WIRELESS MESSAGE FAULT MITIGATION

Systems and Methods for 5G/6G Receivers to Detect, Localize, and Correct Message Faults - Without a Retransmission

David E. Newman and R. Kemp Massengill

Executive Summary

Message faulting is a major challenge for 5G-Advanced and especially 6G, due to increased network crowding and increased pathloss at higher frequency bands. Prior-art methods for error mitigation, such as HARQ and its variants, are costly due to automatic retransmissions and wasteful due to message bloating with FEC bits. Therefore, to assist broadband receivers in recovering weak and noisy messages, methods are proposed herein for extracting information from the received signal waveform and exploiting correlations with message faults [1]. The receiver can perform these diagnostics in real-time to determine which message elements are likely corrupted, and in many cases can also determine the most likely corrected message, thereby avoiding the delays, costs, and energy usage of a retransmission.

The examples are based on uplink messaging because the current development challenge is for reliable reception of user device signals, however the fault mitigation procedures presented herein can be applied equally to downlink, sidelink, backhaul, and non-3GPP communications as well. In each application, the receiver can achieve automatic real-time stand-alone message error correction for improved reliability, reduced latency, and enhanced network efficiency overall.

The Message Reliability Problem

Reception of RF signals has always been a challenge. Wireless signals are inevitably limited to a low energy density, while electromagnetic noise is ubiquitous. The reliability problem is a primary challenge in the current rapidly-changing wireless environment due to (a) network crowding from the exponential increase in wireless device population in the coming years, (b) shorter symbol-times planned for high-numerology encoding, (c) atmospheric attenuation at multi-GHz frequencies, and (d) interference from reflections and diffractions unavoidable at the higher frequencies. The reliability deficit is most critical in the uplink because user devices (mobile phones, IoT gadgets, etc.) generally have much lower transmission power capabilities than base stations and access points, although all wireless receivers are susceptible to the reliability degradation that appears inevitable in next-generation systems.

Error detection and correction methods of the past are not adequate for next-generation networking due to the inherent costs and inefficiencies of legacy methods. In the simplest ARQ, a retransmission is automatically requested whenever a demodulated message disagrees with its embedded error-detection code, thereby wasting a wealth of good information available in the as-received message despite the fault. In soft-combining, the initial signal and a retransmitted signal are combined at the analog level, which results in at best only a $\sqrt{2}$ improvement in SNR - and much worse for non-Gaussian and non-static interference which are typical, thereby piling bad signals on top of good ones. Various flavors of HARQ include FEC bits with each message, unavoidably bloating the message, while also presenting a larger interference target due to the larger size. In other versions, the FEC bits are provided in a second transmission on request - at great cost in latency. Even with extensive FEC data, the receiver often fails to correct the message because the specific faulted message elements cannot be identified, or excessive number of faults, or corruption of the FEC data itself, among other mishaps that frequently occur by this method. Further variations involve complex partial retransmissions with or without puncturing, and all involving greatly increased demands on the receiver processor. As a result, the expected QoS is severely violated, especially for time-critical messages.

Next-generation wireless applications demand a faster, simpler, and more reliable way to identify faulted message elements and to determine the corrected version.

Available Information in the Received Signal

The signal of a corrupted message is rich with information about the faults, hinting at their corrected values. The first step is to determine whether the demodulated message is faulted, by comparing to an embedded error-detection code. Most error-detection codes (such as CRC or a parity construct) are only 16 bits in size and therefore can be added to almost all messages with negligible added cost. If the message is corrupted, the second step is to determine which message element(s) is/are faulted, that is, which ones are demodulated incorrectly due to some kind of signal distortion. The third step is to determine the corrected values of all the faulted message elements.

The likely-faulted message elements can be identified using receiver-based diagnostics on the initially received message signal. For example, the receiver can compare the signal parameters of each message element with (a) a predetermined set of values, (b) an average of the other like-modulated message elements in the same message, and (c) an expected value based on other messages correctly received in the past. More specifically, the receiver can determine a modulation quality of each message element according to the deviation between the amplitude or phase of the message element versus the amplitude or phase of the nearest predetermined modulation state (obtained from a demodulation reference signal near the message). If the modulation scheme is QAM, the receiver can determine the amplitude deviation in both I and Q branches, relative to the nominal modulation levels or relative to an average of like-modulated message elements in the as-received message. In each case, a large deviation indicates a likely faulted message element. In addition, the receiver can calculate an average amplitude or phase for each modulation state in the received message by averaging each message element that has the same decoded value, and then flag message elements that deviate from the message average. The receiver can also determine, from the digitized waveform, a width or variation in the signal amplitude or phase for each message element, or for each branch in QAM. A larger than average variation indicates a high level of noise or interference, and hence a likely faulted message element. In addition, the digitized data can expose a slight but unexpected frequency offset (within the subcarrier bandwidth) of faulted message elements. The data can also reveal a slight but unexpected change in the received power (other than the modulation levels), thereby implicating the affected message elements. The same digitized waveform data can expose interference according to the signal present during transitions between subsequent message elements. Normal transitions are generally smooth and monotonic, whereas fluctuations in the transition zone may indicate interference. As a further test, the receiver can measure the polarization angle of each message element. Any change in the received polarization angle indicates interference, and hence a likely faulted message element. Often the fault indicator in each of these diagnostics may be subtle, but when multiple tests are combined, the accumulated errors in faulted message elements tend to add, while random variations in the good message elements tend to cancel out. Thus a total quality factor may be calculated for each message element by combining the results of the various diagnostics, so that any faulted message elements stand out clearly as outliers in the combined data.

The receiver can also test the demodulation reference(s) used to calibrate the modulation levels of the message, since the demodulation reference can also be subjected to interference. A weak or out-ofspec demodulation reference can cause a good message signal to be incorrectly demodulated, which looks like a message fault but is actually a fault in the demodulation reference. A good receiver can check for this effect using the diagnostics listed above. In addition, the receiver can quantify the noise or interference by measuring the received signal during special blank resource elements with no transmission. Any background signal observed in the nominally blank spaces can then be subtracted from the message waveform to (roughly) compensate for slowly-varying noise or interference.

In addition to the waveform diagnostics, the receiver can check the type and format of the as-received message. Sometimes a fault results in a legal but rarely-seen message type, or a peculiar value that is unexpected for the present application, or other indication besides waveform distortion. The receiver can compare the message type and format and content to a database of previously received (uncorrupted) messages, to further expose likely errors. In many cases, the corrected message version becomes apparent by comparison between the as-received message and the prior messages.

In addition, the receiver can select among a multitude of possible message corrections according to the likelihood of each candidate version based on the waveform results and/or the inferred intent and/or the likely meaning of the message. For example, if the most likely faulted message elements happen to be in the error-detection code, then the error-detection code cannot be used to select which candidate is correct. A similar quandary can arise if there are several faulted message elements. In that case, the receiver can compare each candidate version with previously received messages, inferring the intent or meaning of each candidate version, and thereby weed out the improbable versions first. The receiver can then select the best corrected version according to the waveform diagnostics and the application context.

Artificial intelligence can greatly assist in this process. AI excels at correlating multiple disparate data sources sensitive to different aspects of the problem, and can arrive at the most likely correct version of the message in a single pass. For example, the AI model may take as input the various waveform diagnostic results listed above, the demodulated values, and the history of message types and allowed formats. The AI model can then correlate all of the input factors to immediately determine the most likely corrected message version. The AI model may also report other candidate versions and the likelihood of each, if trained to do so. In addition, the AI model can evaluate the uncertainty in its conclusions, thereby further guiding the receiver in determining what to do. The AI model can also issue special alerts whenever the most-likely version actually has low likelihood, or when two different solutions have similar likelihood, or when the number of likely-faulted message elements exceeds a limit, or other problems that the AI model can reveal and that the receiver should know about.

Receivers implementing the waveform diagnostics on corrupted messages, optionally with assistance of a trained AI program, can localize the likely faulted message elements and their likely corrected values, and thereby rescue the message, in real-time, internally in the receiver, without a retransmission. Receivers with stand-alone fault mitigation technology can thereby enhance communication reliability in the next generation of wireless networks, at no cost in latency or transmission energy, for the benefit of users everywhere.

Waveform Diagnostics

A wide range of data about the received signal is available to the receiver, but is generally ignored or discarded. Some of the more useful fault indicators, readily available to any receiver that digitizes the received waveform, are shown in this section. Base stations, access points, and most user devices such as cell phones and personal computers, can identify faulted message elements by checking these diagnostic parameters in the received signal of each message element symbol-time, and in many cases can determine the most likely corrected version, without a retransmission.

Figure 1 shows the effect of noise or interference which is in-phase with a signal waveform, thereby causing a distortion in the received amplitude. This can result in an erroneous assignment of the modulation state. Figure 2 shows the same signal waveform, but now with interference at 90 degrees phase relative to the transmitted signal, thereby causing a phase shift in the received signal, which can also result in erroneously demodulated message elements.

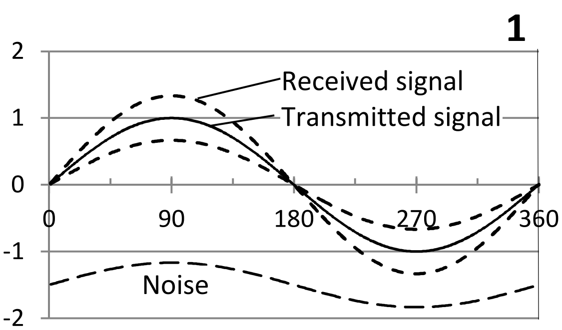


Fig. 1: Amplitude fault caused by noise or interference, added or subtracted in-phase.

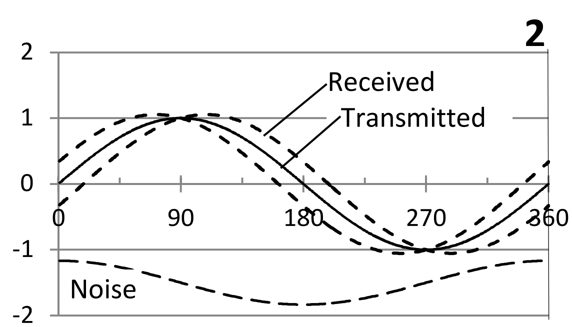


Fig. 2: Phase fault caused by noise or interference with 90-degree phase shift.

Figure 3 shows the waveform of a received signal within one symbol-time, including substantial amplitude variation due to noise or interference. A receiver can measure these amplitude variations of each message element. The message elements with the worst variations are flagged as suspicious. If the message turns out to be faulted (according to the error-detection code) the suspicious message elements are then corrected. Figure 4 shows the distribution of wave amplitudes in the symbol-time of Fig. 3. The width of the distribution is a measure of noise and interference, and hence of suspiciousness.

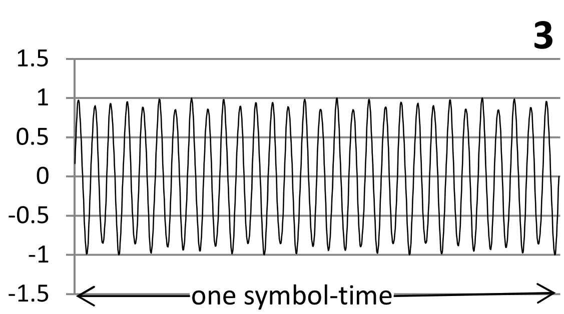


Fig. 3: Random amplitude variations in the received signal, during the symbol-time of one message element.

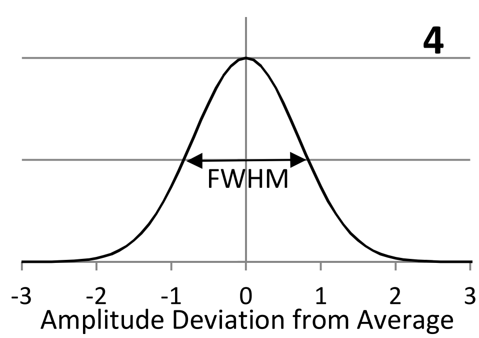


Fig. 4: Distribution of waveform amplitudes within one message element symbol-time.

Figure 5 shows a message element waveform with a "peaking" amplitude variation, due to interference by an intruder signal having a slightly different frequency and phase. Figure 6 shows the resulting amplitude distribution. The amplitude distribution peak is skewed or displaced, relative to the average, as a result of the detuned interference of Fig. 5. The receiver can detect the suspicious message elements according to the increased width of the amplitude distribution or the peak skew as shown.

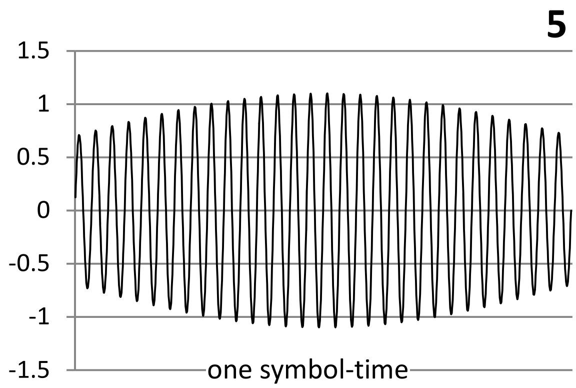


Fig. 5: Peaking amplitude variation due to slightly dephased interference.

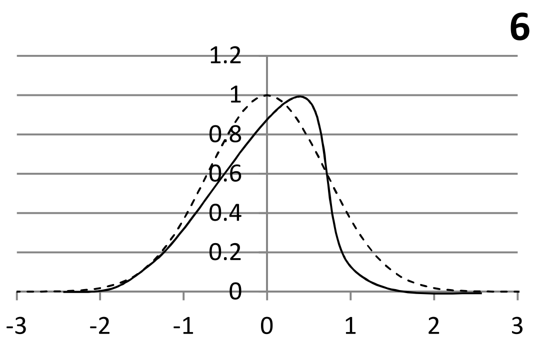


Fig. 6: Skew or asymmetry in the amplitude distribution of the peaking signal shown in Fig. 5.

Figure 7 shows a modulation chart for 16QAM with the I-branch amplitude horizontally and the Q-branch amplitude vertically. The central cross represents zero amplitude. Each point is the nominal modulation of the 16 nominal states. The modulation deviation of a message element is the distance between the as-received I and Q amplitudes of the message element (circle) and the nearest nominal state. Alternatively, the I-branch deviation and the Q-branch deviation may be tallied separately. Message elements with the highest modulation deviations are suspicious. Figure 8 shows a highly schematic layout of a receiver monitoring two orthogonal polarizations in the received signal. Message elements that have an unexpected change in polarization likely include interference and may be faulted.

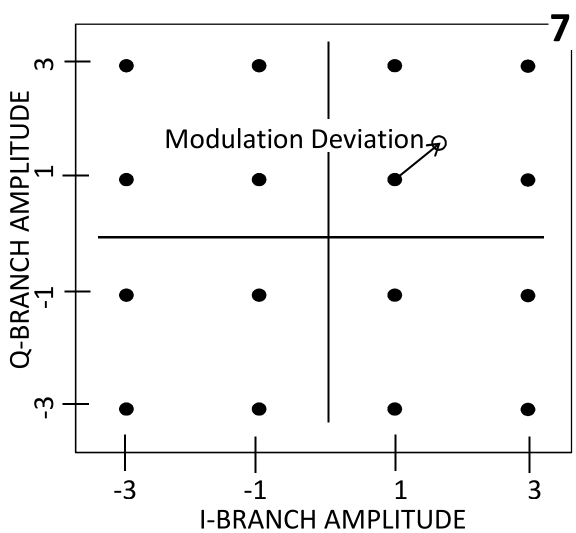


Fig. 7: Modulation chart for 16QAM, with nominal branch amplitudes and one deviated message element.

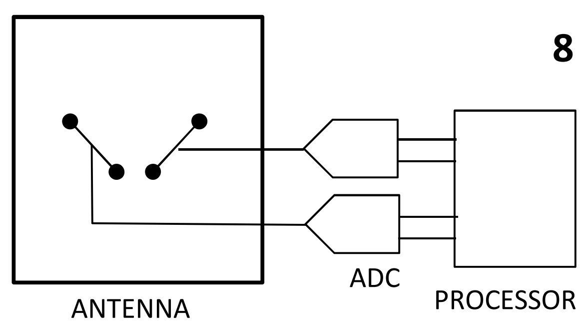


Fig. 8: Polarization receiver detects two orthogonal polarizations in the received signal of each message element.

Figure 9 shows a series of phase measurements representing the as-received phases of the message elements in a QPSK modulated message with carrier avoidance. All the message element phases are consistent with one of the four predetermined modulation phases of 45, 135, 225, and 315 degrees within an expected scatter width of a few degrees. There is a systematic phase offset of a few degrees, due perhaps to noise during a previously received demodulation reference or a slight time drift between the demodulation reference and the message. Nevertheless, the modulation is unambiguous for all of the message elements except for one "outlier" point, which differs substantially from the received phases of the other like-modulated message elements. The receiver therefore flags the outlier as suspicious, even though it is consistent with the predetermined modulation levels, due to the difference between the outlier phase and the average phases of the other like-modulated message elements of the message. If the message turns out to be corrupted, as indicated by the embedded error-detection code, then the receiver can select the outlier as the most likely faulted message element, and can proceed to find the correct value, without wasting time altering the other message elements or requesting an unnecessary retransmission.

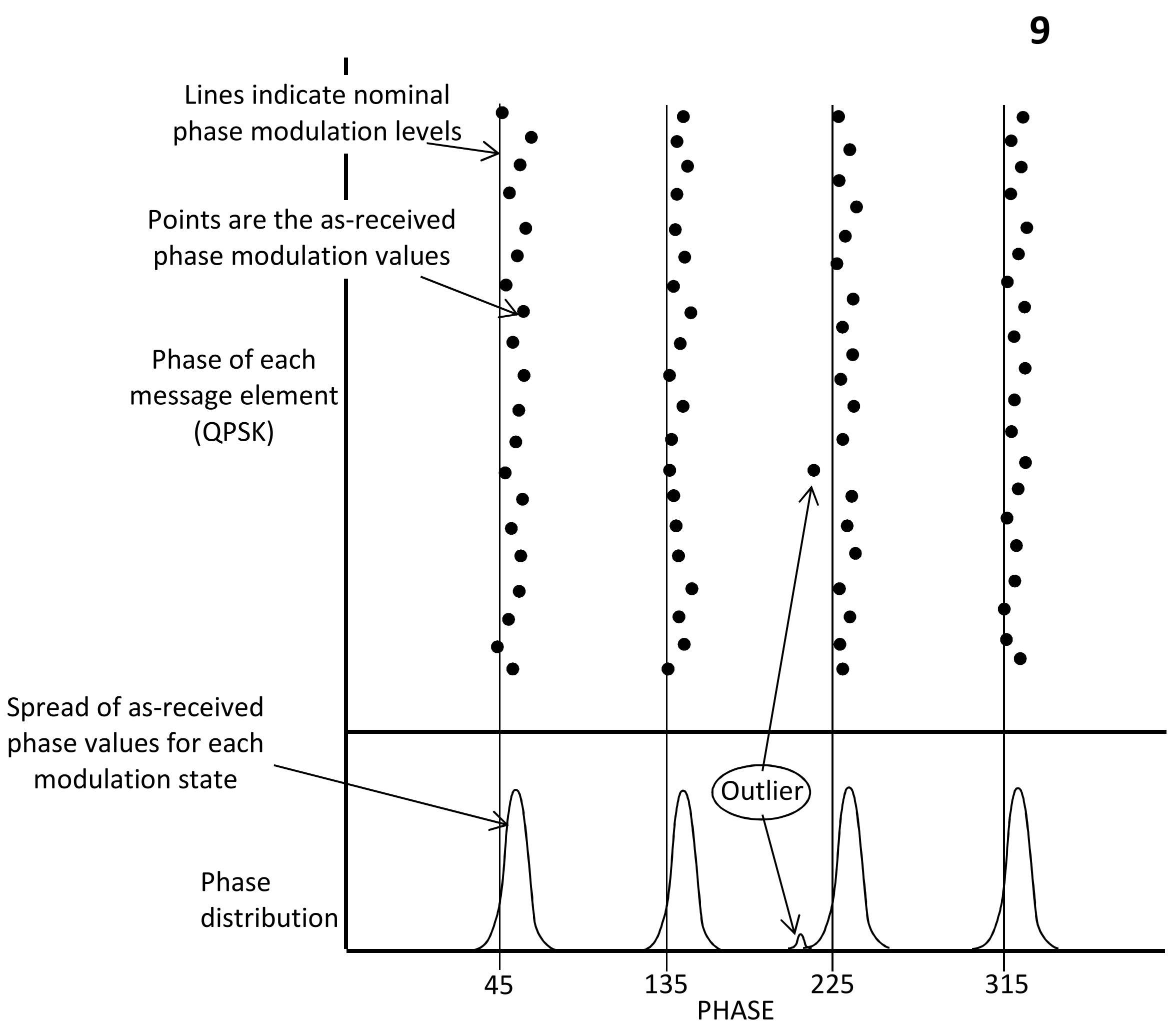


Fig. 9: Deviation relative to average. Receiver measures the as-received modulation of each message element, calculates an average of all the like-modulated values for the message, and flags any message elements that deviate excessively from the average. The "outlier" is a suspicious message element since its modulation differs from the average, even though it is consistent with the nominal modulation level.

Figure 10 shows multiple fault diagnostics evaluated for each message element of a received message, including the modulation deviation of the received I and Q branches of a QAM message, the FWHM width of the I and Q amplitudes, the deviation of the signal amplitude and phase relative to the message average for each message element, and its frequency offset relative to each subcarrier nominal frequency. The last line shows a combination of all the measurements. Although the individual diagnostics cannot definitively identify the faulted message element, it stands out clearly in the combined data.

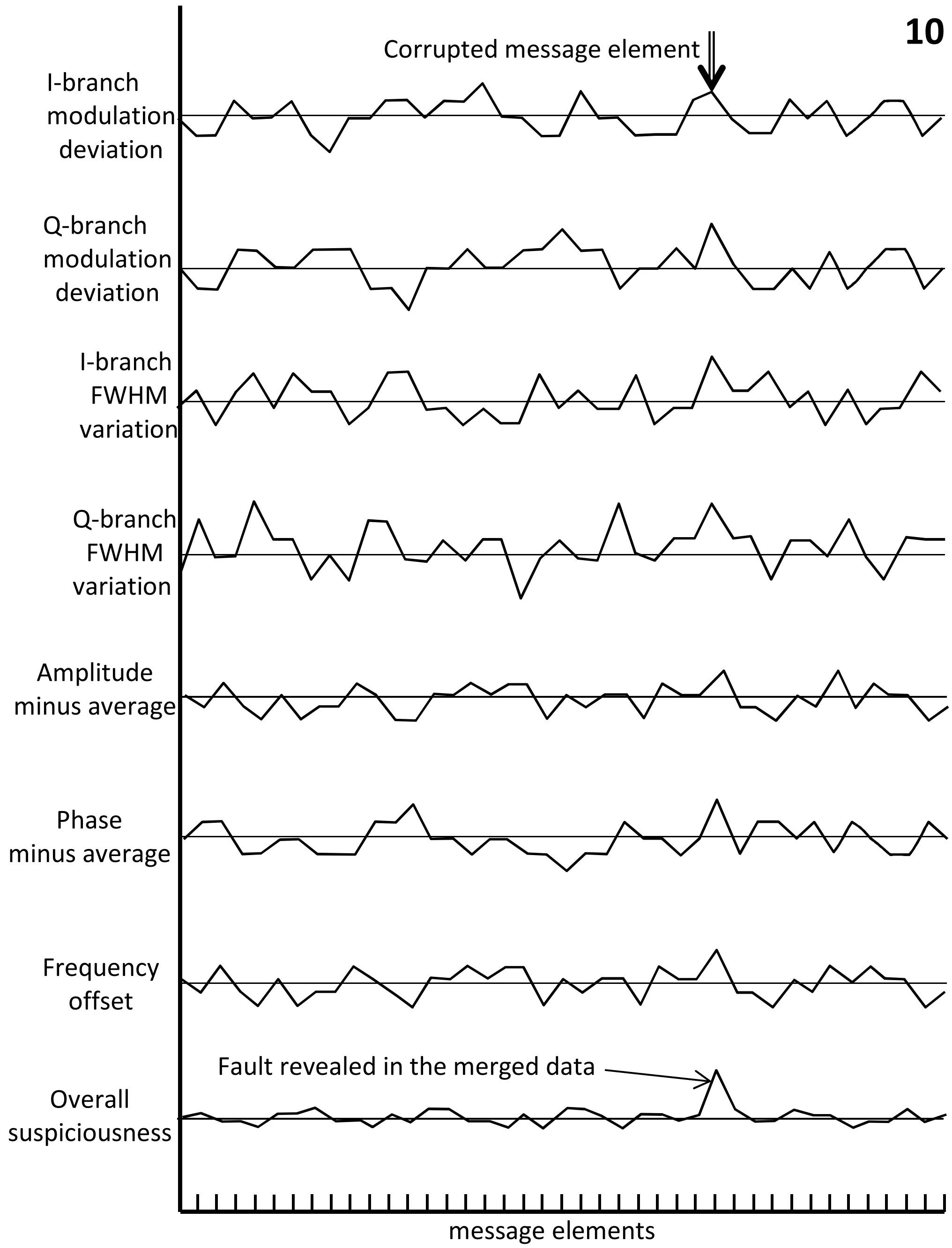
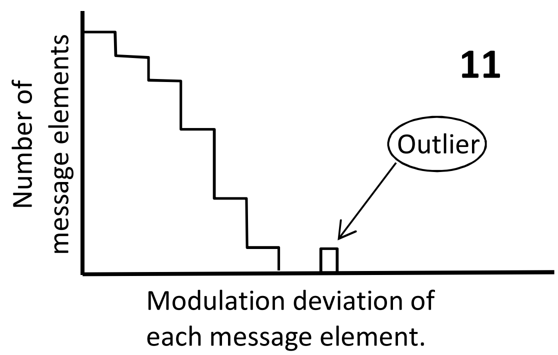
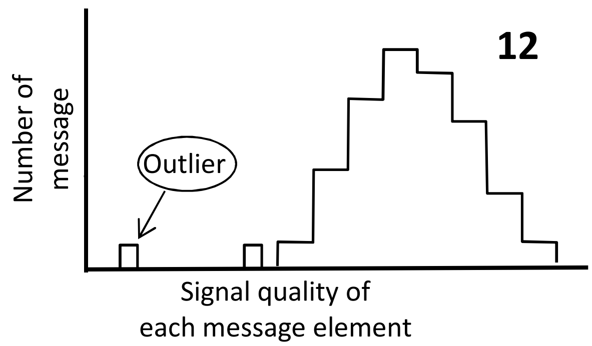


Fig. 10: Combined data reveals which message element is likely faulted. For the unfaulted message elements, measurement variations largely cancel out.

Figure 11 shows a histogram of the message elements in a received message. The number of message elements with a particular value of the modulation deviation are plotted versus the size of the deviation. Most of the message elements have a small deviation, as expected, but one message element has a higher than expected modulation deviation, and therefore may be faulted.

Figure 12 shows a histogram of the overall signal quality of each message element in the message. The signal quality is a combination of all the waveform diagnostic data, inversely related to the "overall suspiciousness". Most of the message elements have a high overall signal quality, but one is a bad outlier which is very likely faulted.





In summary, the receiver can measure numerous features of each message element of the message, including the deviation in amplitude or phase relative to the nominal modulation level or an average of the other message elements, as well as a FWHM width of the amplitude or phase variations within each message element (or branch), plus the polarization angle, inter-symbol transition properties, received power (aside from the the modulation level), and other parameters of the message elements, as well as the demodulation reference(s) used in calibrating the modulation levels of the message. The receiver can then identify specific message elements as suspicious if they deviate substantially from the others in any of these diagnostics. The receiver can also calculate an overall signal quality or suspiciousness of each message element by combining the diagnostic data so that the unfaulted message elements random variations cancel out, while the faulted message elements - with violations in multiple diagnostics - would show up as a substantial deviation from the rest.

The next step, after identifying the likely faulted message elements, is to determine the correct value for each one. If there is only one clearly faulted message element, the receiver can calculate the correct value for it using the error-detection code of the message. Often, however, there are multiple suspicious message elements, or the error-detection code itself may be suspect, or the demodulation reference may be suspect, in which case a far more powerful analysis procedure may be required to infer the correct message. For that, artificial intelligence is the method of choice.

Artificial Intelligence for Fault Identification and Correction

Artificial intelligence (AI) can greatly enhance the recovery of faulted messages. The AI model works by finding correlations between each received message element and the corrected version, based on input data such as the waveform diagnostics summarized above. In a complex problem like message fault correction, multi-parameter conditional correlations are often extremely complex and extremely subtle. But this is exactly the type or problem that AI excels at. In many applications, not dissimilar to the message fault mitigation problem, AI performs at the highest levels, far beyond the capabilities of even the most experienced experts in the field.

To "train" the AI model, numerous examples of faulted messages are needed, along with the corresponding true versions. For each example, certain variable parameters in the model are iteratively adjusted for optimal predictions. After such training, the model can deduce the most likely fault locations in each message, and can inductively determine the corrected message, virtually instantaneously. AI-based fault mitigation concepts can be applied by base station receivers for uplink message reliability, by user devices for downlink reliability, and all other wireless receivers requiring high reception reliability despite crowded or noisy network environments.

Figure 13 shows a schematic of an artificial intelligence model, a neural net in this case, configured to determine which message elements are likely faulted, and the most probable corrected message. The model inputs generally include the waveform data of each message element, the demodulation reference(s) used in demodulating the message, data about current noise and interference (based on measurements during a non-transmission period), the expected type and format and possibly meaning of messages in the application, and anything else that may be correlated with the message faults.

The input data are fed into a series of layers (two shown), each layer consisting of a large number of internal functions or "nodes", linked to the input values or to the output values of the previous layer, and all feeding results into a final output node. The purpose of the internal functions is to find correlations between the various inputs and each possible output value. A high correlation means that when the particular input (or combination of input values) is present, the probability of the corresponding output value is increased, and is decreased if the particular input values are absent. To identify such correlations, it is necessary to first train the model by adjusting the variables, contained in each of the internal functions, using "known" examples in which the correct answer (ground truth) is already known. During training, the ground truth is not given to the model. The model tries to guess whether there is a fault, and if so, where it is, and then tries to determine the corrected version of the message. Only then, after the model does its best calculation and presents the most probable corrected version of the message, the correct answer is then revealed. If the model was right, the current set of variables is "firmed up". If the model was wrong, then some of the internal variables are altered in an attempt to bring the prediction into better agreement with the known answer. The successive adjustment of variables, based on the model's analysis, is termed "supervised machine learning". The training is complete when the model is finally able to predict the locations and corrected values of message faults with high accuracy.

After training, the AI model, or an algorithm derived from it, is used by an actual receiver for message fault mitigation. The model generally produces its output very quickly, in a single pass, requiring much less time than even a single message element's symbol-time, and far less time than any possible retransmission. Hence using AI, the message can usually be recovered, without degradation of the expected QoS or of stringent latency requirements. Fault correction by the receiver also saves a lot of money, time, transmission power, and headaches for the user.

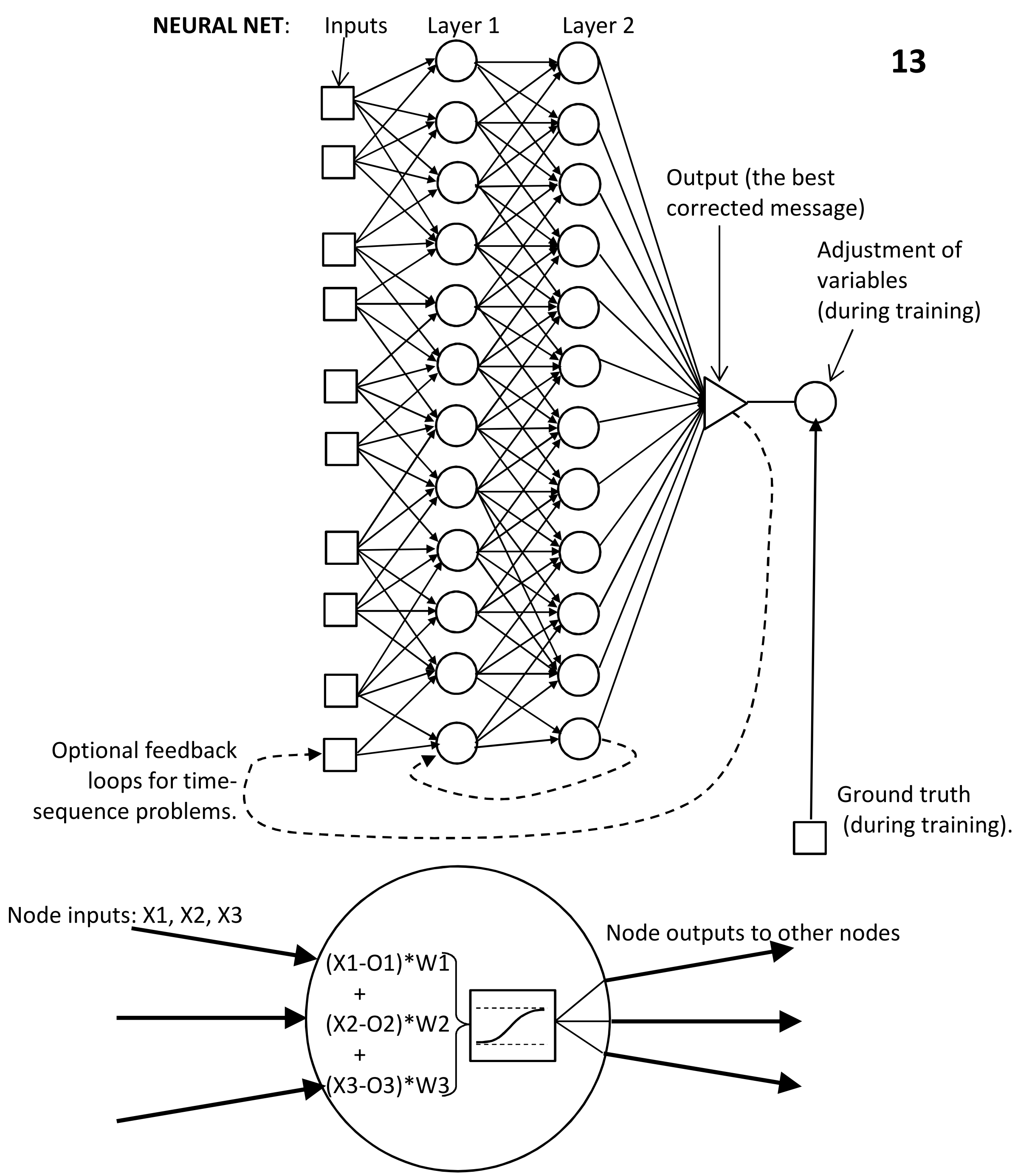


Fig. 13: A neural net. Input values are fed into a series of layers of nodes, or adjustable internal functions. A final result is then accumulated in the output node (triangle). During training, the output is compared to the known correct value (such as the unfaulted version of a message), and the node parameters are adjusted until the predictions become accurate. A single node is also shown expanded, with node-inputs X going into a summation function, followed by a limiting function, with multiple node-outputs leading to other nodes. The variables are the offsets O and the weighting W which are adjusted for each X based on a large number of examples, for optimal correlation with the best predictions. When fully trained, the AI model can then predict which message elements are faulted and their corrected values.

Figure 14 shows the input values and various output options of an AI model trained to recognize faulted message elements and, when possible, to determine the corrected values. Not all input data need be present, and not all output options need be implemented, in each actual implementation. The AI model may also be configured to determine the likely intent or meaning of the received message, despite having one or more faulted message elements, based on format constraints, prior messages of the same type, what would make sense in the current context, and other subjective and inductive factors which the AI model can be trained to exploit. With such an AI model, the receiver can recover faulted messages without a costly retransmission.

14

Input data to the AI model:

-- Amplitude and phase fluctuations - width and skew.
-- Amplitude and phase modulation deviations relative to nominal levels.
-- Amplitude and phase deviations relative to average of like-modulated values.
-- Received amplitude or power (average over the message element).
-- Noise and interference (from demodulation reference and blanks).
-- FEC or CRC or parity code if provided.
-- Demodulation reference(s) used for demodulation of this message.
-- Polarization of each message element.
-- Frequency offset of each message element.
-- Smoothness of transitions between symbols.
-- Historical record of similar messages.
-- Type and format requirements.
-- Other rules and limits governing content of the message or waveform.

↓

Operate the AI model.

↓

AI model outputs:

-- Fault probability of each message element.
-- The most likely corrected message version.

OPTIONAL:
-- Uncertainty of each prediction.
-- Interpretation of most probable message meaning, and its uncertainty.
-- Tabulation of multiple candidate message versions, along with probability estimates for each version and for each message element of each version.

Fig. 14: Input and output parameters of an AI model trained to detect faulted message elements and optionally to determine the most likely corrected value of the faulted message elements.

Conclusion

Wireless message reception reliability is a key factor, perhaps THE key factor, for successfully expanding the available bandwidth into higher frequency regimes, since the other technical challenges appear to be solvable in principle. In this paper, we offer new systems and methods enabling wireless receivers to provide stand-alone, automatic fault mitigation. The receivers can analyze waveform data to identify faulted message elements and to correct them, without a retransmission, in a time short compared to symbol-times and retransmission times. Reception of the next message can therefore proceed without interruption. The received signal of any wireless message is rich with information about each of faulted and unfaulted message elements. This information is generally ignored in prior-art systems but will be essential for success in the coming high-frequency epoch. Base stations and access points can employ the methods to improve uplink reliability, while user devices can obtain improved downlink and sidelink communications at little or no incremental cost.

We are confident that progress in the next-generation FR-2 communications will not be blocked by persistent message faulting, because the systems and methods presented herein show how to locate and correct faulted message elements in real-time, using electronics already available to nearly all wireless receivers. The benefits of faster, better communication will become available to all, in the coming years.

Glossary

ADC	Analog-to-Digital Converter
AI	Artificial Intelligence
ARQ	Automatic Repeat reQuest
CRC	Cyclic Redundancy Code
FEC	Forward Error Correction
FR-2	Frequency Range 2, frequencies 24.25-52.6 GHz.
HARQ	Hybrid Automatic Repeat reQuest
ML	Machine Learning
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
3GPP	Third Generation Partnership Project

A "message element" is a modulated resource element of a wireless message.
A "node" is an internal function, containing adjustable parameters, of an AI model.
"Ground truth" is the known correct answer, such as the unfaulted message, in AI model training.
"Modulation Deviation" is the difference between a transmitted and received message element.
"Deviation from Average" is the difference between a message element and an average of the message.
"I-branch and Q-branch" are orthogonal components of a QAM-modulated message element.
"Amplitude Skew" is the frequency of maximum amplitude minus the subcarrier frequency.
"Frequency Offset" is the received frequency of a message element minus the subcarrier frequency.
"Signal Quality" is a combination of multiple signal diagnostics, inversely related to fault probability.
A "neural net" is an AI model in which layers of linked adjustable functions are combined in an output.

References
[1] Patents on message fault mitigation can be found at www.UltraLogic6G.com.

US Patent	Title
11,387,935	Error Detection and Correction by Modulation Quality in 5G/6G
11,398,876	Error Detection and Correction in 5G/6G Pulse-Amplitude Modulation
11,405,131	AI-Based Error Detection and Correction in 5G/6G Messaging
11,411,795	Artificial-Intelligence Error Mitigation in 5G/6G Messaging
11,418,281	Error Correction by Merging Copies of 5G/6G Messages
11,456,821	Retransmission of Selected PAM-Modulated Message Portions in 5G/6G
11,463,296	Error Correction by Merging Copies of PAM-Modulated 5G/6G Messages
11,469,856	Retransmission of Selected Message Portions in 5G/6G
11,516,065	Identifying Specific Faults in 5G/6G Messages by Modulation Quality
11,522,636	Modulation Quality and Fault Mitigation in 5G/6G
11,522,637	Selection of Faulted Message Elements by Modulation Quality in 5G/6G
11,522,638	Artificial Intelligence Fault Localization in 5G and 6G Messages
11,522,745	Identification and Mitigation of Message Faults in 5G and 6G Communications
11,546,201	Selection of Message Elements based on Modulation Quality in 5G and 6G
11,563,515	Fault Recovery by Selection based on Modulation Quality in 5G/6G
11,616,679	Detection and Mitigation of 5G/6G Message Faults
11,695,612	Method to Locate Faulted Message Elements Using AI in 5G and 6G
11,736,332	Enhanced Fault Correction and Noise Avoidance in 5G/6G Networking
11,770,207	Method for Mitigating Branch-Amplitude Faults in 5G and 6G Messages
11,770,209	Signal Quality Input for Error-Detection Codes in 5G and 6G
11,784,764	Artificial Intelligence for Fault Localization and Mitigation in 5G/6G
11,799,585	Error Correction in 5G and 6G using AI-Based Analog-Digital Correlations
11,811,579	Recovery of Corrupted 5G/6G Messages by Modulation Quality
11,817,950	AI Means for Mitigating Faulted Message Elements in 5G/6G
11,824,667	Waveform Indicators for Fault Localization in 5G and 6G Messages
11,843,468	Fault Detection, Localization, and Correction by 5G/6G Signal Quality
11,848,774	AI-Based Analog-Digital Fault Detection and Localization in 5G/6G
11,848,788	AI-Based Waveform Analysis for Fault Localization in 5G and 6G
11,849,349	Fault Mitigation Using Signal Quality and Error-Detection Codes in 5G/6G