grass valley

# WHITEPAPER

## 3D Distribution

Claude Perron, *Chief Technology Officer, Headend*

May 2010

Broadcasters already possess much of the transmission equipment necessary to conduct 3D trials to the home. However, the Blu-ray Disc Association initiative is further pushing the envelope with its goal of bringing the 3D experience into the living room with full 1080p resolution to each eye. Grass Valley has the encoding and systems-level expertise to provide solutions for today's 3D requirements. The company is also working with its partners to collaborate on several next-generation 3D solutions and is planning live events for 2010 and 2011.

# Contents

# Introduction

Following the successful introduction and adoption of HDTV, the entertainment industry is now investigating what might be the next step in the home viewing experience. Will the future of television be in super high-definition—approaching conditions similar to many state-of-the-art movie theaters? Or, considering the film industry's increase in 3D content output, will a 3D viewing experience carry over into consumers' living rooms? Interest in 3D has waxed and waned several times in the past. Now, however, flat display technology has penetrated home markets and is quite popular with large screens, ready for 3D viewing. 3D's present success in movie theaters could benefit from the 'wow' factor. A handful of broadcasters are conducting their own 3D trials, fishing for positive consumer feedback. The adoption of this new experience in home entertainment is hard to predict as 3D could be perceived as a gimmick. One particular barrier to the adoption of 3D are the glasses required in current systems.

**Figure 1 –** *The future of home entertainment?*

Longer-term solutions without glasses could be multiview solutions with auto stereoscopic displays. However, these solutions require significant investments in the display technology to produce true HD resolutions for both eyes. The industry needs to see the 3D experience adopted on a broad scale to drive this next wave of investment. Broadcasters are currently investing large sums of money to roll out HD, and they are looking at 3D as an evolution, generating requirements such as 2D backward compatibility, gradual introduction of 3D to coexist with 2D, and full 3D HD resolution per eye. We recall that it took nearly 20 years for HD to become commercially successful; so when will the next generation of television technology become a reality?

This whitepaper aims to describe 3D technologies with a focus on transmission. Products may include headend encoders, consumer set-top boxes, and displays. The first portion of the paper describes 3D solutions introduced some time ago with color anaglyphs compatible with existing displays. The second part of the paper, frame-compatible 3D solutions (also requiring glasses), is presented which broadcasters can deploy today using existing digital transmission infrastructures. These, however, require the new 3D displays introduced during the most recent Consumer Electronics Show in January 2010. The last section is devoted to the next step required to propel the industry forward. It calls for a new generation of headend encoders and set-top boxes, one which leverages the full 3D HD resolution per-eye capabilities of today's display technology.

# First Generation 3D: Compatible with Existing Displays

3D was invented in the early 1920s for movie theaters. It was based on a color anaglyph which is the combination of the left and right pictures using complementary color filters. The filters are typically red and blue/green (cyan) and are used to build a composite picture as described in Reference (1).

**Figure 2 –** *Anaglyph picture.*

**Figure 3 –** *Glasses for anaglyph pictures.*

An anaglyph carries the information of the left view in the red color channel and the information from the right view in the green and blue channels of the color image:

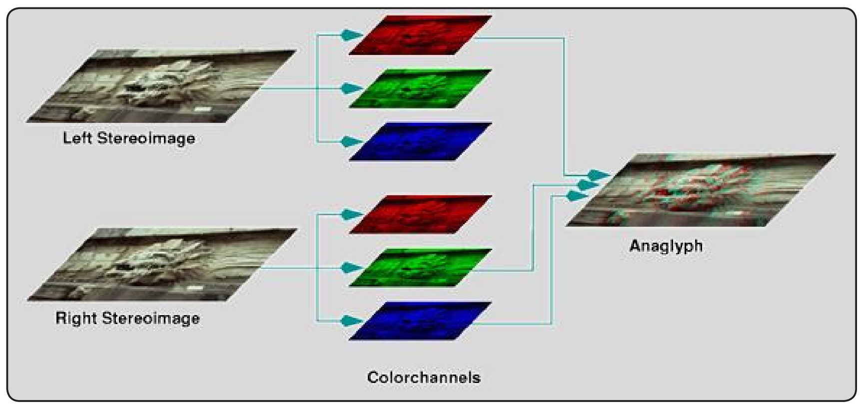



**Figure 4 –** *An anaglyph is created by taking the data from the red color channel of the left image and combining this with the green and blue channels of the right stereo image.*

The red filter in front of the left eye extracts only the information of the left view, and a cyan filter placed in front of the right eye only the information from the right view. Thus, the two eyes actually see the following images:


**Figure 5 –** *Viewing a color anaglyph through appropriate filters separates the 3D information for the left and right eye. The two images are arranged so that one can cross-fuse them. As depicted here, neither the left eye nor the right eye receives any useful color information. Interestingly, we are usually unaware of this effect.*

The two-color anaglyph method has one primary advantage: from the headend encoder to the set-top box, it does not require any specific infrastructure. It also works with existing displays. However, this technique suffers from severe limitations, which most likely explain its limited use until now:

- Poor color fidelity and resolution
- No 3D effects for objects whose color matches with the filters
- Discomfort generated by light disparities between the two eyes
- Ghosting due to the pairing of glass filters and screen (cross talk)

There are ways to optimize the design of the color anaglyph to work around those issues as described in Reference (2) by transforming the color space of the original pictures, however these techniques are merely compromises and offer limited color fidelity.

## Second Generation 3D: Compatible with Existing Transmission Infrastructure

An improved solution is to transmit each of the left and right pictures arranged to look like a single HDTV frame (frame-compatible solutions). This can be accomplished using the following techniques:

a) Arranging them as side-by-side anamorphic (side by side)

640 640

720

720p format

960 960

1080

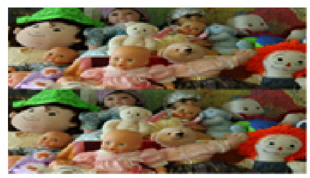
1080i format

b) Assembling left and right as over and under anamorphic (over/under)

1280

360

360

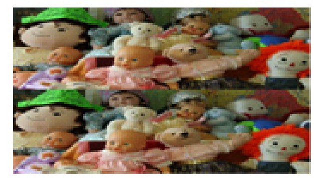


720p format

1920

540

540

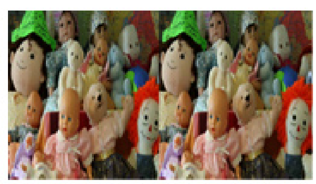


1080i format

c) Or quincunx sub-sampling as a compromise between horizontal and vertical resolution loss, pixels presented side-by-side

640 640

720



720p format

960 960

1080

1080i format

d) By using what is called offset sub-sampling, one pixel left, one pixel right (pixel interleaved)

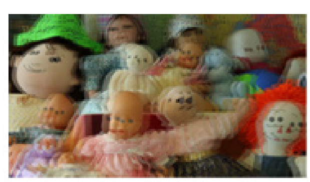
e) Or left and right on alternate lines (line interleaved).

f) Or quincunx sub-sampling as a compromise between horizontal and vertical resolution loss (checker board)

g) Frame or field interleaved

1280

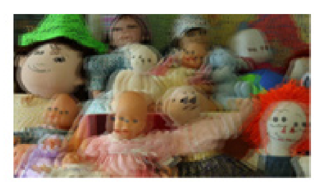
720



720p format

1920

1080



1080i format

With these methods, there is a compromise between quality and bandwidth. When one of these formats is used, there is a loss of resolution in either the vertical, horizontal, diagonal, or temporal planes. The results, however, may not be as poor as one might expect, because the stereopsis process in the brain effectively causes the left signal to lie on top of the right signal, so what you see is a sum of them.

Depending on the way the left and right views are combined for transmission, the composite signal matches more or less with the compression tools used in the transmission chain, with a drawback on the bitrate required. Solutions (a), (b), and (c) maintain spatial consistency, which is fundamental for the encoding process, but reduce the bandwidth by a factor of two for each view. Solutions (d), (e), and (f) create high frequencies due to the disparities between the two views, leading to a penalty in the compression efficiency, but keeps the full resolution where there is no disparity. Solution (g) uses time for splitting the two views with the transmission of odd field or frame for the left view and the even field or frame for the right view, thus preserving the spatial redundancy within each field or frame, relying on the motion compensation mechanism for the interview prediction. Solution (e) for the 1080i interlaced format benefits from the time splitting feature as the left view will be carried on the odd field, and the right view is carried on the even field, thus preserving the spatial redundancy within each field which fits with the field encoding modes of an H.264 encoder.

Different 3D format alternatives have to been tested, using a commercial AVC HD encoder, comparing bit rates for the same PSNR figures using the 1080i format close to broadcasters' operating points:

- Simulcast of left and right views
- Side-by-side
- Over/under
- Line interleaved
- Quincux offset sub sampling (Sensio format)
- Anaglyph

## Second Generation 3D: Compatible with Existing Transmission Infrastructure (cont.)

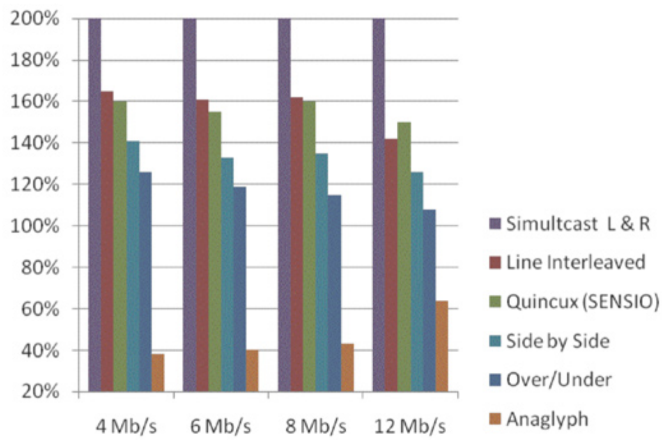



**Figure 10 –** *3D compression efficiency with Grass Valley EM3000 HD Encoder, 1080i format.*

Results of those experiments show that the anaglyph composite signal has the lowest entropy level (20% to 30% of the 3D signal) in line with the perception of poor quality compared to the other techniques.

The line interleaved and quincux methods require 20% to 30% less bitrate than the left and right simulcast solution, maintaining a significant level of entropy in the frame compatible signal as the dropped part of the 3D signal spectrum carries little energy. The pure horizontal or vertical decimation styles lead to a more severe entropy reduction of the 3D signal with a highest reduction being for the over/under format as aspect ratio is less appropriate (i.e., vertical decimation by a factor of two cuts more signal energy).

The quincunx sub sampling style represents a good compromise in preserving the horizontal and vertical definition while keeping the temporal resolution required for video material like sports. Experiments have been done with the Sensio Encoder in front of the Grass Valley EM3000 H.264 encoder. Sensio and Grass Valley are partnering to fine tune the H.264 encoder algorithms in order to optimize together the transmission chain.

To insure backward compatibility with legacy set-top boxes that support both 720p/1080i formats, a 3D ‘flag’ in the stream is required to notify the STB to modify the signal to the viewing modes:

- 2D mode: Upscale one view to be displayed full-screen. The side-by-side and over/under formats may be the easiest scenarios as all decoder ICs in deployed set-top boxes have horizontal and vertical up sampling capabilities. This means legacy STB require specific software be downloaded in order to be able to display a 2D picture from of a 3D stream.
- 3D mode: the set-top box is transparent, the 3D split processing is done in the display.

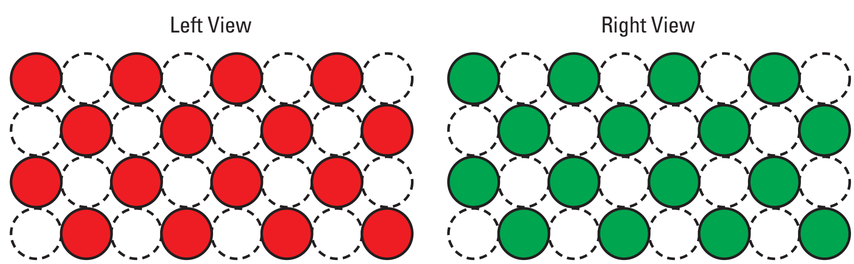



**Figure 11 –** *Quincux offset subsampling.*

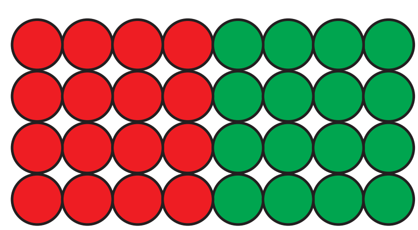


**Figure 12 –** *Transmitted frame.*

# Third Generation 3D: Full HD Resolution

With the Blu-ray Disc Association announcing its final 3D specifications in December 2009 (3) and display technology becoming available for 3D full resolution per eye, broadcasters must plan their 3D to coincide with the deployment of next-generation set-top boxes. Double HD or 1080p decode capabilities offer the opportunity to achieve Blu-ray-like 3D resolution streams to the home, which is particularly important for sports content. This new wave of consumer devices will require a new interface to the display. HDMI 1.4 version (4), revealed in June 2009, includes the definition of super high-definition formats up to 4k x 2k, of common 3D formats and resolutions up to 1080p, and supports many 3D techniques:

- Full side-by-side
- Half side-by-side
- Field alternative
- Line alternative
- Left + depth
- Left + depth + Gfx + Gfx depth

For reasons of backward compatibility, 1080p is anticipated to be broadcast using the Scalable Video Coding (SVC) format which itself is an extension of the ITU-T H.264 Advanced Video Coding (AVC) standard with the spatial scalability feature: the Base Layer that provides backward compatibility with today's HD broadcast streams encoded using AVC streams, and the Enhancement Layer that describes the incremental details toward the full 1080p resolution. SVC works in a similar way to AVC, with inter layer prediction possibilities on top of the standard AVC prediction modes (5):

- Inter layer Intra prediction
- Inter layer Motion Vector prediction
- Inter layer Residual prediction

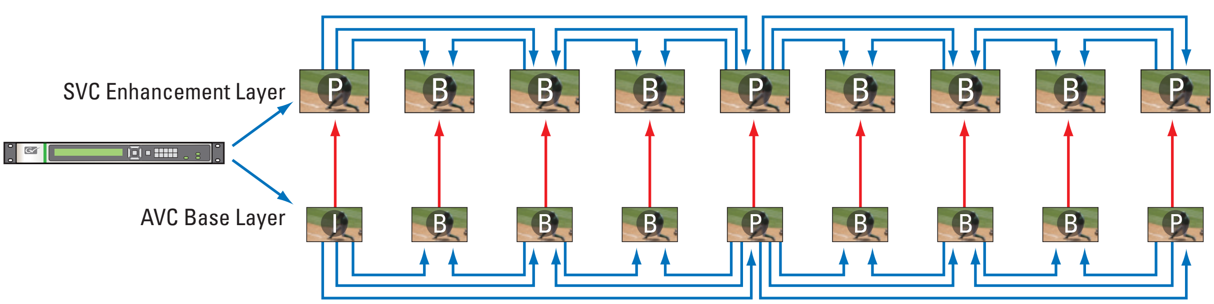



**Figure 13 –** *SVC encoder scheme.*

# Third Generation 3D: Full HD Resolution (cont.)

Headend encoders or decoder ICs in set-top boxes with SVC capability have approximately twice the encode or decode capability: the Base Layer encoder and the Enhancement Layer encoder or decoder. It is anticipated these devices will have Multiview Video Coding (MVC) capability as well. MVC is the revision to ITU-T Rec. H.264 which contains enhancement extensions to support 3D, and was approved in March 2009 (6).

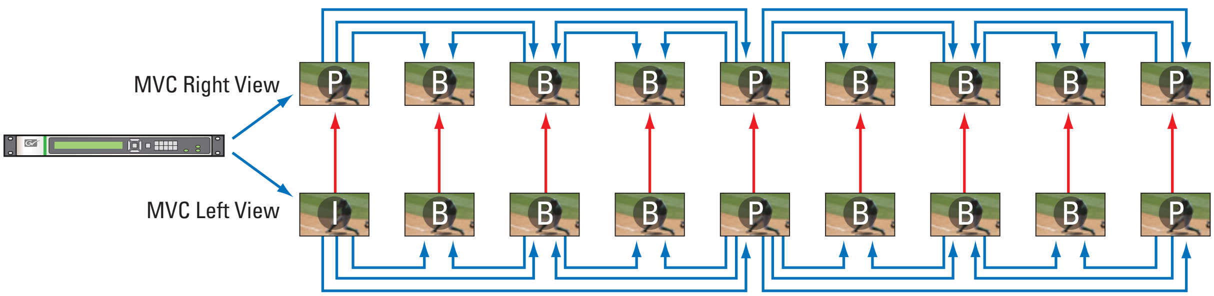



**Figure 14 –** *MVC encoder scheme.*

Instead of an inter layer prediction mode as in SVC, there is an inter view prediction mode which works very much like motion prediction. The primary difference between the two is the use of disparity vectors replacing motion vectors. The overhead required over 2D is between 65% and 100%, depending on source material. The overhead increases when the quality of encoding increases.

Using the JMVC model developed by the Joint Video Team, tests (7) show 75 to 100% overhead with 40 dB PSNR and 65 to 98% with 33 dB PSNR. This limited performance can be explained by the difference in statistics between motion vectors (in general one peak) and disparity vectors (several peaks corresponding to different objects at different depths) and the temporal motion prediction scheme using a linear model that fits with motion vectors but is not as accurate when predicting disparity vectors. It has been observed that the bitstream overhead can be attributed to the encoding of large residuals with revealed regions between the two views. As expected with predictive coding, performance varies with content. MVC can provide great bandwidth savings over the simulcast approach when, for example, the block-based translational temporal prediction does not match well (typically content with large zoom or rotation).

Paving the way for 3D at home, the Blu-ray 3D specification calls for encoding 3D video using MVC.

# Display Aspects

As stated above, any display device can be used with the 3D anaglyph solution. For the second and third generation of 3D solutions, new displays are required. The display industry is going to great lengths to market these products. During the most recent Consumer Electronics Show in Las Vegas, several options were demonstrated, all claiming HD resolution per eye, if not yet 1080p per eye:

- New 3D displays with passive polarized glasses and polarizers in front of standard screens for left and right views. By nature these are reducing the resolution by a factor of two. They also have some issues with crosstalk on top, depending on the viewing angle, though mainly for LCD-based displays (parallax between the display panel and the micro-polarizers with a substrate between the two elements. The problem can be minimized when left and right views are line interlaced so that lateral head movement is no longer affected by parallax.)

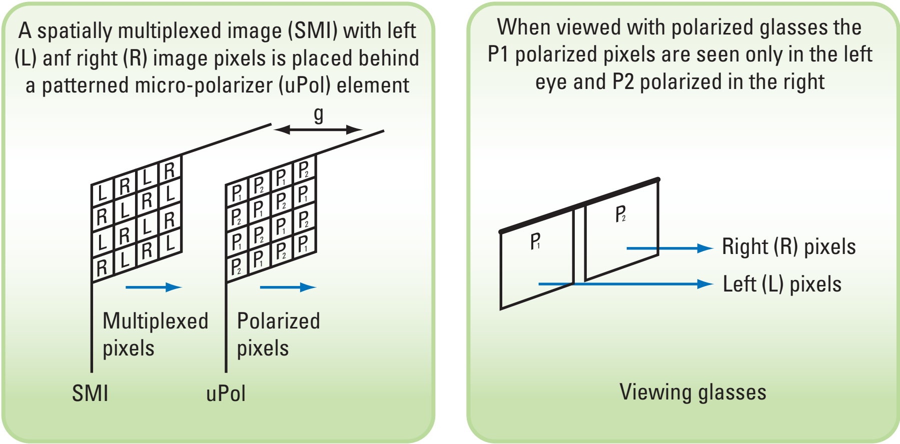



**Figure 15 –** *3D with polarized glasses.*

- New display devices with double frame rates (or higher), combined with either active glasses synchronized by an infrared transmitter to drive the shutter in front of each eye, or with passive polarized glasses with more sophisticated displays having switched polarized screen covers. This is the only display option that complies with the Blu-Ray Disc Association's full HD per eye requirement in the short term. Together with new 3D Blu-ray players, this is the technology that the major industry players will use to introduce 3D HDTV to the market place in 2010.

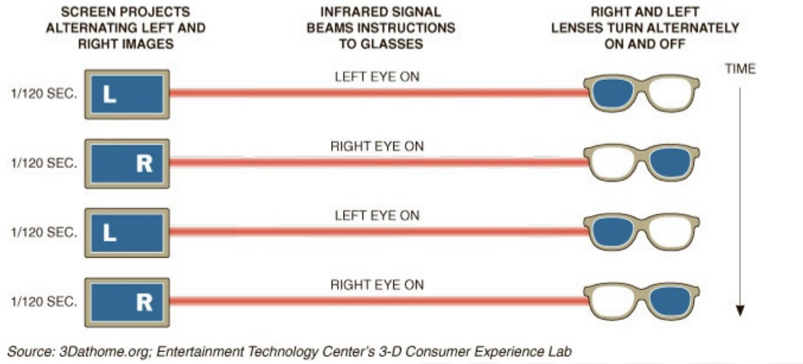



**Figure 16 –** *Active shutter system showing display and glasses corresponding for proper sequencing of left eye and right eye viewing.*

- Auto stereoscopic displays are the only displays that can offer a 3D viewing experience without glasses. However they do not provide HD resolution today as the resolution of the panels must be divided by the number of views. A significant gap in display technology is required (a factor of 10 in resolution if not more) to get the full HD resolution for an increased number of views versus what exists today (nine views). It could be a decade until glasses-free 3D television will be ready for the mass market.

# Conclusion

Broadcasters today are able to use their existing infrastructure and encoder/receiver products to conduct their 3D trials. The Blu-ray initiative is pushing the industry forward to take the 3D experience to consumers' living rooms with full HD 1080p resolution to each eye. As a leading encoder manufacturer, Grass Valley can provide 3D solutions for today's transmission infrastructure as well as serve as a source for future projects. The company is actively pursuing opportunities with its partners to explore next steps in the 3D transmission chain, evaluating next-generation technologies with live events planned in 2010 and 2011.

# References

(1) http://axon.physik.uni-bremen.de/research/stereo/color_anaglyph/

(2) http://www.david-romeuf.fr/3D/Anaglyphes/TCAnaglypheLSDubois/TransformationCouleursPourAnaglyphe.html

(3) http://www.businesswire.com/portal/site/home/permalink/?ndmViewId=news_view&newsId=20091217005371&newsLang=en

(4) http://www.hdmi.org/download/press_kit/PressBriefing_HDMI1_4_Final_083109.pdf

(5) SVC tools performances analysis for next generation HDTV 1080p50/60. Vincent Bottreau & Christophe Chevance, Thomson internal document

(6) http://www.itu.int/rec/T-REC-H.264-201003-P/en

(7) Stereo Coding with MVC, Walter Gish & Christopher Vogt

**SALES**

Local and regional sales contacts can be found by visiting
www.grassvalley.com/sales

**SUPPORT**

Local and regional support contacts can be found by visiting
www.grassvalley.com/support

grass valley

© Copyright 2010 Thomson. All rights reserved. Grass Valley is a trademark of Grass Valley, Inc. All other tradenames referenced are service marks, trademarks, or registered trademarks of their respective companies. Specifications subject to change without notice.

CDT-4122M