**CLAIMS:**

1. A method for decoding a video stream, the method comprising:

storing previously decoded pictures of the video stream in a picture buffer, including a plurality of first pictures of a same temporal sub-layer, the plurality of first pictures including at least one sub-layer reference picture for predicting a current picture of the video stream, the current picture and the plurality of first pictures being of the same temporal sub-layer;

identifying a network abstract layer (NAL) unit type of a picture of the plurality of first pictures, and determining whether the picture is a sublayer non-reference (SLNR) picture based on the NAL unit type identified;

removing, based on the picture being determined as the SLNR picture, the SLNR picture from the picture buffer;

determining whether a value of a temporal sub-layer of a second picture is greater than a predetermined value, wherein the previously decoded pictures stored in the picture buffer comprises the second picture being the sub-layer reference picture, comprising:

comparing the predetermined value to a value corresponding to a highest temporal sub-layer identification number, wherein the determining whether the value of the temporal sub-layer of the second picture is greater than the predetermined value occurs based on the predetermined value being determined to not equal the value corresponding to the highest temporal sub-layer identification number;

removing the second picture from the picture buffer based on determining that the value of the temporal sub-layer of the second picture is greater than the predetermined value, comprising:

providing an identifier for the second picture, based on determining that the value of the temporal sub-layer of the second picture is greater than the predetermined value; and

removing the second picture from the picture buffer based on the identifier; and

predicting the current picture using one or more of the at least one sub-layer reference picture that is stored within the picture buffer after removing the SLNR picture and/or the second picture from the picture buffer.

2. The method of claim 1, further comprising:

providing an identifier for the picture determined to be the SLNR picture, based on the picture being determined to be the SLNR picture, wherein

the removing comprises removing the SLNR picture from the picture buffer based on the identifier.

3. The method of claim 2, further comprising:

forming a reference picture list including an entry for each of the plurality of first pictures, wherein

the providing the identifier comprises providing the identifier in the entry of the reference picture list that corresponds to the picture determined to be the SLNR picture.

4. The method of claim 1, further comprising:

determining whether the current picture is an intra random access point (IRAP) picture;

determining whether a flag indicates that there is no output of a random access skipped leading (RASL) picture; and

determining whether to set a respective identifier for each reference picture stored in the picture buffer, the respective identifier for the each reference picture indicating whether the each reference picture is to be removed from the picture buffer, based on whether the current picture is determined to be the IRAP picture and whether the flag is determined to indicate that there is no output of the RASL picture.

5. The method of claim 1, wherein

the value of the temporal sub-layer of the second picture is greater than a value of the temporal sub-layer of the plurality of first pictures that are stored in the picture buffer.

6. The method of claim 1, further comprising:

removing pictures from the picture buffer that are not referred to by a reference picture list, based on the pictures not being referred to by the reference picture list.

7. A decoder for decoding a video stream, the decoder comprising:

memory configured to store computer program code; and

at least one processor configured to access the computer program code, and operate as instructed by the computer program code to perform a method of any of claims 1 to 6.

8. A non-transitory computer-readable medium storing computer instructions that, when executed by at least one processor, cause the at least one processor to perform a method of any of claims 1 to 6.