–	–	–	–	5
The Bank Job	45	11	8	9	27	3	3	–	5	5
Big Fat Gypsy Weddings	63	14	9	15	3	–	3	–	–	5

Source: GfK/ITV Vision Panel Data from 1 January – 30 April 2012, based upon programmes available from 18 channels behaviour is the strongest indicator – not surprisingly. If viewing history can be incorporated into a social media profile, however, while traditional promotions are still the main drivers to viewing, it is interesting to see how different programme genres have more of an online presence than others.

3.1 What do the audience actually want from social media?

In preparation of this paper, we conducted a short survey amongst a representative sample of the online population, and among those working in the broadcast industry (either content providers, platform or technology suppliers), drawn from Red Bee Media, to understand how social media is perceived, how it is used – and, interestingly, how closely aligned the two groups are.

One of the most striking differences between the two groups was the difference in devices used to access the internet. Just over one in ten (13%) of the consumers surveyed regularly use tablets to access the internet. In comparison, well over half (58%) of the professionals used tablets regularly. This difference in usage is reflected in perceptions of how social media is being used – with industry professionals using their personal experience as a means of benchmarking what they think the population in general is doing. This is explored in more detail below.

Of the 500 consumers surveyed, one in three had 'liked' a TV programme. This figure was lower than for brands or products, and significantly lower than likes for posts by friends or family.

Research has shown, and indeed it is pretty self-evident, that an emotional bond will increase the likelihood of social media activity, to seek out the means to express and affirm the attachment. No surprise therefore that TV programmes do not attract as many likes as our friends.

More interesting was the reaction to Twitter – and particularly the gap that exists again between consumers and professionals. Consumers are very polarised in their views towards Twitter. When asked to comment on what they thought about Twitter in general, one in three commented positively, one in three commented negatively and one in three were ambivalent. Nevertheless, one in ten consumers who had either tweeted or read a tweet about a TV programme felt more involved as a result. However, amongst all the professionals, four in five thought it made the viewer feel more involved.

Generally, those in the industry feel that consumers are more connected with programmes because of the advent of Twitter.

Perhaps more interesting was the reaction to adding social media elements to all content. One in two consumers were indifferent to the concept – with one in four thinking that there should be an element for 'some programmes – not all' and a similar number completely against the idea. However, amongst those in the industry, three in four thought that there should be a social media element for some programmes and only one in five indifferent to the idea.

3.2 How is social media impacting on viewing behaviour?

The increasing proliferation of screens in the living room (TV, laptop, tablet, mobile) has led to a rising interest in knowing how the different devices are being used and how interlinked they are, and how social media activity complements television viewing.

In Germany, GfK runs the Media Efficiency Panel (MEP) which measures TV viewing, online usage using meters, and purchase behaviour using barcode scanners. From this panel of over 16,000 households, we are able to measure the time spent online while watching television.

The results show clearly that among those households with online access at home and television, dual screen use is a familiar activity, with over three quarters of households showing some evidence of parallel use of online and television during a month (one minute or more of simultaneous use), and growing.

While this parallel use represents a relatively small and declining share of all television time, it represents a growing 34% of all time spent online at home: more than one third of our time online at home is in front of the television, showing clearly how important the internet is becoming as a valuable companion to our viewing.

We are able to drill down to the individual sites and applications used during these dual-use sessions.

Not surprisingly, Google and Facebook are the sites most frequently visited, both in total and during dual-use sessions.

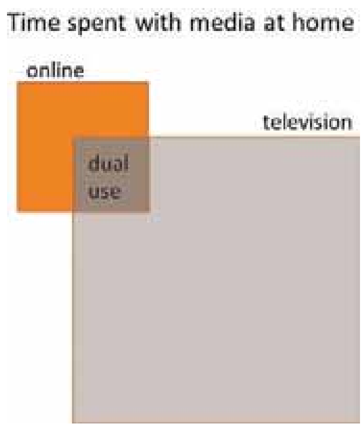


Figure 3 *Time spent by activity (GfK Media Efficiency Panel, 2012)*

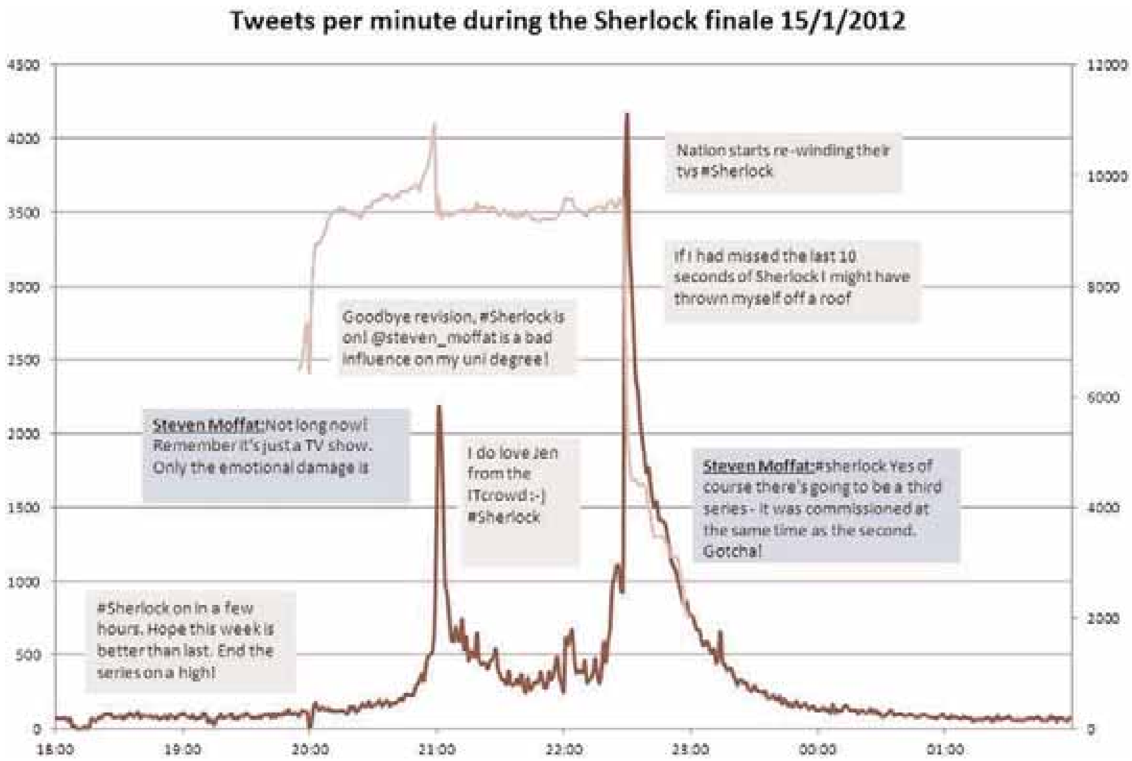


Figure 4 *Twitter activity during Sherlock*

If we compare the percentage of all visits to each site during dual-use sessions, and compare this to the percentage of all visits to each site in total, we can produce an index of relative likelihood to visit a site. Here we see that while 8% of all visits were to google.de, only 7.3% of visits during dual use were to google.de, delivering a dual-use index of 94. Facebook delivers a dual-use index of 110, suggesting that there is a slightly higher propensity to socially engage during television viewing sessions than overall.

What is also of note is that the type of sites showing the highest dual-use index are games websites – spielmit.com, Farmville and the rest. This suggests that we may spend some of our shared viewing time entertaining ourselves while our friends or family enjoy watching the programme more. And among younger viewers, the capacity to play games, watch TV and engage socially is a well-known stereotype, which we see evident in the data.

3.3 Complementary dual use

Within the dual use sessions, the question remains whether viewers are texting or commenting on what they are watching – or is any social interaction completely unrelated to the main screen in front of them?

To explore this question, in order to accumulate higher volumes of social data than can be derived from the MEP, we turn to Red Bee Media's monitoring of Twitter, coupled with an analysis of BARB, the TV industry currency in the UK.

Fig. 4 overlays audience data with twitter data during the final episode of *Sherlock* broadcast on BBC1 in 2012. It is interesting to note how the Twitter peaks correspond with the audience peaks – and that during the programme itself, the activity drops away substantially, suggesting high engagement with the programme itself.

4 Conclusions

Social media has provided a new means for viewers to interact and engage with their favourite programmes. It is striking how great a share of time online at home is as an accompaniment to television viewing. It is also clear that social media serves different purposes for different programmes at different times, and understanding this dynamic is important to ensure that resources are allocated efficiently and in the right areas. As Facebook 'likes' are not time specific, different measures of consumer engagement may be more important for media brands than for consumer brands, at least while our 'likes' don't fully represent our viewing histories. Facebook nevertheless is clearly a highly relevant platform for conversation about media brands, with its use disproportionately high during dual use sessions. Updating Facebook pages with new content between broadcasts helps to ensure programme buzz is maintained between episodes. Twitter activity more closely reflects viewing, demonstrating its role as a TV companion.

The viewers themselves see social media as providing new ways of engaging with programmes but not necessarily needed for all programmes. Understanding the different patterns of social media interaction will help develop a strategy that complements the main viewing experience, and ultimately creates a more engaged consumer. With a growing share of companion use, the opportunity for social media and content discovery tools to play an increasingly important role in shaping our viewing choice is evident.

5 Acknowledgments

The authors would like to thank colleagues at GfK NOP Media, at GfK Panel Services Deutschland, and at Red Bee Media for their contributions to this work. They would also like to thank the International Broadcasting Convention for permission to publish this paper.

The results shown in Fig. 2 are reproduced with the permission of ITV.

6 References

[1] comScore Media Metrix October 2011

[2] IPA/ Future Foundation Delphi Stage 1

[3] BARB April 2012 adults aged 16+

Selected content from IBC2012

ISSN 2041-5923

Impact of delivery ecosystem variability and diversity on internet video quality

A. Ganjam P. Pappu I. Stoica J. Zhan H. Zhang
Conviva, San Mateo, CA 94403, USA
E-mail: hzhang@conviva.com

Abstract: In this paper, we study the impact of the variability and diversity of multiple internet delivery ecosystem components (CDNs, ISPs, devices, video players, streaming protocols) on the quality of internet video. Our study is based on a data set that is unique in two aspects: 1. *client-side instrumentation*: this provides critical insights into video quality as the video has traversed all ecosystem components before reaching the client; 2. *large-scale*: our data set spans scores of popular sites, billions of streams, hundreds of thousands of video objects, and hundreds of millions of unique viewers. We show that there is large variability and diversity in each of the many internet video delivery ecosystem components, which can significantly impact the quality of video that is delivered to the end users. We present a set of architectural principles that are essential to optimise video quality in the presence of variability and diversity.

1 Introduction

The internet, as a distribution mechanism for video, presents huge opportunities for publishers and consumers. In particular, the internet's *global reach* enables publishers to connect with consumers anywhere without the geographical constraint of traditional physical access networks such as cable, broadcast, and satellite networks. Content publishers such as Netflix, HBO, BBC, and CNTV have taken advantage of this by launching internet video services in multiple markets around the globe. In addition, the internet allows consumers to access content via multiple devices including PCs, smart phones, pad devices, and TVs, i.e. 1st screen, 2nd screen, and 3rd screen. Finally, internet video enables publishers to introduce rich social, interactive, and on-demand features.

To realise these benefits, a number of sophisticated technologies are used to implement a complex internet delivery system. Fig. 1 shows the typical end-to-end delivery pipeline for internet video delivery. In a nutshell, once the video is captured, it is encoded and published to one or more video origin servers. From there, the video is either pushed or pulled by content distribution networks (CDNs) and delivered over the internet to client devices, where it is decoded and played back.

If any of the modules in this pipeline experience issues, the quality will be impacted. The probability of failure is exacerbated by the huge diversity of devices, protocols, and systems that take part in the delivery, and the fact that different components are run and managed by different administrative entities. For example, typically, a content publisher uses different CDNs and ISPs to deliver video to different devices and players that are all built and managed by different vendors.

In this paper, we study the impact of the variability and diversity of multiple internet delivery ecosystem components on the quality of internet video. In particular, we examine the following key components: CDNs (e.g., Akamai, Limelight, Level 3), ISPs (e.g. ATT, Comcast, BT, Telefonica), consumer devices (e.g. PC, iPhone, iPad, Xbox), software video players (e.g. Brightcove, OSMF, Platform), and streaming protocols (e.g. RTMP, Akamai HD1, HDS, HLS, SmoothStreaming). Our study is based on a large data set that has been gathered with a highly scalable data collection system over billions of streams, hundreds of thousands of video objects, and hundreds of millions of unique viewers. The system consists of two parts. 1. a client-resident instrumentation library in the video player; and 2. a data aggregation and processing service that runs in data centers. The client library gets

Figure 1 *Encoding to playback end-to-end video delivery chain*

loaded when internet users watch video on an instrumented site. The library listens to events from the video player and additionally polls for statistics from the player. Because the instrumentation is on the client side we are able to collect very high fidelity, raw data that captures the video quality after video has traversed all the ecosystem components. We collect and process 1TB of data on average per day from various content providers over a diverse spectrum of end users, video content, internet service providers, content delivery networks, and geographical regions.

2 Performance metrics and methodology

For the analysis in this paper, we first compute the performance for each video session. While there are several metrics that can be used to characterise the performance of a video session, we focus on (a) buffering ratio, which is defined to be the total amount of time a video player is spent in a re-buffering state (play-out buffer in empty and no video is showing on the screen) divided by the total video session duration; (b) the average of the bit rate of the session. Most of the video sites today implement adaptive streaming technology in which the video is encoded in multiple bit rates and a video player can dynamically pick a bit rate.

We pick these two per session metrics for the following reasons. First, as shown in an earlier study [3], buffering ratio is the most important metric with respect to impact on user engagement. Second, as users demand more HD quality content, the goal of adaptive streaming technology is to have a user to view the content at the highest bit rate video subject to the capacity availability of ISP, CDN, and device at the moment. In general, it is difficult to achieve both lower buffering ratio and high average bit rate simultaneously.

We compare two instances of one ecosystem component by computing the aggregate metrics for sessions that pass through the components. For example, when we compare the performance of two CDNs, we use the average buffering ratio across all the sessions that are streamed from each of the two CDNs.

3 Variability of CDN and ISP performance

Today video content publishers use a variety of CDNs, including global CDNs, regional CDNs, operator CDNs, and in-house CDNs. Each of these CDN types varies in its reach, peak performance, support for delivery protocols and other characteristics. While most CDNs offer standard streaming protocols, some emphasise proprietary protocol– further increasing diversity.

- **Global CDNs:** These CDNs are the largest, provide the widest coverage, and support most of the standard protocols and content protection technologies. However, their delivery quality can vary on a per country basis, as the amount of investment in the infrastructure and ISP peering varies.

- **Regional CDNs:** These CDNs focus on providing service in certain geographic regions (e.g. Northern Europe). They are not as large as global CDNs, and may focus on specific market segments (e.g. live events).

- **Operator CDNs:** These CDNs are run by ISPs. While they are typically cheaper than other CDNs, their delivery is restricted to the ISP's users. Operator CDNs are fairly new to the market, and have not built the level of reliability and support capabilities that Global and Regional CDNs have. The streaming protocols and content protection schemes they support may vary across operators.

- **In-house CDNs:** These can range from a set of servers to a fully distributed CDN infrastructure and are operated by the content publisher. They are feature limited, as they specifically target the needs of the content publisher.

ISPs are indispensable links in the video distribution chain. Unfortunately, the performance of the ISPs can vary dramatically, owing to an array of factors, such as different communication technologies, service models, and peering strategies. For example, wireless ISPs tend to offer less stable and lower capacity connections than their wired counterparts. Even two wired ISPs can have different performance profiles as they use different datalink technologies (e.g. DSL vs. cable modems). Using the same technology does not guarantee the same streaming quality either, as some ISPs may use traffic limiting, which can significantly impact video quality. Finally, direct peering with top CDNs enables ISPs to provide better quality by avoiding transit through other networks.

Fig. 2 illustrates the variability in performance across both CDNs and ISPs. Each column corresponds to an ISP, and each stacked bar within a column corresponds to a CDN. The height of each bar within a column shows the percentage of time for which that CDN has outperformed the other two. There are two points worth noting. First, the

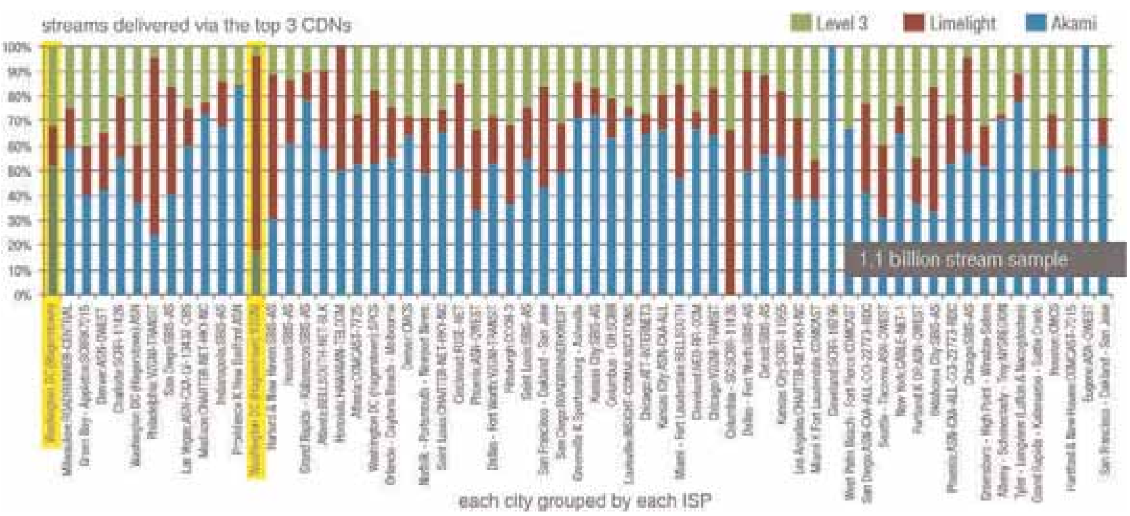


Figure 2 *Variability of CDN performance across ISPs*

performance of every CDN can vary widely from ISP to ISP. Second, there is no single CDN that outperforms the other two for all ISPs. In particular, while in many cases the blue CDN outperforms everyone else, there are quite a few cases in which the green CDN provides the best performance.

Next, we illustrate the variability of CDN performance in time. Fig. 3 plots the percentage of streaming errors over a two month period for three CDNs. Again, there is no single best CDN, and at different moments of time there are different CDNs performing the best or the worst. This suggests there is a significant potential for improving the streaming quality by dynamically picking the best CDN at any given time. While in general the performance of the CDNs is not correlated, there is one moment of time (around 1/5–22/5) when all CDNs experience spikes in the streaming errors. This corresponds to a high profile event that overloaded the origin causing failures on all CDNs.

4 Diverse devices and video players

When it comes to devices and video players, we are witnessing an even higher heterogeneity than in the case of CDNs and ISPs. A typical content publisher supports a wide array of devices and players to enable video viewing 'anytime and anywhere' on PCs, phones, and TVs. Devices can vary significantly in capabilities (e.g. support for adaptive bit rate, support for content protection etc.), performance (e.g. memory footprint of apps, CPU usage), and in the development environment (e.g. programming language). Table 1 shows the capability along three key dimensions across several popular devices.

Differences in streaming protocol and content protection usually mean the content publisher must prepare multiple copies of content. Differences in development language and player development frameworks mean the content publisher must build different applications. In addition, lack of standardised support for streaming protocols or content protection on a device leads to proliferation of third party solutions or lack of adoption. Android is a good example of the first case. Android only supports HLS in the standard player starting from version 3, which is only used on tablets. A majority of Android phones in the market at version 2.x and up require a third party solution to play HLS content. This quickly resulted in several third party solutions for Android, further complicating the delivery chain. A good example of the second scenario where the platform is not adopted is HTML5 video. While HTML5

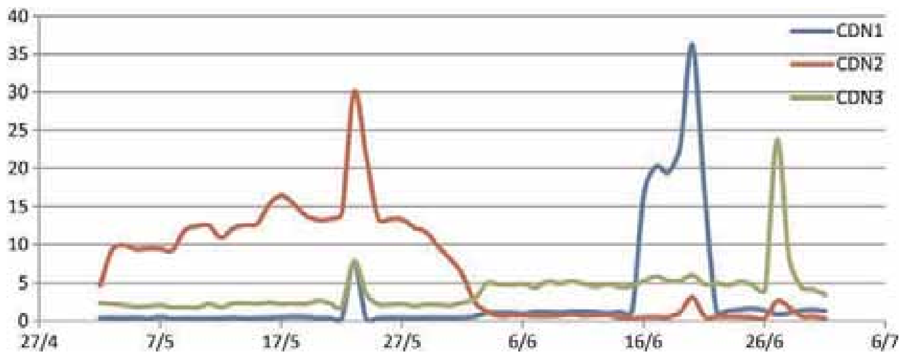


Figure 3 *Performance variability of CDN performance (streaming error %) across two month period*

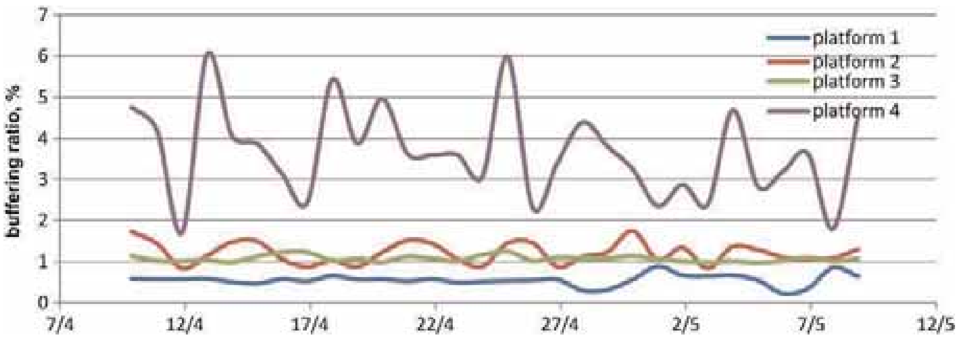


Figure 4 *Performance variability (buffering ratio) among iOS vs. Flash vs. Xbox vs. Android players*

video is gaining significant adoption in general, it lacks adoption by premium content publishers owing to the lack of content protection.

Performance of different platforms also varies significantly, depending on the maturity of the platform, protocols and technologies available on the platforms, etc. Fig. 4 shows that for the same content provider, the buffering ratio on four different platforms (i.e. iOS, Flash, Xbox, and Android) is quite different from each other. Buffering ratio represents the percent of time viewers spend in a re-buffering state.

The quality of the picture shown to viewers also varies between players on the same platform. Fig. 5 shows the percentage of views with over 1 Mbit/s average bit rate on the five different Flash-based video players. All five video players are production players for a major premium video website. All five implement adaptive streaming with multiple video bit rates. However, they differ in their adaptive streaming algorithm: when to switch a bit rate up or down and which bit rate to switch to. As discussed earlier, it is fundamentally difficult to achieve both high average bit rates and a lower buffering ratio. One way to avoid buffering is to be very conservative in switching to higher bit rates, thus resulting in a lower average bit rate.

5 Streaming protocols

Content publishers use a variety of streaming protocols to accommodate the diversity of devices and video formats. Streaming protocols fall into two categories: 1. stateful client-server protocols; and 2. stateless HTTP-based protocols. Examples of stateful protocols are RTMP (developed by Adobe for Flash), MMS (developed by Microsoft for Windows Media Player), and RTSP. Since these protocols require server-side support, CDNs need to deploy custom servers to support these protocols (e.g. Adobe's FMS or Microsoft's Media Servers). In contrast, stateless HTTP-based protocols chunk the video in small pieces that are distributed as regular files through the existing HTTP infrastructure. Examples of such protocols are Apple's HLS, Adobe's HDS, and Microsoft's Smooth Streaming. While these protocols can leverage the existing HTTP servers for distribution, they typically incur a higher end-to-end latency.

Table 1 Heterogeneity in device capability

Device / platform	Supported streaming protocols	Supported content protection	Development language
PC / Flash	RTMP, HDS, HTTP progressive download	URL Tokenisation, Flash Access DRM, SWF Verification, Protocol encryption	ActionScript
iPhone / iOS	HLS, HTTP progressive download	Encryption, Apple DRM and some third party DRMs	Objective-C
Xbox	SmoothStreaming HTTP progressive download	PlayReady DRM	C#
PC / Silverlight	SmoothStreaming, MMS, HTTP progressive download	PlayReady DRM, URL Tokenisation	C#
HTML5 Video	HLS, HTTP progressive download	No standard content protection	Javascript
Android	HTTP progressive download, HLS (3.x and later)	Encryption, some third party DRMs	Java
Roku	HLS, HTTP progressive download	Encryption	Brightscript

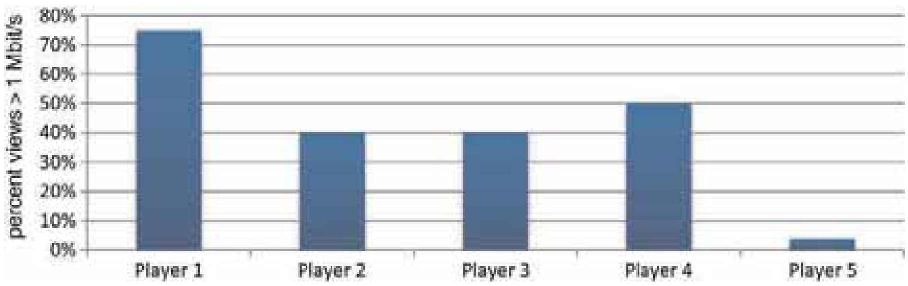


Figure 5 *Performance variability between video players in Flash (percentage of HD streams)*

The diversity of these protocols increases the complexity of the video distribution ecosystem, especially as different devices use different streaming protocols. For example, today, the common streaming protocol on PCs is RTMP, on iOS devices is HLS, and on Microsoft's devices is Smooth Streaming. Not surprisingly, the quality provided by these protocols can vary significantly. In particular, the buffering ratio varies from 1% to as high as 4.6%, while the percentage of views with the average bit rate higher than 1 Mbit/s varies from 17.5% to 56.9%. While there are standardisation efforts to migrate to MPEG-DASH, we expect this fragmentation to continue for the next few years.

6 Architectural principles for optimising video quality in the presence of variability and diversity

As has been discussed in the previous sections, the many components in the internet video distribution ecosystem exhibit large variability and diversity that can result in significant degradation of video quality. As internet video is becoming better monetised and is increasingly consumed on large screens, consumers' expectations for high quality will continue to increase. In the meantime, while we expect continuous improvement of technology in all components of the internet distribution ecosystems, the variability and diversity will take different forms but will remain key challenges for a predictable, consistent, and high quality video experience as demanded by both content publishers and consumers.

To insulate content publishers and consumers from this continuously evolving variability and diversity, powerful service-layer software is needed to optimise video quality in the presence of sophisticated delivery ecosystems with diverse and variable components. The detailed design of such service-layer software is beyond the scope of this paper. However, we do believe that any such service-layer software should incorporate the following architectural principles:

- The system should take advantage of the diversity and variability of the delivery ecosystem to solve the quality problem. For example, since there are usually independent failure modes among different CDNs, servers, and ISPs, content should be distributed via multiple CDNs. Each client video player should retrieve video from multiple servers in different CDNs via diverse internet paths. This concept of taking advantage of diversity to achieving predicable and consistent quality is a powerful technique used to solve other computer science problems. For example, the modern highly available internet services are built on top of large number of servers, each of which does have performance variability and failure modes. The traditional architecture of having a single CDN to distribute video is inadequate

- The optimisation mechanism should be placed in the client-side, NOT the server-side, to enable the client to dynamically select bit rate, CDN, server, or internet path. The traditional architecture of using a server-side mechanism is inadequate

- The optimisation should be continuous throughout the duration of a video session. The traditional architecture of DNS re-direction and DNS load balancing at the beginning of a session is inadequate

- The optimisation should be dynamic and based on the real-time condition of various ecosystem components such as ISPs and CDNs. The traditional architecture of using DNS re-direction and DNS load balancing based on stale network state information is inadequate

- The optimisation should be specific for each individual video player. The traditional architectural of optimisation based on average network state information is inadequate

7 Conclusions

The desire to achieve 'anytime, anywhere' video distribution poses great challenges for internet video publishers. A publisher has to deal with multiple encoder formats and profiles, CDNs, ISPs, devices, and a plethora of streaming protocols and video players to ensure accessibility of video to all viewers. We have shown that there is significant variability in each of these diverse components, and together this variability and diversity present huge challenges to delivering a consistent, predictable, and high-quality user experience. We argue that this quality challenge can be addressed by taking advantage of the diversity. In particular, the architectural principles are the following: 1. multiple CDNs should be used for video distribution, 2. the optimisation mechanism should be placed at the client instead of server, and 3. the optimisation should be continuous throughout the duration of a video session, based on the real-time performance of ecosystem components, and specific to each individual video player.

8 References

[1] Global Internet Phenomena Report: Spring 2011, http://www.sandvine.com/news/global_broadband_trends.as

[2] Cisco Visual Networking Index: Forecast and Methodology, 2009-2014, http://tinyurl.com/3p7v28

[3] DOBRIAN F., AWAN A., STOICA I., SEKAR V., GANJAM A., JOSEPH D., ZHAN J., ZHANG H.: 'Understanding the impact of video quality on user engagement'. Proceedings of SIGCOMM 2011, Toronto, Canada,

[4] ALI S., MATHUR A., ZHANG H.: 'Measurement of commercial peer-to-peer live video streaming'. Workshop on Recent Advances in Peer-to-Peer Streaming 2006, Waterloo, Canada

[5] HEI X., LIANG C., LIANG J., LIU Y., ROSS K.W.: 'A measurement study of a large-scale P2P IPTV system', *IEEE Trans. Multimedia*, 2007, **9**, (8), pp. 1672–1687

[6] MAHIMKAR A., GE Z., SHAIKH A., WANG J., YATES J., ZHANG Y., ZHAO Q.: 'Towards automated performance diagnosis in a large IPTV network'. SIGCOMM 2009, Barcelona, Spain

[7] SRIPANIDKULCHAI K., MAGGS B., ZHANG H.: 'An analysis of live streaming workloads on the internet'. Proceedings of IMC 2004, Taormina, Italy

[8] H.Y., *ET AL*.: 'Inside the bird's nest: measurements of large-scale live VoD from the 2008 Olympics'. Proceedings of IMC 2009, Chicago, IL, USA

Selected content from IBC2012

ISSN 2041-5923

ISDB-Tmm mobile multimedia broadcasting in Japan

T. Ohya[1] *T. Morizumi*[1] *T. Kondo*[2] *Y. Miyashita*[2] *C. Kamise*[1,2]

[1]*mmbi Inc., Midtown Tower 38F, 9-7-1 Akasaka, Minato-Ku, Tokyo 107-6238, Japan*
[2]*Japan Mobilecasting, Inc., Midtown Tower 37F, 9-7-1 Akasaka, Minato-Ku, Tokyo 107-6237, Japan*
E-mail: ohya@mmbi.co.jp

Abstract: A new multimedia broadcasting service for mobile terminals was launched in April 2012 in Japan, using vacant spectrum from 207.5 MHz to 222 MHz after the shutdown of analogue terrestrial TV broadcasting. The system uses ISDB-Tmm technologies with various new functionalities for the enhanced multimedia broadcasting services, such as file-based broadcasting, high-quality video and audio coding, advanced use of metadata, and interaction with communications channels. We developed an ISDB-Tmm based broadcasting station and installed a nationwide 14.5 MHz bandwidth single frequency network in Japan. This paper describes the new technologies employed in the ISDB-Tmm system, the overview of the new services, and also the overview of the broadcasting network and its field research results.

1 Introduction

In Japan, shutdown of the analogue terrestrial TV broadcasting was completed at the end of March 2012. Using the vacant spectrum from 207.5 MHz to 222 MHz (VHF high band), a new multimedia broadcasting service for mobile terminals, named 'Mobacas', was launched on April 1st 2012.

The system uses ISDB-Tmm (integrated services digital broadcasting for terrestrial multimedia) that is based on ISDB-T (integrated services digital broadcasting – terrestrial) [1] and has enhanced multimedia broadcasting capabilities, such as file-based broadcasting, high-quality video and audio coding, advanced use of metadata, interaction with communications channels etc. The physical layer transmission system is compatible with ISDB-T, and has excellent performance even under severe wireless channel conditions of mobile environments.

We developed a new ISDB-Tmm broadcasting system that is capable of both real-time and file-based broadcasting, and also installed a nationwide single frequency network of 14.5 MHz bandwidth.

This paper describes the new technologies employed in the ISDB-Tmm system, an overview of the new services, and also an overview of the broadcasting network and its field research results.

2 ISDB-Tmm services

As shown in Fig. 1, ISDB-Tmm provides two types of multimedia broadcasting services: 1. high-quality real-time broadcasting offering video, audio, and data, and 2. file-based broadcasting that can temporally store various combinations of video, audio, images, text, and data. The file-based broadcasting enables access to content and services without the user having to care about broadcast schedules or place of use.

As the ISDB-Tmm requires receiving terminals to have communications capabilities, such as 3G mobile or WiFi, it can provide coordinated broadcasting and communications services as shown in Fig. 2. The broadcasting channels can transmit large amounts of content, and the communications channels can be used to purchase licenses, to receive the list of recommended content, to post messages and to interact with communications-based services, such as social networking services (SNS). When content cannot be completely transmitted over the broadcasting channel, the remaining portion of the content can also be obtained using communications channels.

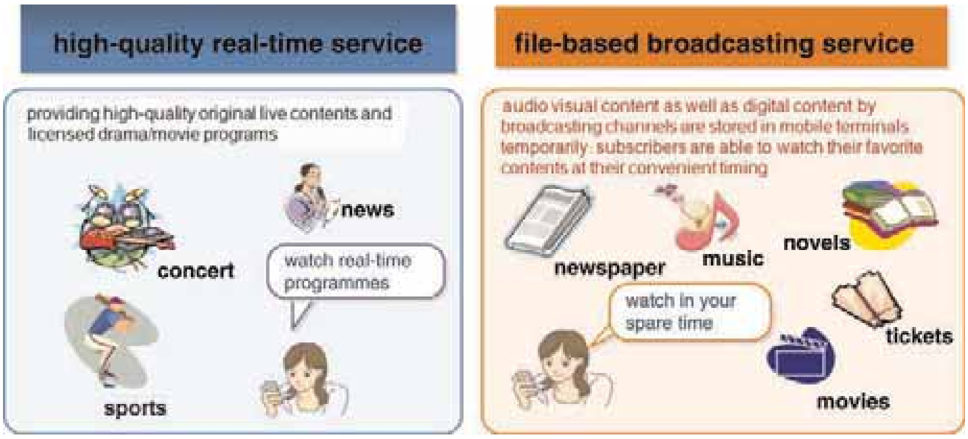


Figure 1 *Two types of ISDB-Tmm services*

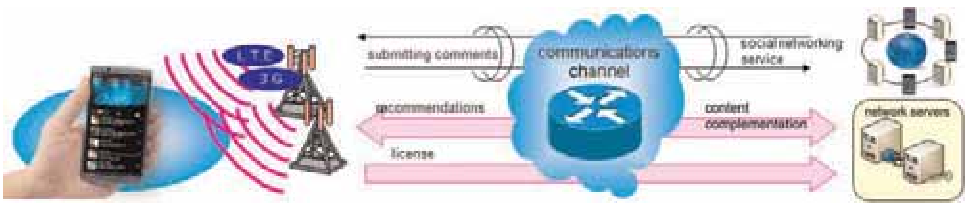


Figure 2 *Simultaneous use of communications and broadcasting channels*

Figure 3 *Example of collaboration between telecommunications and real-time broadcasting service*

An example of the coordinated broadcasting and communications service is illustrated in Fig. 3. The viewers share comments through SNS services on the same monitor as the broadcasting programme while watching sports games, concerts, live programmes etc., which enables viewers to maximise their interest and sympathy by viewing experience.

3 ISDBT-mm technologies

ISDB-Tmm is based on ISDB-T, a terrestrial digital broadcasting standard, so – similar to terrestrial digital TV, OneSeg and ISDB-Tsb (ISDB-T sound broadcasting) [2] digital radio – ISDB-Tmm employs orthogonal frequency division multiplex (OFDM) transmission technology and can operate on an single frequency network (SFN). It offers excellent reception performance in mobile environments and is resilient against multipath propagation.

3.1 Transmission system

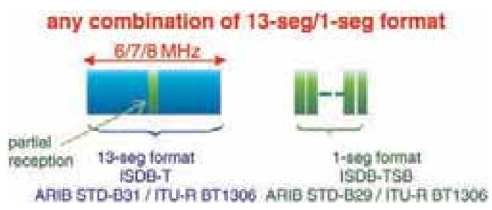
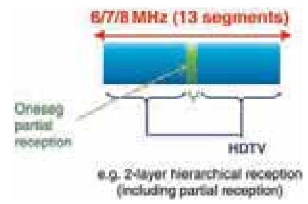
The ISDB-Tmm transmission system [3] can utilise any combination of the 13-segment format, which has a 6/7/8 MHz bandwidth and is compatible with ISDB-T, and 1-segment format, which has 1/13th the bandwidth of the 13-segment format and is compatible with OnSeg and digital radio as shown in Table 1. The remaining parameters are shared with ISDB-T and ISDB-Tsb. In Japan, 14.5 MHz – from 207.5 MHz to 222 MHz – is allocated to ISDB-Tmm. Two 13-segment formats and seven 1-segment formats can fit into the bandwidth as shown in Fig. 4.

3.2 Multiplexing

Fig. 5 shows the ISDB-Tmm protocol stack. The protocol stack for the real-time service is designed to be compatible with ISDB-T, considering the implementation to the OneSeg capable mobile phones. Moreover, since long-duration interleaving is good for large file transmissions, the file-based broadcasting service uses the AL-FEC (application layer – forward error correction) on FLUTE (file delivery over unidirectional transport) [4], in addition to the radio physical layer error correction.

Real-time services and file-based broadcasting services are packetised and multiplexed into an MPEG-2 transport stream (TS). The transmission bit rate can be instantaneously shared between two services.

Table 1 ISDB-Tmm transmission parameters

	ISDB-Tmm (ARIB STD-B46)	(Ref.) ISDB-T (digital terrestrial TV, Oneseg)
Spectrum usage		
Bandwidth/segment (Bws)	6000/14 = 428.57 kHz, 7000/14 = 500 kHz, 8000/14 = 571.428 kHz	
Number of radiated carriers/ segment	108 (Mode 1), 216 (Mode 2), 432 (Mode 3)	
Modulation	OFDM (DQPSK, QPSK, 16QAM, 640AM)	
Forward error correction	Outer code: Reed Solomon (204, 188) Inner code: Convolutional Code (7/8, 5/6, 3/4, 2/3, 1/2)	
Guard interval duration	1/4, 1/8, 1/16, 1/32 of active symbol duration	
Transmissionframe duration	204 OFDM symbols	

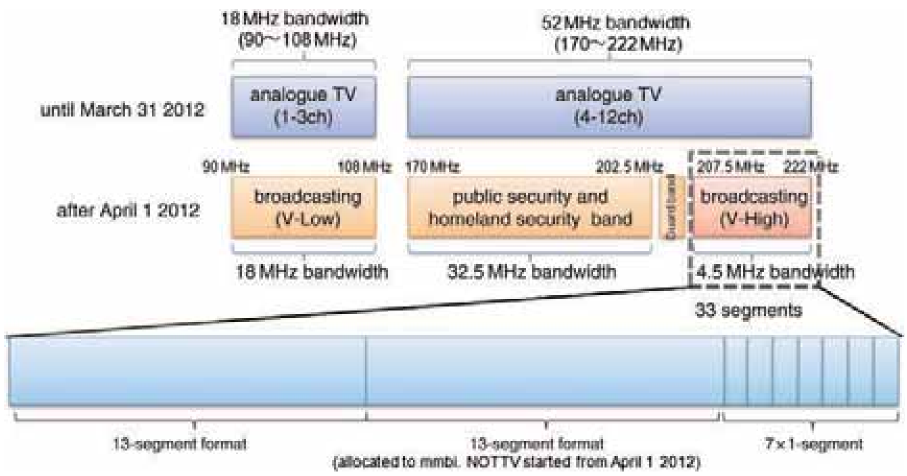


Figure 4 *ISDB-Tmm spectrum allocation in Japan*

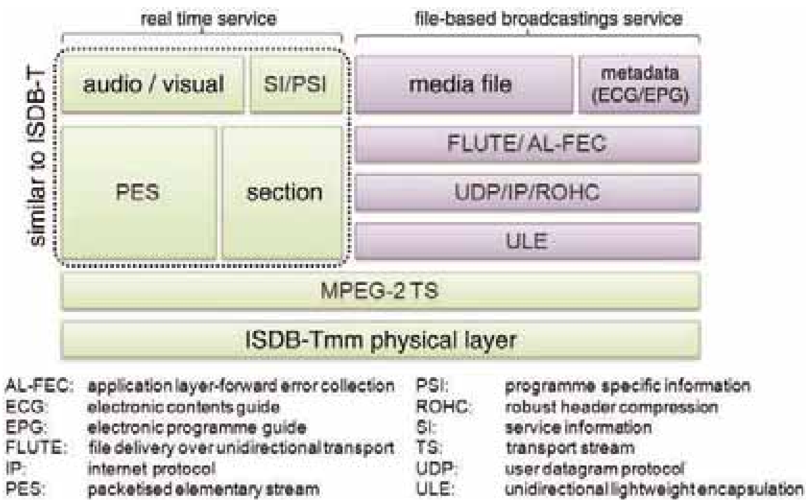


Figure 5 *ISDB-Tmm protocol stack*

Table 2 Media coding for real-time broadcasting

Video coding	Codec	ITU-T H.264/MPEG-4 AVC
	Profile	Baseline or Main
	Level (maximum)	3.0
	Resolution (maximum)	720×480 (525SD)
	Frame rate (maximum)	30 fps
Audio coding	Codec	MPEG-2 AAC + SBR + PS, Surround
	Number of channel	Monaural/Stereo/ 5.1ch
	Sampling rate	48 kHz, 44.1 kHz, 32 kHz, 24 kHz, 22.05 kHz, 16 kHz

Table 3 Media coding for file-based broadcasting

Video coding	Codec	ITU-T H.264/MPEG-4 AVG
	Profile	High
	Level (maximum)	4.2
	Resolution (maximum)	1920×1080 (1080HD)
	Frame rate (maximum)	60 fps
Audio coding	Codec	MPEG-2 AAC + SBR + PS, Surround MPEG-4 ALS/SLS (option)
	Number of channel	Monaural/Stereo/ 5.1ch
	Sampling rate	48 kHz, 44.1 kHz, 32 kHz, 24 kHz, 22.05 kHz, 16 kHz

3.3 Media coding

For real-time broadcasting, audio and visual media can be transmitted in the same way as in conventional digital TV broadcasting. As shown in Table 2, ISDB-Tmm has high-quality media coding, including multi-channel audio, using ITU-T H.264 and MPEG Audio. File-based broadcasting is capable of sending large media files encoded at higher bit rates, such as H.264 high profile level 4.2 as in Table 3.

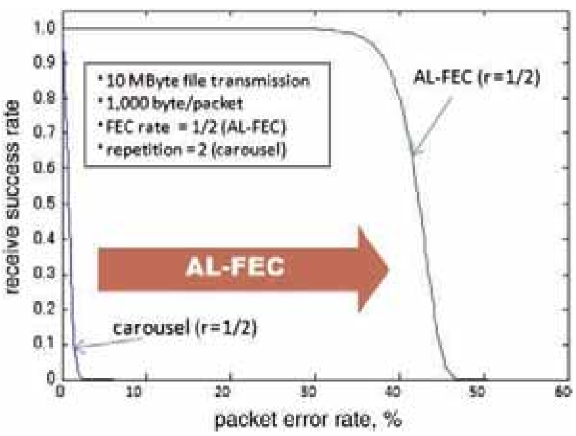


Figure 6 *High efficiency file transmission by AL-FEC*

3.4 File-based broadcasting

For the efficient file-based transmission of large size contents over error prone broadcast channels, the ISDB-Tmm employs low density parity check (LDPC) based application layer forward error correction (AL-FEC), and achieves higher reception probability than conventional carousel type repetitive transmission. Fig. 6 shows the advantage of the AL-FEC against the conventional carousel type repetition under the same transmission overhead (coding rate = 1/2).

3.5 CAS/DRM

The conditional access system (CAS) is used to restrict access to the real-time service, and the DRM (digital rights management) system is used by the file-based service to enable restricted playback of encrypted content that is stored in the terminal. In both cases, the contents are broadcasted through the broadcasting channel encrypted by advanced encryption standard (AES) 128-bit algorithm, and the license information is transmitted through the communications channel as shown in Fig. 7.

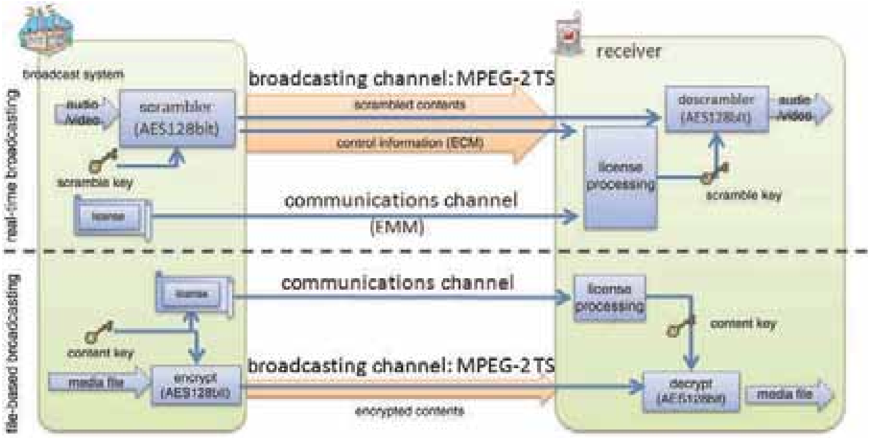


Figure 7 *CAS/DRM*

Table 4 ISDB-Tmm related standard documents

Standards	Title
ITU-R BT-1306	Error correction, data framing, modulation and emission methods for digital terrestrial television broadcasting
ITU-R BT-1833	Broadcasting of multimedia and data applications for mobile reception by handheld receivers
ITU-R BT-1888	Basic elements of file-based broadcasting systems
ARIB STD-B10	Service information for digital broadcasting system
ARIB STD-B25 Part 4	Conditional access system specifications for digital broadcasting
ARIB STD-B32	Video coding, audio coding and multiplexing specifications for digital broadcasting
ARIB STD-B38	Coding, transmission and storage control for server type broadcasting
ARIB STD-B45 Part 2	Contents download for digital broadcasting system
ARIB STD-B46	Transmission system based on connecting segments for terrestrial mobile multimedia broadcasting
ARIB STD-B53	Receiver for multimedia broadcasting
ARIBTR-B33	Operational guidelines for terrestrial mobile multimedia broadcasting by transmission system based on connected segments

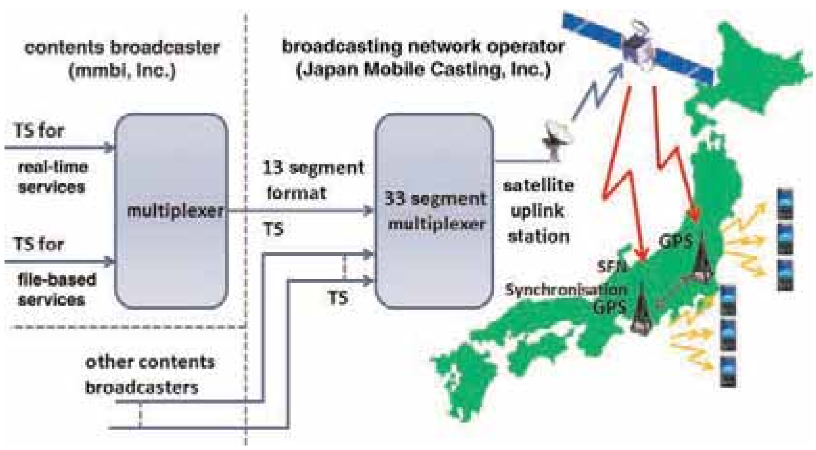


Figure 8 *ISDB-Tmm broadcasting network*

Figure 9 *ISDB-Tmm antenna*

Figure 10 *25 kW, 14.5 MHz ISDB-Tmm transmitter*

3.6 Standardisation

The ISDB-Tmm related standards and the operational guideline, listed in Table 4, are published by the International Telecommunication Union – Radiocommunication sector (ITU-R) and the Association of Radio Industries and Businesses (ARIB).

4 ISDB-Tmm broadcasting network

The operation licences for Japanese multimedia broadcasting are split into a single 'broadcasting network operator' licence and multiple 'contents broadcaster' licences. The broadcasting network operator multiplexes transport streams from contents broadcasters either in 13-segment format or in 1-segment format, and distributes the multiplexed 33-segment format stream to the transmitters through satellite links as shown in Fig. 8. The OFDM segments are connected without guard band to form a single 33-segment spectrum. We developed high-efficiency amplifiers capable of simultaneous handling of the 14.5 MHz OFDM signal (Figs. 9 and 10).

For the efficient use of the spectrum, a nationwide single frequency network (SFN) is developed. The transmitters are equipped with GPS receivers for the SFN synchronisation.

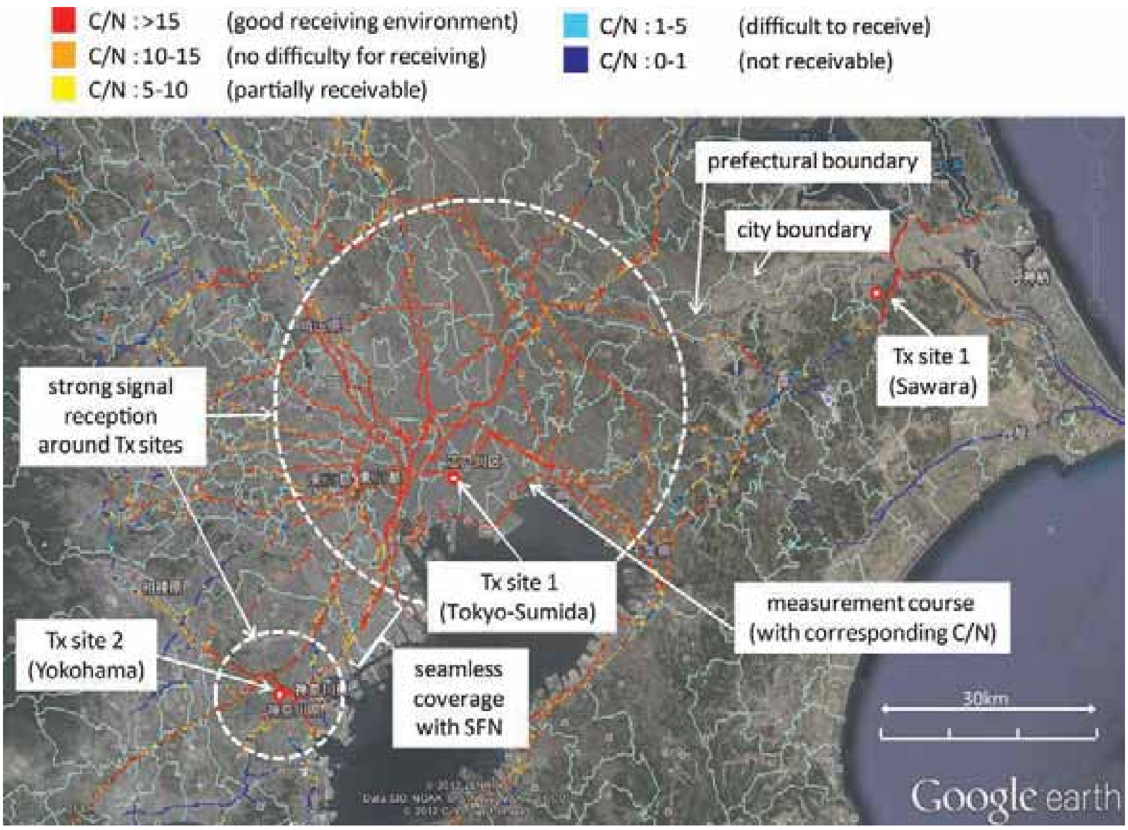


Figure 11 *Field measurement results in Tokyo*

4.1 Field experiment

We conducted a field measurement of the ISDB-Tmm radio propagation using three transmitters (Tokyo-Sumida, Yokohama and Sawara) installed in the Tokyo metropolitan area in April 2012. For the purpose of efficient field experiments, we developed a smartphone-type radio strength measurement system that can record signal strength and GPS position at the same time. Fig. 11 shows the result of the measurement along major streets. The Yokohama transmitter and the Tokyo-Sumida transmitter are located 35 km in distance from one another, and have single frequency area overlap in between. It was confirmed that the area is covered by multiple transmitters efficiently.

5 Conclusion

This paper gave an overview of the ISDB-Tmm mobile multimedia broadcasting technology. ISDB-Tmm is based on ISDB-T and is easily implemented with OneSeg and ISDB-Tsb mobile digital TV receivers in the same handy phone terminal. It has superior transmission characteristics in mobile environments. It provides high-quality audio visual media transmissions, based on H.264 and AAC, for both real-time and file-based broadcasting services, and also provides the capability of interacting with communications-based services.

An ISDB-Tmm network was developed using 14.5 MHz connected OFDM transmitters, which synchronised with each other to form a nationwide single frequency network. A field experiment was conducted in Tokyo using three transmission sites. It was confirmed that the wide area was covered seamlessly with a single frequency network.

A commercial ISDB-Tmm service, named 'Mobacas,' was launched in Japan in April 2012, with about 60% of household coverage. It is expected to expand to over 90% of households until the end of fiscal year 2015.

6 References

[1] Association of Radio Industries and Businesses (ARIB). 'Transmission system for digital terrestrial television broadcasting'. ARIB STD-B31

[2] Association of Radio Industries and Businesses (ARIB). 'Transmission system for digital terrestrial sound broadcasting'. ARIB STD-B29

[3] Association of Radio Industries and Businesses (ARIB). 'Transmission system for terrestrial mobile multimedia broadcasting based on connected segment transmission'. ARIB STD-B46

[4] Internet Engineering Task Force (IETF). 'FLUTE – file delivery over unidirectional transport'. RFC 3926

Selected content from IBC2012

ISSN 2041-5923

New equipment for 22.2 multichannel sound production and reproduction

M. Okano[1] *K. Matsui*[2] *H. Okubo*[2] *K. Hamasaki*[2]

[1]Engineering Department, NHK (Japan Broadcasting Corporation), 2-2-1 Jinnan, Shibuya-Ku, Tokyo 150-8001, Japan
[2]NHK Science & Technology Research Laboratories, 1-10-11 Kinuta, Setagaya-Ku, Tokyo 157-8510, Japan
E-mail: okano.m-ek@nhk.or.jp

Abstract: This paper describes the latest developments in SHV equipment, namely, a live mixer for 22.2 multichannel sound, a 22.2 multichannel sound processor for headphones, and a loudspeaker-array frame for a flat-panel display. The mixer has a 3D panning function and a GUI to visualise positional information related to the audio in a 3D space. The mixer's features enable simple control of 3D panning even in live situations. The headphone processor makes it possible to listen to 22.2 multichannel sound with ordinary headphones by using binaural audio technology. In addition, the headphones' processor implements the world's first 3D head-tracking function. The loudspeaker-array frame can localise sound images on the display and to the side and back of the viewer by using technology to simulate sound propagation characteristics of such images. It makes it possible for the listener to enjoy 22.2 multichannel sound without having to be surrounded in a room with 24 individual loudspeakers.

1 Introduction

NHK has been developing Super Hi-Vision (SHV), a next-generation television system that will convey a far stronger sense of reality than existing audio and visual systems do. SHV has a 7,680 × 4,320-pixel video image that has 16 times the resolution of current Hi-Vision (HDTV) images and a 22.2 multichannel sound system that can reproduce 3D spatial sound.

The 22.2 multichannel sound system can reproduce the immersive sense of reality. So far, the system has been used to record cultural heritage and historical events so that they can be shared with future generations.

2 22.2 Multichannel sound system

The 22.2 multichannel sound system was designed to exceed all existing multichannel sound systems in sound quality and to offer a highly immersive sense of presence and reality.

The channel names and labels of this system have been standardised [2], as shown in Fig. 1. The system consists of 24 discrete channels distributed in three layers. The top layer has nine channels that are placed near the ceiling. The middle layer has ten channels at the height of the listener's ears or at the centre of the screen. The bottom layer has three channels, including two low-frequency-effects (LFE) channels (which are usually placed on the floor).

3 22.2ch sound live mixer

The 22.2 multichannel sound mixer – for live productions of big sporting events like the Olympics Games – has a redundant configuration for the sake of reliability. A special feature of this mixer is its 3D panning function [3].

Panning of sound in a 3D space means making adjustments on the x-, y-, and z-axes, and the difficulty of doing such an operation has thus far made it impossible to arrange sounds arbitrarily in a 3D space in a live situation. To alleviate this problem, we developed a 3D panning controller for enabling simple panning even in live situations and a GUI for visualising positional information related to audio in a 3D space. Fig. 3 shows photos of the controller and the 3D panning display on the mixer. The controller has a joystick for intuitive operation; sound sources can be moved in an x-y plane (with the z-axis position fixed) simply by moving the joystick. Having a

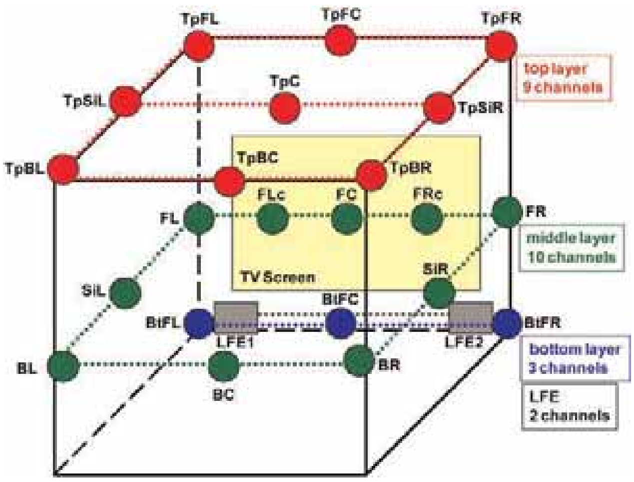


Figure 1 *Channel name and label of 22.2 multichannel sound*

Figure 2 *22.2ch sound live mixer*

visual indication of a sound's location makes it easy to perform 3D panning. Furthermore, it is possible to control the panning in regard to divergence and rotation in the same manner as stereo or surround sound.

In addition, all functions that users could want in a sound live mixer, such as a 22.2ch sound monitor controller, an auxiliary feed function, and down-mixing functions to 5.1 surround and 2.0 stereo sound, are integrated into a single unit. Thanks to these features, the time needed to set up the mixer is short. For instance it would take much longer to set up a system composed of a conventional 5.1 surround sound mixer and a digital audio workstation. Overall, this means programmes featuring 22.2ch sound will be relatively easy to produce.

4 Binaural headphone processor for 22.2 multichannel sound

The headphone processor for monitoring multichannel sound productions (Fig. 5) makes it possible to listen to 22.2ch sound on ordinary headphones. The processor uses binaural reproduction – a head-related transfer function (HRTF)-based signal processing technique [4]. The HRTF describes how a given sound wave reaches the entrance of a listener's ear canal through reflection and diffraction on the head, pinna, and torso. In binaural reproduction, the audio signals of each of the 22 channels, excluding the two LFE channels, are convolved with left and right ear HRTFs for loudspeakers in specific directions to simulate their propagation characteristics at both ears. A 22×2 convolution is therefore required.

Fig. 6 shows the appearance of the HRTF measurement setup. Loudspeakers were mounted on a semicircular frame at 20° increments from $-40°$ to 90°. The subject was seated on a rotating stool. Measurements corresponding to all directions were performed by rotating the subject in relation to the loudspeaker arrangement for 22.2ch sound. Even HRTFs measured with a dummy head make it possible to sense movement of sound images in a 3D space to some degree. However, because the HRTF varies in accordance with the shape of the subject's head, pinna and torso, it is preferable to perform measurements on every user.

In-head localisation of sound images (lateralisation) is a critical problem when it comes to headphones. To enable proper localisation, we implemented the world's first 3D head-tracking function in the headphone processor. The sound image should move as the listener's head moves up/down and left/right. Such head movements are detected

Figure 3 *3D panning controller and GUI*

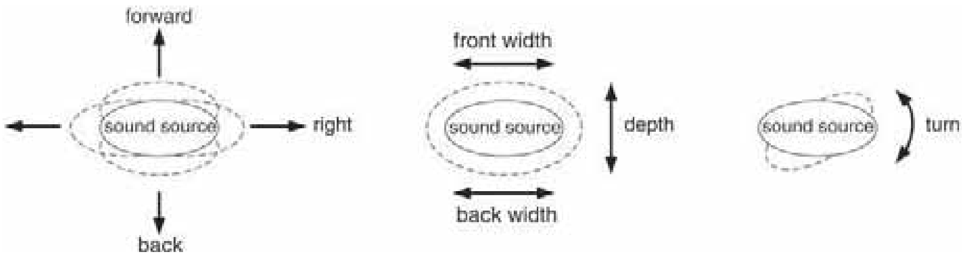


Figure 4 *Behaviour of individual sources*

Figure 5 *Headphones with 3D positional sensor (upper) and binaural headphone processor (lower)*

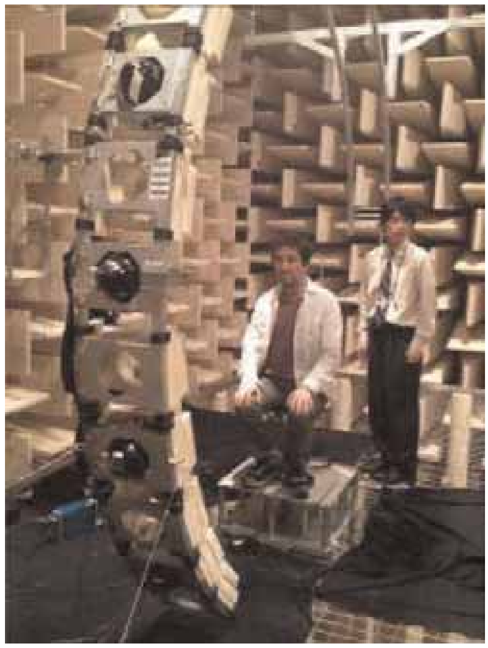


Figure 6 *HRTF measurement in anechoic chamber*

using sensors mounted on the headphones, and this data is used to make the sound image move with the head. However, up/down and left/right panning has to be performed on a 24×2 sound signal, which is a vast

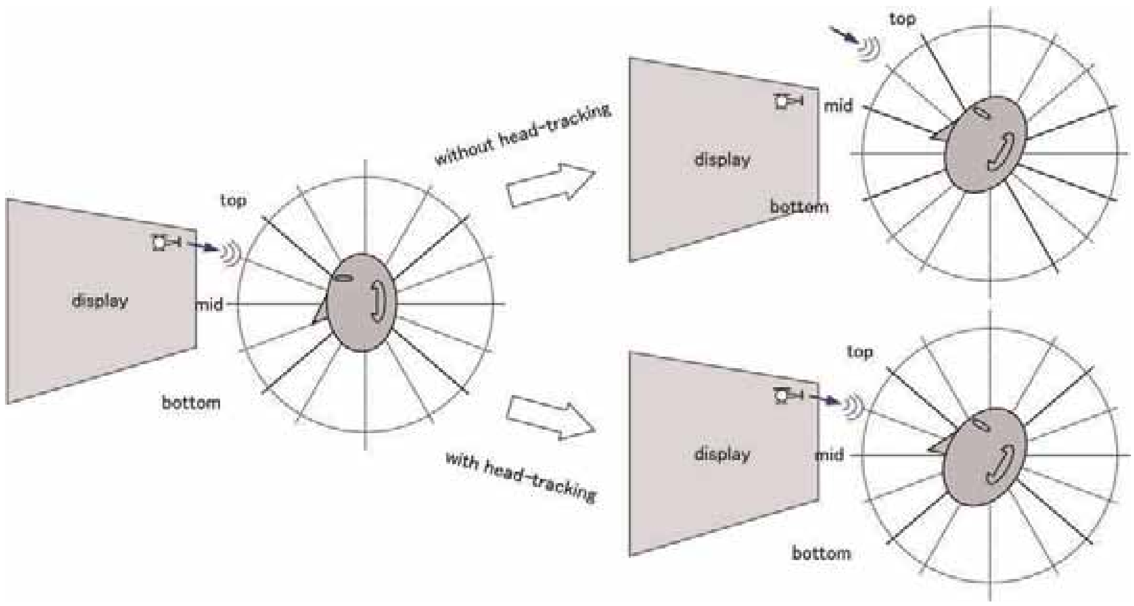


Figure 7 *Effect of 3D head-tracking system*

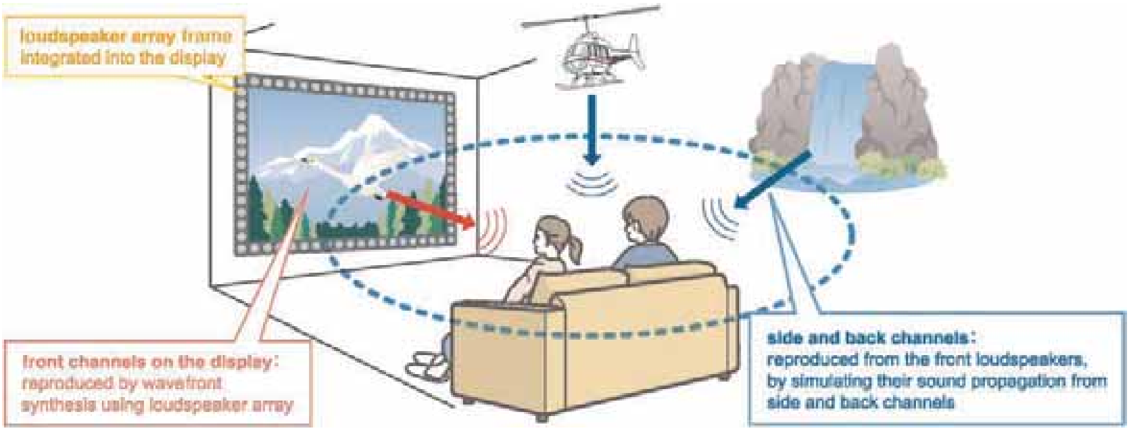


Figure 8 *Concept of 22.2ch sound reproduction using LAF*

amount of data. With the developed headphone processor, adjustments (such as thinning out the data depending on the orientation) were performed, and practical values were derived.

5 Loudspeaker array frame

We have been investigating several home reproduction methods for 22.2 multichannel sound. In particular, we devised a reproduction method using a loudspeaker array frame (LAF) integrated in a flat panel display for medium-sized living rooms (12 m^2) [5]. The concept of the system is illustrated in Fig. 8. The LAF consists of many small loudspeaker units installed on the sides of the display, and it reproduces sounds in all directions by using two different sound reproduction principles: wavefront synthesis (WFS) and binaural reproduction.

WFS is derived from Huygens' Principle, which states that any wavefront can be synthesised as a superposition of elementary waves. In accordance with this principle, we calculate the sound pressure at the points of the respective loudspeaker units and drive them to synthesise the wavefronts of the sounds from the front channels on the display where conventional loudspeakers cannot be placed [5].

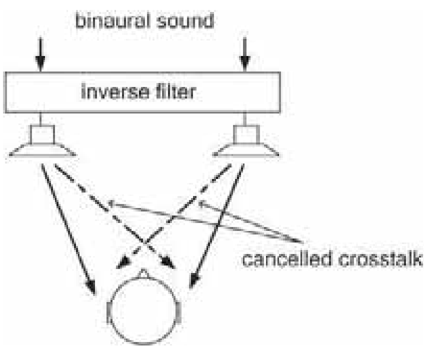


Figure 9 *Principle of binaural reproduction over loudspeakers*

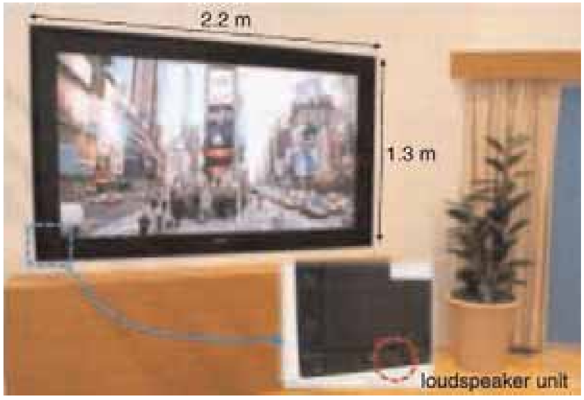


Figure 10 *LAF for 85 inch SHV LCD*

Binaural reproduction is used to imitate sound propagation from the side and back channels. The concept is almost the same as that of the binaural headphone processor except for the existence of crosstalk, undesired sound propagation from loudspeakers to contralateral ears. To cancel the crosstalk, audio signals are inverse filtered before being sent to the loudspeakers (see Fig. 9). Since inverse filters for crosstalk cancellation are generally acausal, we divided them into all-pass and minimum-phase components and modelled them individually [6].

6 Conclusions

The 22.2ch sound live mixer, which is an upgrade of a conventional live sound mixer with a 3D panning function, supports efficient live mixing and sound production. The binaural headphone processor for 22.2 multichannel sound with 3D head-tracking capability makes it possible for sound images to follow the movements of the head accurately. Note that without such headphones, it was impossible to verify the audition of 22.2 multichannel sound recorded in the field, making efficient TV production out of the question. Turning to the home front, placing a loudspeaker array frame around a flat-panel display is a convenient way to realise 22.2ch sound reproduction in a living room. It allows viewers to enjoy the superior features of 22.2 multichannel sound without having to install 24 individual speakers in a room.

NHK will spread the word about SHV while continuing its development of the whole broadcast pipeline from programme production to the viewing environment.

7 Acknowledgments

The authors thank the staff of Lawo AG, Fourbit Inc., and Foster Electric Co., Ltd for their help in developing the equipment, and the staff of the NHK production operations center for their help in making adjustments to the equipment.

8 References

[1] HAMASAKI K., NISHIGUCHI T., OKUMURA R., NAKAYAMA Y., ANDO A.: 'A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)', *SMPTE Motion Imaging J.*, 2008, **117**, (3), pp. 40–49

[2] SMPTE ST2036-2-2008, Ultra High Definition Television – Audio Characteristics and Audio Channel Mapping for Program Production

[3] ANDO A., HAMASAKI K.: 'Sound intensity based three-dimensional panning'. AES 126th Convention, AES Convention Paper 7675, Munich, Germany, 2009

[4] MATSUI K., NAKAYAMA Y., HAMASAKI K.: 'Development of binaural headphone processor for 22.2 multichannel sound'. Annual Technical Conf. & Exhibition of 2011 SMPTE, Hollywood, CA, USA, 2011

[5] SUGIMOTO T., MATSUI K., OKUBO H.: 'A loudspeaker array frame reproducing 22.2 multi-channel sound for super hi-vision flat panel display'. Session of 2012 NAB SHOW, Las Vegas, NV, USA, 2012

[6] MATSUI K., ANDO A.: 'Binaural reproduction of 22.2 multichannel sound over loudspeakers'. AES 129th Convention, AES Convention Paper 8272, San Francisco, CA, USA, 2010

ISSN 2041-5923

Interview – Kazuya Kitamura

Both IBC and the IET have a focus on encouraging young people into and within the industry. As part of this, each year *The Best of IET and IBC* features the paper chosen as the best young professional contribution in the conference. The winner for IBC2012 is ‘Development of 33 megapixel 120 Hz CMOS image sensor and experimental colour camera system’; before we present the paper here is an interview with lead author and winner Kazuya Kitamura.

Tell us a bit about yourself and what you do

I am a principal research engineer in the NHK Science and Technology Research Laboratories (STRL), Tokyo, Japan. NHK is Japan’s only public broadcaster, financed by receiving fee. I received BEng and MEng degrees from Chiba University, Chiba, Japan, in 1998 and 2000, respectively. In 2000 I joined NHK, and from 2000 to 2003 I worked at the Fukuoka regional broadcasting station as a television engineer. Since 2003, I have been engaged in the research and development of ultra-high-speed charge-coupled devices (CCD) and ultra-high definition CMOS image sensors with high frame rate in the Imaging & Storage Device Division of the NHK STRL.

What is your paper about?

We have been researching and developing a next-generation television system called Super Hi-Vision which is an advanced broadcasting service with enhanced picture and sound quality that provides a stronger sense of presence for the viewer.

The video system needs 7680 × 4320 pixels and a frame frequency of 120 Hz. To achieve this extremely high resolution and high frame rate video system, we developed a 33 megapixel CMOS image sensor operating with a frame frequency of 120 Hz. Generally, the larger the number of pixels of the image sensor, the more difficult it is to achieve high-speed operation. For the Super Hi-Vision, the output data rate will exceed 47 Gbit/s. To meet the challenges of this high-speed operation, we developed a high-speed analogue-to-digital converter circuit and a signal output circuit and succeeded in outputting the 33 megapixel images with 120 Hz, which is double the frame frequency of the previous image sensors (60 Hz) used for Super Hi-Vision. We also developed an experimental three-chip colour camera system for 120 Hz Super Hi-Vision and confirmed that it captures fast-moving subjects more clearly than Super Hi-Vision with 60 Hz image sensors.

What interests you about this area of work?

I am excited about helping to create a future video system through research and development of the image sensor. Super Hi-Vision is designed to produce video that looks ultra-realistic to the human eye, so that when people watch the television they might feel as if they are really there in the scene. I think that the image sensor is the most important device in determining image quality.

How would you like to see this area develop and what do you think the challenges will be?

Super Hi-Vision is an ultimate 2D television system, so I think the next step will need to be research and development for a 3D television system. A 3D system would require very large amounts of image data, which will be tough for image sensors to collect. Therefore, the main challenge will be increasing the sensitivity of the image sensor because the high-resolution and high-speed image sensor needs to have higher sensitivity. This is because a high-resolution sensor decreases the pixel size and a high-frame-rate sensor decreases the exposure time. I suppose the technologies that accumulate photons effectively and detect the image signal with low noise will be the most important factors in researching the image sensor.

Is this the first paper you have submitted to IBC, and have you been to the conference before?

This is my second IBC paper. The first was for an ultra-high-speed camera that could capture images at up to 1 million frames per second. I presented it in the cutting edge session of IBC2007. I also exhibited the camera and demonstrated fascinating slow motion images of a bursting water balloon.

What will you be doing at the conference this year?

I will exhibit the 120 Hz Super Hi-Vision experimental colour camera in the Future Zone. NHK will also exhibit Super Hi-Vision images from the London Olympic games. I am looking forward to meeting the many attendees and discussing the 120 Hz images and Super Hi-Vision images of the London Olympics with them.

Selected content from IBC2012. This paper was voted best young professional contribution

ISSN 2041-5923

Development of 33 megapixel 120 Hz CMOS image sensor and experimental colour camera system

K. Kitamura[1,2] *T. Watabe*[1,2] *H. Shimamoto*[1] *H. Ohtake*[1] *S. Kawahito*[2] *T. Kosugi*[3] *T. Watanabe*[3] *T. Yanagi*[4] *H. Kikuchi*[5] *T. Yoshida*[4] *N. Egami*[1]

[1]*NHK (Japan Broadcasting Corporation), 1-10-11 Kinuta, Setagaya-ku, Tokyo 157-8510, Japan*
[2]*Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu 432-8011, Japan*
[3]*Brookman Technology, Inc., 125 Daikumachi, Naka-ku, Hamamatsu 430-0936, Japan*
[4]*Hitachi Kokusai Electric Inc., 32 Miyuki-cho, Kodaira-shi, Tokyo 187-8511, Japan*
[5]*Link Laboratory Inc., 2-35-28 Higashi-koigakubo, Kokubunji-shi, Tokyo 185-0014, Japan*
E-mail: kitamura.k-hi@nhk.or.jp

Abstract: A 33 megapixel CMOS image sensor operating with a frame frequency of 120 Hz and an experimental colour camera system using the sensors were fabricated and tested as a first step toward developing 'full-spec' Super Hi-Vision, which has double the frame frequency of the previous Super Hi-Vision.

The diagonal length of the effective image area of the fabricated sensor is only 25 mm, and the sensor is equipped with new two-stage pipe-lined cyclic analogue-to-digital converters and parallel digital readout circuits to achieve high-speed operation for the 120 Hz frame frequency progressive scanning.

The experimental results revealed that the sensor could run at a frame frequency of 120 Hz with power consumption of only 2.5 W and that the colour camera system with the sensors could obtain clearer motion images with 33 megapixel 12-bit resolution, which promises to open the way to full-spec Super Hi-Vision.

1 Introduction

Super Hi-Vision is a next-generation ultra-HDTV system that is designed to provide viewers an enhanced sense of presence and a sense of realness. The spatial resolution of the Super Hi-Vision is 7680 × 4320 pixels, which achieves a pixel structure undetectable to the human eye with a viewing angle around 100 degrees [1]. The frame frequency of the Super Hi-Vision has been 60 Hz progressive [2-3], but recently NHK has started studying improvements in motion portrayal by subjectively assessing the motion images, and has proposed the frame frequency of 120 Hz for the full-spec Super Hi-Vision [4].

A major challenge for the image sensor in full-spec Super Hi-Vision is the simultaneous fast operation speed and low power consumption. The most common way to increase the operation speed is to increase the current of the circuits in the image sensor. However, if we operate it at high speed, the image sensor will consume too much power, causing the temperature and noise to increase. Another way to increase the operation speed is to use high-speed analogue-to-digital (ADC) architecture in the image sensor.

2 Image sensor

To increase operation speed without increasing the power consumption, we developed a multi-stage cyclic ADC architecture and a parallel readout operation in the image sensor.

2.1 Structure and operation

The structure of the image sensor is shown in Fig. 1. The total pixel array consists of 7808 × 4336 pixels, including the effective pixel count of 7680 × 4320 pixels, which is the full resolution for Super Hi-Vision, and all the pixels must be read out within 1/120 second. The pixel size is 2.8 × 2.8 μm, which is decided by considering the sensitivity and picture quality needed for a broadcasting camera. Each pixel adopts a buried photodiode structure for low noise and an on-chip micro lens for better sensitivity. Above and below the pixel array are gain amplifiers, column-parallel ADCs, and data readout circuits. The signal from the pixel is read out for upper and lower directions in every other column. Each column has an ADC, meaning a total of 7808 ADCs in a horizontal line, and all ADCs convert the pixel signal into 12-bit digital data together in the horizontal scanning period. The digital data are output from low-voltage differential signalling (LVDS) drivers.

2.2 ADC architecture

The ADC used in this image sensor is a column-parallel ADC, and the conversion time of the ADC must be less than one horizontal scanning period, which is decided by multiplying the frame frequency (120 Hz) and the number of vertical lines (4336), and is thus 1.92 μs. There are several ADC types for large pixel-count image sensors, such as single slope, successive approximation, and cyclic methods.

In those ADC types, the cyclic ADC has a short A/D conversion time and is suitable for use in high-speed operation [5]. Fig. 2*a* shows a block diagram of a conventional single-stage cyclic ADC circuit. The cyclic ADC consists of the amplifier circuits, sub-ADC, and a digital-to-analogue converter (DAC). When the switch is on, the input signal is sampled, amplified, and returned to the input. This cyclic operation generates digital data from the most significant bit (MSB) to the least significant bit in a sequence. For example, if 12-bit resolution is required, the 11 cycles of the cyclic operation are needed. However, for the full-spec Super Hi-Vision specifications, even with the cyclic ADC, the operation speed of the conventional cyclic ADC is still not met.

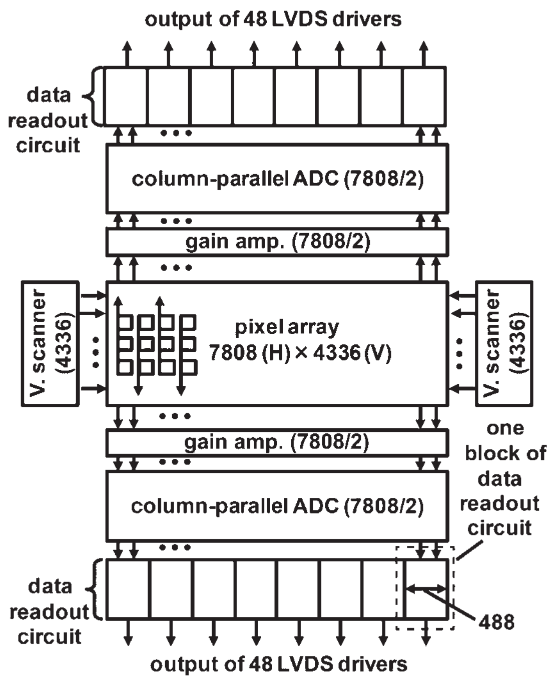


Figure 1 *Sensor structure*

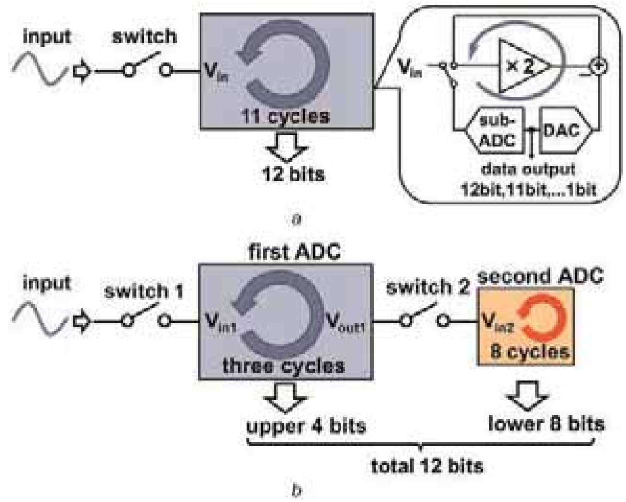


Figure 2 *Block diagrams of:*

a Conventional cyclic ADC
b Two-stage cyclic ADC

To overcome the limit of the cyclic ADC operation speed, we proposed the multi-stage cyclic ADC [6]. The multi-stage structure decreases the number of requirements for the operation speed in each ADC, because the multi-stage cyclic ADCs can divide the ADC process into plural ADCs, and the multi-stage ADCs are operated in parallel using pipeline operation. For example, when we use a two-stage cyclic ADC for 12-bit resolution and the first ADC outputs N-bit (here N is smaller than 12), the first and second ADCs need only (N-1) and (12-N) cyclic operations, respectively. In this case, both the first and the second ADCs need fewer than 11 cyclic operations, and requirements for the ADCs operation speeds are relaxed. When the operation speed in each ADC decreases, the power consumption decreases, because current conditions of transistors or amplifiers in the circuit are relaxed.

Fig. 2*b* shows an example of the multi-stage cyclic ADC, which is for 12-bit resolution and is divided into two stages. In this case, we decided the bit separation number by simulating the power consumption, and it shows four bits for the first ADC and eight bits for the second ADC to be the lowest power consumption.

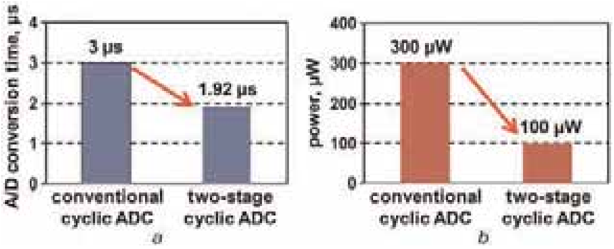


Figure 3 *Comparison of:*
a Conversion time
b Power of new and conventional cyclic ADC

Fig. 3 compares the measured A/D conversion time and the power consumption of the two-stage cyclic ADC with those of the previous single-stage cyclic ADC [5]. The A/D conversion time with 12-bit precision was as short as 1.92 µs and covered the required conversion time. The power consumption was only one third that of the previous cyclic ADC.

2.3 Data readout circuit

The required data output capability of the data readout circuit is over 47 Gbit/s, which is the multiplication of the frame frequency (120 Hz), pixel count (7680 × 4320 pixels), and bit depth of each pixel (12 bits). Dividing the readout circuit and output from parallel ports decreases the readout speed effectively. However, the larger the paralleling number, the larger the power consumption.

We divided the readout circuit into 16 blocks. Fig. 4 shows a block diagram of one block of the data readout circuit. The circuit contains a data transfer path between CML drivers and receivers. As the CML signal is transferred as a differential signal, if the speed is too fast, the difference voltage (eye pattern voltage) is low and jitter occurs in the receivers. If the speed is too slow, the difference voltage stays high, but the numbers of CML drivers and receivers increases, causing the total power consumption to increase. We simulated the balance of the paralleling number and the difference voltage. Fig. 5 shows the results. The difference voltage went down when the speed (data rate) exceeded 150 Mbit/s. The power consumption depending on the paralleling number increased when the data rate was inferior to 100 Mbit/s. From these results, we found that the data rate of 133 Mbit/s and the paralleling number 32 were the best.

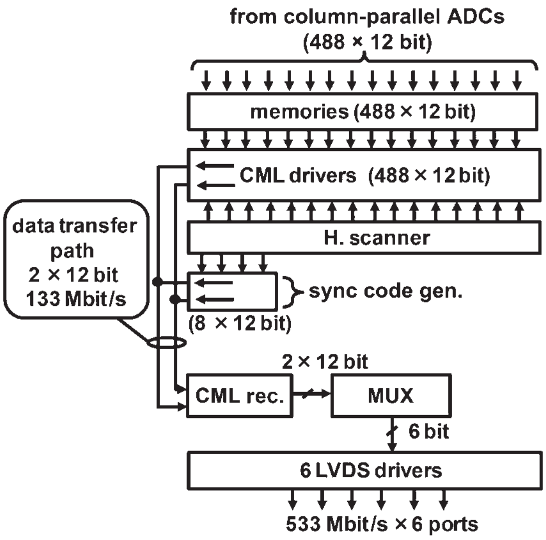


Figure 4 *Block diagram of one block of data readout circuit*

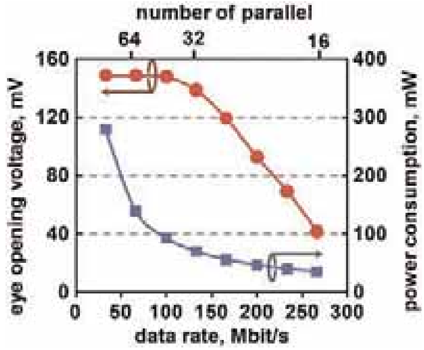


Figure 5 *Eye pattern voltage and power consumption of CML circuits at various data rates*

The data from the CML receivers are multiplexed four times so that the transfer speed is converted from 133 to 533 Mbit/s. Each block has six LVDS drivers, and the image sensor has 96 drivers in total. As a result, the aggregate sensor output data rate is 51.2 Gbit/s (533 Mbit/s × 96 ports).

2.4 Specifications

Table 1 compares the specifications of the developed image sensor with those of the conventional 60 Hz Super Hi-Vision image sensor [7]. The developed image sensor has double the frame frequency but smaller power consumption. The diagonal length is also smaller, which is advantageous when the lens and other optical devices are downsized. Fig. 6 shows the appearance of the image sensor.

Table 1 Specification of image sensor

	This work	Conventional
Effective pixel count	7680 × 4320	7680 × 4320
Frame frequency	120 Hz	60 Hz
Bit depth	12 bits	12 bits
Scanning system	progressive	progressive
Diagonal length	25 mm	33 mm
Power consumption	2.5 W	3.7 W

Figure 6 *Appearance of image sensor*

3 Experimental colour camera system

To test the developed CMOS image sensor and capture the moving image with a frame frequency of 120 Hz with 33 megapixel resolution, we developed a three-chip colour test camera system using the image sensors.

3.1 Structure

A block diagram of the camera system is shown in Fig. 7. The camera head consists of three image sensors for Red, Green, and Blue, a prism, and three headboards. The 96 parallel output signals from each image sensor are integrated in the headboard. In the headboard, image data of 7680 × 4320 pixels, 120 Hz, and 12 bits are allocated to the high-definition serial-digital-interface (HD-SDI) format. The image data are then output to a camera-control-unit (CCU). The CCU performs camera processing such as pixel realignment from the image sensor output to the display format, fixed-pattern-noise (FPN) cancellation, signal level and gain control, gamma correction, detail enhancement, and so on. The appearance of the camera head is shown in Fig. 8.

For the image display, as a 7680 × 4320 pixel 120 Hz display has not been developed yet, we cropped 3840 × 2160 pixel images from 7680 × 4320 pixel images and displayed them on a 3840 × 2160 pixel 120 Hz projector.

3.2 Experiments

We performed shooting experiments using the camera. Fig. 9 shows an example of a captured and reproduced colour image. We also shot images of a moving object by moving an object to the right and capturing it. Fig. 10 shows the reproduced results. When the frame frequency was 60 Hz, motion blur was observed (Fig. 10*a*), because the speed of the object generates a motion blur. When the frame frequency was 120 Hz, on the other hand, the motion blur was smaller than that at 60 Hz. From the results, we confirmed that the frame frequency of 120 Hz improves the image quality of Super Hi-Vision motion images.

We measured the characteristics of the experimental three-chip colour camera system. The sensitivity was F-stop 4.8 at a light intensity of 2000 lux, and the signal-to-noise

Figure 8 *Appearance of camera head*

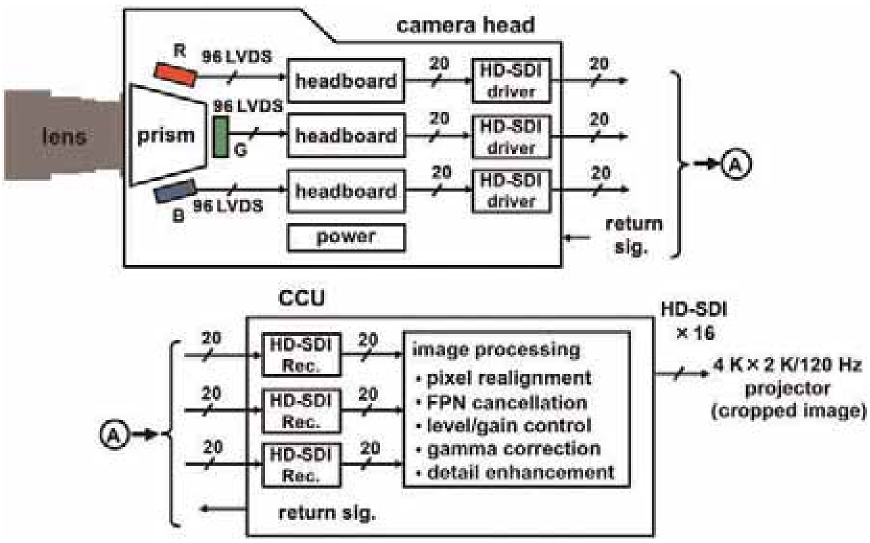


Figure 7 *Block diagram of experimental camera*

Figure 9 *Example image*

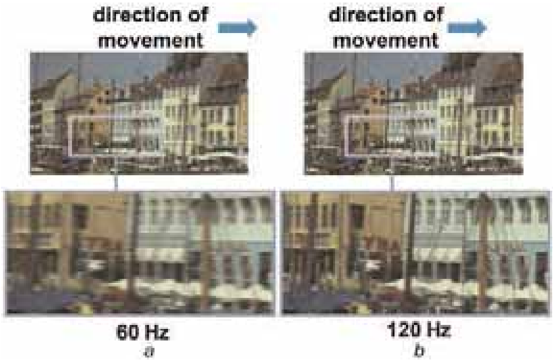


Figure 10 *Example image of moving object with frame frequencies of:*

a 60 Hz
b 120 Hz

Table 2 Camera specifications

Parameter	Values
Colour type	three 33 M pixel CMOS
Sensitivity	2000 lux F4.8
S/N *	51.2 dB

*S/N was measured at 1/16 extracted HDTV resolution image data

ratio (S/N) was 51.2 dB. Note that the S/N was measured at one sixteenth extracted HDTV resolution image from the camera system. The camera specifications are listed in Table 2.

4 Conclusions

This paper described the world's first CMOS image sensor for full-spec Super Hi-Vision. The two-stage pipelined cyclic ADC architecture and high-speed digital readout circuit design enabled the image sensor to operate at 120 Hz. We also developed the first experimental colour camera system for 120 Hz Super Hi-Vision using three CMOS image sensors. Our test shooting confirmed that the frame frequency of 120 Hz improves the image quality of Super Hi-Vision motion images. These image sensors and the prototype camera system announce the dawn of the next-generation ultra-HDTV system, and full-spec Super Hi-Vision will open the way to the future of television systems.

5 References

[1] SUGAWARA M., KANAZAWA M., MITANI K., SHIMAMOTO H., YAMASHITA T., OKANO F.: 'Ultrahigh-definition video system with 4000 scanning lines', *SMPTE Motion Imaging Journal*, 2003, **112**, (10 & 11), pp. 339–346

[2] ITU-R Rec. BT.1769, 2007. 'Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange.' International Telecommunications Union. Geneva

[3] SMPTE 2036-1-2009, 'Ultra high definition television—image parameter values for program production.' www.smpte.org

[4] NISHIDA Y., MASAOKA K., SUGAWARA M., OHMURA K., EMOTO M., NAKASU E.: 'Super Hi-Vision system offering enhanced sense of presence and new visual experience'. Proceedings of 2011 International Broadcasting Convention, Amsterdam, The Netherlands, 2011

[5] PARK J.H., AOYAMA S., WATANABE T., ISOBE K., KAWAHITO S.: 'A high-speed low noise CMOS image sensor with 13-b column-parallel single-ended cyclic ADCs', *IEEE Trans. Electron. Devices*, 2009, **56**, (11), pp. 2414–2422

[6] WATABE T., *ET AL.*: 'A 33Mpixel 120fps CMOS image sensor using 12b column-parallel pipelined cyclic ADCs'. ISSCC Dig. Tech. Papers, San Francisco, CA, USA, 2011, pp. 388–389

[7] HUANG S., *ET AL.*: 'A 2.5 inch, 33Mpixel, 60 fps CMOS image sensor for UHDTV application'. in Proc. 2009 International Image Sensor Workshop, Bergen, Norway, 2009, pp. 308–311

ISSN 2041-5923

Introduction to *Electronics Letters*

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 software used in a wide range of applications. The fields of research covered are relevant to many aspects of the broadcasting industry 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.

All three of the papers chosen this year have also appeared as features in the free magazine style news section that was added to Electronics Letters in 2010. 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. A version of the associated news article precedes each of the papers.

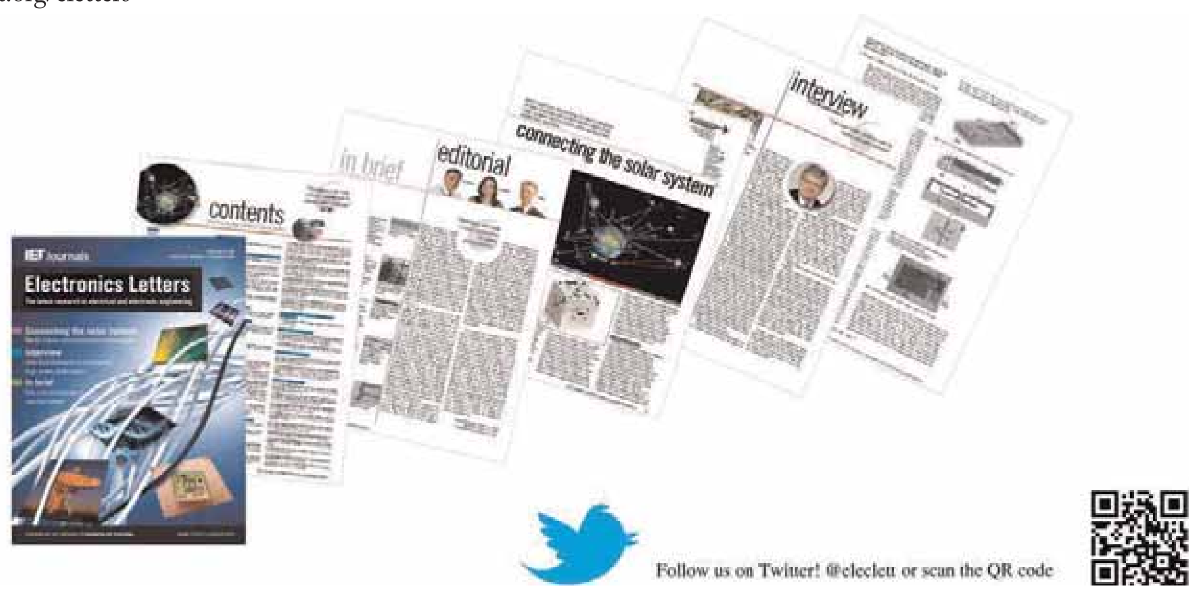
This year we have two papers from the arena of antenna design. One reports work on antennas that can be integrated into clothing for mobile applications, and the other looks at a way to reduce the size of MIMO antenna arrays to gain the advantages of multiple antenna systems in a more compact unit. The third paper reports work from Fujitsu that reduces the power requirements of lasers used in optical communication transmitters.

A common thread in all three papers is the improved distribution of Web content, which as Nick and Martin observe in their editorial (page 3) is sure to have a major impact on the continuing development of broadcasting media and technology.

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



ISSN 2041-5923

Antennas zip it up

The first zip-based antennas have been created by researchers at the University of Rennes in France. In their design the zip behaves like a monopole antenna and it can be tuned using the slider on the zip's metal teeth to define the length of the antenna and therefore the resonant frequency. The prototype demonstrates that it could be used in WiFi applications at 2.4–2.7 GHz.

Above The team at the Institute of Electronics and Telecommunications of Rennes (IETR) with some of their wearable antennas.
From left to right: S. Collardey, A-C. Tarot, M. Mantash, and K. Mahdjoubi

A useful accessory

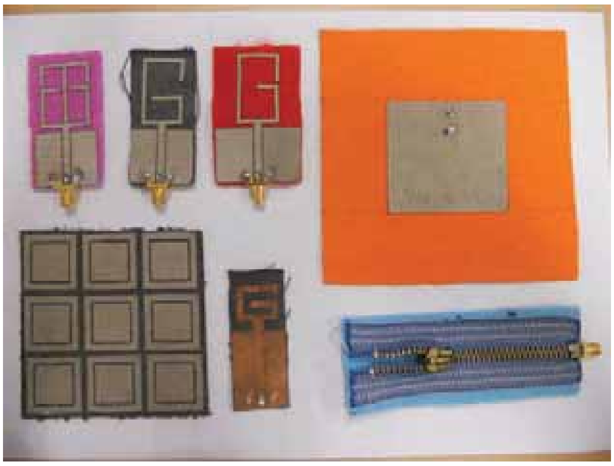
Many types of antennas that can be integrated into clothing have appeared in the last few years. They have many potential uses – such as security systems, monitoring the body for medical diagnosis or during exercise, transport fare payment and hospital patient health management – but only a few wearable antennas are available commercially at the moment for mainly military applications.

Wearable antennas are most commonly from the family of microstrip patches, planar-inverted F-antennas or monopoles over EBG/AMC reflectors, and they offer the advantages of being flexible, low in cost, low maintenance and reasonably robust. As they form a relatively large area over the surface of the cloth, they can also offer more services and availability than mobile or miniature terminals.

The most attractive wearable antennas so far are based on accessories like buttons and belts as the conductive element is already integrated into the clothing. Textile-based antennas have been more challenging to develop as the textile needs to be metallised, which complicates the process, and also leads to conductivity and loss issues. Textiles also present other challenges such as not being waterproof.

Shape-shifter

The researchers at the Institute of Electronics and Telecommunications of Rennes (IETR) have been working on wearable antennas, and more specifically textile antennas, for the last three years. In this work they focused on the zip as they wanted to develop reconfigurable or tunable antennas based on a common and easy-to-use accessory.



Above Some of the textile antennas developed at IETR including a G-shape antenna, an EBG based on a rectangular cell, a patch antenna and a zip

'Considering the zip slider as a switcher, one can use it to connect or de-connect parasitic elements to change the antenna shape in order to modify, for example, the radiation or polarisation pattern.' The same technique can be used to add or subtract lumped elements to control, for example, the filtering effect or the impedance matching for microwave circuits or antennas. The zip's teeth constitute a periodic structure and can be considered as a corrugated or periodically loaded (dielectric) transmission line in high frequency (millimetre-wave) ranges. It can therefore be used as a waveguide for Goubau waves. It can also be considered as a metamaterial line exhibiting left-handed behaviour and backward wave propagation.

The prototype zip antenna described in this issue of *Electronics Letters* was created by combining the zip with an SMA connector and sewing it to a felt substrate with a piece of electro-textile underneath to create the ground plane. The biggest challenge that the researchers faced was matching the antenna to 50 Ω; as the zip width (teeth length) is fixed by the manufacturer, this was done instead by careful choice of the thickness and permittivity of the backing material.

Having demonstrated the first prototypes of the zip antennas at 2.5 GHz, the team will next be working on a prototype piece of clothing, such as a denim jacket, to test the performance of the zip antenna in an operational context.

A solution now

These zip antennas have been developed as part of a French national project 'METAVEST' (metamaterials for smart garments), for which Anne-Claude Tarot, one of the IETR researchers, is the co-ordinator. The project has four academic and mobile communication company partners, and IETR's contribution is through the development of textile antennas in the WiFi bands, and the evaluation of on-body propagation channels.

There has been rapid progress in the fabrication technologies of conductive fibrous materials, and there is also increasing demand for wireless communications in smart clothing systems, so the potential application of wearable antennas continues to increase. The advantage of the IETR team's zip antennas is that they do not require any new technology and are able to reply immediately to the demands for wearable antennas. The researchers hope they can therefore help to more rapidly grow the markets and industries of communicating objects and smart clothing.

Selected content from *Electronics Letters*

ISSN 2041-5923

Wearable monopole zip antenna

M. Mantash A.C. Tarot S. Collardey K. Mahdjoubi

Institute of Electronics and Telecommunications of Rennes (IETR), University of Rennes 1, Rennes, France
E-mail: mohamad.mantash@univ-rennes1.fr

Abstract: The design of a novel textile monopole antenna based on a zip at 2.5 GHz is presented. A prototype is manufactured and characterised based on return loss and radiation pattern. The bandwidth and radiation characteristics obtained make the zip antenna suitable to be used for WiFi applications in the band from 2.4 to 2.7 GHz.

1 Introduction

In recent years, wearable antennas [1] have attracted much attention. There are many examples of antennas being integrated in garments [2–4] but probably the more appealing ones are those which are focused on accessories like buttons [5], belts [6] and other items. The main advantage is that conductive elements are already integrated in the clothes, and then additional metallisation is not required.

For this reason, in this reported work, a zip is used to realise a monopole antenna. This element is very often present in garments, often in pockets and it can be an easy way to hide antennas. Preferably a pocket zip will be used for practical considerations. This novel zip antenna can be used in WiFi communication systems. The geometry of the zip antenna is firstly described. The zip antenna is manufactured using felt as dielectric and electro-textile as the conductor. Theoretical and experimental return losses are compared to those of a conventional textile monopole antenna. Also, the radiation pattern of the manufactured prototype has been measured in an anechoic chamber.

2 Zip antenna geometry

The geometry of the zip monopole antenna is shown in Fig. 1. The zip is sewn on the upper side of a 2 mm-thick felt fabric. The ground plane is realised with Shieldit electro-textile in the back side (dimensions Wg by Lg). The felt fabric was characterised and exhibits a relative dielectric permittivity $\epsilon_r = 1.22$ and loss tangent $tg\delta = 0.016$. The connected zip monopole length above the ground plane, Lz, determines the resonance frequency and so it is designed to be $\lambda g/4$. The antenna zip was optimised to operate at 2.5 GHz. The dimensions are given in Table 1. The antenna is then realised (Fig. 2) and measured. Moreover, the zip is replaced by a printed line to accomplish a conventional monopole antenna [7, 8] on the felt fabric to be taken as reference (Fig. 1 right).

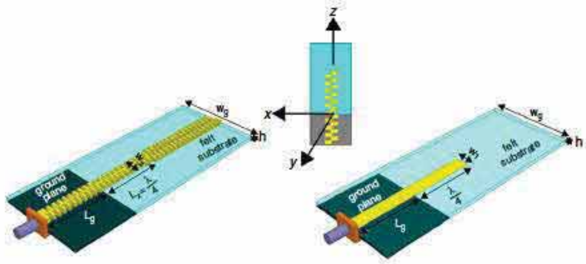


Figure 1 *Antennas: zip geometry (left) and monopole geometry (right)*

Table 1 Zip antenna dimensions

Prototype	Dimensions, mm			
	Lz	Wz	Lg	Wg
ZIP	27	6	–	–
Ground plane	–	–	30	40

3 Results

The simulated return loss for the zip monopole antenna is compared to that of the conventional monopole antenna in Fig. 3. A slight frequency shift can be observed because the

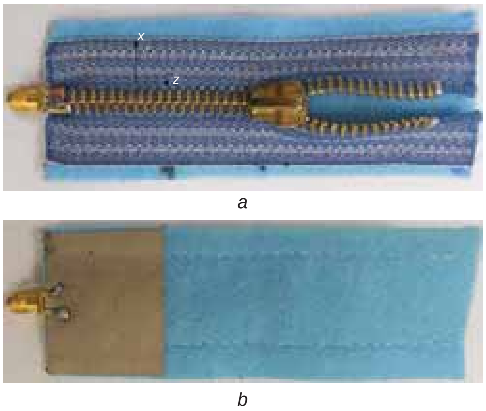


Figure 2 *Antenna prototype*

a Zip antenna upper side
b Zip antenna back side

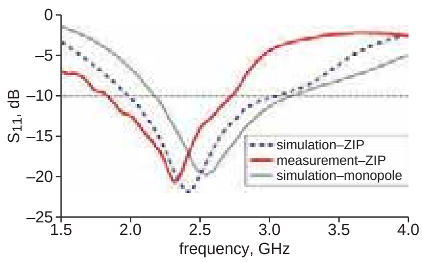


Figure 3 *Return loss simulation and measurement for zip antenna and simulation for conventional microstrip monopole antenna on felt fabric*

electrical length of the zip antenna is longer than that of the conventional monopole. It can be concluded that the zip antenna works like a conventional monopole.

The measured return loss of the zip antenna is compared to the simulated one. There is good agreement between simulation and measurement and a bandwidth of 1 GHz is obtained in both cases. The measured resonance frequency is slightly shifted down in frequency compared to simulation owing to manufacturing tolerances using the electro-textile and soldering the connector.

To complete the prototype characterisation, radiation pattern measurements have been carried out using a spherical near-field antenna test system (SATIMO SG32). From the radiation pattern cuts measurements it can be concluded that the zip antenna tends to be omnidirectional in the H-plane, almost equivalent to a monopole. The simulated radiation pattern cut in the E-plan for both a conventional monopole and the zip antenna are also very similar, as shown in Fig. 4. The measured radiation pattern cut in the E-plane of the manufactured prototype for the zip antenna is plotted in Fig. 4.

Finally, the measured realised gain of the antenna at 2.5 GHz is close to 0 dB. The gain could be improved by

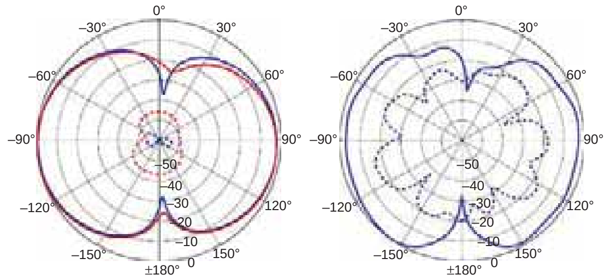


Figure 4 *Radiation patterns at 2.5 GHz (normalised, in dB)*

E-plane: zip antenna (red) and monopole antenna (blue) simulated for comparison (left); zip antenna measured (right)
—— co-polarisation
– – – cross-polarisation

increasing the ground plane size and/or by using a metal with higher conductivity.

4 Conclusion

The design of a novel textile monopole antenna based on a zip for WiFi applications has been presented. A prototype has been manufactured at 2.5 GHz and characterised based on return loss and radiation pattern measurements in an anechoic chamber. Results show that the behaviour of the zip antenna is equivalent to the conventional monopole one. The prototype presented in this Letter has several advantages such as compact size, a planar feature, flexiblility, easy to disguise and integrate in a garment, tunable, low cost and simple practical implementation.

5 Acknowledgments

This work has been supported by the 'Agence Nationale de la Recherche' (ANR), France, under grants ANR-09-VERS-003 for the METAVEST project and also by the 'Images et Réseaux' cluster. The authors thank J. Sol for help in anechoic chamber measurements.

6 References

[1] ZHU S., LANGLEY R.: 'Dual-band wearable textile antenna on an EBG substrate', *IEEE Trans. Antennas Propag.*, 2009, **57**, (4), pp. 926–935

[2] LOCHER I., KLEMM M., KIRSTEIN T., TROSTER G.: 'Design and characterization of purely textile patch antennas', *IEEE Trans. Adv. Packag.*, 2006, **29**, (4), pp. 777–788

[3] GUPTA B., SANKARALINGAM S., DHAR S.: 'Development of wearable and implantable antennas in the last decade: a review'. Proc. 2010 Mediterranean Microwave Symp., (MMS), Guzelyurt, Cyprus, August 2010, pp. 251–267

[4] VALLOZZI L., VANDENDRIESSCHE W., HENDRIK R., HERTLEER C., SCARPELLO M.L.: 'Wearable textile GPS antenna for integration in protective garments'. Proc. Fourth European

Conf. on Antennas and Propagation, Barcelona, Spain, April 2010, pp. 1–4

[5] SANZ-IZQUIERDO B., HUANG F., BATCHELOR J.C.: 'Covert dual-band wearable button antenna', *Electron. Lett.*, 2006, **42**, (12), pp. 668–670

[6] SANZ-IZQUIERDO B., BATCHELOR J.C.: 'A dual band belt antenna'. Proc. Int. Workshop on Antenna Technology: Small Antennas and Novel Metamaterials, (iWAT), Chiba, Japan, March 2008, pp. 374–377

[7] GARG R., BHARTIA P., BAHL I., ITTIPIBOON A.: 'Microstrip antenna design handbook' (Artech House, Boston, London, 2001)

[8] BALANIS C.: 'Antenna theory: analysis and design' (John Wiley & Sons, Inc., New York, 2005)

ISSN 2041-5923

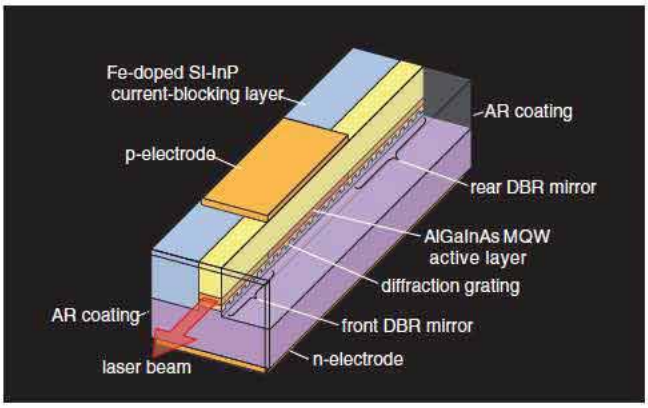
Mirror, mirror

Researchers from Fujitsu Laboratories Ltd in Japan have continued their work towards creating low-power high-speed optical transmitters by reducing the peak driving current in their distributed-reflector (DR) lasers. By using distributed-Bragg-reflector mirrors and anti-reflection coatings in the 1.3 μm AlGaInAs MQW DR lasers, they were able to achieve 25.8 Gbit/s direct modulation with very low driving currents at temperatures up to 85°C.

Heavy traffic

With data traffic increasing rapidly – for example, by 30–40% each year in Japan – there has been much global research activity into high-speed optical transmitters. The IEEE standardised 10 Gigabit Ethernet (GbE) in 2005 and then 100 GbE in 2010, which represents a ten-fold increase in capacity over 5 years. Recently, 100 GbE transponders have been commercially developed, and research beyond 100 GbE has now begun.

Alongside increasing the capacity of optical transmitters, reducing the increasing power consumption is also a major goal. For short-reach transmission, electro-absorption-modulator-integrated lasers (EMLs) have very good high-speed performance characteristics, but the power consumption is still too high.



Above: Fujitsu's DR lasers have integrated distributed-Bragg reflector mirrors on both sides of the short active region to reduce the threshold gain for lasing

Directly modulated semiconductor lasers, which operate without the need for thermoelectric cooling, offer a lower power consumption, but their performance at high speeds needs to be developed as it is limited mainly by the relaxation oscillation frequency and the parasitic capacitance. The Fujitsu team are therefore working on this as part of the 'Next Generation High-Efficiency Network Device Project' organised by the 'New Energy and Industrial Technology Development Organsiation' in Japan.

On reflection

In a conventional distributed-feedback (DFB) laser, shortening the cavity length causes an increase in the threshold gain for lasing and degrades the performance of the laser at high temperatures. In their previous work, the Fujitsu researchers created 'distributed-reflector' lasers in which distributed Bragg reflector mirrors are integrated at the back or on both sides of the active region to help reduce the threshold gain, even when the active region is reduced in length. Anti-reflection coatings on both facets of the lasers also stopped the facet-phase problem normally seen in conventional DFB lasers. With this design the researchers are able to get close to 100% singlemode lasing.

In their work in this issue of *Electronics Letters* the researchers demonstrate how their laser cavity design is able to achieve high speed and low power simultaneously. They were able to demonstrate 25.8 Gbit/s direct modulation with a driving current of 30 mA at 25°C and 38.9 mA at 85°C using their 1.3 μm AlGaInAs MQW DR lasers. This driving current is significantly better than the previous value using DMLs of 60 mA or higher at high temperatures, and it can allow operation with a third less driving power compared to EMLs.

Searching high and low

To reduce the driving current even further, the researchers will need to increase both the relaxation oscillation frequency and its efficiency against the drive current. They can do this by further shortening the length of the active region, but they will have to find ways of balancing this with making the front DBR mirror longer to avoid an increase in the threshold gain, and reducing the current leakage that results from an increase in the driving voltage and laser resistance.

It is a significant challenge to realise a reduction in power consumption and an increase in the transmission capacity simultaneously. The team will continue to improve the performance of their DR lasers, and they hope that higher capacities may also come from digital signal processing in future transmitters and receivers.

Selected content from *Electronics Letters*

ISSN 2041-5923

Uncooled, low-driving-current 25.8 Gbit/s direct modulation using 1.3 μm AlGaInAs MQW distributed-reflector lasers

M. Matsuda[1,2] *T. Simoyama*[1,2] *A. Uetake*[1,2] *S. Okumura*[3] *M. Ekawa*[1,2] *T. Yamamoto*[3]

[1]*Fujitsu Laboratories Ltd and Fujitsu Limited, 10-1 Morinosato-Wakamiya, Atsugi 243-0197, Japan*
[2]*Photonics Electronics Technology Research Association (PETRA), 1-20-10 Sekiguchi, Bunkyo-ku, Tokyo 112-0014, Japan*
[3]*Fujitsu Laboratories Ltd, 10-1 Morinosato-Wakamiya, Atsugi 243-0197, Japan*
E-mail: matsuda.manabu@jp.fujitsu.com

Abstract: Uncooled, low-driving-current 25.8 Gbit/s direct modulation with clear eye-opening up to 85°C is demonstrated using 1.3 μm AlGaInAs multiple quantum well (MQW) distributed-reflector lasers. The peak driving currents were as low as 30.0 and 38.9 mA at 25 and 85°C, respectively.

1 Introduction

Continuous increase in data traffic has been accelerating development of high-speed optical transmitters. Since directly modulated semiconductor lasers have an advantage in low power consumption, attempts to increase their modulation speed have been actively studied [1–5]. There are several reports on uncooled 25 Gbit/s direct modulation using 1.3 μm-wavelength lasers [2–5] and the maximum operating temperature was up to 100°C [5]. However, the peak driving currents were still high and 60 mA or more is required at high temperatures.

To realise high-speed optical modules with low power consumption using directly modulated lasers, higher resonant oscillation frequency (f_r) under low driving current is required. To obtain higher f_r, it is necessary to reduce the volume of the active region and to increase the differential gain of the active region. To meet these requirements, we recently introduced the concept of distributed-reflector (DR) lasers [6] to short-cavity lasers. A short AlGaInAs quantum well DFB active region for larger differential gain with integrated DBR mirrors on both sides of the active region provides high f_r and we have already reported uncooled 40Gbit/s direct modulation using our DR lasers [7, 8]. This cavity structure is also effective for stable singlemode operation owing to avoiding the influence of facet phases of a DFB grating, in contrast to conventional DFB lasers with the cavity of AR–HR (antireflection – high-reflection) facet coatings.

In this reported work, we investigated the reduction of the driving current for uncooled 25.8 Gbit/s direct modulation using 1.3 μm AlGaInAs MQW DR lasers, and we achieved low driving current less than 40 mA at peak up to 85°C.

2 Fabricated devices

We have fabricated DR lasers that integrate distributed-Bragg-reflector (DBR) mirrors on both sides of the short active region, in order to reduce the threshold gain for lasing for a short-cavity structure. The length of the DFB active region is as short as 100 μm. The rear DBR mirror is set at 100 μm, and acts as a high-reflectivity coating in the conventional DFB lasers. The front DBR mirror is set at 50 μm. A longer front DBR mirror provides higher frequency response at the low driving current up to high temperatures owing to larger feedback to an active region. The grating is formed continually over an entire waveguide.

The coupling coefficient of the grating is set at 210 cm^{-1} to obtain sufficient reflectivity. The active region consists of compressively-strained AlGaInAs quantum wells of 12 layers for larger differential gain, and the DBR mirror regions have lattice-matched AlGaInAs passive waveguides. The mesa stripe including the active region and the DBR mirrors is buried with Fe-doped InP semi-insulating buried-heterostructure (SI-BH) layers. The buried waveguide also contributes to reducing the volume of the DFB active region. Both facets of the lasers are also covered with antireflection coatings. Therefore, stable singlemode lasing is achieved, avoiding the facet-phase problem.

3 Basic characteristics

The threshold current (I_{th}) of this laser was 3.6 mA and the maximum output power was over 15 mW at 25°C. I_{th} at 50, 70 and 85°C was 5.5, 8.1 and 11.0 mA, respectively. We obtained optical power higher than 5 mW up to 85°C. We measured the dependence of f_r on the square root of the driving current (I) from the relative intensity noise (RIN) spectrum. Owing to the short-cavity structure and a large feedback to the active region, a very high resonant frequency of over 25 GHz was obtained at 25°C with injection current of 40 mA, as shown in Fig. 1. Even when the temperature increased to 85°C, a still high f_r value of 17.2 GHz was attained at injection current of 60 mA. The slopes of resonant frequencies were as high as 4.04 and 3.27 GHz/mA$^{1/2}$ at 25 and 85°C, respectively. Fig. 2 shows the small-signal frequency response at the driving current of $I_{th} + 25$ mA from 25 to 85°C. Wide −3 dB bandwidths of 17.5 GHz were achieved even at 85°C under the low driving current of 36 mA.

4 Modulation characteristics

We carried out high-speed direct modulation experiments at 25.8 Gbit/s with the non-return-to-zero (NRZ) signal having a $2^{31}-1$ pseudorandom bit sequence (PRBS). We changed and reduced the applied bias current (I_b) and

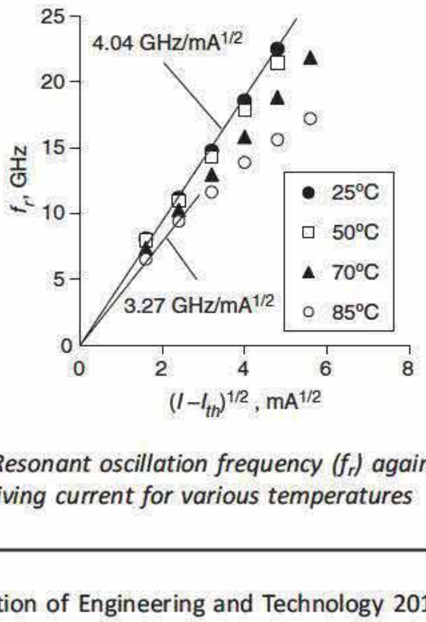


Figure 1 *Resonant oscillation frequency (f_r) against square root of driving current for various temperatures*

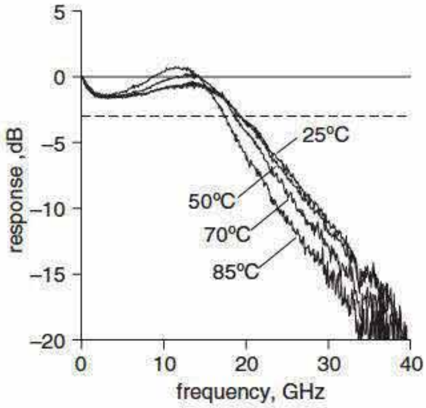


Figure 2 *Small-signal frequency response at driving current of $I_{th} + 25$ mA for various temperatures*

modulation amplitude while maintaining the dynamic extinction ratio of 5 dB, and we set the amplitude voltage as low as 1.04 $V_{p\text{-}p}$. Fig. 3 shows the 25.8 Gbit/s eye diagrams at 25, 50, 70 and 85°C. The modulation current was 21.8 $mA_{p\text{-}p}$ and the centre bias currents at 25, 50, 70 and 85°C were 19.1, 21.5, 24.6 and 28.9 mA, respectively. We observed that the eyes clearly opened up to 85°C. The longer (50 μm-long) front DBR mirror caused large damping in frequency response, as shown in Fig. 2, and also led to clearer eye-opening by suppressing overshoot. The peak driving currents (I_{peak}) during modulation at 25, 50, 70 and 85°C were as low as 30.0, 32.4, 35.5 and 39.8 mA, respectively. The peak driving current below 40 mA in 25 Gbit/s direct modulation with dynamic extinction ratio of 5 dB was the lowest, to the best of our knowledge [3–5]. Fig. 4 shows lasing spectra during 25.8 Gbit/s modulation at the same conditions as measuring eye diagrams, for each temperature. We confirmed stable singlemode oscillation under direct modulation up to 85°C. The sidemode suppression ratios were over 43 dB.

These results show that the AlGaInAs-MQW DR laser structure is effective for uncooled high-speed transmission with low driving current under stable singlemode operation.

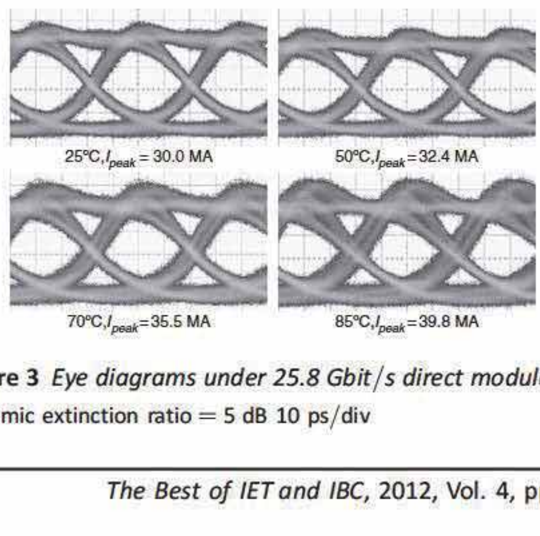


Figure 3 *Eye diagrams under 25.8 Gbit/s direct modulation*
Dynamic extinction ratio = 5 dB 10 ps/div

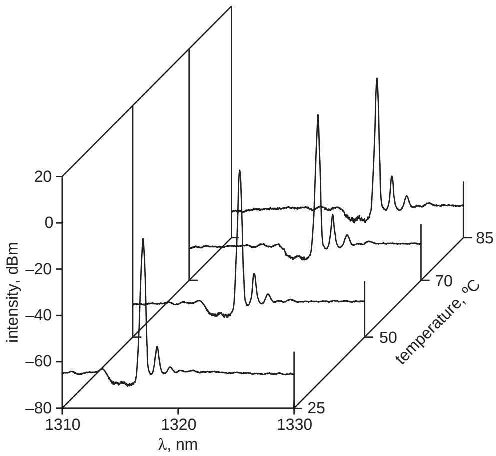


Figure 4 *Lasing spectra under 25.8 Gbit/s direct modulation*

Driving conditions are same as in Fig. 3

5 Conclusions

We have demonstrated low-driving-current 25.8 Gbit/s direct modulation with clear eye-opening using 1.3 μm AlGaInAs MQW DR lasers up to 85°C. The peak driving currents during modulation with dynamic extinction ratio of 5 dB were as low as 30.0 and 39.8 mA at 25 and 85°C, respectively. Stable singlemode oscillation up to 85°C with sidemode suppression ratios of over 43 dB was achieved under modulation. These results show that the 1.3 μm AlGaInAs-MQW SI-BH DR lasers are promising for future low-power-consumption high-speed transmitters.

6 Acknowledgment

A part of this work belongs to the 'Next-generation High-efficiency Network Device Project', which PETRA contracted with the New Energy and Industrial Technology Development Organisation (NEDO).

7 References

[1] NAKAHARA K., TSUCHIYA T., KITATANI T., SHIOTA K., TANAKA S., AOKI M., NAOE K., SASADA N., YAMAUCHI S., SHISHIKURA M., HATANO T., MOROHASHI N., INOUE H.: '40-Gb/s direct modulation with high extinction ratio operation of 1.3-μm InGaAlAs multiquantum well ridge waveguide distributed feedback lasers', *IEEE Photonics Technol. Lett.*, 2007, **19**, pp. 1436–1438

[2] OTSUBO K., MATSUDA M., TAKADA K., OKUMURA S., EKAWA M., TANAKA H., IDE S., MORI K., YAMAMOTO T.: 'Uncooled 25-Gbit/s direct modulation of semi-insulating buried-heterostructure 1.3-μm AlGaInAs quantum-well DFB lasers', *Electron. Lett.*, 2008, **44**, pp. 631–633

[3] TADOKORO T., YAMANAKA T., KANO F., OOHASHI H., KONDO Y., KISHI K.'Operation of a 25-Gb/s direct modulation ridge waveguide MQW-DFB laser up to 85°C', *IEEE Photonics Technol. Lett.*, 2009, **21**, pp. 1154–1156

[4] FUKAMACHI T., ADACHI K., SHINODA K., TSUJI S., KITATANI T., TANAKA S., AOKI M.'Recent progress in 1.3-μm uncooled InAlGaAs directly modulated lasers'. 22nd IEEE Int. Semiconductor Laser Conf., Kyoto, Japan, 2010, (ThB1)

[5] FUKAMACHI T., ADACHI K., SHINODA K., KITATANI T., TANAKA S., AOKI M., TSUJI S.'Wide temperature range operation of 25-Gb/s 1.3-μm InGaAlAs directly modulated lasers', *IEEE J. Sel. Top. Quanum Electron.*, 2011, **17**, pp. 1138–1144

[6] SHIM J.-I., KOMORI K., ARAI S., ARIMA I., SUEMATSU Y., SOMCHAI R.: 'Lasing characteristics of 1.5 μm GaInAsP-InP SCH-BIG-DR lasers', *IEEE J. Quantum Electron.*, 1991, **27**, pp. 1736–1745

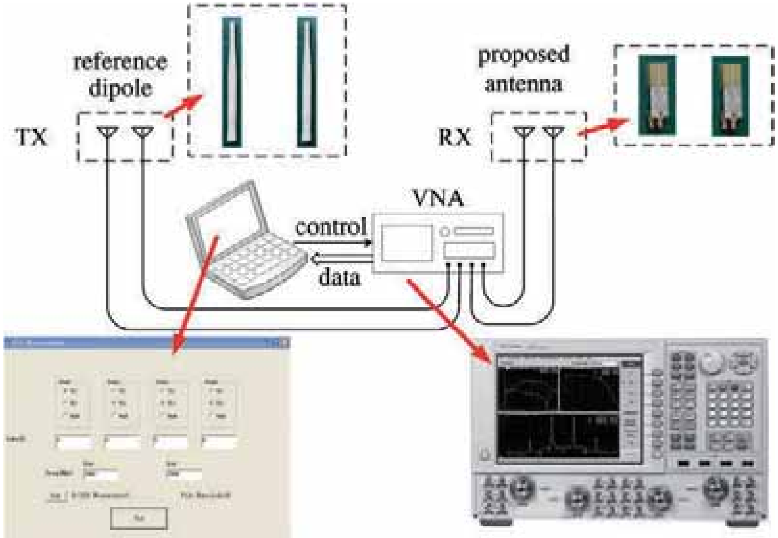
[7] YAMAMOTO T., UETAKE A., OTSUBO K., MATSUDA M., OKUMURA S., TOMABECHI S., EKAWA M.: 'Uncooled 40-Gbps direct modulation of 1.3-μm-wavelength AlGaInAs distributed reflector with semi-insulating buried-heterostructure'. 22nd IEEE Int. Semiconductor Laser Conf., Kyoto, Japan, 2010, (ThB3)

[8] SIMOYAMA T., MATSUDA M., OKUMURA S., UETAKE A., EKAWA M., YAMAMOTO T.: '40-Gbps transmission using direct modulation of 1.3-μm AlGaInAs MQW distributed-reflector lasers up to 70°C'. Optical Fiber Communication Conf. and Expo. and The National Fiber Optic Engineers Conf., Los Angeles, CA, USA, 2011, (OWD3)

ISSN 2041-5923

A little distance

Researchers at Tsinghua University have taken a 3D approach to MIMO antennas and have built a 2.4 GHz MIMO antenna for WLAN applications that is much smaller than previous designs but still has high isolation and good omnidirectional coverage.



Above: The testbench system is composed of a Vector network analyser, transmit and receive antennas, a computer and RF cables.

Two reference dipoles are utilised as the transmit antennas (TX), and two test antennas are used as the receive antennas (RX). The 2×2 antennas are connected to the 4 ports of the VNA

The isolation issue

The spectrum efficiency and transmission quality of wireless communication systems are improved through the use of multiple antennas at both the transmit and receive ends. MIMO antennas are being widely researched at the moment as they enable the use of the spatial degree of freedom in a wireless multipath channel, and they form a key technology for future wireless communication systems such as 3GPP, LTE, WiMAX 802.16, IEEE 802.20 and IMT-Advanced.

At the moment MIMO antennas are only adopted in some large static base stations. This is because in these space-separated antenna arrays, nearly half of the wavelength is needed to achieve acceptable isolation; about 15 dB for most situations. In modern communication systems, the space is limited in both the base stations and the mobile terminals, and as the antennas shrink in size, the mutual coupling between the elements increasingly restricts the performance. The design of antennas for a space-limited MIMO system is therefore still a challenge.

A couple of challenges

Two previous methods have been used to reduce the size. First, decoupling circuits with lumped components reduce the mutual coupling between elements, but the space between two antennas can only be reduced to slightly less than half a wavelength, high loss is introduced, and it is a complicated arrangement. Secondly, two antenna elements can be arranged orthogonally, and antennas with different polarisations can be arranged together to save on space. However, the size of the orthogonal array is still large, and cannot be decreased without deterioration of the mutual coupling, and the feeding network is difficult to design in a compact structure.

The group from the State Key Lab of Microwaves and Communications at Tsinghua University has focused on MIMO antenna design for about the last 20 years. To address the latest requirements of MIMO systems, they have designed different 2.4 GHz dual-polarised antennas. Having previously worked on orthogonal antenna arrangements, they took a different approach in their Letter which appeared in the last issue (23) of *Electronics Letters*, and designed a co-located MIMO antenna using a monopole and a slot element.

Space saving

In order to reduce the volume of the antenna, the researchers folded the ground of the monopole to create a 3D column structure. This gave them the dimensions of $50 \times 16 \times 16$ mm^3 ($0.4\ \lambda_0 \times 0.128\ \lambda_0 \times 0.128\ \lambda_0$, where λ_0 is the wavelength in free space).

They were able to enhance the isolation to over 24 dB by creating a perfectly symmetrical structure from the monopole and the slot.

Another merit of this design is the omnidirectional radiation pattern for both polarisations, which is required by omnidirectional coverage of base stations. For a typical dual-polarised MIMO antenna design, the antenna is unidirectional or bidirectional due to the large ground. The Tsinghua team showed that their folded ground design also achieves a good omnidirectional coverage and it can be mounted in portable access points with better performance than dipole antennas of single polarisation.

As the performance of MIMO antennas is dependent on their environment, the team next plan to measure the channel capacity of a 2×2 MIMO system in different indoor scenarios, for which they have designed their own measurement system.

Above: The Tsinghua group have recently developed several 2.4 GHz dual-polarised antennas.

Left to right: Two reference dipole with single polarisation; dual-polarised slot antenna with bidirectional radiation patterns; same antenna as previously but with a loop element rather than a slot element; and the co-located MIMO antenna

Selected content from *Electronics Letters*

ISSN 2041-5923

Dual-polarised monopole-slot co-located MIMO antenna for small-volume terminals

Y. Li Z. Zhang J. Zheng Z. Feng

State Key Lab of Microwave and Communications, Tsinghua University, Beijing, People's Republic of China
E-mail: zjzh@tsinghua.edu.cn

Abstract: A dual-polarised monopole-slot co-located MIMO antenna is proposed for 2.4 GHz wireless local area network (WLAN) applications. The antenna consists of a monopole with its ground folded and a compact coplanar waveguide feed, providing the vertical and horizontal polarisations. The antenna is designed for a volume-limiting MIMO system due to its compact dimensions of $50 \times 16 \times 16$ mm^3 ($0.4\lambda_0 \times 0.128\lambda_0 \times 0.128\lambda_0$) and high isolation better than -24 dB. The performance of the proposed antenna has also been measured, including the S-parameters, the radiation patterns and the gain, which agree well with the simulation results.

1 Introduction

For modern wireless communication technology, the multi-input multi-output (MIMO) system is a promising solution to provide high channel capacity. Multiple antennas are utilised at both sides of the transmitter and receiver to provide multi-channel transmission. However, at least half of the wavelength must be required for isolation between two adjacent elements, making the overall system very large. In a volume-limiting system, it is difficult to mount such a spatially extended antenna array.

To save space between two elements, the dual-polarised antenna has been widely studied and adopted in the MIMO system. The dual-polarised antenna with high isolation is able to take the place of two antennas with single polarisation. The feasibility of dual-polarised antennas in the MIMO system has been validated for indoor and outdoor scenarios [1–3]. In recent papers [3–8], different patterns of dual-polarised antennas have been proposed, such as the patch [4, 5], slot [6, 7] and loop [8]. As discussed in [3], the dimensions and isolation are two key issues of dual-polarised MIMO systems. Low mutual coupling between modes can be achieved by adopting a high-isolated feeding structure as in [4–8]. However, the overall dimensions are dictated by the pattern of the radiating element. The required resonating length of the patch [4, 5] or slot [6, 7] is half of the wavelength for each polarisation. The dimensions can be decreased by using a loop antenna [8], the circumference of which is a wavelength. However, the mutual coupling increases with reducing volume. It is a challenge to reduce the antenna dimensions while maintaining port isolation.

In this Letter, a 3D design of a dual-polarised MIMO antenna system is proposed to achieve even smaller dimensions. Dual polarisations are provided by a monopole-slot co-located antenna at the operating frequency of 2.4 GHz for WLAN applications. The proposed antenna consists of a ground-folded monopole with a compact CPW feed with the overall dimensions of $50 \times 16 \times 16$ mm^3 ($0.4\lambda_0 \times 0.128\lambda_0 \times 0.128\lambda_0$, λ_0 is the wavelength in free space). The vertical polarisation is provided by the monopole and the horizontal polarisation is provided by the slot of the CPW, with the port isolation lower than -24 dB. The reflection coefficient, radiation patterns and gain are also measured. Because of the compact volume, the proposed antenna is suitable for volume-limited portable access points, where the antennas of [3–8] are unable to mount.

2 Antenna design and configuration

The geometry of the proposed antenna is shown in Fig. 1. As shown in Fig. 1*a*, the antenna is made of FR4 substrate ($\varepsilon_r = 4.4$, $\tan\delta = 0.01$), with thickness of 1 mm. A monopole and the CPW feed are printed on the front side. The CPW is connected to a 50 Ω microstrip line on the

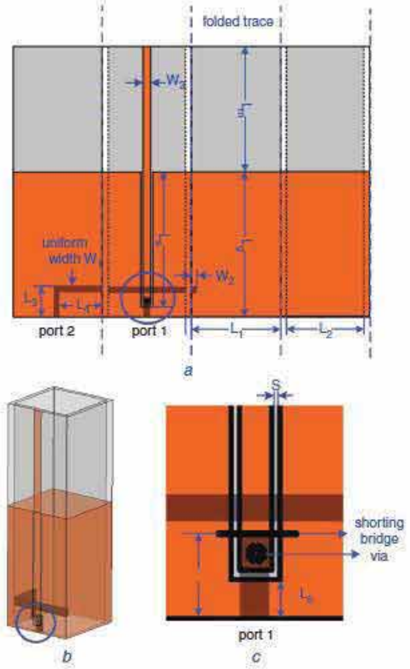


Figure 1 *Geometry of proposed antenna*

a Planar view
b 3D view
c Detailed view of shorting bridge

back side through a via, as shown in Fig. 1*c*. A shorting bridge is soldered on the front side. Another 50 Ω microstrip line on the back side positions across the CPW. By folding the ground along the folded trace, a centre-hollow cuboid antenna was built and is illustrated in the 3D view of Fig. 1*b*. As a result, the ground and the monopole are on the outer side and the microstrip lines are in the inner side. The overall dimensions are $50 \times 16 \times 16\ mm^3$, much more compact than the dimensions in the reference papers. When fed through port 1, the antenna operates at the vertical polarisation mode. The CPW feed structure is also designed to 50 Ω by tuning the parameters of W_3 and S. When the antenna is fed through port 2, the energy is coupled from the microstrip line to feed the slot of the CPW. Horizontal polarisation is excited in the slot. By tuning the position of the shorting bridge L_6 and the parameters of W_2, L_3 and L_4, good impedance matching is achieved. The lengths of the monopole and the CPW are approximately a quarter of a wavelength on the substrate. The key parameters are optimised by using the Ansoft High Frequency Structure Simulator (HFSS) software. The detailed optimised values of each parameter are listed in Table 1.

Table 1 Detailed dimensions of proposed antenna

Parameter	L_1	L_2	L_3	L_4	L_5	L_6	L_s
Value, mm	16	14	6.5	6.05	1.5	3	25.5
Parameter	L_m	L_g	S	W_1	W_2	W_3	
Value, mm	23	27	0.2	1.9	1	1.9	

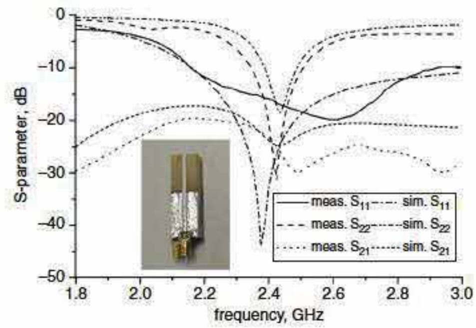


Figure 2 *Simulated and measured S-parameters of proposed antenna*

The isolation between the vertical mode and the horizontal mode is the key issue in the dual-polarised antenna design. The currents on both sides of the slot are in the same direction for the vertical mode and in the opposite direction for the horizontal mode. The amplitudes of the current on the slot sides are almost the same. Therefore, good orthogonality can be achieved between the two modes.

3 Experimental results

A prototype of the proposed dual-polarised antenna was built and measured. The simulated and measured S-parameters, including the reflection coefficient and isolation, are shown in Fig. 2. The measured −10 dB impedance bandwidths are 2.16–2.92 GHz and 2.35–2.51 GHz for vertical and horizontal modes, respectively, both covering the desired WLAN band of 2.4–2.48 GHz. In this band, the port isolation is better than −24 dB.

The measured radiation patterns at 2.44 GHz for both polarisations are shown in Fig. 3, compared with the

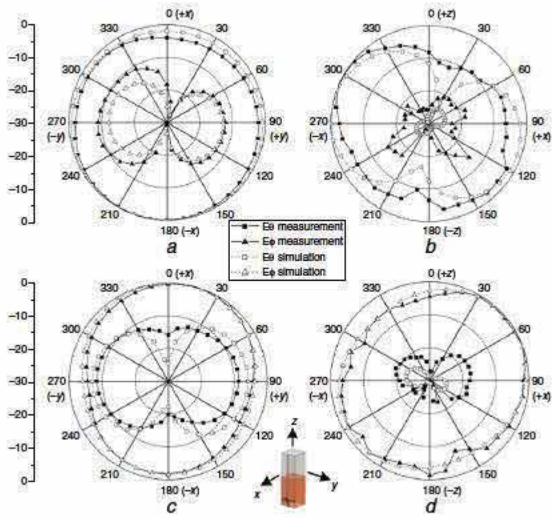


Figure 3 *Simulated and measured radiation patterns of proposed antenna at 2.44 GHz*

a *x-y* plane fed through port 1
b *x-z* plane fed through port 1
c *x-y* plane fed through port 2
d *x-z* plane fed through port 2

simulated results. For the vertical polarisation, a nearly omnidirectional pattern is shown in the azimuth plane. Owing to asymmetrical structure of the monopole and the ground, a tilt pattern is shown in the elevated plane. For the horizontal polarisation, the electric field propagates along the folded ground to the back side, and a nearly bidirectional pattern is shown in the azimuth plane. The measured gains are better than 2.2 and 0.8 dBi for the vertical and horizontal polarisations separately. The efficiency of the horizontal mode is a bit lower owing to the cavity effect of the folded ground.

4 Conclusion

A dual-polarised monopole-slot co-located MIMO antenna is proposed for volume-limited portable terminals. The vertical polarisation is radiated by the monopole with a folded ground fed by a CPW. The horizontal polarisation is radiated by the slot of the CPW. The overall dimensions of the proposed antenna are $50 \times 16 \times 16\,\text{mm}^3$, much more compact than the reference design. The isolation is lower than -24 dB in the 2.4 GHz WLAN band. Because of the advantages of compact dimensions and good isolation, the proposed MIMO antenna has potential application for volume-limited portable systems.

5 Acknowledgments

This work is supported by the National Basic Research Program of China under contract 2010CB327402, in part by the National High Technology Research and Development Program of China (863 Program) under contract 2009AA011503, the National Science and Technology Major Project of the Ministry of Science and Technology of China 2010ZX03007-001-01, and Qualcomm Inc.

6 References

[1] MOLINA-GARCIA-PARDO J.-M., RODRIGUEZ J.-V., JUAN-LLACER L.: 'Polarized indoor MIMO channel measurements at 2.45 GHz', *IEEE Trans. Antennas Propag.*, 2008, **56**, (12), pp. 3818–3828

[2] ERCEG V., SAMPATH H., CATREUX-ERCEG S.: 'Dual-polarization versus single-polarization MIMO channel measurement results and modeling', *IEEE Trans. Wirel. Commun.*, 2006, **5**, (1), pp. 28–33

[3] LI H., XIONG J., YING Z., HE S.L.: 'Compact and low profile co-located MIMO antenna structure with polarisation diversity and high port isolation', *Electron. Lett.*, 2010, **46**, (2), pp. 108–110

[4] CHIOU T.-W., WONG K.-L.: 'Broad-band dual-polarized single microstrip patch antenna with high isolation and low cross polarization', *IEEE Trans. Antennas Propag.*, 2004, **50**, (3), pp. 399–402

[5] WONG H., LAU K.-L., LUK K.-M.: 'Design of dual-polarized L-probe patch antenna arrays with high isolation', *IEEE Trans. Antennas Propag.*, 2004, **52**, (1), pp. 45–52

[6] LEE C.-H., CHEN S.-Y., HSU P.: 'Isosceles triangular slot antenna for broadband dual polarization applications', *IEEE Trans. Antennas Propag.*, 2009, **57**, (10), pp. 3347–3351

[7] LI Y., ZHANG Z., CHEN W., FENG Z., ISKANDER M.F.: 'A dual-polarization slot antenna using a compact CPW feeding structure', *IEEE Antennas Wirel. Propag. Lett.*, 2010, **9**, pp. 191–194

[8] LI Y., ZHANG Z., CHEN W., FENG Z., ISKANDER M.F.: 'Dual-mode loop antenna with compact feed for polarization diversity', *IEEE Antennas Wirel. Propag. Lett.*, 2011, **10**, pp. 95–98