Bit-Significances of Source-Coded Data: Analysis and Applications

Norbert Goertz

Inst. für Nachrichtentechnik und Hochfrequenztechnik, TU Wien, Gusshausstrasse 25/389, A-1040 Wien
E-Mail: `norbert.goertz@nt.tuwien.ac.at`
Web: `www.nt.tuwien.ac.at`

Abstract

The Logarithmic Bit-Significance Ratio (*S*-value) is a universal measure to quantify the significances of the bits in a communication system. We define the *S*-value and discuss relations to the well-known *L*-values that are used for soft-in/soft-out decoding at the receiving end. We argue that the *S*-values are the transmitter-based counterpart of the *L*-values, and we motivate the idea of soft-in/soft-out data processing at the transmitter by demonstrating practical benefits in a simple digital communication system.

1 Introduction

Data bits in digital communication systems are either produced by a quantiser, which may well be part of a complicated multimedia source codec, or they represent information that is discrete in time and amplitude by nature, e.g. bits describing a compressed version of text by a Huffman code. The transmission of digital data is often carried out by "frames" of limited size that are separately encoded; this concept is driven by restrictions on the delay and the complexity. A data bit at a particular position within a frame may sometimes be very important while in other frames it is less significant.

It was shown in [1] that large performance improvements can be achieved by allocating transmit energy to the bits according to their "significances". In this paper we extend a new concept presented in [2]: we analyse a universally applicable bit-significance measure, the logarithmic bit-significance ratio (*S*-value). We discuss the definition of the *S*-values for some practically relevant cases and we demonstrate possible performance gains using the *S*-values for transmit-energy allocation in binary modulation.

2 Logarithmic Bit-Significance Ratio

2.1 *S*-Values of Quantiser Bits

Basic Concept: In Figure 1 we investigate the bit significances that result from scalar quantisation of a source signal x. We consider two realisations "A" and "B" of the source signal that both are quantised by the (scalar) reproducer values $y_0,...,y_7$.

The sample A is quantised by y_1 as it is the nearest neighbour[1] of x among the reproducer values, and the corresponding bit vector "001" is transmitted. At the decoder output an error in the middle bit will *not* lead to much higher distortion than the reproduction without any bit errors, because the reproducer value y_2, corresponding to the bit-combination "011", is also close to the input sample A.

The situation is different when we consider the sample B. Although B is also quantised by the reproducer value y_1, an error in the middle bit will now lead to much higher

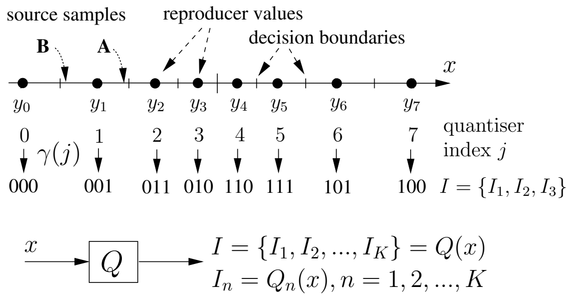

Figure 1: Three-bit ($K = 3$) scalar quantiser with Gray bit-mapping $\gamma()$.

distortion. This time, however, an error in the rightmost bit will cause only a small increase of the mean squared error. Hence, we have bits with variable significances, depending on the location of the unquantised source samples.

This argument can be extended to vector quantisers as illustrated by Figure 2: we have chosen a two-dimensional

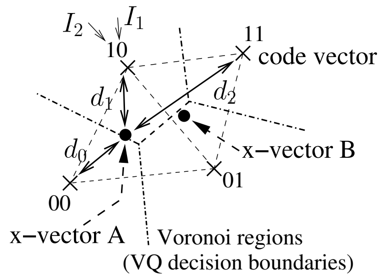

Figure 2: Two-dimensional vector quantiser

source-vector A that is located close to one of the Voronoi region boundaries. Although the input vector A lies closest (with distance d_1) to the code vector with the bit string "10", the distance d_0 to the code vector with the bit string "00" is only little larger, so the leftmost bit I_2, which is the only one different between the bit labels of both code vectors, is not really significant as $d_1 \approx d_0$. For input vector A, the bit I_1 (rightmost bit) is much more significant, as $d_2 \gg d_0$. If the input x-vector was located close to another Voronoi region boundary, I_2 may be the more and I_1 the less significant bit, i.e., the bit significances change dependent on the location of the source vector x. The input x-vector B represents a more complicated case, in which not a single bit is insignificant or not: it is rather insignificant if we chose the bit *string* "01" or "10". This is a consequence of the code vector bit mapping being not at real Gray mapping. However, there are indeed single bits that are more significant than others, depending on the input vector, and there are more complicated dependencies between the significances of groups of bits, and this argument is true as a long as there are Voronoi regions with decision boundaries and bit strings that encode the code vectors, no matter what

[1] For simplicity we use the common (mean) squared error to measure "quality"; other quality criteria would apply in a straightforward way, by appropriately modifying (1) below.

ISBN 978-3-8007-3300-2

the vector dimension is. Therefore, the principle will apply to any vector quantiser of finite dimension. For simplicity we will only consider scalar quantisers below.

Formal Definition of S-Values: We assume that $I = \{I_1, I_2, \ldots, I_K\}$ denotes a K-bit vector (so $I_n \in \{0,1\}$, $n = 1,2,\ldots,K$) that represents the quantiser index j. We obtain the bit-vector I from the quantiser index j by the bit-mapping $\gamma()$, i.e., $I = \gamma(j)$; in Figure 1 we use a Gray mapping as an example.

With the input source sample x given, we quantify by

$$d'(I_l = \xi \,|\, x) \doteq (x - y_{\gamma^{-1}(I)})^2 \qquad (1)$$

the distortion that is produced, when a particular value $\xi \in \{0,1\}$ is enforced on the bit position I_l, $l \in \{1,2,\ldots,K\}$. In (1), $\gamma^{-1}(I)$ denotes the inverse bit mapping producing a quantiser index j and y_j is the quantiser reproducer value (or code vector in case of vector quantisation). The bit vector $I = \{I_1, I_2, \ldots, I_K\}$ is given by

$$I_n = \begin{cases} Q_n(x), & n \neq l \\ \xi, & n = l \end{cases} \quad \text{for} \quad n = 1,\ldots,K\,. \qquad (2)$$

The notation $Q_n(x)$ corresponds to the output bit number n of the quantiser (including the bit mapping $\gamma()$) given the continuous-valued source sample x. In other words: in (1) we use the reproducer value for reconstruction that has the same bit vector as the nearest neighbour of the input sample x, only excluding the bit position l for which we enforce the bit value $\xi \in \{0,1\}$ that is specified on the left-hand side of (1).

Using (1) we define the normalised distortion

$$d(I_l = \xi \,|\, x) \doteq \frac{d'(I_l = \xi \,|\, x)}{d'(I_l = 0 \,|\, x) + d'(I_l = 1 \,|\, x)} \qquad (3)$$

for which the following holds by definition:

- $d(I_l = \xi \,|\, x) \in [0,1]$ $\forall x$ and $\xi \in \{0,1\}$
- $d(I_l = 0 \,|\, x) + d(I_l = 1 \,|\, x) = 1$ $\forall x$.

The significance of a reconstruction of I_l with some bit value ξ is "inversely" proportional to the (normalised) distortion (3) corresponding to this value of ξ. Therefore, we define the *significance* by

$$q(I_l = \xi | x) \doteq 1 - d(I_l = \xi | x). \qquad (4)$$

As $d(I_l = \xi | x) \in [0,1]$ we obtain $q(I_l = \xi | x) \in [0,1]$ as well. Now we define the *logarithmic bit-significance ratio* (S-value) as the natural logarithm of the bit significances (4) for a "0" and a "1" bit:

$$S(I_l \,|\, x) \doteq \log \frac{q(I_l = 0 \,|\, x)}{q(I_l = 1 \,|\, x)}\,. \qquad (5)$$

The S-value measures how significant a particular bit position is. If the magnitude $|S(I_l \,|\, x)|$ is zero, this means it doesn't matter if the bit is reconstructed as a "0" or a "1". If the magnitude is large, the bit is very significant and the sign[2] of $S(I_l \,|\, x)$ determines whether the reconstruction of the bit should be "0" (when $S(I_l \,|\, x) > 0$) or "1" (when $S(I_l \,|\, x) < 0$). It should be noted that the S-value depends on the "current" source signal sample x.

[2] The definition (5) is such that the interpretation of the sign is consistent with the well-established log-likelihood ratios [3] used for soft-in/soft-out *de*coding at the *receiver side*.

Example: Figure 3 shows, for each bit of an optimal scalar quantiser for a Gaussian source with Gray bit mapping (see illustration in Figure 1), the S-values that are obtained from (5) and (1) for $-3 < x < 3$. The locations of the reproducer values are indicated by "×" and the bit mappings used for the transmission are given as well. Figure 3 confirms that the S-values indeed measure the significances of the bits: the S-values are large in magnitude if the x-value is close to a reproducer value and the bits that change at the decision boundaries in the middle between two reproducer values have S-values close to zero. A positive sign of the S-value indicates that the reconstruction of this bit should be "0" while a negative sign indicates a "1".

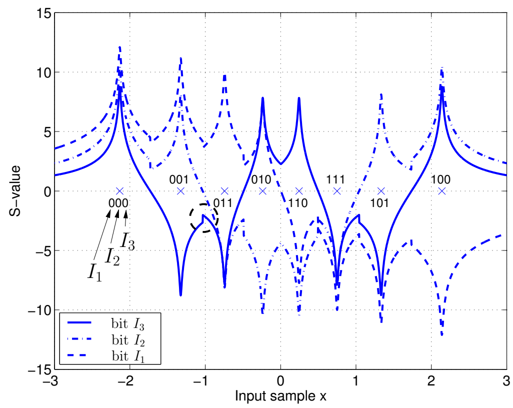

Figure 3: S-values $S(I_l), l = 1,2,3$, for a three-bit scalar Lloyd-Max quantiser for a Gaussian source; the quantiser reproducer values are marked by "×", a Gray bit mapping for binary transmission is included as well.

A closer inspection of Figure 3 reveals discontinuities of the S-value curves that occur at the decision boundaries, in the "middle" between two reproducer values[3], for those bits that do *not* change as we change from one reproducer to its neighbour when running along the x-axis. One such discontinuity that occurs for the S-value curve of the (right-most) bit I_3 is marked in Figure 3 by a dashed circle and is further investigated in Figure 4. At the decision boundary

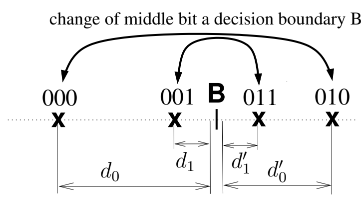

Figure 4: Discontinuity of the S-value curve of bit I_3.

B we observe for bit I_3 the S-values

- from the left: $S(I_3|x - \varepsilon) = \log \frac{d_1}{d_0}$
- from the right: $S(I_3|x + \varepsilon) = \log \frac{d_1'}{d_0'}$

with ε positive and small. As for $\varepsilon \to 0$ we have $d_1 = d_1'$ but $d_0 > d_0'$, the S-value $S(I_3|x)$ is discontinuous at the decision boundary B. This effect goes away for a uniform quantiser with Gray mapping but it is perfectly normal for non-uniform quantisers and/or non-Gray bit mappings.

[3] This location of the decision boundaries is a direct consequence of using the mean-squared error as a distortion measure.

2.2 S-Values for Binary Data

In binary-data transmission, where data is discrete in time and amplitude by nature, varying significances of a particular bit-position in different data packets may arise, e.g., when variable-rate lossless source coding is used in text compression: some bit errors might cause symbol insertions and deletions, while an error in exactly the same bit position in another block may only cause a single symbol error at the decoder output: hence, we observe variable bit-error sensitivities of the data bits. Another example is “in-band” signaling in mobile radio, with the signaling information being much more important than “normal” data.

We assume that we can tolerate some bit-error probability $p_l < 0.5$ in the reconstruction ξ of a bit I_l with the “correct” realisation $i_l \in \{0,1\}$ (that is known at the transmitter side). We define a distortion measure according to

$$d(I_l = \xi | i_l) \doteq \begin{cases} p_l & i_l = \xi \\ 1 - p_l & i_l \neq \xi \end{cases} \quad \text{with} \quad \xi \in \{0,1\} \, ; \tag{6}$$

the interpretation is as follows: the value of the distortion measure is $1 - p_l$ if the reconstruction of I_l is different from the (correct) bit-value i_l, i.e., it is smaller than the Hamming distance “1”, as we are willing to tolerate some bit-error probability p_l. As we want to define a *normalised* distortion, the value for a correct reconstruction ($\xi = i_l$) must have the value p_l, as only then both alternatives in (6) add up to one.

The normalised significance $q(I_l = \xi | i_l)$ of a reconstruction of I_l with some bit value ξ is inversely proportional to the distortion (6) so we define:

$$q(I_l = \xi | i_l) \doteq 1 - d(I_l = \xi | i_l) \, . \tag{7}$$

Similar as in (5) we define the *logarithmic bit-significance ratio* (S-value) according to

$$S(I_l \,|\, i_l) \doteq \log \frac{q(I_l = 0 \,|\, i_l)}{q(I_l = 1 \,|\, i_l)} \, . \tag{8}$$

If we insert (6) and (7) into (8) we obtain

$$S(I_l \,|\, i_l) = \mathrm{sign}(1 - 2 \cdot i_l) \cdot \log \frac{1 - p_l}{p_l} \quad \text{for} \quad p_l < 0.5 \tag{9}$$

with $i_l \in \{0,1\}$. Note that again the sign of $S(I_l \,|\, i_l)$ determines the bit-value a reconstruction of I_l should have, while the magnitude $|S(I_l \,|\, i_l)| = \log \frac{1-p_l}{p_l} > 0$ describes the significance of the bit. If the tolerable error probability is $p_l = 0.5$, we obtain $(1 - p_l)/p_l = 0$, i.e., the bit has no significance at all, while the value of $(1 - p_l)/p_l$ is very large when p_l is close to zero.

2.3 Remarks

From a comparison of (3) and (6) we observe that we can equivalently write any normalised distortion for a quantiser bit as a distortion of a data bit according to

$$p_l \doteq \min\big(d(I_l = 0 \,|\, x), d(I_l = 1 \,|\, x)\big) \, , \tag{10}$$

with p_l being used in (9) to compute the S-value. Therefore, we don’t need to distinguish between different types of data (or code) bits, i.e., by the definition of the S-value we achieve a very useful abstraction. As the origin of the bits is no longer relevant to us, we drop the condition on a source sample x or a data bit i_l below, as long as there is no risk of confusion.

3 Application of S-Values

3.1 Energy Allocation in Binary Modulation

After computing the S-values they can, e.g., be used to allocate transmission energy. For simplicity we will assume binary phase shift keying. The allocated energy E_l for the transmission of a bit I_l is related to the power P_l by $P_l = E_l/T$, with T the constant bit transmission period.

A simple heuristic approach is to distribute the energy, with the S-Values as weighting factors, according to

$$E_l = \frac{|S(I_l)|^\alpha}{\frac{1}{K}\sum_{l=1}^{K} |S(I_l)|^\alpha} \cdot E_s \tag{11}$$

with $\alpha > 0$; E_s is the given *average* energy for each bit and K is the number of bits for which the joint energy allocation is to be applied. An important property of the rule (11) is that $E_l = 0$ if $|S(I_l)| = 0$, i.e., no energy is allocated to insignificant bits. Another essential property of (11) is that the average energy stays the same, i.e., $\frac{1}{K}\sum_{l=1}^{K} E_l = E_s$.

This allocation of energies will not be “optimal” in any strict sense. If, e.g., quantiser bits are transmitted, one could, in principle, state an optimisation problem and try to solve it by variational techniques. The drawback of this approach is that the problem can usually only be solved numerically with high complexity [1]. Hence, our goal with the definition of S-values is to state a feasible general means of how to quantify bit significances independent of their origin and to perform an energy allocation in the sense of a good practical solution that, once the S-values are known, no longer depends on the specific details of the source data. The most suitable allocation rule for a particular application (e.g., the parameter α in (11)) can be determined by simulations.

If data bits are transmitted and their number (i.e., the bit rate) is fixed but the channel cannot carry this amount of information, it is impossible to reconstruct the data bits at the receiver without errors. If the channel quality is very low it might even be impossible to reconstruct them with the given tolerated bit error rates. In the latter case the energy allocation rule will achieve a best-effort solution, which means that significant data bits might exhibit a bit error rate that is larger than actually desired but it will still be lower than for the less significant bits.

3.2 Capacity Considerations

It is well-known that for binary phase-shift keying modulation over an additive white Gaussian noise channel the modulation symbols should have equal probability and that the same power should be applied for each channel use. The concept of variable power according to the S-value is, therefore, contradictory to achieving capacity. It should be noted, however, that our primary goal is to maximise transmission *quality* which, e.g., in the case of the scalar quantiser is *not* high data rate on the channel but rather high reconstruction quality in the source signal domain. Nevertheless, capacity-aspects are interesting and it is important to know “how much we lose” due to using “non-ideal” source coding (which, among other things, involves bits of different significance) for direct binary transmission.

We investigate this issue in Figure 5. We consider a binary-input AWGN channel (BI-AWGN) and we alternate in every other channel-use between the two energy levels $\beta \cdot \frac{E_s}{N_0}$ and $(2-\beta) \cdot \frac{E_s}{N_0}$ for transmission, with $0 \leq \beta \leq 1$.

ISBN 978-3-8007-3300-2

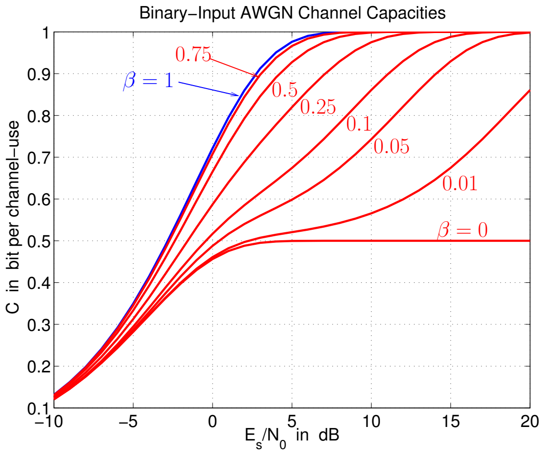

Figure 5: Loss in capacity in binary transmission over an AWGN channel due to unequal transmit power.

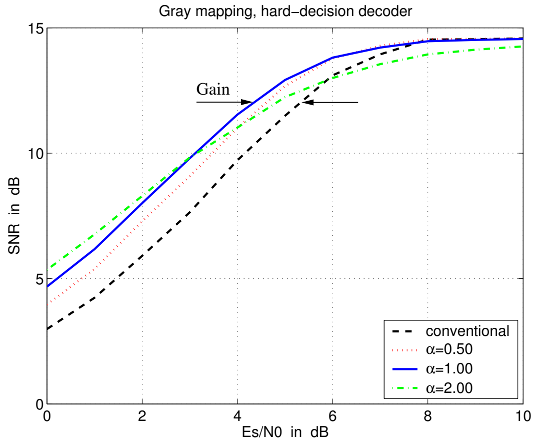

Figure 6: Source SNR vs. average E_s/N_0; energy allocation by S-values with $\alpha = \{0.5, 1.0, 2.0\}$; comparison with a conventional transmitter; Gray mapping; conventional hard-decision receiver.

The total capacity equals $C = \frac{1}{2}(C_1 + C_2)$, i.e., it is the simple average over the two capacities that occur for the two power levels and the average "power" used is still E_s/N_0. Note that for $\beta = 1$ we obtain the standard BI-AWGN capacity for $\frac{E_s}{N_0}$, as then the two power levels are the same.

From Figure 5 we observe a loss in capacity if we use non-constant power (i.e., $\beta < 1$) on the channel, particularly for $E_s/N_0 = 0....20$dB. The loss in capacity is, however, small for $0.5 < \beta < 1$. Hence, if the data bits are not equally important we will be willing to live with some loss of capacity and trade it off against better source reconstruction quality. It should be noted that capacity can usually only be achieved for large block size and highly complex coding algorithms, while the benefits of unequal power allocation will normally be stronger for limited block size that is the more practical case caused, e.g., by constraints on delay and complexity.

3.3 Binary Transmission of Quantiser Bits

In Figure 6 we present some performance results for quantisation and transmission of uncorrelated Gaussian source samples x. The source samples are scalar Lloyd-Max quantised with a fixed rate of $K = 3$ bits per sample. The transmission of the quantiser bits is carried out over a BI-AWGN channel with the average "channel SNR" E_s/N_0.

As a performance measure we use the signal-to-noise ratio (SNR) between the source samples x and their reconstructions $\tilde{x}$ at the decoder output. We compare the conventional transmitter (without source-adaptive energy allocation) with three schemes using S-values for energy allocation with $\alpha = \{0.5, 1.0, 2.0\}$ in (11). A simple hard-decision decoder was used, i.e., bit decisions were taken from the signs of the sampled matched-filter outputs. The resulting bit vectors were converted into quantiser indices via the inverse bit mapping $\gamma^{-1}()$ and the corresponding reproducer values were used as the source reconstructions $\tilde{x}$. Across the whole range of channel SNRs (x-axis) we observe a gain of about 1 dB in E_s/N_0, which directly turns into a gain in average transmit power. The energy distribution with $\alpha = 1.0$ in (11), i.e., linear averaging, turns out to be a good choice, and the peak energy level generated by (11) will be around twice the average energy E_s (measurement result). If a conventional soft-decision receiver [4] is used that, of course, is not aware of the source-adaptive energy allocation at the transmitter, similar results are obtained, but the performance is generally about 2dB better in terms of transmit energy. It should be noticed that, for the results in Figure 6 (hard decision decoder), neither at the transmitter nor at the receiver *any* knowledge is required about the (average) channel-SNR (E_s/N_0).

4 Conclusions

We discussed and analysed the logarithmic bit-significance ratio (S-value) as a systematic measure for the significance of bits. We demonstrated by application in binary transmission of quantiser bits that S-values can be practically useful, despite the fact that their use, e.g., for transmit energy allocation, may incur a loss in channel capacity.

References

[1] N. Görtz and E. Bresch, "Source-adaptive power allocation for digital modulation," *IEEE Communications Letters*, vol. 7, pp. 569–571, Dec. 2003.

[2] N. Goertz, T. Eriksson, and J. B. Anderson, "Logarithmic bit-significance ratio: Definition, calculation rules and examples," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, (Seattle, Washington, USA), pp. 2869–2873, July 2006.

[3] J. Hagenauer, "The turbo principle: Tutorial introduction and state of the art," in *Proceedings International Symposium on Turbo Codes & Related Topics*, pp. 1–11, Sept. 1997.

[4] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 1662–1670, Sept. 2001.

ISBN 978-3-8007-3300-2