MERIDIAN: WHITE PAPER

The Boxx Meridian wireless uncompressed HD transmission system with a unique video-modem approach:

A technical overview

Abstract

Delivering uncompressed HD content wirelessly has become the Holy Grail of wireless video transmission. The Broadcast industry is always interested in wireless transmission options to allow more flexibility, creativity and safety. However, conventional standards, such as COFDM are not capable of delivering such high video rates to achieve these objectives.

This paper introduces a new video-modem approach for wireless video which bridges this gap. This paper will describe a new approach which uses joint-source channel coding to optimize the wireless modem for video delivery. The paper will explain how this video optimization enables more than a 10X improvement when compared to the traditional data-modem method for wireless video, thereby enabling wireless delivery of very high uncompressed video rates with high reliability.

UK Head Office +44 (0)845 643 2874 • Fax: +44 (0)207 900 6350 • www.boxx.tv • sales@boxx.tv

Introduction

The success of COFDM and other wireless standards show that there is widespread acceptance of wireless technology. Productions like camera operators to have the flexibilty of an easy to use and install wireless environment. However, existing wireless standards such as COFDM are not capable of delivering the high video rates required for high-quality video connectivity without large amount of compression and delay. Something more is needed in order to achieve a high quality HD wireless trasmission with extremely low delay.

High-quality wireless video transfer is a significant technical problem. One aspect is the high video rate requiring a communication channel with enough bandwidth and signal-to-noise ratio (SNR), i.e., enough capacity. But more so, the wireless channel is unstable and unpredictable. Its characteristics change rapidly; due to fading and interference, its SNR and capacity vary considerably. Forward error correction, buffers and re-transmissions can compensate for these problems. This is far from ideal in video (and audio) connectivity, where the transfer must be done in real time with as little delay as possible and the high fidelity must remain intact throughout the transfer.

Background from Information Theory

How to transfer an information source to a destination? More specifically, consider an HD audio/video stream over a noisy, fading wireless channel. From the information theory point of view, a communication channel has a capacity, denoted C, that specifies the maximal rate (in bits per second) that can be sent reliably over the channel. C depends in general on the specific channel characteristics. Yet a good yardstick providing a simple approximated formula is the famous Shannon capacity of the additive while Gaussian noise (AWGN) channel:

$$C = W \log (1+SNR) \text{ bits per second}$$

where W is the channel bandwidth (in Hz) and the logarithm is base 2.

The source information content is also measured in bits per second. If the HD video source is simply quantized to say, 8 bits per pixel per color, then 720 p requires about 1.3Gbps, 1080i about 1.5Gbps, and 1080p about 3Gbps. This is probably an overestimate of the information content since by lossless compression this amount of bits can be reduced by, say, a factor of two. Furthermore, if more distortion is allowed, the bit rate needed to represent the video source within that distortion is further reduced.

The source information content is also measured in bits per second. If the HD video source is simply quantized to say, 8 bits per pixel per color, then 720 p requires about 1.3Gbps, 1080i about 1.5Gbps, and 1080p about 3Gbps. This is probably an overestimate of the information content since by lossless compression this amount of bits can be reduced by, say, a factor of two. Furthermore, if more distortion is allowed, the bit rate needed to represent the video source within that distortion is further reduced.

The information theory quantity R(D), the rate-distortion function of the source, defines the minimal bit-rate that is needed to represent the source within a specified distortion, D. It is common to measure D by the average square error between the original source and its representation, as compared to the source dynamic range. This measure, called peak signal to noise ratio (PSNR), is for 8-bit representation of the video:

$$PSNR = 10 \log_{10} (255)^2/D \text{ dB}$$

While R(D) of a video source is not easily specified and depends on the video content and the correlation between the pixels and the frames, for low D (i.e., high PSNR), it can be assumed that the fine quality video samples are independent. At this operating point, a bit per samples is required to reduce the error magnitude by a factor of two (or achieve 6dB better PSNR). 720p video has 55.3 Mega pixels per second or 55.3x3=~166 M samples per second, 1080i has 62.2 Mega pixels or 186.6 Mega samples per second, and 1080p has 124.4 Mega pixels or 373.2 Mega samples per second. In other words, for high-quality video, it requires 166, 186.6 and 373.2 Mbps respectively for improving the video quality by 6dB at the high PSNR region.

The best attainable performance in transferring a source over an information channel using any communication scheme is given by Shannon's celebrated "source-channel coding theorem," asserting that this best performance (minimum distortion) of maintains:

$$R(D) = C$$

i.e., the source is distorted by an amount, D, that cannot be better than R-1(C). As channel capacity depends on the channel signal-to-noise ratio (CSNR), to assess the performance, one has to draw a graph of the PSNR as a function of the channel capacity or the CSNR.

Shannon's theorem statement is intuitively understandable. C measures the maximal bit rate that can be transferred reliably over the channel, while R(D) is the minimal bit rate required to represent the source up to a distortion, D. Furthermore, Shannon's theorem suggests a solution given by the separation principle: The optimal performance can be achieved by first compressing the source to R bits-per-second, inflicting a distortion D on the source, then (assuming that R is smaller than the channel capacity, C) sending these bits without error over the channel. This is the common approach. However, while this approach can (theoretically) achieve optimal performance, it cannot do so in practice at reasonable complexity and delay for reasons discussed below.

An alternative approach is joint source-channel coding (JSCC), where there is no separation between source coding and channel coding. This approach leads to the video-modem architecture. As described below, this approach fits the varying and unpredictable nature of the wireless channel and can attain (close to) the optimal performance via a robust, cost-effective solution that is adaptive to the varying channel capacity.

Feasibility assessment for the 5GHz, 20MHz MIMO channel

The expression above allows checking the feasibility of any proposed solution to deliver HD video over a channel. Let us then check The Boxx Meridians solution: Is the channel capacity indeed greater than the source rate-distortion function at the desired low distortion? Can it provide a reliable wireless video connectivity that can cover the whole home?

What is the capacity? The Meridian solution uses a wireless MIMO channel over 5GHz Wi-Fi band, with four transmit antennas and 20Mhz bandwidth. The wireless MIMO channel capacity, when the channel matrix is H (known to the receiver) and with Gaussian noise is:

$$C = \log_2 \det(I+\rho HH^H) \quad \text{bits/sec/Hz}$$

where ρ is the CSNR per transmit antenna, H is the channel matrix, and the superscript H indicates the complex conjugate. If nR, the number of antennas at the receiver, is greater or equal nT, the number of transmit antennas, and the channel matrix is non-singular, the capacity formula is approximated by:

$$C \approx n_T \log_2 (1+\rho(H)) \quad \text{bits/sec/Hz}$$

where ρ(H) is an average effective CSNR that depend on the specific fading channel matrix H, which in turn depends on the environment and the distance. Thus, the capacity is not constant but random, changing according to the varying effective CSNR. With n_T=4 and 20MHz bandwidth, the capacity can vary from about 800Mbps for 30dB CSNR to about 500Mbps for 18dB CSNR, and it can get down to 250Mbps for ~10dB CSNR, in high fading situations.

Consider now the source. As discussed above, it is hard to assess the HD video rate-distortion function since the video is composed of many different images – some require a very small rate (like a uniform color image) and some require a higher rate. Yet, as confirmed by the fact that the 8-bit HD samples can be losslessly compressed by about a factor of two, it can be assumed that the HD (1080i or 720p) rate distortion value is below ~700 Mbps at that 8-bit distortion level. If a higher distortion is allowed, a lower rate is sufficient to represent the source. Following reports on the performance of compression algorithms the HD source requires about 100-200 Mbps for representing the HD source at better than 40dB PSNR. Note that an assessment for the rate-distortion value of a 1080p source is twice the assessment of the 1080i source.

Summarizing the above, for 720p/1080i, the 20MHz channel used by Meridian has a capacity of 250-800 Mbps, which is higher than that HD rate-distortion value at video quality of 40dB PSNR. However, it emphasizes the importance of being close to the Shannon bound and utilizing the optimal performance. The JSCC approach can do it “on-the-fly,” while the separation approach struggles and, as noted above, requires a reliable transmission of hundreds of Mega bits per second in order to improve the source quality (PSNR) by 6dB

UK Head Office +44 (0)845 643 2874 • Fax: +44 (0)207 900 6350 • www.boxx.tv • sales@boxx.tv

The Joint Source-Channel Coding (JSCC) approach

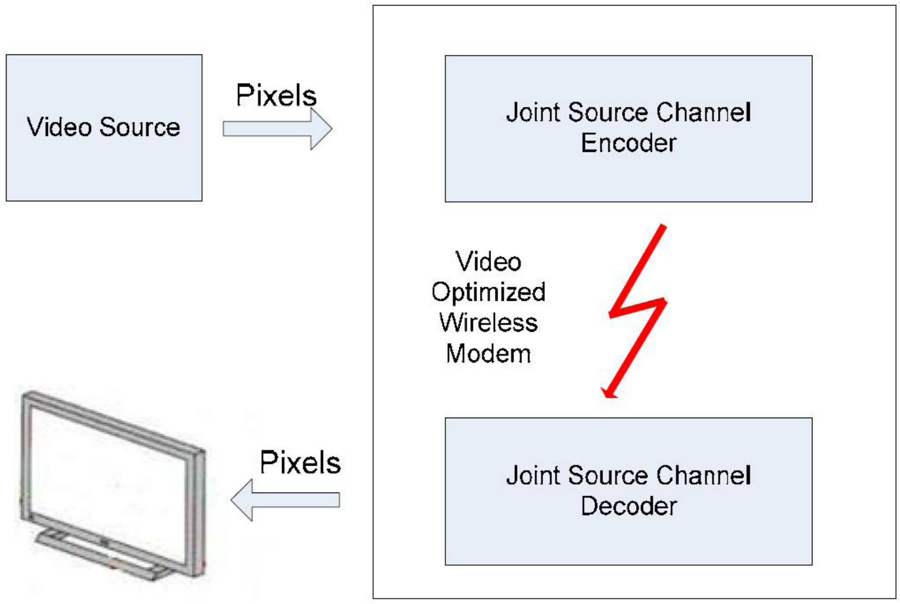

Figure 2: System design based on joint source-channel coding

The JSCC is composed of three elements:

1. Video processing and representation to prioritize the video components according to their importance. Some examples of prioritization include:

    a. The most significant bits (MSBs) of a pixel are more important than the least significant bits (LSBs).

    b. Lower spatial frequencies are more important than higher frequencies.

    c. The luminance component is more important than the chrominance components.

2. Unequal error protection (UEP) to encode the most significant bits of the important components by better (lower rate) error correcting code, than the least significant bits of the less important components.

3. Combination of modulation and UEP to generate the proper constellation in the channel signal space. For example:

    a. Use coarser constellation for the important components and finer constellation for the less important components.

    b. Use “noisier” frequency bands for the less important components.

    c. Use space-time coding for better protection of the important components.

This process translates the video pixels, with a very low latency of several image lines, to modulated symbols, e.g., orthogonal frequency division modulation (OFDM) symbols. A block diagram of a joint source-channel coding is depicted in Figure 3.

Figure 3: JSCC

Advantages of JSCC over traditional approach

The approach, based on joint source-channel coding, has several profound advantages over the traditional approach. First, since the video components and their bit representation are not equally important, JSCC uses unequal error protection (UEP) while traditional systems provide equal protection to all bits. Thus, in traditional systems the most significant bits of the important components are not protected enough, while the least significant bits of the less important video components are protected too much (and consequently waste channel resources). Depicted in Figure 4.

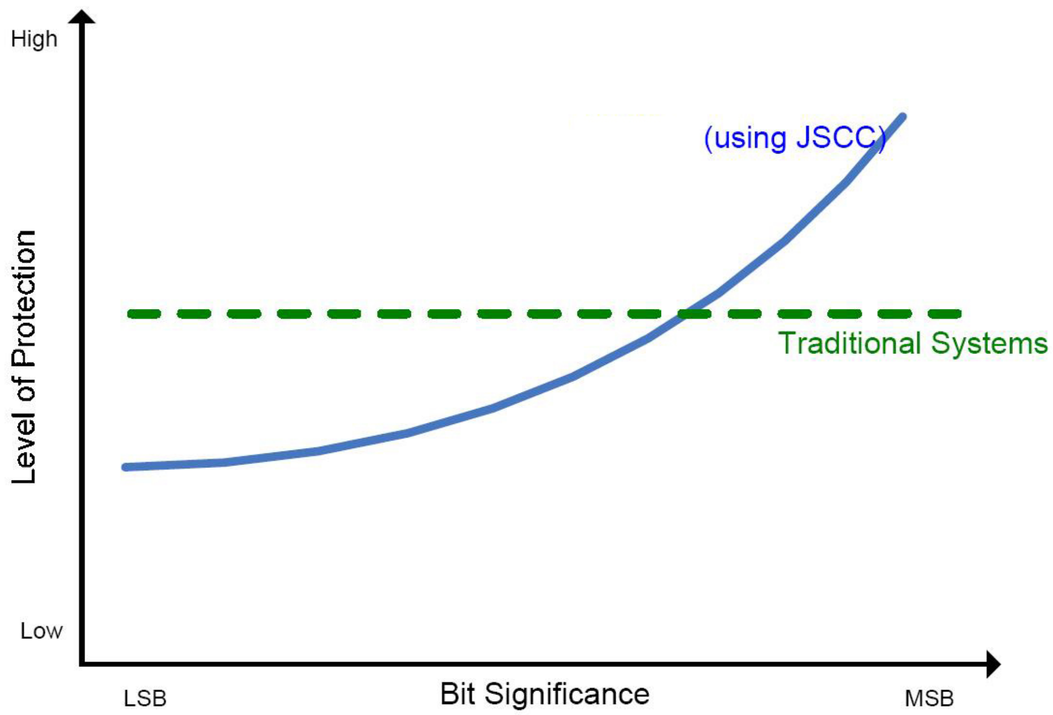

Figure 4: Unequal error protection vs. equal protection

The JSCC approach enables a better utilization of the available channel capacity, even when it is varying. Traditional systems should work at a rate that is below the worst-case channel capacity, since otherwise the video communication will not be reliable. By allocating whatever available capacity to send information that is less sensitive, JSCC utilizes capacity almost to its fullness. It should be noted that some traditional systems can "average out" short periods when the capacity decreases abruptly, but this requires either forward error correction or large buffers leading to high complexity, delay, and reduces the video quality resulting in more artefacts. Other systems adapt to the varying channel characteristics by using feedback. But feedback is cumbersome and requires complex return channel. Furthermore, since the wireless channel can change quickly, by the time the transmitter gets the feedback information, the channel has been changed. The comparison in channel utilization is described in Figure 5.

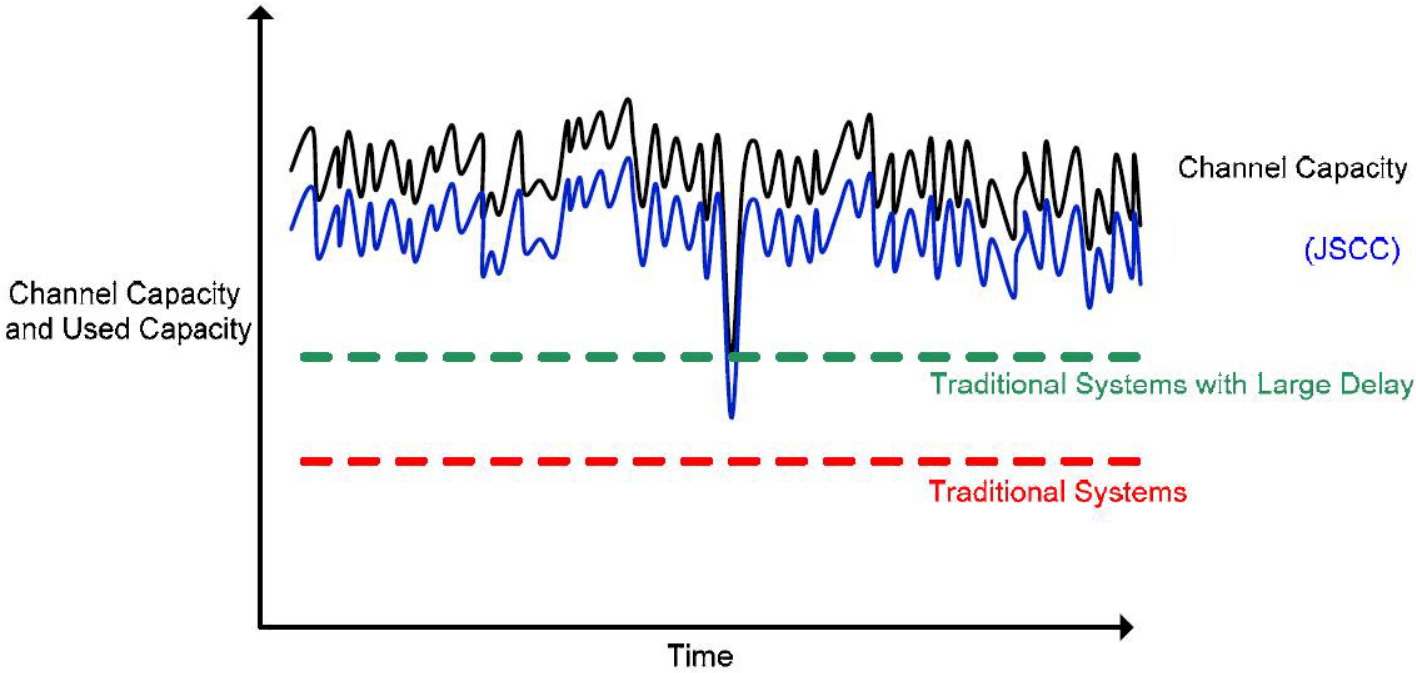

Figure 5: Utilization of channel capacity

The main advantage of JSCC over traditional systems lies in its ability to gracefully adapt to the varying channel capacity and the channel SNR. Traditional systems suffer from a threshold effect – They must guarantee a minimal SNR, otherwise the entire communication fails. To lower the threshold, traditional systems reduce the rate, e.g. by using deeper compression. But then the quality degrades. In any case, traditional systems always have a "quality ceiling," where the picture quality cannot improve, even if the channel becomes better, over the pre-defined quality associated with the worst-case design. The JSCC approach does not have a threshold SNR, and the picture quality improves when channel conditions improve. The video quality of traditional and JSCC in varying channel conditions is depicted in Figure 6.

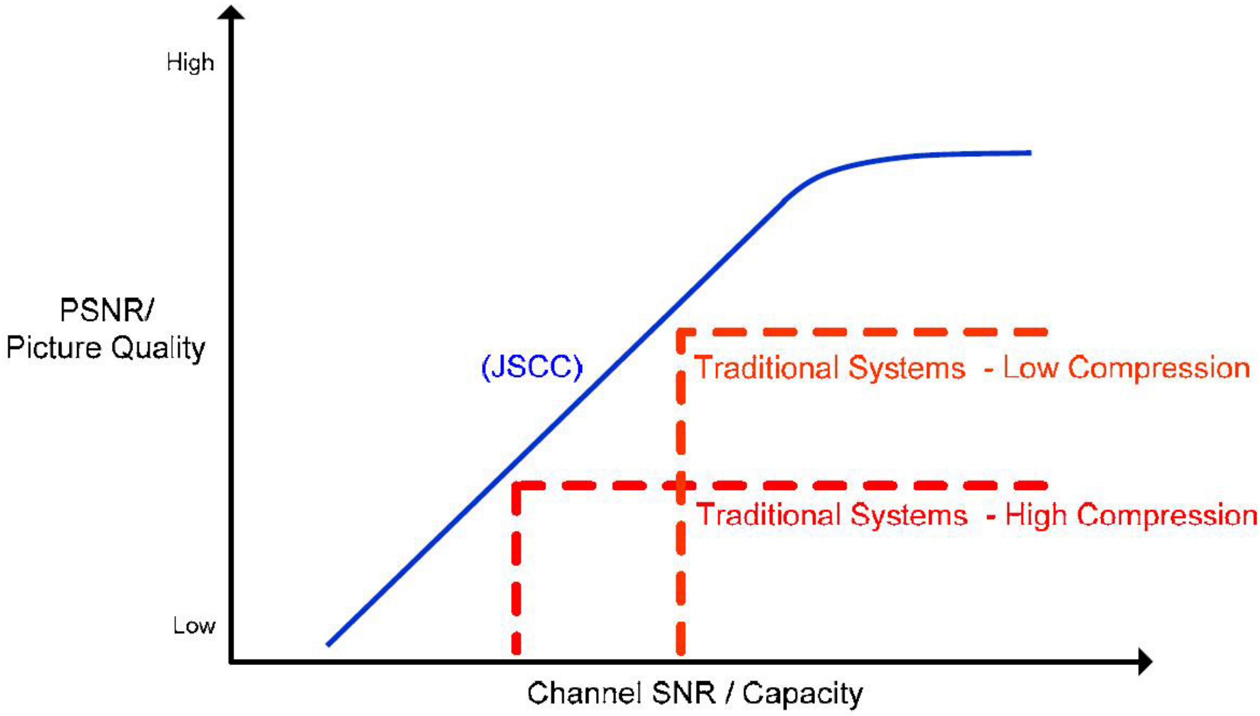

Figure 6: Graceful adaptation vs. threshold/ceiling effects

The CSNR is mostly between 20-30dB for the video-modem used by Meridian, allowing a PSNR of over 40-45dB for almost perfect reconstruction. But even if the CSNR occasionally drops, the resulting PSNR is still good, while the traditional system will then completely break down.

Robustness and complexity

Robustness is a highly important feature for the wireless video connectivity solution. Following the above, JSCC is much more robust to the varying characteristics of the wireless channel than the traditional approach, and so it is much less sensitive to obstacles and distance variations. Figure 7(a) shows a typical graph representing the probability that in a given instance, the channel supports a bit rate that is greater than B. To ensure an operating system, the traditional approach must work at a point where the attainable bit rate is achieved with 99.99 percent (otherwise, forward error is needed or there will always be re-transmissions). As can be seen from this typical graph, this rate can be, say, 5 times less than the average capacity and more than 10 times the maximal capacity [see Figure 7(b)]. Even at that work point, there is always the chance (say, 0.01 percent) that the channel will not even support the required bit rate, and so the system based on the separation principle will collapse and result in large errors, especially if the data had to be compressed beforehand to fit the available bit rate at the work point. On the other hand, the JSCC approach allows use of the instantaneous capacity and so it enjoys a much higher capacity, which can be translated to better video quality [as can be seen in Figure 7(c)], yet it does not have to worry about the chance that the capacity and available bit rate drops below some specific value.

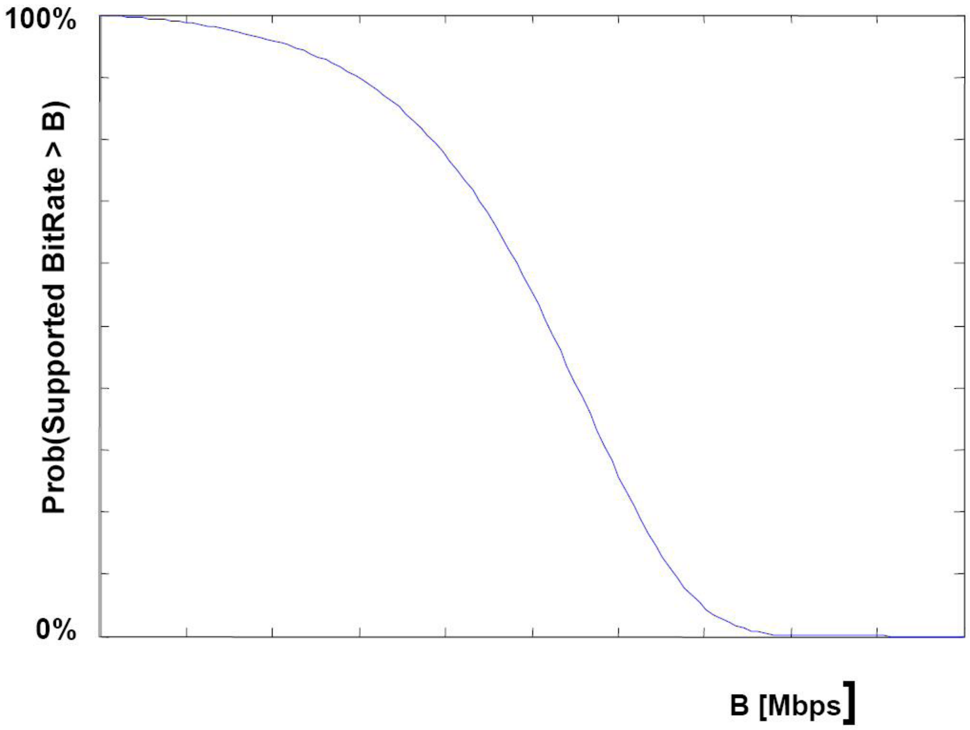

Figure 7(a): The Wireless Fading Channel – Probabilistic behaviour

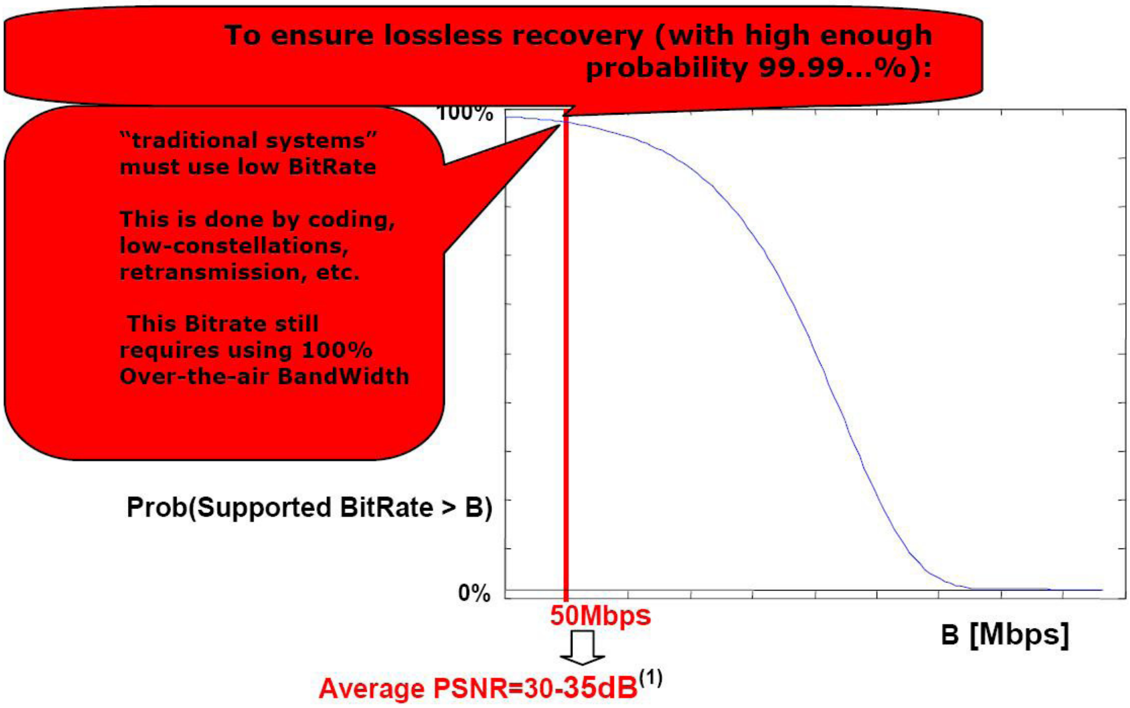

Figure 7(b): Work point of traditional systems

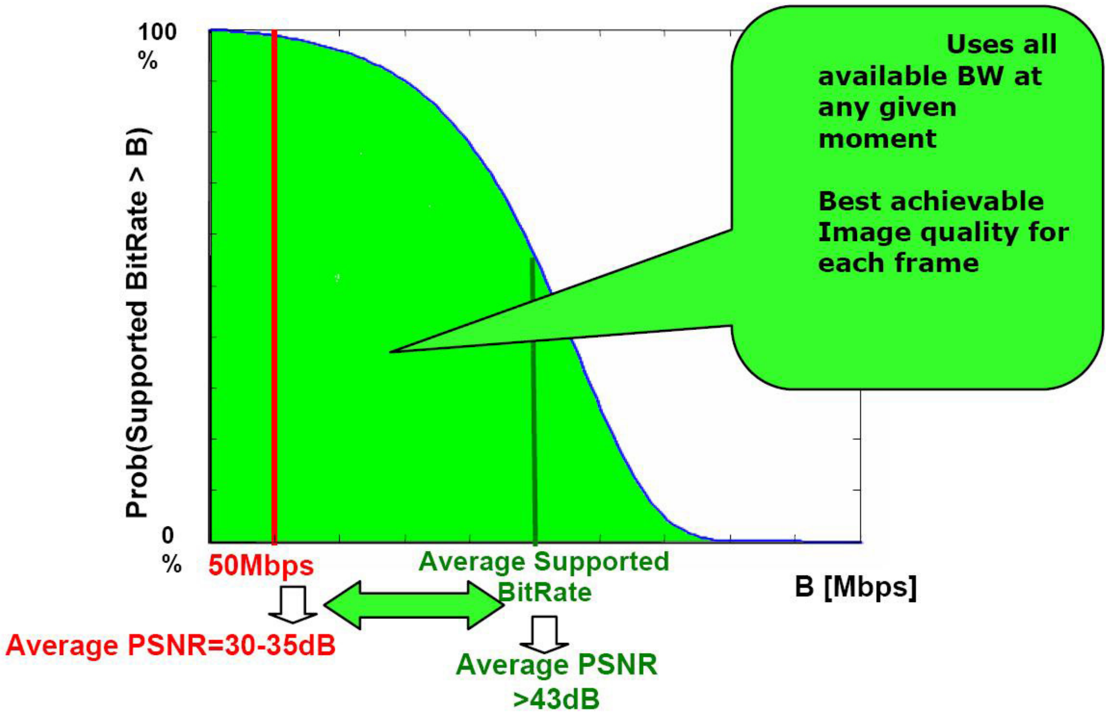

Figure 7(c): JSCC performance – PSNR is better by 8-13dB than traditional systems over the same channel

The JSCC performance and robustness are attained with much less computation and memory complexity and much simpler system architecture than the traditional systems. JSCC does not require complex compression. It does not require forward error or large buffers for compression and for modem re-transmissions. Its latency is virtually zero (less than 1 ms). It can work essentially without compromising the communication link and the video quality and allows a natural point-to-multipoint system architecture. Its built-in robustness and adaptability allow the video modem based on JSCC to achieve a 10 times improvement over traditional systems!