1. Introduction

The JPEG 2000 family of standards provides a large set of capabilities that can be used to address the needs of almost any image compression task. The complete set of capabilities, including resolution scalable, quality scalable and region accessible coding of images with up to 16384 components, up to 37 bits per sample and dimensions up to $4 \times 10^9$, is addressed by the finely embedded block coding algorithm defined in Rec. ITU-T T.800 | ISO/IEC 15444-1, identified here as the J2K1 block coder. The high degree of quality scalability offered by the J2K1 block coder enables highly responsive remote interactive browsing of large image sources over low bandwidth lossy or lossless communication channels, which can be achieved using the technology defined in Rec. ITU-T T.808 | ISO/IEC 15444-9 (a.k.a. JPIP). For applications that require very high processing throughput, the “high throughput” (HT) block coding algorithm defined in Rec. ITU-T T.814 | ISO/IEC 15444-15 is particularly beneficial, allowing overall encoder and decoder throughput increases on the order of 10:1 on common platforms.

Codestreams that use the HT block coding algorithm are known as HTJ2K codestreams, as defined in Rec. ITU-T T.814 | ISO/IEC 15444-15. Moreover, it is possible to use either the HT or the J2K1 block coding algorithm within an HTJ2K codestream, and the choice of block coder can even change on a block-by-block basis. However, for most applications it is desirable to use a restricted class of HTJ2K codestreams, known as HTONLY codestreams, for which a decoder can be sure of encountering only encountering HT block bit-streams.

While JPEG 2000 offers two different block coding algorithms, namely HT and J2K1, their coded representations are highly compatible. Both algorithms involve a set of coding passes that progressively refine the coded representation of a block of subband samples, at a rate of one magnitude bit-plane per three passes. In both cases, the coding passes are known as Cleanup, Significance Propagation (abbreviated as SigProp here) and Magnitude Refinement (abbreviated as MagRef here). In fact, it is always possible to reversibly transcode any J2K1 block bit-stream into a corresponding HT block bit-stream and vice-versa. Although HT block bit-streams are not fully embedded, so that the quality scalability feature of JPEG 2000 is largely sacrificed in exchange for speed, this reversible transcoding feature allows the full quality scalability associated with a J2K1-based codestream to be recovered on demand from an HT-based codestream.
JPEG 2000 codestreams are organised as a sequence of “J2K packets,” not to be confused with network transport packets. The organization of code-blocks into precincts, precincts into a sequence of $Q$ packets corresponding to $Q$ global quality layers, and packets into coding pass contributions that are encoded within a packet header are all identical for both J2K1- and HT-based codestreams. Nonetheless, the construction and parsing of packet headers for HTJ2K codestreams involves some specific considerations, notably “placeholder passes” and “multiple HT Sets.” These concepts are explained further in Section 4.

The primary purpose of this document is to provide guidelines for HTJ2K decoder implementations, in regard to the parsing of packet headers and the use of placeholder passes and multiple HT Sets, to the extent that they are encountered. A second objective of this document is to explain how placeholder passes and multiple HT Sets can be used to preserve quality layer boundaries while transcoding between J2K1- and HT-based codestreams. Finally, this document provides guidance for HTJ2K encoders in regard to the use of placeholder passes and multiple HT Sets.

For simplicity, this document addresses only the case of HTONLY codestreams, for which a decoder can be sure of encountering only HT block bit-streams, as explained earlier. Another important class of HTJ2K codestreams is known as SINGLEHT, where a SINGLEHT codestream provides at most one HT Set for each code-block. The HTONLY and SINGLEHT restrictions are advertised in the CAP marker segment near the start of an HTJ2K codestream, so that decoders can assess their ability to process all content within the codestream. A decoder that supports only codestreams belonging to both the HTONLY and SINGLEHT classes still needs to address the possibility that placeholder passes are encountered within packet headers, so the guidelines in this document are highly relevant even in this simplified setting.

2. Notation used in this document

- $M_b$ is the maximum number of bit-planes in a given code-block, as used in Rec. ITU-T T.800 | ISO/IEC 15444-1
- $P_b$ is the number of missing bit-planes for code-block $b$, which is the same as $P$ from Annex B of Rec. ITU-T T.814 | ISO/IEC 15444-15 and also the same as the number of zero bit-planes from Annex B of Rec. ITU-T T.800 | ISO/IEC 15444-1
- $P_b^{\text{phld}}$ is the same as the symbol P0 from Annex B of Rec. ITU-T T.814 | ISO/IEC 15444-15
- $T_b$ is the same as the set T from Annex B of Rec. ITU-T T.814 | ISO/IEC 15444-15
- $Q$ is the number of quality layers in a codestream, which is also the number of packets for each precinct
• \( W_b^q \) is the total number of coding passes contributed to code-block \( b \) from packet \( q \)
• \( C_b^q \) is the number of distinguishable contributions to code-block \( b \) from packet \( q \)
• \( C_b \) is the total number of distinguishable contributions to code-block \( b \) from all packets
• \( c_q \) is the index of the first contribution to code-block \( b \) within packet \( q \), if there is one
• \( Z_b^c \) is the number of coding passes associated with contribution \( c \)
• \( L_b^c \) is the number of bytes associated with contribution \( c \)

3. J2K packet header parsing in general

A JPEG 2000 codestream consists of a sequence of J2K packets, prepended by a codestream main header and a first tile-part header, and optionally interspersed with additional tile-part headers. A codestream with \( Q \) quality layers has \( Q \) packets for each precinct, although a decoder might need to be prepared to handle truncated codestreams, and each precinct is a defined collection of code-blocks from a single tile-component-resolution – i.e., one resolution level within one component of a tile.

Conceptually at least, the packet headers of a precinct are parsed sequentially until all \( Q \) packets have been parsed or there are no more packets available for the precinct – this might happen if the codestream has been truncated. Following this sequence a packet header parser recovers the following quantities for each code-block \( b \) in the precinct, regardless of whether the codestream is HT-based on J2K1-based.

1. A sequence of \( C_b \) ordered pairs \((Z_b^c, L_b^c)\), for \( 0 \leq c < C_b \), describing a sequence of distinguishable contributions from code-block \( b \) to the compressed representation, where \( Z_b^c \) is the number of coding passes in contribution \( c \) and \( L_b^c \) is the number of bytes (length) of the contribution.
2. If \( C_b \neq 0 \), the number \( P_b \) of missing bit-planes for code-block \( b \).

This information is sufficient for the purpose of decoding and reconstructing the image; however, in some cases it is helpful to keep track of the quality layer index \( q_b^c \in \{0,1,\ldots,Q-1\} \), of the J2K packet containing contribution \( c \).

In order to recover these quantities, the packet header parser performs tag-tree decoding and follows all other procedures described in Rec. ITU-T T.800 | ISO/IEC 15444-1. When parsing the header for packet \( q \), the parser first recovers the number of new coding passes \( W_b^q \) that are contributed by the packet to code-block \( b \). This value \( W_b^q \) is the sum of the \( Z_b^c \) values associated with all contributions \( c \) to code-block \( b \) that are found in the packet; that is,

\[
W_b^q = \sum_{c : q_b^c = q} Z_b^c
\]
The packet header parser uses $W_b^q$ to determine the number of contributions $C_b^q$ to code-block $b$ that are found in the packet, decoding $C_b^q$ length values $L_b^q$. Note that the total number of contributions from all packets to code-block $b$ is

$$C_b = \sum_{q=0}^{0-1} C_b^q$$

A key step here is the determination of $C_b^q$. While the packet parsing procedure itself is generic, the method for determining $C_b^q$ depends upon a set $\mathcal{T}_b$ that describes how the bit-stream for code-block $b$ is partitioned into segments.

Specifically, $\mathcal{T}_b$ is a sequence of coding pass indices $z_b^s$, such that segment $s$ consists of coding passes $z \in (z_b^{s-1}, z_b^s)$, for each $s = 0, 1, 2, \ldots$. Here, $z_b^{-1} = 0$ is always implied, so that segment 0 consists of passes 0 to $z_b^0$. That is, $z_b^s$ is the zero-based index of the last coding pass in segment $s$.

It is worth clarifying the relationship between contributions $c$ and segments $s$ associated with a code-block bit-stream. Formally, the coding passes associated with a code-block are first partitioned into segments, as described by $\mathcal{T}_b$. Then as each packet $q$ contributes $W_b^q \geq 0$ new coding passes to code-block $b$ these coding passes are associated with one or more segments. In this way a packet might contribute only some of the coding passes associated with a segment, so that the segment is split across multiple packets. Each contribution consists of the coding passes contributed by one packet that belong to one segment. Thus, when segments are split across multiple packets, the total number of contributions $C_b$ is larger than the total number of segments.

$\mathcal{T}_b$ depends upon the block coder type (HT or J2K1) and other block coder mode flags found in COD and/or COC marker segments within the relevant main or tile-part headers. Moreover, for HT block bit-streams, $\mathcal{T}_b$ actually depends on the point at which the first non-zero length contribution $L_b^c \neq 0$ is discovered for code-block $b$. The next section develops this point in detail.

4. Placeholder passes, HT Sets and HTJ2K packet header parsing

As mentioned earlier, HT and J2K1 block coding algorithms have the same fundamental set of coding passes, identified here as Cleanup, SigProp and MagRef. For the HT block coder, these coding passes are grouped into “HT Sets,” where each HT Set commences with an HT Cleanup pass. With the possible exception of the last HT Set, each HT Set for a code-block also consists of one SigProp pass and one MagRef pass, in that order; in general, no HT Set contains more than one coding pass of each type.
While multiple HT Sets may possibly be found within the codestream for a given code-block, a decoder only needs to decode at most one HT Set for each block. By contrast, a J2K1 block decoder generally needs to be prepared to process many coding passes. This is because a single HT Cleanup pass encodes all of the information that is encoded by both the corresponding J2K1 Cleanup pass and all earlier J2K1 coding passes. Nonetheless, all coding pass boundaries for J2K1 are also notionally available in the HT case, and an HT encoder might potentially generate encoded content for each pass before deciding which ones should be included within the final codestream.

The magnitude bit-plane associated with an HT Cleanup pass that is processed by the decoder is determined by the symbol $S_{blk}$, as defined in Rec. ITU-T 814 | ISO/IEC 15444-15, while the number of coding passes to be processed is denoted by the symbol $Z_{blk}$, which is not more than 3 since only one HT Set needs to be processed. $S_{blk}$ can be understood as the number of magnitude bit-planes that are more significant than the one associated with the HT Cleanup pass, out of the total number of magnitude bit-planes $M_b$ that are available within the relevant sub-band. Thus, $S_{blk}$ can be as small as 0 and as large as $M_b - 1$. The value of $S_{blk}$ for code-block $b$ is given by

$$S_{blk} = (P_b + P_b^{phld}) + S_{skip}.$$ 

Here, $P_b$ is the number of missing bit-planes recovered by packet header parsing, as explained earlier, $P_b^{phld}$ is a number of “placeholder bit-planes,” that is also recovered by packet header parsing, and $S_{skip}$ denotes the number of initial available HT Sets that the decoder chooses to skip over.

$S_{skip}$ is necessarily zero unless a code-block contains multiple HT Sets; when there are multiple HT Sets, $S_{skip}$ can be chosen freely by the decoder, but specific guidance on how to make this choice is provided in Section 5.

From a decoding perspective, it makes no difference how the quantity $(P_b + P_b^{phld})$ is partitioned into missing bit-planes $P_b$ and placeholder bit-planes $P_b^{phld}$. However, the partition provides a mechanism for preserving all relevant structural properties from a quality scalable J2K1-based codestream. Specifically, transcoding a J2K1-based codestream to an HT-based codestream should leave $P_b^{phld} + 1$ equal to the total number of J2K1 Cleanup passes whose information is collapsed into the first (usually the only) HT Cleanup pass in the transcoded representation, as discussed further in Section 6.

The first $3P_b^{phld}$ coding passes recovered for an HT code-block by packet header parsing are known as “placeholder passes.” These placeholder passes are counted amongst the contribution pass counts $Z_b^c$. 

For an HT code-block $b$, the segment partition $T_b$ actually depends upon the number of placeholder bit-planes $P_b^{\text{phld}}$. Specifically,

$$ T_b = \{3P_b^{\text{phld}}, (3P_b^{\text{phld}} + 2), (3P_b^{\text{phld}} + 3), (3P_b^{\text{phld}} + 5), (3P_b^{\text{phld}} + 6), (3P_b^{\text{phld}} + 8), \ldots \}. $$

This is an area of potential confusion for implementors, since $P_b^{\text{phld}}$ depends on where the first non-zero length contribution occurs within the $(Z_b^e, L_b^e)$ contribution descriptors, while $T_b$ is needed to recover these contribution descriptors during packet header parsing. Nonetheless, ambiguity is avoided by one very constraint that is imposed upon an HTJ2K codestream by Rec. ITU-T.814 | ISO/IEC 15444-15; specifically:

"The packet as defined in Rec. ITU-T.800 | ISO/IEC 15444-1 that contains the first HT Cleanup coding pass for a code-block shall include only one HT Cleanup coding pass."

To see how this constraint allows a decoder to unambiguously recover the value of $P_b^{\text{phld}}$ along with all contribution descriptors $(Z_b^e, L_b^e)$, recall that while parsing the header of packet $q$, the parser first recovers the total number of coding passes $W_b^q$ contributed by the packet to code-block $b$. Then if $W_b^q \neq 0$, there must be at least one contribution $c_q$, whose length $L_b^q$ can be recovered before deciding the number of contributions $C_b^q$; here $c_q = \sum_{i=0}^{q-1} C_b^i$ is the index of the first contribution from packet $q$. Then the following cases exist:

- If $W_b^q \neq 0$ and $L_b^q = 0$ and no earlier packet has contributed any bytes to block $b$, then packet $q$ makes only $C_b^q = 1$ contribution, with $Z_b^{c_q} = W_b^q$ passes. These $Z_b^{c_q}$ passes are all placeholder passes.
- If $W_b^q \neq 0$ and $L_b^q \neq 0$ and no earlier packet has contributed any bytes to block $b$, then $L_b^q$ must be the length of the first HT Cleanup pass; therefore $\sum_{c=0}^{c_q} Z_b^c$ must be equal to $3P_b^{\text{phld}} + 1$. According to the constraint quoted above, this packet may contain only one HT Cleanup pass, and so there are only 3 possibilities:
  a. $\sum_{c=0}^{c_q-1} Z_b^c + W_b^q = 3P_b^{\text{phld}} + 1$ – then $C_b^q = 1$ and $Z_b^{c_q} = W_b^q$
  b. $\sum_{c=0}^{c_q-1} Z_b^c + W_b^q = 3P_b^{\text{phld}} + 2$ – then $C_b^q = 2$ and $(Z_b^{c_q}, Z_b^{c_q+1}) = \{W_b^q - 1, 1\}$
  c. $\sum_{c=0}^{c_q-1} Z_b^c + W_b^q = 3P_b^{\text{phld}} + 3$ – then $C_b^q = 2$ and $(Z_b^{c_q}, Z_b^{c_q+1}) = \{W_b^q - 2, 2\}$

In all three cases, the values of $P_b^{\text{phld}}$ and $C_b^q$ are unambiguously determined and if $C_b^q = 2$, the length $L_b^{c_q+1}$ of the second contribution is parsed from the J2K packet header, as described in Rec. ITU-T.800 | ISO/IEC 15444-1. If $W_b^q \neq 0$ at least one earlier packet has provided a non-zero length contribution to code-block $b$, then the segment partition $T_b$ is already known, because $P_b^{\text{phld}}$ has been recovered previously. Therefore the number of contributions $C_b^q$ from packet $q$ to code-block $b$ is readily determined, along with the $Z_b^e$ values for $c_q \leq c < c_q + C_b^q$, which
allows the lengths $L^c_b$ of these contributions also to be parsed from the packet header, following the methods defined in Rec. ITU-T T.800 | ISO/IEC 15444-1.

After a packet that contains the first HT Cleanup pass for a code-block $b$, there are no constraints on the number of coding passes that can be contributed to code-block $b$ within subsequent packets, apart from the constraint that $Z_b = \sum c Z^c_b$ cannot exceed $3(M_b - P_b - 1) + 1$, which is common to both J2K1 and HT block bit-streams. Each additional HT Cleanup pass that is contributed by subsequent packets belongs to an additional HT Set and each such additional HT Set, other than the last one, contains additional HT SigProp and HT MagRef passes.

As explained earlier, each HT Cleanup pass encodes all of the information associated with earlier coding passes for a block $b$, so that blocks with multiple HT Sets (and hence multiple Cleanup passes) could introduce substantial redundancy. However, non-initial HT Cleanup passes for a block are allowed to have zero length. When this happens, all coding passes in the corresponding HT Set are required also to have zero length, so that the HT Set is empty; such an empty HT Set can be understood as a "placeholder set," which can be used to preserve quality layer boundaries within an HT block bit-stream while reducing redundancy.

In general, placeholder passes and placeholder sets allow the mapping $q^b_z$ between each coding pass index $z$ and quality layer $q$ to be preserved for each block $b$ via the information found in packet headers, regardless of how many non-empty HT Sets a content generator chooses to include in the codestream. If at most one non-empty HT Set is included for each code-block, then there is no redundancy in the codestream, and there are also no placeholder sets for any code-block, since the first HT Set is non-empty by definition. In this case placeholder passes alone are sufficient to preserve the information associated with quality layer boundaries. What this means is that code-blocks with multiple HT Sets are only useful for preserving quality layer boundaries that lie between two non-empty HT Sets, in which case the representation necessarily contains redundancies. Nonetheless, a decoder that claims to be capable of processing HTJ2K codestreams that are not SINGLEHT codestreams needs to be prepared for the possibility that all HT Sets after the first one might be empty.

5. HTJ2K decoder guidelines

All HTJ2K decoders must be capable of handling placeholder passes that might be encountered during packet header parsing. There are no conditions under which a decoder can assume that $p^b_{b\text{phld}}$ is zero.

In the event that a decoder encounters a code-block for which multiple non-empty HT Sets are available, it is recommended that the decoder process the last such HT Set, in preference
to an earlier one. This is because later HT Sets correspond to higher precision representations of the code-block sample data, except in the case of placeholder sets of course.

There might, however, be some conditions under which a decoder chooses to process a lower precision HT Set; for example:

- If the decoder is unable to process a higher precision HT Set, due to implementation precision limitations, it might choose to process a lower precision non-empty HT Set for the same code-block, if one is available.
- If the codestream is received progressively, a decoder might first reconstruct a lower quality representation of the code-block based on any available non-empty HT Sets and later refresh the decoded result using higher precision non-empty HT Sets, if any become available.

6. Reversible transcoder guidelines: J2K1 to HT

Any J2K1-based codestream can be transcoded to an HT-based codestream in a reversible way, meaning that no information is lost and the HT-based codestream can even be transcoded back to a J2K1-based codestream, preserving all original quality layer boundaries. This section describes the recommended way to do this, wherein all J2K1 coding passes up to and including the last available Cleanup pass for a code-block are collapsed into a single HT Cleanup pass, which is followed by at most one HT SigProp pass and at most one HT MagRef pass. In this way, only one HT Set is included in the transcoded codestream for each code-block.

Specifically, let \( z_0 \) be the index of the last Cleanup pass in a J2K1 code-block \( b \), so that the total number of coding passes for the block satisfies \( Z_b \in \{ z_0 + 1, z_0 + 2, z_0 + 3 \} \). Then \( P_b^\text{phld} \) is set to \( p \). Equivalently, the number of placeholder passes is set to \( z_0 \), meaning that all coding passes prior to the last Cleanup pass in the J2K1 block bit-stream become placeholder passes in the transcoded representation.

After decoding the J2K1 block bit-stream, the transcoder encodes the one HT Cleanup pass with index \( z_0 \), meaning that

\[
S_{bk} = P_b + P_b^\text{phld},
\]

where \( P_b \) is the number of zero bit-planes recovered while parsing packet headers from the original codestream.

Additionally, if \( Z_b = z_0 + 2 \), the transcoder encodes an HT SigProp pass with index \( z_0 + 1 \), while if \( Z_b = z_0 + 3 \), the transcoder encodes an HT SigProp pass and an HT MagRef pass, with indices \( z_0 + 1 \) and \( z_0 + 2 \), respectively.
The transcoder outputs the same number of packets $Q$ for each each precinct that were found in the original J2K1-based codestream. In order to preserve all quality layer boundaries, each packet $q$ should contribute exactly the same number of coding passes $W^q_b$ to block $b$ in both the original and transcoded codestreams.

7. **HTJ2K encoder guidelines**

Encoders should include at most one HT Set within the codestream for each code-block, unless they have a good reason to do otherwise. One possible reason to generate multiple HT Sets is presented below, although this necessarily involves the introduction of redundant encoded content into the codestream.

Moreover, when an encoder includes at most one HT Set within the codestream, it should identify the codestream as having the HTSINGLE attribute via the CAP marker segment that appears after the SIZ marker segment at the start of the codestream.

Encoders should use $P_b^{phld} = 0$ for each code-block $b$, unless they have a good reason to do otherwise. In particular, non-zero values for $P_b^{phld}$ are only needed to preserve quality layer boundaries within the codestream. Some applications might potentially use $P_b^{phld}$ to trigger custom behaviour, so it is safest for encoders to avoid unintentionally triggering any unexpected behaviour by always using $P_b^{phld} = 0$.

Encoders can use placeholder passes and even multiple HT Sets to introduce quality layer boundaries into an HT block bit-stream representation. The main reason for doing this is to facilitate later transcoding operations, including transcoding operations that produce HT-based and J2K1-based codestreams.

For example, an image or video content server application that needs to offer codestreams encoded at a multitude of bit-rates could be orchestrated by encoding each image or video frame at the highest of the target bit-rates, while including quality layer boundaries within each codestream that correspond to each of the lower target bit-rates of interest. The lower quality representations can then be recovered on-demand by transcoding.

There are two ways in which an encoder can facilitate later transcoding. In the first approach, the encoder includes at most one HT Set for each code-block, using placeholder passes to record the quality boundaries associated with lower quality layers; this is no different from the way in which a J2K1 to HT transcoder preserves quality boundaries from an original J2K1-based codestream. Such codestreams contain $Q > 1$ quality layers and can be transcoded to a lower quality representation corresponding to the first $Q' < Q$ quality layers, by first decoding each code-block and subsequently encoding one HT Cleanup pass, having the largest index $z_0 = 3p$ within those coding passes that are contributed by the first $Q'$ packets of the original
codestream; additionally, the transcoder should encode the HT SigProp pass with index $z_0 + 1$ and the HT MagRef pass with index $z_0 + 2$, to the extent that these are contributed also by the first $Q'$ packets. While this approach does not avoid the need for decoding and re-encoding of the code-blocks, it does avoid the need for a transcoder to perform the non-trivial task of identifying the specific coding passes that should be generated for each code-block during re-encoding; this information has been embedded within the codestream.

The second way in which an encoder can facilitate transcoding is by including multiple non-empty HT Sets, such that the encoded content associated with multiple quality layers is already present and need not be regenerated during transcoding. This necessarily renders the original codestream redundant, which might not be desirable. As mentioned earlier, this case in which a codestream contains redundant information is the only case in which multiple HT Sets are needed for a code-block, whether empty or otherwise.