PHASE NOISE MITIGATION

Systems and Methods for 5G/6G Receivers to Measure and Correct Phase Faults at High Frequencies

David E. Newman and R. Kemp Massengill

Executive Summary

A key strategy for enhancing wireless throughput in 5G-Advanced and especially 6G is to expand into increasingly higher frequency ranges, which will enable wider bandwidths and higher information density in the transmitted waveforms. However, phase-noise message faulting becomes predominant at higher frequencies due to clock drift in the transmitter and receiver, and interference in heavily congested networks. Multiple innovative methods will be needed to enable sufficient phase-noise mitigation before reliable communications are achieved in the planned FR2 frequency range. Presented herein is an advanced modulation and noise-cancellation demodulation scheme that eliminates the narrow phase margins of QAM. Also shown are improved demodulation references that provide both amplitude calibration and phase correction, and new phase-tracking signals and methods that provide non-CPE phase compensation in real-time, at little or no cost in transmission power and resource usage.

Phase Noise and Signal Waveforms

Wireless messages are conventionally modulated in QAM (quadrature amplitude modulation) because it makes signal processing easy for the receiver. However, QAM has three major disadvantages for high-frequency communications. First, the phase noise margins are quite limited, relative to other modulation schemes, resulting in phase faults. Second, the number of amplitude modulation levels is lower than many alternatives, limiting throughput. Third, QAM cannot support non-square modulation charts, greatly limiting the ability to counter observed fault types. Forward error-correction (FEC) bits added to the message often fail to correct the problem, and greatly inflate the message size.

Figure 1 shows the constellation chart of 16QAM. Points indicate the 16 modulation states, representing multiplexed orthogonal I and Q branches. Each branch is a sinusoidal component of the waveform signal, with a branch amplitude of ±1 or ±3 arbitrary units. The central cross shape is zero amplitude. Negative amplitudes represent 180-degree phase reversal. A particular state in the upper right, for example, has the maximum positive branch amplitude in both I and Q branches (arrow).

In transit, however, each indicated state propagates not as a pair of branch signals, but rather as a single sine wave whose amplitude is given by the length of the arrow ($3\sqrt{2}$ arb. units), and the phase as indicated by the arc (45 degrees). Thus the transmitted waveform incorporates both branches in a single phased waveform.

1

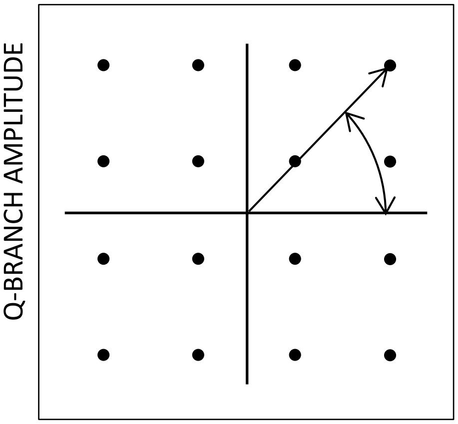


Figure 2 shows the effect of phase noise on a QAM transmission. Points again show the transmitted states, while the gray blobs show the as-received distributions, smeared out by phase noise. As long as the spread is less than the phase separation between states, the receiver can still demodulate the message correctly. But in 16QAM, unfortunately, several of the states are separated from their neighbors by only 36.9 degrees, as indicated by the arc. Also note that the modulation states lie on just three amplitude circles, which limits the ability of the network to maintain throughput. In the depicted case, the phase noise is almost large enough to cause overlap. At slightly higher frequencies, the phase spreading would be even worse, resulting in numerous message phase faults.

2

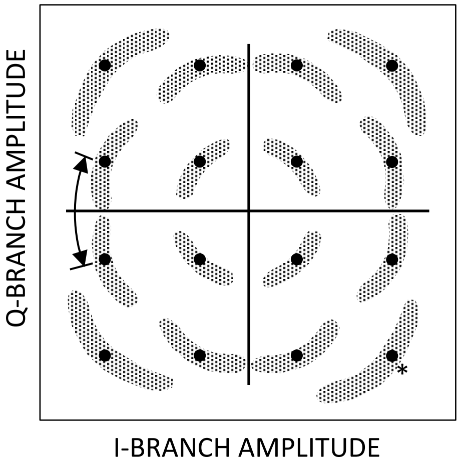
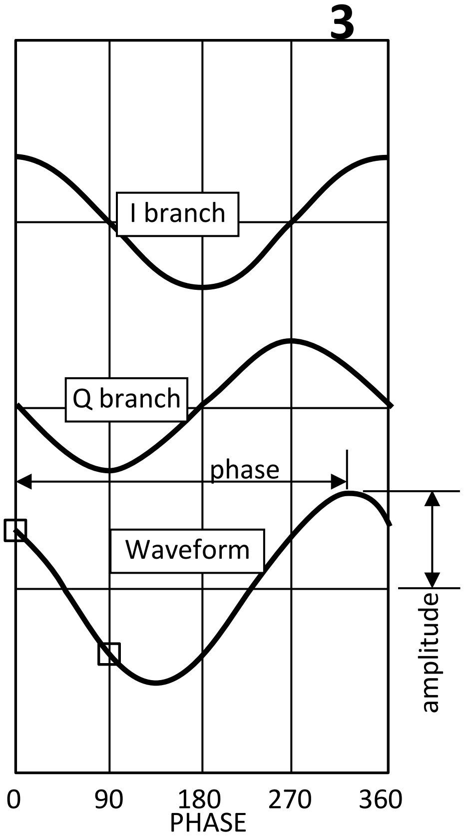


Figure 3 shows how a receiver can process a QAM signal, by separating the received waveform into its orthogonal I and Q branch components. The I branch amplitude is indicated by the first small box, which is where the Q branch passes through zero. Likewise, the Q branch amplitude is where the I branch is zero, as indicated by the second box. In this case, the amplitude of the I branch is the maximum positive branch amplitude, and the amplitude of the Q branch is the maximum negative branch amplitude. Hence this signal corresponds to the modulation state in the lower right corner of Fig. 2, marked with an asterisk. Together, the I and Q branches comprise the received waveform. The phase of the received waveform (measured from its positive crest) is about 315 degrees in this case, and its amplitude equals $\sqrt{2}$ times the maximum branch amplitude, as indicated.

3



The receiver can reconstruct the as-received waveform from the measured I and Q branch amplitudes using common algorithms. For example, the waveform amplitude equals $\sqrt{(I^2+Q^2)}$, the square root of the sum of the branch amplitudes squared. The waveform phase equals the arctangent of Q/I. The receiver can also calculate the branch amplitudes from the received amplitude and phase, specifically the received amplitude times the sine and cosine of the received phase. Thus the receiver can determine the waveform amplitude and phase either by separating the I and Q branches first, and then calculating the waveform amplitude and phase from the I and Q amplitudes, or alternatively by digitizing the full waveform as it arrives. In either case, the receiver can determine the amplitude and phase of the as-received waveform.

Noise and interference do not attack the I and Q branches directly; they attack the transmitted waveform as a whole. Any effects on the branch amplitudes are derivative. Hence, mitigation efforts can be focused on the phase and amplitude of the overall waveform, instead of the artificially-derived branch amplitudes.

Increasing Phase Margins in QAM

4

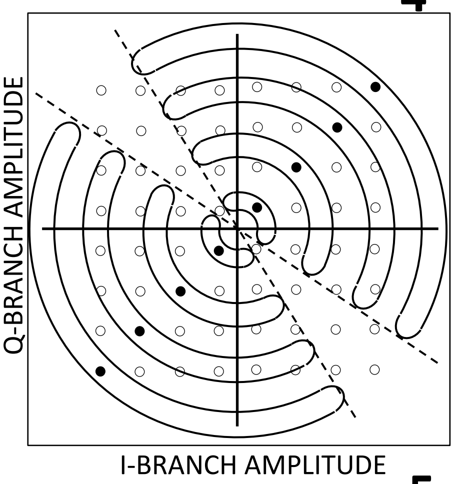


Figure 4 shows a modified QAM scheme for combating phase noise. Only eight modulation states are actually used (solid points with I=Q). The phase acceptance regions are shown as loops. The unused 16QAM states are shown open. This scheme provides a full 180-degree phase margin, a substantial improvement over 16QAM, but at the cost of half the modulation states, so all messages would be twice as long. Nevertheless, in situations with excessive phase noise, the trade-off may be beneficial. As an option, the receiver can define exclusion zones (dashed) such that any message element in which the QAM signal is within one of the exclusion zones is automatically considered faulted.

5

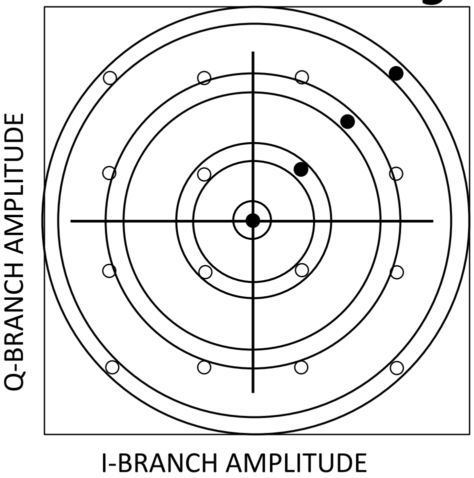


Figure 5 shows an even more noise-tolerant modulation scheme. Each state has a full 360-degree phase margin, and the acceptance regions are complete circles as shown. However, there are only four modulation states. The allowed modulation states have I=Q, positive only. Advantageously, the allowed amplitude levels are equally spaced in this example. As an option, one of the amplitude levels can be zero amplitude as shown (a blank resource element) which the receiver can recognize as one of the modulation states when detected within a message. This modulation scheme is completely insensitive to phase noise at any frequency. However, messages are four times as long as in regular 16QAM since there are only four modulation states.

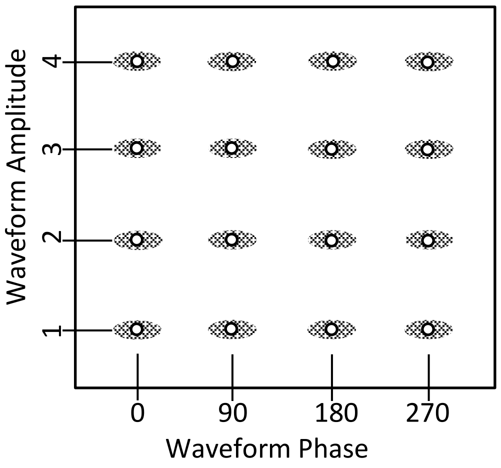
These and the previous examples show that regular 16QAM is not the ideal modulation scheme for mitigating phase noise, due to the rigid branch phase rule and the unequal separation between waveform amplitude levels. Extreme mitigation measures, such as those shown in Fig. 4 and 5, can increase phase noise margins, but they are expensive in terms of message size, latency, and throughput. Therefore, a more flexible modulation scheme may be required for high-frequency communications.

Amplitude-Phase Modulation

A multiplexed amplitude and phase modulation scheme with 16 allowed states can be compared to 16QAM. Amplitude-phase modulation includes "polar" modulation, as well as related versions that provide the versatility needed for high frequency applications, as indicated below. In particular, amplitude-phase modulation encodes information in the overall transmitted waveform amplitude and phase, instead of the I-Q components as in QAM, thereby enabling the phase dimension to be used for optimizing throughput and reliability. In addition, amplitude-phase modulation readily accommodates non-square modulation schemes, for mitigations tailored to extant fault types. Message faults are readily diagnosed according to fault types with amplitude-phase modulation, leading to logical mitigation strategies, as detailed below.

Figure 6 shows the modulation chart for amplitude-phase modulation of the transmitted waveform, with 16 states. Although this looks similar to 16QAM, the meaning is quite different. The phase levels are equally separated by a full 90 degrees (unlike QAM) thereby providing much larger, and uniform, phase margins. There are four amplitude levels instead of three, and they are equally spaced (unlike QAM). The noise distribution is shown in gray, on the same scale as Figure 2. There is no close overlap because the phase margins are much larger. In addition, the additional, equally-spaced amplitude levels can provide higher throughput, especially in congested networks.

6



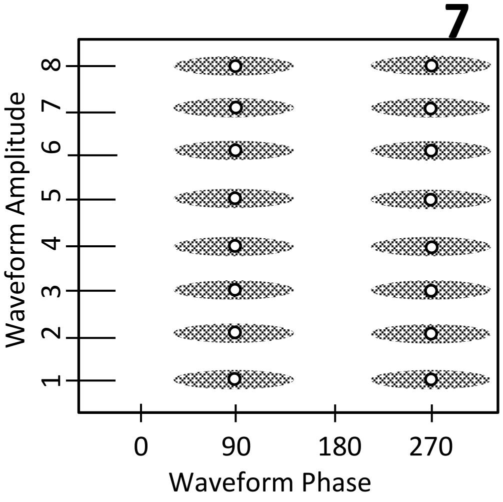
Networks can exploit amplitude-phase modulation without changing the hardware of either the transmitter or the receiver. The transmitter can modulate according to the amplitude-phase chart as shown, and transmit a waveform signal with the programmed phase, as usual. The receiver can separate the I and Q branches with normal signal processing, to determine the I and Q branch amplitudes, as usual. But then, the receiver can calculate the "sum-signal" amplitude and phase before demodulating. The calculated sum-signal reconstructs the received waveform from the I and Q amplitudes. The receiver can then compare the sum-signal amplitude and phase to a set of predetermined amplitude levels and phase levels, acquired from a demodulation reference proximate to the message, and thereby demodulate the message unambiguously despite noise. Due to the wider phase margins, phase faulting can be substantially reduced, relative to 16QAM, and higher overall throughput can be provided in crowded locations, due to the increased number of amplitude levels.

Advantages of Non-Square Modulation Tables

In high-frequency environments, phase faults are expected to dominate over amplitude faults. Therefore, both reliability and throughput can be optimized by adjusting the number and spacing of the phase and amplitude modulation levels. This option is not available to QAM because the I and Q branches are geometrically equivalent, prohibiting non-square constellations. The important point is that noise does not attack the branches directly - it alters the amplitude or phase of the received waveform, and does so in measurable ways. This can be exploited to correct the message elements in amplitude-phase modulation, but is difficult to accomplish in QAM because phase noise inevitably scrambles the I and Q branches, in addition to the narrower phase margins.

Figure 7 shows a non-square amplitude-phase modulation scheme, with two phase levels and eight amplitude levels, 16 states in all. Unlike 16QAM, however, every state has a full 180-degree phase margin. This wide phase margin easily accommodates even a greatly expanded phase noise, as illustrated here. Eight amplitude levels are feasible so long as a demodulation reference is provided close to the message (such as prepended to the message itself) for proximate amplitude

7



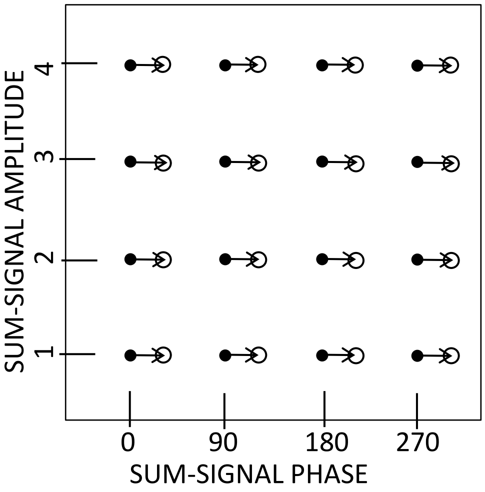
recalibration. Other non-square examples have 3 or 5 equally spaced phase levels, and as many amplitude levels as the noise environment allows.

Mitigation of Phase and Amplitude Noise

A receiver can subtract noise effects only if it can measure them. For example, the transmitter can provide a periodic "blank" resource element with no transmission, and the receiver can measure the noise amplitude and phase received in the blank resource element. Optionally, the transmitter can transmit a predetermined amplitude and phase signal in a second resource element, which the receiver can use to calibrate the amplitude and phase levels of the modulation scheme.

Figure 8 shows how CPE phase noise can alter the received signal of every modulation state in an amplitude-phase modulation scheme. CPE is due to clock drift, so every modulation state is altered by the same phase rotation angle. If the message uses amplitude-phase modulation, the receiver can readily mitigate CPE phase noise by measuring the noise waveform in the blank resource element, determine the phase rotation angle, and then rotate each message element phase in the same symbol-time back by the phase rotation angle.

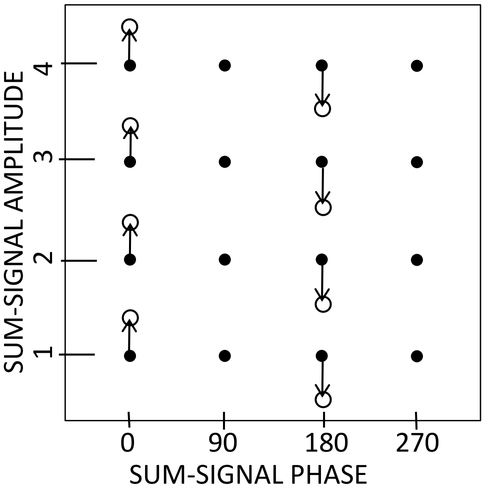
8



Non-CPE phase noise is due to interference and other sources that affect each subcarrier and each symbol-time, especially in crowded networks. In that case, the best defense against phase noise is to include lots of embedded single-branch demodulation signals in the message (see below), and to use a modulation scheme that provides the widest phase margins achievable.

Figure 9 shows how a receiver can negate additive amplitude noise in messages using amplitude-phase modulation. Based on interference between the noise signal detected in one or more blank resource elements proximate to the message, the receiver can determine the proper amplitude correction for each waveform phase of the modulation scheme separately, and apply the corrections to the message element signals before demodulation. In this model, noise signals are treated as synchronous with the subcarrier, but with an arbitrary amplitude and shifted in phase by a random amount. The receiver determines how the current noise interferes with each modulation state, and then applies a positive or negative amplitude correction as shown, for each message element. For example, at 0 degree phase the noise interferes constructively, and at 180 degrees it interferes destructively as shown. The receiver can apply the reverse adjustment after measuring the amplitude and phase of the noise signal proximate to the message.

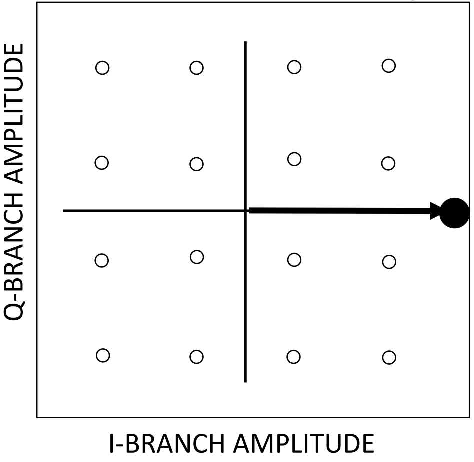
9



In QAM modulated messages, the default response is HARQ, automatically requesting a retransmission of the message or its FEC bits when the message is corrupted. This unavoidable limitation of QAM is due to the ambiguity introduced when the I and Q branches are scrambled by time/frequency-dependent noise. With amplitude-phase modulation, on the other hand, the receiver can measure and correct for noise as shown, often avoiding a retransmission entirely.

Single-Branch Reference Signal

10A



A new compact demodulation reference is presented, which provides both phase-tracking and amplitude calibration, while occupying only one resource element. The new demodulation references can be placed proximate to each message or embedded throughout the message, at low cost. Figure 10A shows such a signal, in the context of a QAM modulation scheme. The transmitter emits a high amplitude signal in the I branch and zero transmission in the Q branch (or vice-versa), termed a "single-branch reference" signal. The I branch in this case is transmitted at √2 times the maximum branch amplitude. The receiver then measures any received signal in the Q branch, thereby determining the phase rotation. Because the Q amplitude is nominally zero, very high sensitivity and precise correction can be obtained.

10B

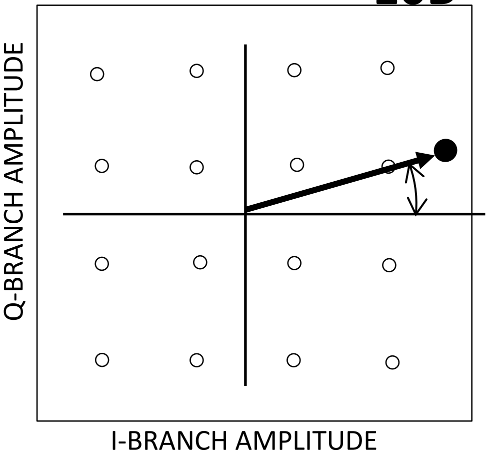


Figure 10B shows the single-branch reference signal as-received including phase noise. The phase is rotated, resulting in energy in the Q branch, slightly reduced in the I branch. The receiver can detect the non-zero Q branch amplitude with high sensitivity, and quantify the phase rotation angle, as shown by the arc. (In other versions, the I branch is zero and Q is powered.) The receiver can also recalibrate the amplitude scale according to the magnitude of the sum-signal amplitude. Since most receivers already determine the I and Q branch amplitudes natively, the phase noise and amplitude offset can be quantified during signal processing instead of post-processing adjustments, enabling real-time mitigation. The single-branch references are not confused with regular data, because none of the modulation states has zero or near-zero power in one branch, and none have zero or near-zero degrees overall phase.

Base stations can provide the single-branch references at predetermined locations on the resource grid throughout the downlink and unscheduled symbol-times. The base station and the user devices can also provide single-branch references at the beginning of each uplink and downlink message, and/or embedded within each message at various intervals, thereby enabling the receiver to track any rapidly changing or frequency-dependent noise conditions. For example, interference in dense networks, and weak signals due to propagation losses, may be mitigated without a retransmission by correcting the message elements in real-time according to the single-branch references.

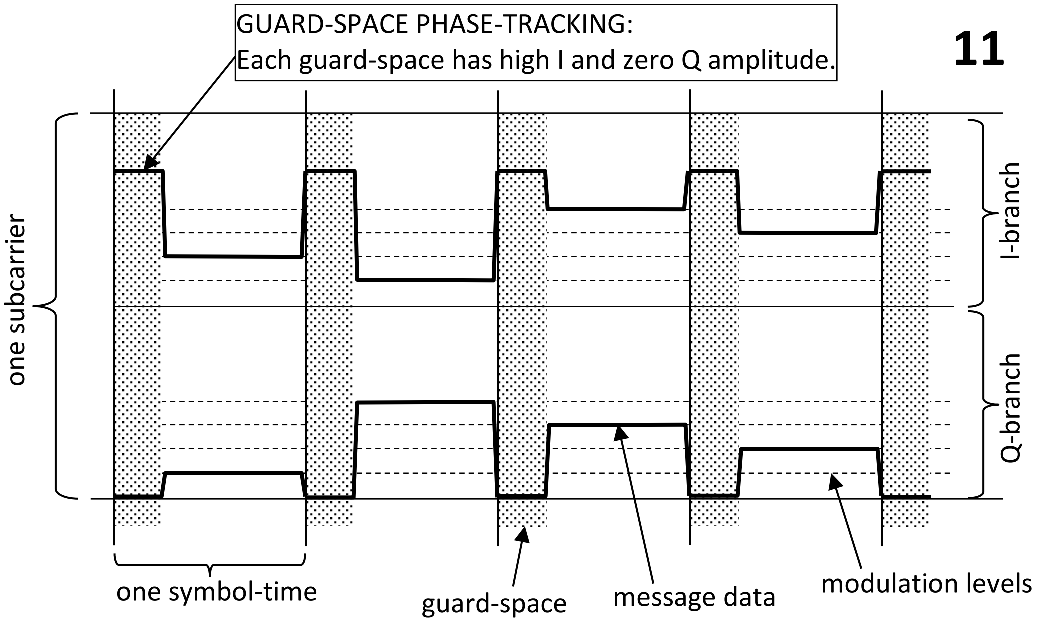
As a further option, message data can be modulated in the amplitude of the single-branch references instead of a predetermined amplitude. Since every single-branch reference would then carry message data in one modulation parameter, it would then cost only one-half of a resource element, and the message length would then be increased only slightly, while still providing a precision phase-tracking and calibration relative to the amplitude levels. When embedded at numerous places throughout the message, single-branch phase-tracking signals thereby enable mitigation of non-CPE noise and interference. Since the tiny single-branch signals consume little or no additional transmission power, they can enhance the reliability of high frequency communications at negligible cost.

Guard-Space Phase-Tracking Signals

For even higher frequencies in FR-2 and non-CPE environments, multiple demodulation references will be needed in and around each message. To achieve sufficient throughput in crowded network locations, the embedded references must be configured to occupy as little resource territory as possible. An advantageous way to meet these conflicting pressures is a single-branch reference signal inserted in the guard-space of each resource element. The message is thereby saturated with phase-tracking and amplitude-tracking signals, enabling extremely localized noise calibration and mitigation.

Each symbol-time of a message is conventionally divided into a guard-space followed by the message data. The guard-space generally holds a "cyclic prefix" which is a copy of the last portion of the message data. The cyclic prefix helps a base station accommodate a range of message arrival times, provides inter-symbol isolation, and provides "circularity" which simplifies certain signal processing tasks. The single-branch guard-space references can provide all the same features, with slight adjustment of procedures. User devices should be required to adjust their transmission times for proper synchrony with the base station upon initial access, thereby obviating the need for cyclic prefix to accommodate sloppy timing. Secondly, the guard-space reference provides the same inter-symbol isolation as a cyclic prefix of the same size. And circularity can be provided by simply widening the transform window to include the guard-space of the following symbol-time, since they are both identical.

Figure 11 shows an example of guard-space phase-tracking signals. In each guard-space of each symbol-time of each subcarrier, the transmitter emits an I-branch with amplitude √2 times the maximum branch amplitude level of the QAM scheme, and a Q-branch with zero amplitude (that is, a waveform with zero degrees phase and a predetermined amplitude). The receiver can measure the received amplitude in the Q-branch in the guard-space, and thereby determine the phase rotation angle with high sensitivity, as well as an amplitude calibration according to the square root of the sum of the I and Q amplitudes squared. The receiver can then correct each message element in amplitude and phase (or, if preferred, adjust the modulation levels to include the effects of noise). Importantly, both amplitude calibration and phase-noise mitigation are provided, for each message element, costing zero resource elements and zero additional transmission power.



Mitigating Faults According to Type

A network can respond to persistent message faults proactively, by determining the primary fault type and switching to another modulation scheme that ameliorates that type of fault. For example, the network (or a user device cooperating with the network) can determine whether amplitude faults or phase faults are more prevalent in received messages, and then recommend switching to a second modulation scheme that provides wider noise margins in the problematic parameter. In many cases, the network can compensate by adding more levels in the non-faulting modulation parameter, to keep the throughput nearly unchanged. For example, if phase faults are excessive, the network can change to a second scheme with fewer phase levels and wider phase margins to prevent the phase faulting, and add one or more additional amplitude levels to compensate, but only if the rate of amplitude faulting is not high.

A receiver can diagnose fault types by comparing a faulted message with a subsequent non-faulted version, noting which message elements differ and how they differ. The receiver can tally adjacent-amplitude, adjacent-phase, or non-adjacent faults, wherein an adjacent-phase fault is a message element shifted by just one phase level, and likewise for adjacent-amplitude. A non-adjacent fault is off by multiple amplitude or phase levels. If adjacent-amplitude faults are dominant, the receiver can switch to a modulation scheme with fewer amplitude levels spaced farther apart. The receiver may also request more phase levels if the phase faulting is not too bad. On the other hand, if the dominant fault type is non-adjacent, the faulting is probably due to strong bursty noise, in which case adjusting the noise margins probably won't work. A better solution may be to increase the number of amplitude and phase levels, to shorten the messages in an attempt to avoid the noisy intervals. If that doesn't work, try increasing the transmission power, or change frequency bands.

Figure 12 shows how a receiver can diagnose message faults by comparing faulted message elements with a subsequent uncorrupted version, using amplitude-phase modulation. If the sum-signal amplitude of the faulted and unfaulted message elements differ by one amplitude level, it is an adjacent-amplitude fault, and likewise for adjacent-phase faults (solid arrows). In a non-adjacent fault, the sum-signal is off by two or more levels (dashed arrow).

If a corrected copy of the message is not available, the receiver may attempt to diagnose the faults by measuring how far each message element's sum-signal amplitude and phase differ from the closest predetermined modulation state of the modulation scheme. The resulting distribution (such as Figure 6) reveals problems related to insufficient noise margins. The receiver can then tally the amplitude and phase deviations to determine the dominant fault type, and select a better modulation scheme.

12

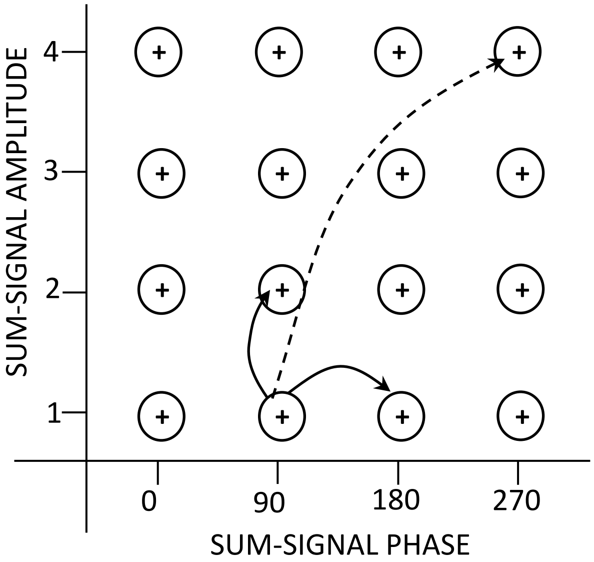


Figure 13 shows how a receiver can determine whether adjacent-amplitude, adjacent-phase, or non-adjacent faulting is present in a message. Each message element is indicated according to the modulation level of its sum-signal amplitude and phase, such as "A2-P3" indicating the second amplitude level multiplexed with the third phase level. The faulted message has four faults in this example. By noting whether the two versions differ in amplitude or phase, the receiver can tally the various fault types. For example, the first fault is an adjacent-amplitude fault because the sum-signal amplitude changed from the correct value of A4 to the faulted value of A3. Continuing in this manner, the receiver can build up tallies of the various fault types, and then decide what to do.

13

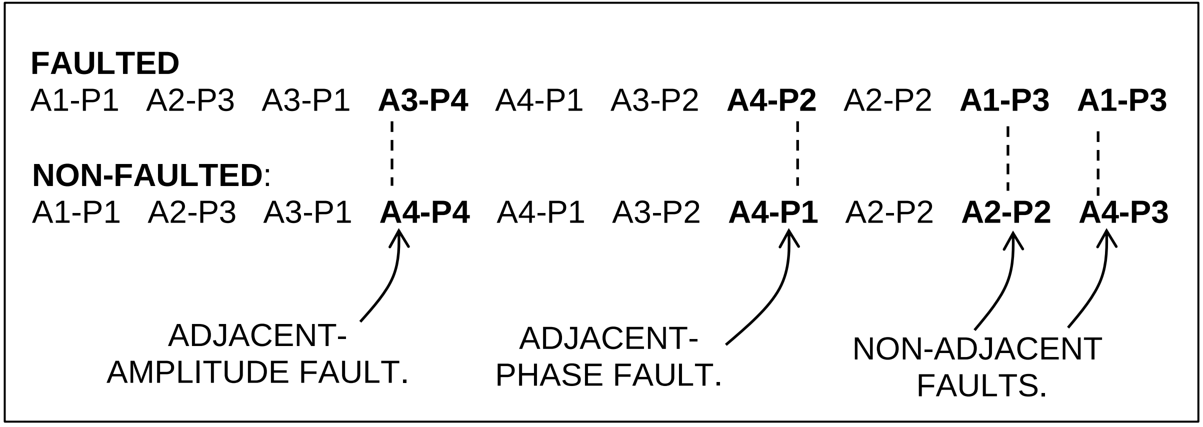


Figure 14 shows how a receiver can respond to message faults of various types. As mentioned, a receiver can mitigate adjacent-amplitude faults by changing to a modulation scheme with fewer, spaced-apart amplitude levels, and adjacent-phase faults by reducing the number of phase levels. To maintain high throughput, it may be possible to increase the number of levels in the opposite parameter. If both adjacent-amplitude and adjacent-phase faults are excessive, it may be necessary to temporarily switch to a slower modulation scheme for sufficient reliability. If non-adjacent faults predominate, due to strong bursty noise or interference, then it may be advantageous to increase the number of amplitude and phase levels, thereby reducing the size of messages, in an attempt to evade the noise episodes. If this still doesn't work, it may be necessary to increase transmission power levels temporarily, although the competing transmitter may do the same, or switch back to FR1 until the noise subsides.

Network operators need to know whether current message fault rates are due to amplitude or phase noise, and whether non-adjacent faulting is present, since these factors influence the optimal response strategies.

14

FAULT PATTERN		MITIGATION OPTIONS
ADJACENT-AMPLITUDE	→	REDUCE NUMBER OF AMPLITUDE LEVELS.
ADJACENT-PHASE	→	REDUCE NUMBER OF PHASE LEVELS.
NON-ADJACENT FAULTS	→	INCREASE NUMBER OF AMPLITUDE AND PHASE LEVELS, INCREASE POWER.

Conclusion

Phase noise is not an insurmountable problem, even at FR2 frequencies. Detailed herein are improved methods and modulation procedures that will enable 5G-Advanced and 6G to reach the speed and volume demanded by next-generation users. Multiplexed amplitude-phase modulation can provide a reliable 90 degrees of phase margin, which is more than twice that of corresponding QAM, while matching the speed, throughput, information density, and ease of reception of QAM. In addition, compact and versatile single-branch phase-tracking demodulation references can provide localized phase and amplitude calibration. The base station can provide single-branch references and blank resource elements at specific times for CPE negation and general background estimation. Further single-branch reference signals may be embedded within each user message for uplink noise mitigation, at negligible resource cost. Receivers can readily diagnose fault types according to the amplitude or phase deviation relative to the unfaulted message element, and can then apply appropriate modulation changes to improve reliability thereafter.

These innovations can enable wireless network operators to improve operational efficiency by minimizing message faulting, avoiding unnecessary retransmissions, and enabling greater geographical range despite pathloss attenuation and network crowding. We encourage 3GPP and other standards organizations to include these solution options in future releases. Users will be grateful for the improved reception reliability and speed.

Glossary

FR1 and FR2 are frequency ranges. FR1 is 7.125 GHz and below. FR2 is 24.25 GHz and up.

CPE (Common Phase Error) refers to phase noise caused by clock drift, resulting in the same phase rotation for all subcarriers in a particular symbol-time or OFDM symbol.

FEC (Forward Error-Correction) bits, added to a message or subsequently transmitted, may sometimes help correct faults, but at substantial resource cost and delay.

HARQ (Hybrid Automatic Repeat Request) is a procedure for requesting a retransmission of a faulted message instead of fixing it.

OFDM (Orthogonal Frequency-Division Multiplexing) refers to a combined signal from many message elements, all at the same time and slightly different frequencies, superposed on a single complex waveform. The receiver can separate the individual subcarrier components, and demodulate to read the message data.

QAM (Quadrature Amplitude Modulation) is a modulation scheme in which the received waveform is broken into two orthogonal I and Q branch signals, each branch separately amplitude modulated.

A "message element" is a modulated resource element of a wireless message.

"Amplitude-phase modulation" includes any modulation scheme in which message data is encoded directly in the modulated amplitude and phase of the overall waveform transmitted.

An "overall waveform" is the as-transmitted sine wave, characterized by a transmitted amplitude and phase.

A "received waveform" is the as-received sine wave, characterized by a received amplitude and phase, which may be altered from the transmitted wave by amplitude noise and phase noise.

"Non-square modulation" is a modulation scheme involving two multiplexed modulation parameters, in which the two parameters have different numbers of modulation levels.

"Interference" occurs when a received signal is the sum of two different transmitted signals.

A "sum-signal" is a calculated amplitude and phase, derived from the I and Q amplitude values.

A "sum-signal amplitude" is the square-root of a sum of the I and Q amplitudes squared.

A "sum-signal phase" is the arctangent of the Q/I amplitude ratio.

3GPP (Third Generation Partnership Project) is the primary wireless standards organization.

References
[1] Patents on phase noise mitigation can be found at www.UltraLogic6G.com.

US Patent	Title
11,558,236	Single-Branch Reference for High-Frequency Phase Tracking in 5G and 6G
11,601,150	Demodulation for Phase-Noise Mitigation in 5G and 6G
11,637,649	Phase-Noise Mitigation at High Frequencies in 5G and 6G
11,671,305	Extremely Compact Phase-Tracking 5G/6G Reference Signal
11,736,320	Multiplexed Amplitude-Phase Modulation for 5G/6G Noise Mitigation
11,777,547	Phase-Tracking Demodulation Reference and Procedure for 5G and 6G
11,777,639	How to Maximize Phase-Noise Margins in 5G and 6G
11,799,707	Guard-Space Phase-Tracking Reference Signal for 5G and 6G Networking
11,811,565	Demodulation Using Two Modulation Schemes in 5G and 6G
11,996,973	Scheduling Single-Branch Phase-Tracking References in 5G and 6G