How Many Bits Are Copied in a JT65 Transmission?

Joe Taylor, K1JT

Recent evidence suggests that some in the EME community are confused about what is actually copied over the radio path when using the JT65 communication protocol. Although the full technical details of JT65 have been published elsewhere,[1] and source code for the program WSJT is openly available,[2] those resources may seem overly technical for non-specialists in communication theory. This brief note is intended to help WSJT users and nonusers alike to understand what happens during the processes of transmitting, receiving, and decoding a JT65 signal. I have taken special care to use language, units of measurement, and terms of reference familiar to amateurs using traditional weak-signal coding and modulation techniques.

User Information, Encoding, and Channel Symbols

Standard JT65 messages contain 72 bits of user information—typically two 28-bit callsigns,[3] a 15-bit grid locator, and one bit to indicate the message type. A Reed Solomon (63,12) error-correcting code translates the 72 message bits into 63 six-bit "channel symbols." Thus, every transmission includes $6 \times 63 = 378$ information-carrying bits and has a redundancy ratio of $378/72 = 5.25$. It is important to understand that the user message is not transmitted in its "natural" sequence of syllables or characters (as it would be in normal speech, Morse code or ASCII data, for example). Instead, the user information is mathematically encoded so that it is spread throughout the entire sequence of 63 symbols. This fact has profound implications when comparing JT65 with traditional modes. First, the distribution of redundant information is designed so as to maximize the potential recovery of the full message content, even when the signal-to-noise ratio (SNR) is so low that many symbols from the transmitted sequence are lost in the noise. Second, a brief signal dropout of a few seconds will have quite different effects in JT65 compared with a traditional sequential mode. In a sequential mode, the dropout will simply result in the loss of those characters that were being transmitted at the time. But in JT65 encoding, the brief loss of signal will not affect any part of the message in particular; the impact is statistical, affecting recoverability of the entire message. These fundamental differences make it difficult to directly compare the performance of JT65 and traditional sequential modes.

---

[1] J. Taylor, K1JT, "The JT65 Communications Protocol", *QEX*, September-October 2005, pp. 3–12.

[2] Source code can be downloaded from `http://developer.berlios.de/projects/wsjt/`.

[3] The number of licensed amateur radio operators is less than 2^{22} (roughly 4.2 million), so in principle 22 bits would be enough to encode a callsign. JT65 uses 28 bits for this task, however, so I adopt the latter number throughout this paper.

After a JT65 message has been encoded into its 63 channel symbols, the symbols are transmitted one by one using 64-tone frequency shift keying (FSK). A 65th audio tone, two tone-intervals below the lowest data tone, is used to facilitate accurate synchronization of time and frequency between transmitter and receiver. The overall modulation scheme can thus be described as 65-FSK. Half of the transmitted energy is devoted to the essential task of synchronization: during each 46.8 s transmission, 63 sync-tone intervals are inserted at specified locations among the 63 data tones, following a prescribed pseudo-random pattern.

Pictorial representations of the spectrograms of two JT65 transmissions are presented in Figure 1. Time increases from left to right in the figure, and frequency from bottom to top of each panel. Each black dot corresponds to a transmitted tone. For each transmission, 63 of the tones convey the encoded channel symbols, while the remaining 63 (along the bottom of each panel) are the sync tone. The upper spectrum represents the message "VK7MO K1JT FN20", while the lower one is "VK7MP K1JT FN20". *Only one character in the message has changed, but 52 of the 63 channel symbols are different.* This fact illustrates the extraordinary power of the RS(63,12) code used in JT65: the sequences of channel symbols for *any* two of the 2^{72} possible JT65 messages can never have more than 11 symbol values in common. This is the reason that complete messages can be received exactly as sent, even when many channel symbols have been corrupted or lost in the noise.

[Figure 1 near here.]

Analysis of a Received JT65 Signal

WSJT begins the analysis of a JT65 signal by identifying the sync tone and using it to establish its frequency offset (relative to the nominal 1270.5 Hz) and the time offset caused by propagation delay and computer clock errors. Spectral analysis is then carried out for each of the 63 intervals containing the information-carrying channel symbols. This process yields measurements of signal power for each symbol, divided into 64 frequency bins. The bin with largest power yields the most probable value of the symbol being received at that moment; the second largest corresponds to the second most probable symbol value, and so on, down to the smallest. Thus, the information conveyed by each symbol is partly contained in its most probable value, but additional information is contained in a series of other possible values, with progressively decreasing probabilities of being correct. Each correctly received symbol conveys 6 of the 378 transmitted bits.

Study of Figure 1 should make it clear that there is no obvious correspondence between individual characters in a message and particular values or locations in the encoded sequence of channel symbols. As stated earlier, all of the message information is mixed together and mathematically distributed over the entire sequence. A correctly received message may result from as few as 12 channel symbols received with high confidence, or all 63 symbols received with relatively low individual confidence, or any of a wide range of possibilities between these two extremes. The spectra of JT65 signals received over the air will differ from the idealized ones shown in Figure 1 because they include random noise and fading signals. The statistical properties of signal and noise variations will determine the levels of confidence assigned to individual symbol values.

Hard and Soft Symbols

As an educational exercise for myself, and to help answer for everyone the question posed in the title, I have carried out a series of explicit measurements of the number of correctly copied channel symbols in JT65 transmissions. One thousand simulated transmissions were generated at each SNR between 0 and -40 dB in 2500 Hz bandwidth, in 1 dB steps. These transmissions were received and analyzed, and the number of correctly copied symbols determined for each one. The averaged results for each signal level are plotted as the solid curve in Figure 2. The probability of correct symbol reception depends only on signal-to-noise ratio; it is independent of details of any decoding algorithms that might be used subsequently, in a program like WSJT, to convert the raw channel symbols into callsigns or other user information.

[Figure 2 near here.]

Up to this point, discussion of the reception of channel symbols has been limited to "hard" decisions: a symbol's value has been taken as the index of its frequency bin with the highest measured power. However, significant additional information is contained in the actual values of the power measurements, which can be used to indicate which symbols are the most reliable and to produce "soft symbol" probabilities. At low SNR, many symbols that were not correctly received as the most probable value will be correct as the second most probable, and others as the third, fourth, and so on. Relatively few correct symbol values will be found among the lowest-ranking probabilities. To provide quantitative examples of this type of information, the dashed and dotted curves in Figure 2 illustrate the average number of received symbols for which the correct value was found in the top 2, 4, or 8 of the 64 spectral power measurements. An ideal decoding procedure takes advantage of all soft-decision information through the full range of probabilities.

Decoding the Message

Figure 2 shows that at SNRs down to -20 dB, JT65 signals transfer an average of more than 39 correct hard-decision channel symbols from transmitter to receiver. With the Reed Solomon (63,12) code used in JT65, 38 correct symbols are always enough to guarantee complete reception of a 72-bit user message with high confidence, even if only a hard-decision decoder is used. If the locations of symbols likely to be unreliable are known—those during a signal fadeout, for example—these can be flagged as "erasures," and a smaller number of correct symbols will then suffice for decoding. Better still, true soft-symbol information may be used. WSJT is packaged with a compiled version of the Koetter-Vardy algebraic soft-decision decoding algorithm,[4] which takes partial advantage of probabilistic information on the most probable and second most probable values of every symbol. The KV algorithm can decode single JT65 transmissions down to about −25 dB, again with very high confidence in the result.

As can be seen in Figure 2, at still lower signal levels the number of correctly copied channel symbols decreases. The hard-decision number falls to 9.3 (the equivalent of two JT65 callsigns) slightly below −28 dB. This signal level is close to the typical lower limit for decoding by the JT65 Deep Search (DS) algorithm, which will be described next.

An ideal JT65 decoder would create two-dimensional arrays like those depicted in Figure 1 for all possible messages, and would cross-correlate each one of them with the corresponding soft-symbol array for the received signal. Any JT65 message different from the one actually transmitted would produce 11 or fewer matching channel symbols, whereas the correct message would exhibit (soft) matches for all 63. The correlation procedure would make use of the full soft-information content, and a quantitative indication of the resulting confidence level would be provided for any decoded message. Unfortunately, an exhaustive search of every possible JT65 message is not computationally feasible. However, since the list of active EME stations is no more than a few thousand, and since activity patterns change rather slowly with time, it is perfectly feasible for WSJT to carry out an exhaustive search for callsigns in that relatively short list. Such a procedure is the basis of the JT65 Deep Search algorithm.

When using the DS decoder, WSJT forms a large number of hypotheses about what message a JT65 signal *might* contain, based on the receiving station's call and the contents of a callsign database. Each hypothesis is tested to see whether the received symbols match those of the hypothetical encoded message. This algorithm can be programmed to result in the high-confidence decoding of any one of the hypothesized messages, but no others. An exhaustive search is the *optimal* decoding strategy, in the sense of achieving the best possible decoder performance. When practicable, it is usually the method of choice.

Levels of Confidence

The number of hypothetical messages tested during a JT65 Deep Search is typically between 14,000 and 20,000. The decoder's task is to determine whether one of these messages matches the transmitted one to some specified level of confidence, or to establish that *none* of them does. To illustrate how this process works, Figure 3 shows examples of correlations

---

[4] R. Koetter and A. Vardy, "Soft-Decision Algebraic Decoding of Reed Solomon Codes," in *IEEE Transactions on Information Theory,* **49**, 2809–2825, 2003.

obtained when doing the simulations reported in this paper. The top panel is for signal-to-noise ratio −26 dB, and the bottom for −28 dB. Dots represent correlations of the noisy, fading signals against the message actually transmitted, for 1000 different simulations at each signal level. Crosses show correlations against incorrect messages selected at random from the group of all possible JT65 messages. It is easy to see that with a threshold set at about 3 correlation units, the decoder can make sure that nearly all transmissions at −28 dB and higher will be decoded correctly and that there will be very few false decodes. Recent versions of WSJT produce numerical confidence levels on an arbitrary 0 to 10 scale monotonically related to the scale used in Figure 3. Roughly speaking, level 3 on WSJT's scale implies moderately high confidence, and anything over 6 implies high or very high confidence. In normal practice the operator will have other relevant information available, so WSJT requires the operator to make all final decisions about valid copy.

Summary and Conclusions

To provide succinct answers to the question posed in the title, a subset of the measurements obtained in the simulations is summarized in Table 1. Entries are included for SNRs between −18 and −28 dB, at 2 dB intervals, and for convenience the results are quoted in channel symbols, bits, and equivalent 28-bit callsigns. Even at the lowest usable signal levels, around −28 dB, the number of hard-decision bits correctly copied over the radio path exceeds the number required to convey two callsigns. The numbers in Table 1 are conservative lower limits, because they are based on hard decisions only. Soft-decision information adds significantly to the totals and further enhances the sensitivity as well as the levels of confidence that can be assigned to decoded messages.

Table 1. Hard-decision channel symbols copied over the air, per JT65 transmission.

SNR (dB)	Channel symbols	Bits	Equivalent callsigns
−18	46.9	281	10.1
−20	39.6	237	8.4
−22	31.9	191	6.9
−24	23.1	139	4.9
−26	15.5	93	3.3
−28	9.6	58	2.1

The information presented here should lay to rest any fears that when using its Deep Search decoder, JT65 might not transfer enough channel symbols over the radio path to satisfy traditional requirements for valid QSOs. It is certainly true that at the lowest usable signal levels, between −25 and −28 dB on the WSJT scale, some prior information about active stations is required for the Deep Search decoder to succeed. Of course, prior information of this type is advantageous when trying to make a difficult contact or identify a weak station calling CQ, using *any* mode, for exactly the same reason—it gives the decoder, consciously or otherwise, some idea of what message content to look for. Helpful lists of "good calls" have been a part of contesting and weak signal amateur communications for decades. It has never been considered illegitimate to possess or make use of such information in the process of copying a weak signal, as long as the copy is truly accomplished. Figure 2 and Table 1 show that this condition is easily met by JT65.

In an article recently published[5] in *DUBUS*, Klaus von der Heide, DJ5HG, presented an analysis of the decoders used in WSJT with the goal of promoting an objective discussion of what constitutes a minimum valid QSO. Using the formalism of information theory, he showed that the KV decoder meets a "strict" definition of a QSO, while the DS decoder meets a "dynamic" definition. The difference between the two is that the strict definition requires that all necessary information be treated as unknown at the start of a QSO, while the dynamic definition acknowledges that some information (such as one's own callsign) may be known, and still other information might be found in a list. Traditional practice has always accommodated the inevitable knowledge of one's own callsign, and has also accepted that lists and schedules do not invalidate QSOs. A mode like JT65 that uses compressed, structured messages cannot account for individual parts of a message such as characters in callsigns, because they are not transmitted as such. One can still insist, however, that a sufficient number of channel symbols be copied over the air, and that operator confidence in the decoded message meets a suitably high standard. The measurements described in this paper show that the JT65 Deep Search decoder passes these tests easily—and therefore that it, too, produces valid QSOs in terms of traditional amateur practice.

The article by DJ5HG quite properly calls attention to the fact that the number of distinct messages decodable by the Deep Search algorithm is no larger than several times the length of the callsign database. Information theory says that if all hypothetical messages constructed from that database are assumed equally likely, the quantity of transferred user information can be taken as the base-2 logarithm of the number of decodable messages. The resulting number of bits—approximately 14.2 for the default Deep Search setup of WSJT—is a useful and valid measure of the previously unknown information transferred when such a message is decoded. In a scheduled QSO, whatever the modulation mode, the number of transferred, previously unknown bits may be even smaller. However, these measures of the quantity of transferred information should not be confused with the number of probabilistically evaluated, information-carrying symbols or bits conveyed over the radio

---

[5] K. von der Heide, DJ5HG, "Minimal QSOs and Their Validity," *DUBUS*, **35**, No 1, pp. 38–53, 2006.

path from transmitter to receiver. The latter quantities—those illustrated in Figure 2 and Table 1 of this paper—are roughly analogous to characters or other fragments copied from a marginal CW signal. They carry proportionally more weight in JT65, however, because of the strong error-correcting code and because the process of synchronization fixes the exact location of each symbol within the transmitted sequence. Signal strength variations provide the JT65 decoders with important probabilistic information about which received symbols are the most reliable. Together, these factors lead to the very high level of QSO integrity that is achieved with JT65.

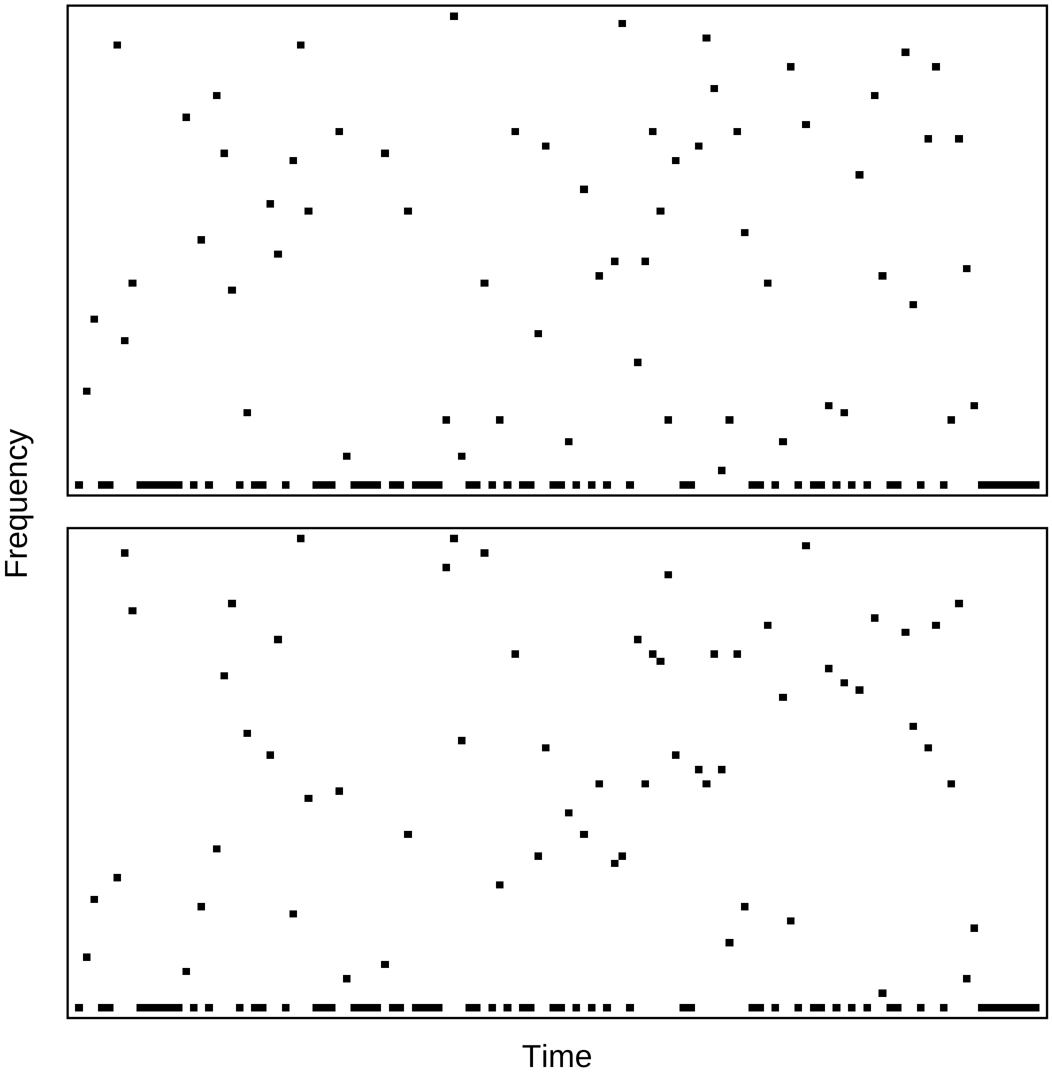

Fig. 1.— Idealized spectrograms of the JT65 transmissions for "VK7MO K1JT FN20" (top) and "VK7MP K1JT FN20" (bottom). The horizontal axis represents 46.8 s of time; the vertical axis of each panel represents frequency over the range of 65 tones used by JT65. Black marks indicate transmitted tones, corresponding to the encoded channel symbols. The pseudo-random pattern along the bottom of each panel is the synchronizing tone. Note that although the two messages are nearly identical, the patterns of channel symbols are almost entirely different (except for the pattern of the synchronizing tone, which is always the same).

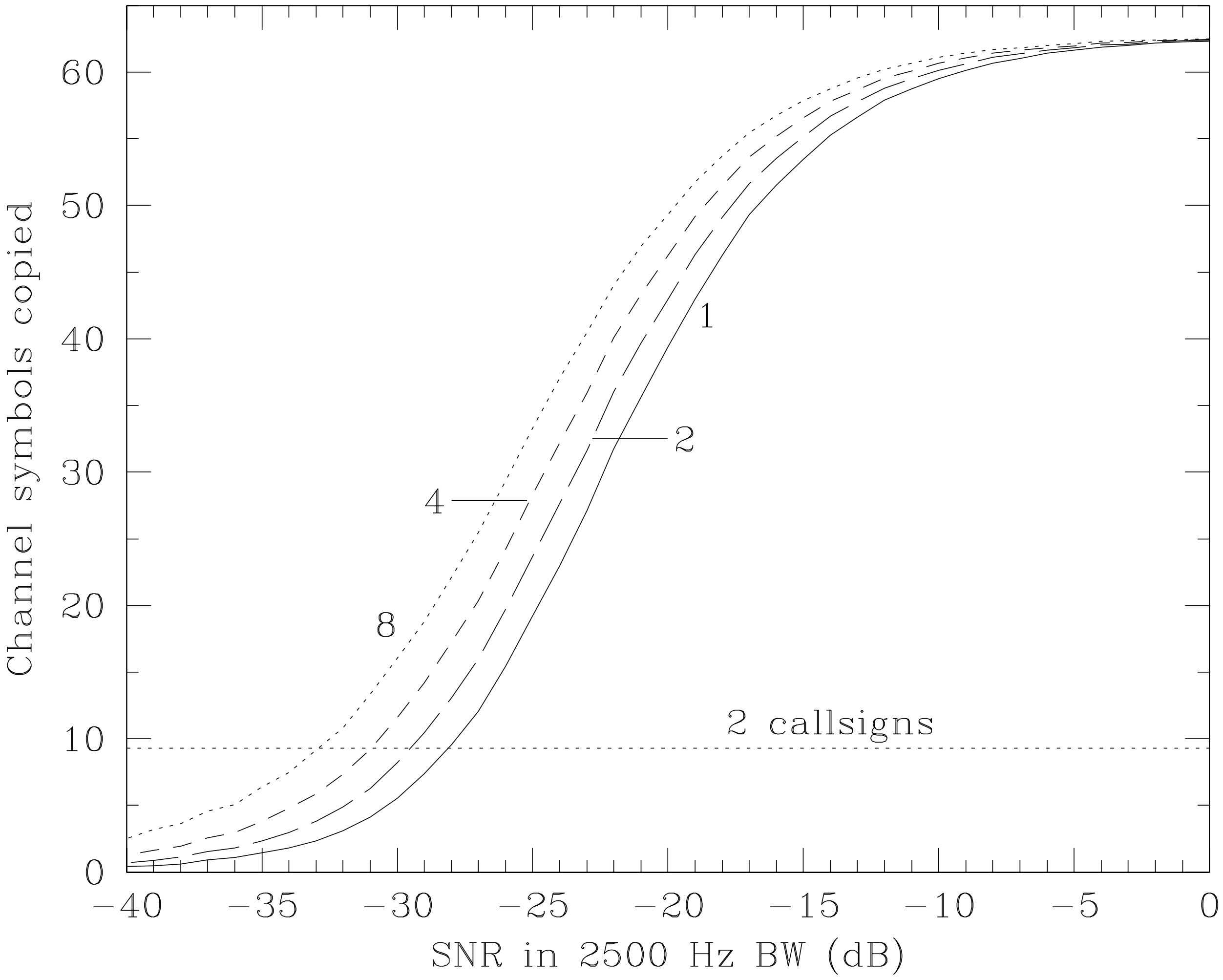

Fig. 2.— Measurements of the number of correctly copied channel symbols over a simulated radio path, plotted as a function of SNR on the WSJT scale and assuming 64-FSK modulation and noncoherent detection. Generated JT65 signals were degraded by Rayleigh fading and by additive white Gaussian noise. The solid curve (labeled "1") gives hard-decision results; the remaining curves provide some indication of the soft-decision information by showing the average number of times that the correct symbol fell in the top 2, 4, or 8 of the 64 measurements of spectral power. Adjustments were made to the curves to remove the biasing effects of accidentally correct symbol values. The dotted horizontal line shows the equivalent number of bits in two 28-bit callsigns.

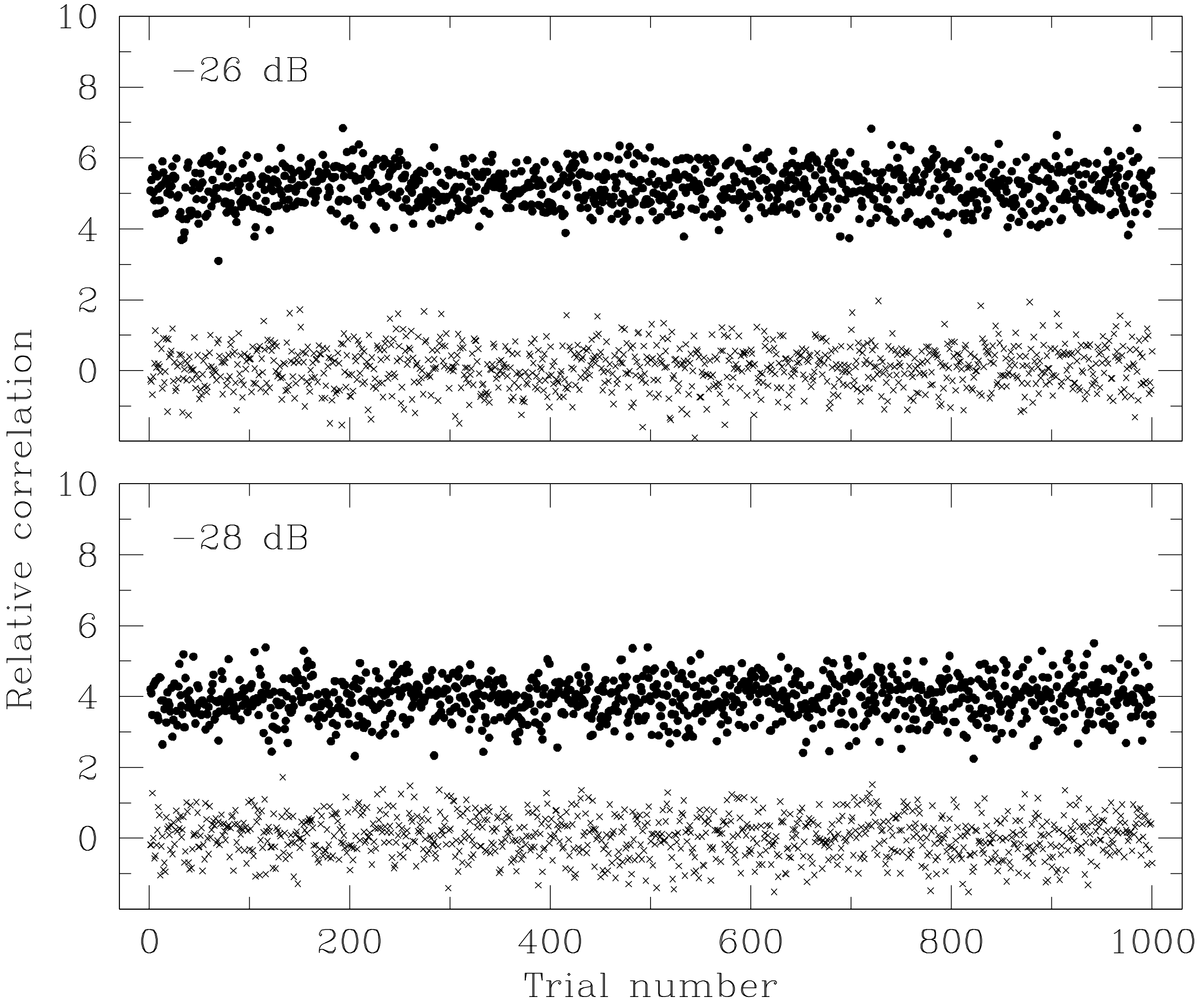

Fig. 3.— Examples of correlation values produced by the Deep Search decoder for four different groups of 1000 simulated transmissions. The signal-to-noise ratios were set at −26 dB (upper) and −28 dB (lower), on the WSJT scale. Dots represent correlations against correct messages, while crosses represent incorrect messages. A decoding threshold set at about 3.0 would assure that nearly all transmissions above −28 dB will be decoded correctly, with few decoding errors.