LAYERED ENCRYPTION TECHNIQUES
FOR DCT-CODED VISUAL DATA

Encryption schemes for multimedia data need to be specifically designed to protect multimedia content and fulfill the security requirements for a particular multimedia application. For example, real-time encryption of an entire video stream using classical ciphers requires heavy computation due to the large amounts of data involved, but many multimedia applications require security on a much lower level (e.g. TV news broadcasting). In this context, several selective encryption schemes have been proposed recently which do not strive for maximum security, but trade off security for computational complexity. The (historically) first and most numerous attempts have been made to secure DCT-based multimedia representations, among them the selective encryption of MPEG streams has attracted the most attention. This has been accomplished by encrypting I-frames only, by manipulating motion vector data, or by manipulating coefficients. In case a selective encryption process requires a multimedia bitstream to be parsed in order to identify the parts to be subjected to encryption, the problem of high processing overhead occurs in general. For example, in order to selectively protect DC and large AC coefficients of a JPEG image (as suggested by some authors), the file needs to be parsed for the EOB symbols 0x00 to identify the start of a new 8×8 pixels block (with two exceptions: if 0xFF is followed by 0x00, 0x00 is used as a stuffbit and has to be ignored and if AC63 (the last AC-Coefficient) not equals 0 there will be no 0x00 and the AC coefficients have to be counted). Under such circumstances, selective encryption will not help to reduce the processing demands of the entire application. A possible solution to this problem is to use the visual data in the form of scalable bitstreams. In such bitstreams the data is already organized in layers according to its visual importance and the bitstreams do not have to be parsed to identify the parts that should be protected by the encryption process. There exist several possibilities how to organize MPEG data into base and enhancement layers and it is not clear which variant is most suited for the selective encryption application. In this work we systematically investigate the different possibilities how to organize DCT-coded visual data into several quality layers and we experimentally compare the respective applicability to the selective encryption application.

Mark M. Fisch 1,+*, Herbert Stögner* 1*, and Andreas Uhl* 1,2

1 School of Telecommunications & Network Engineering,
Carinthia Tech Institute Primoschgasse 8,
A-9020 Klagenfurt, AUSTRIA

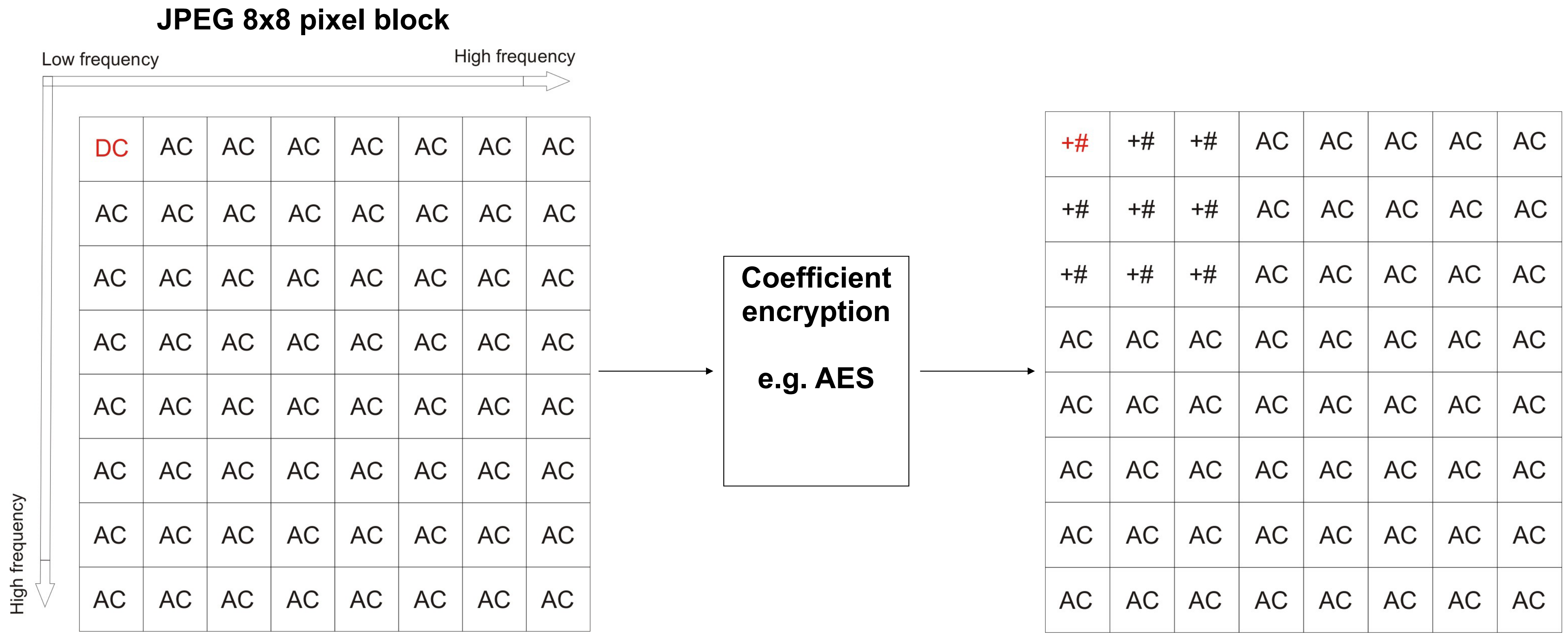
2 Department of Scientific Computing, Salzburg University
Jakob-Haringerstr.2, A-5020 Salzburg, AUSTRIA
e-mail: uhl@cosy.sbg.ac.at

+ Mark M. Fisch is an artifical name representing a group of students working on this project in the framework of the Multimedia I laboratory (winterterm 2003/2004): H. Fischer, C. Gattringer, M. Mauritsch, M. Oberwasserlechner, C. Probst, M. Schauer, F. Schmidt, M. Schuster, C. Sturmer.

This work has been partially supported by the Austrian Science Fund, project no. 15170.

Spectral selection

The first scan contains the DC coefficients from each block of the image, subsequent scans may consist of a varying number of AC coefficients, always taking an equal number from each block.



JPEG 8x8 pixel block

DC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC

Coefficient encryption

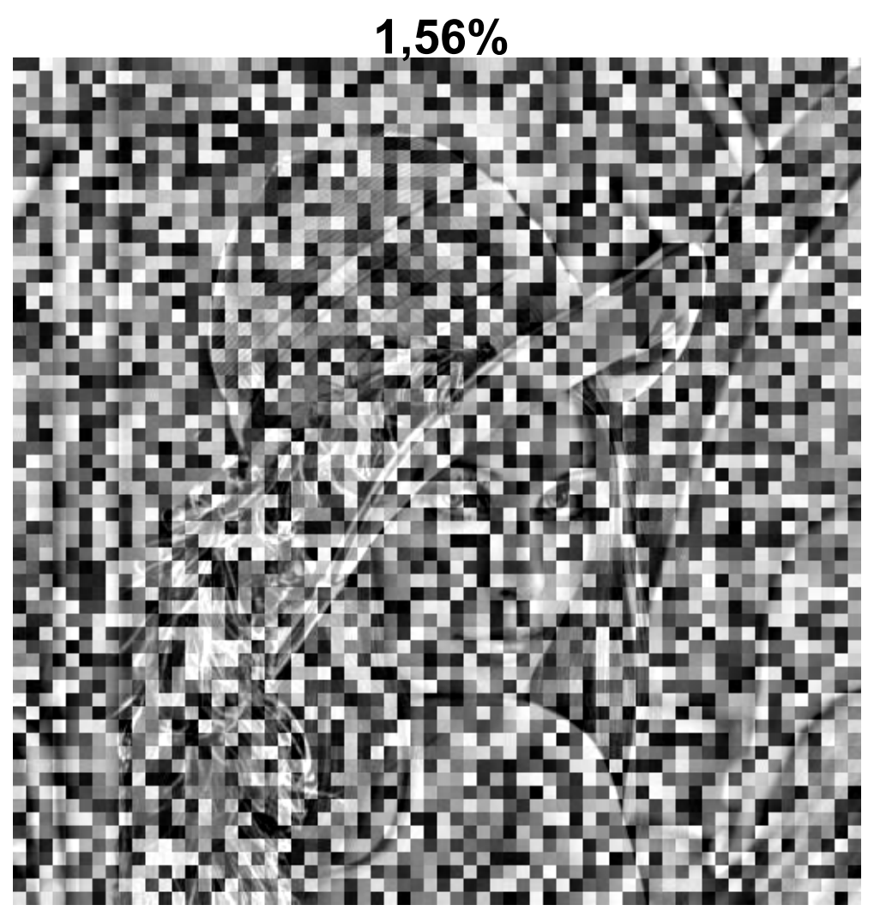
e.g. AES

+#	+#	+#	AC	AC	AC	AC	AC
+#	+#	+#	AC	AC	AC	AC	AC
+#	+#	+#	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC
AC	AC	AC	AC	AC	AC	AC	AC

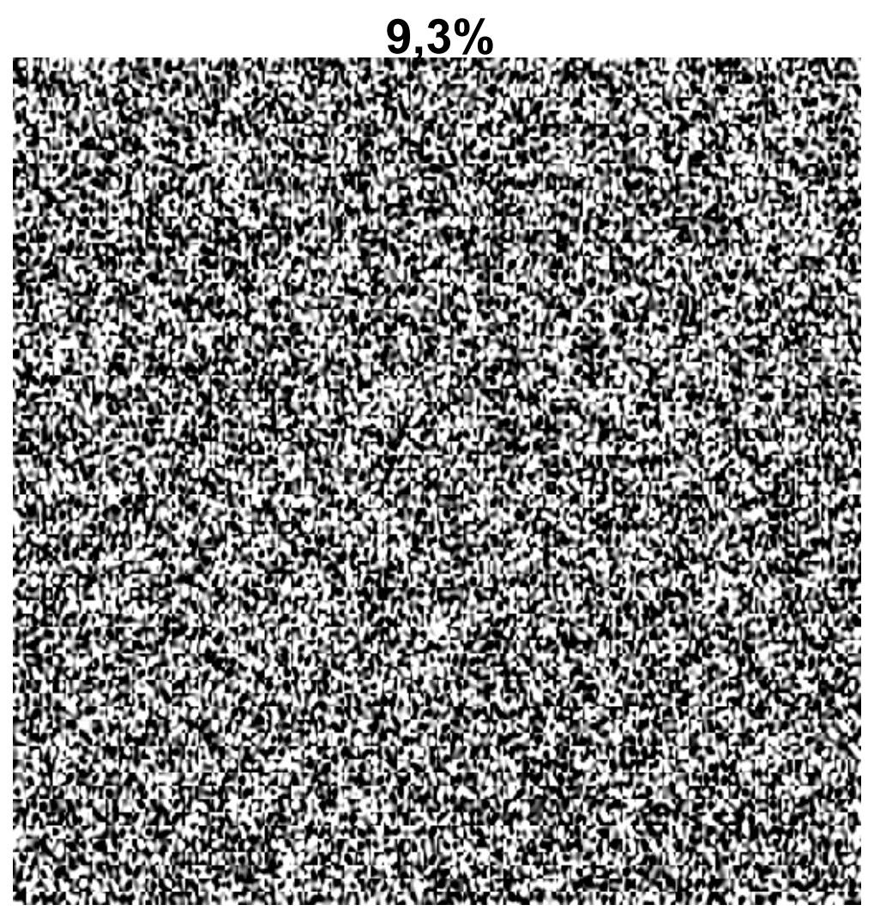
Example

Original Image: Lena.jpg

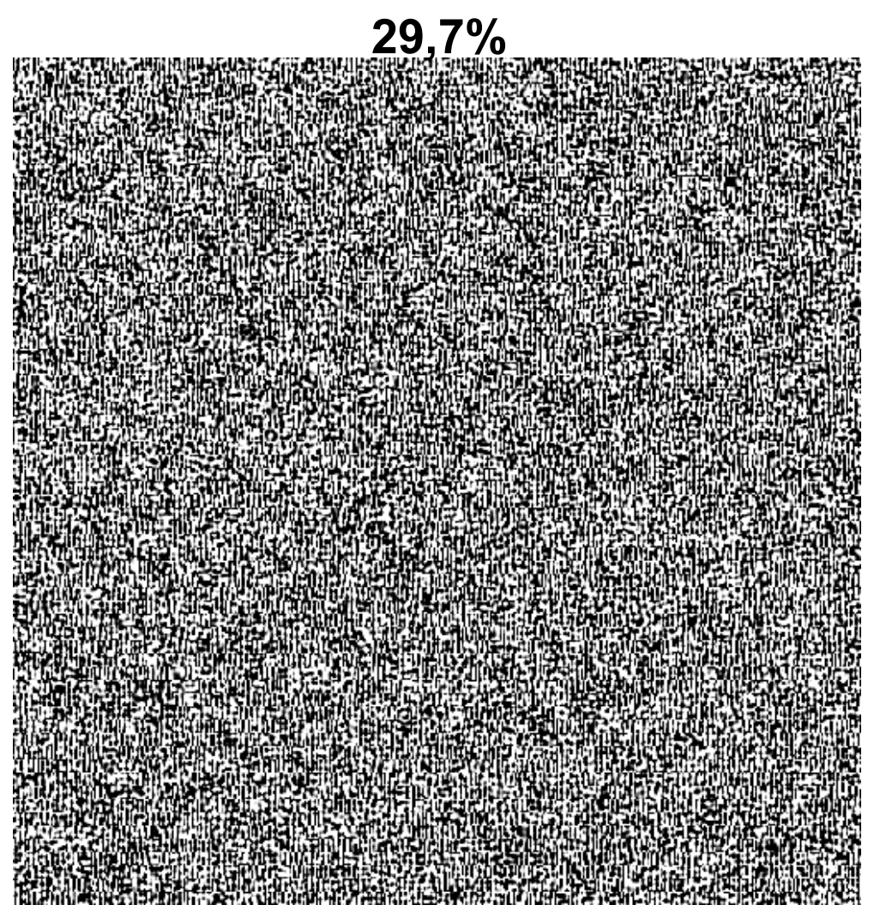
JPEG Baseline coded



DC coefficients encrypted

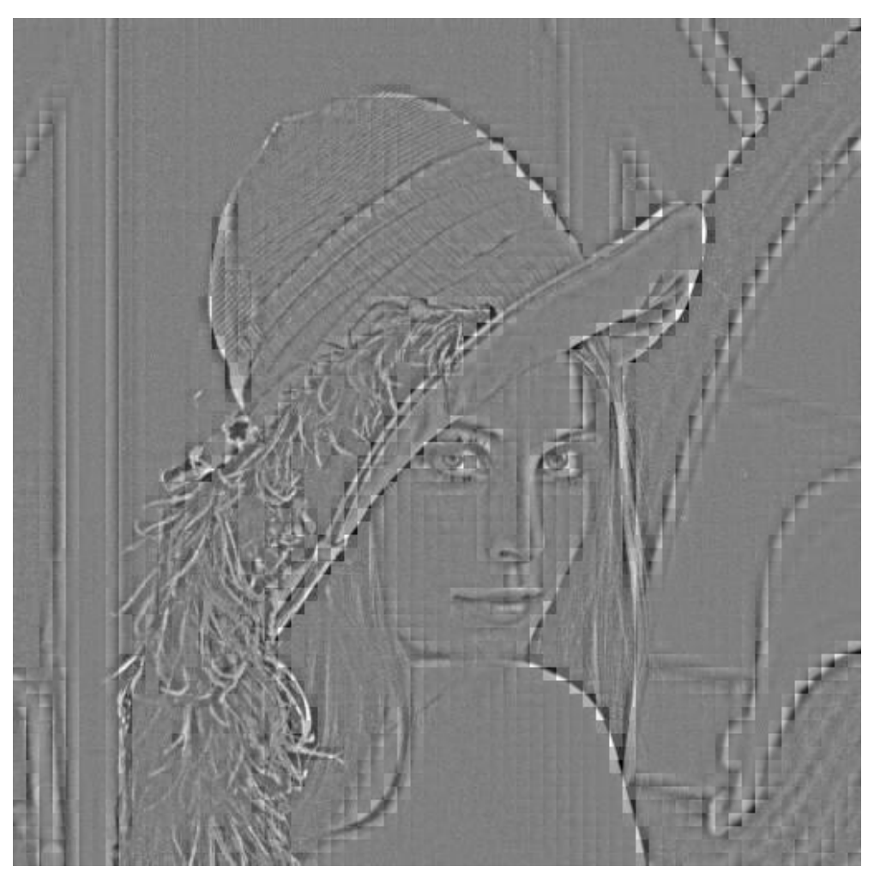


DC and 6 AC coefficients

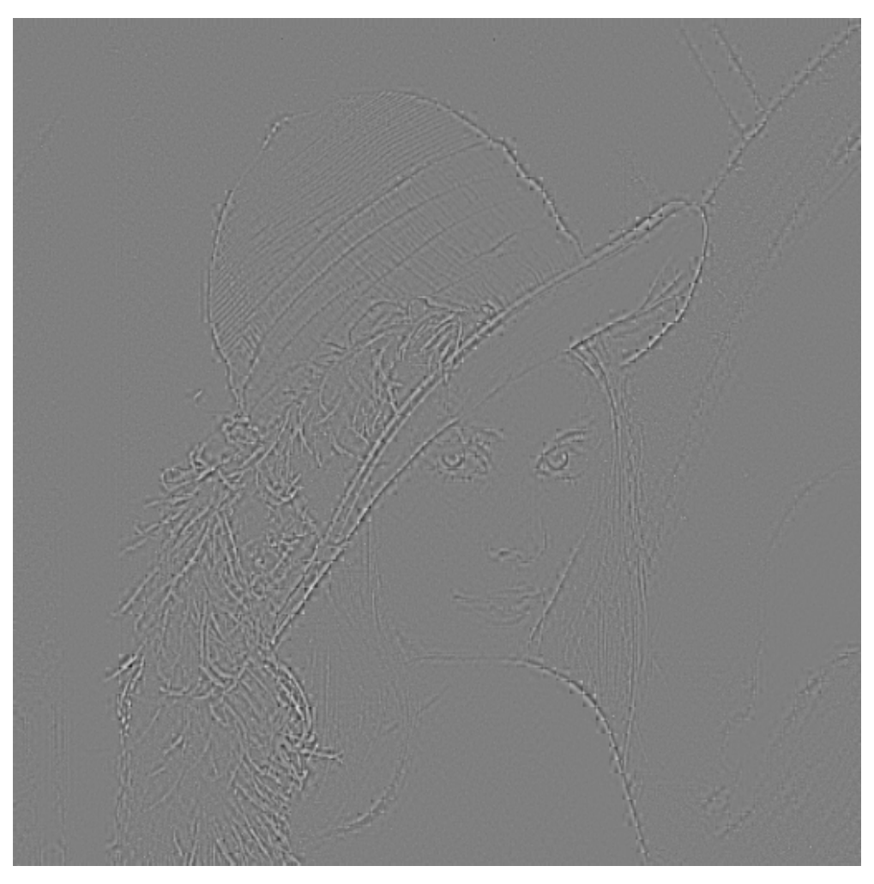


DC and 19 AC coefficients

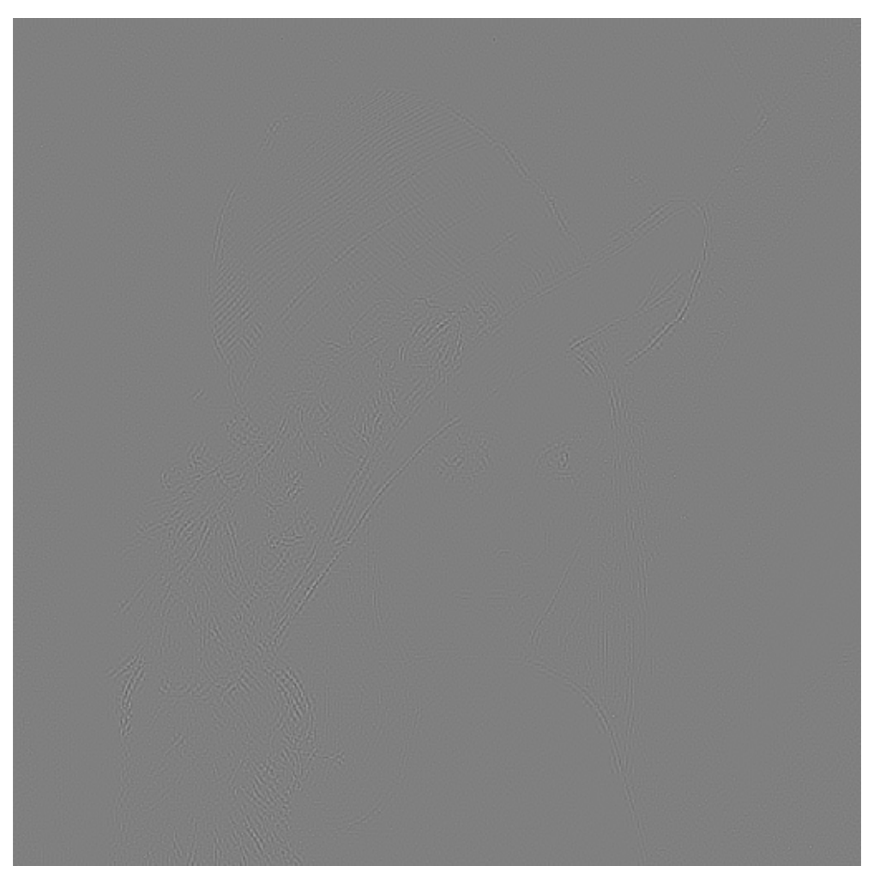
noise cancelling: replacement attack



DC coefficients set to mean



DC coefficients set to mean



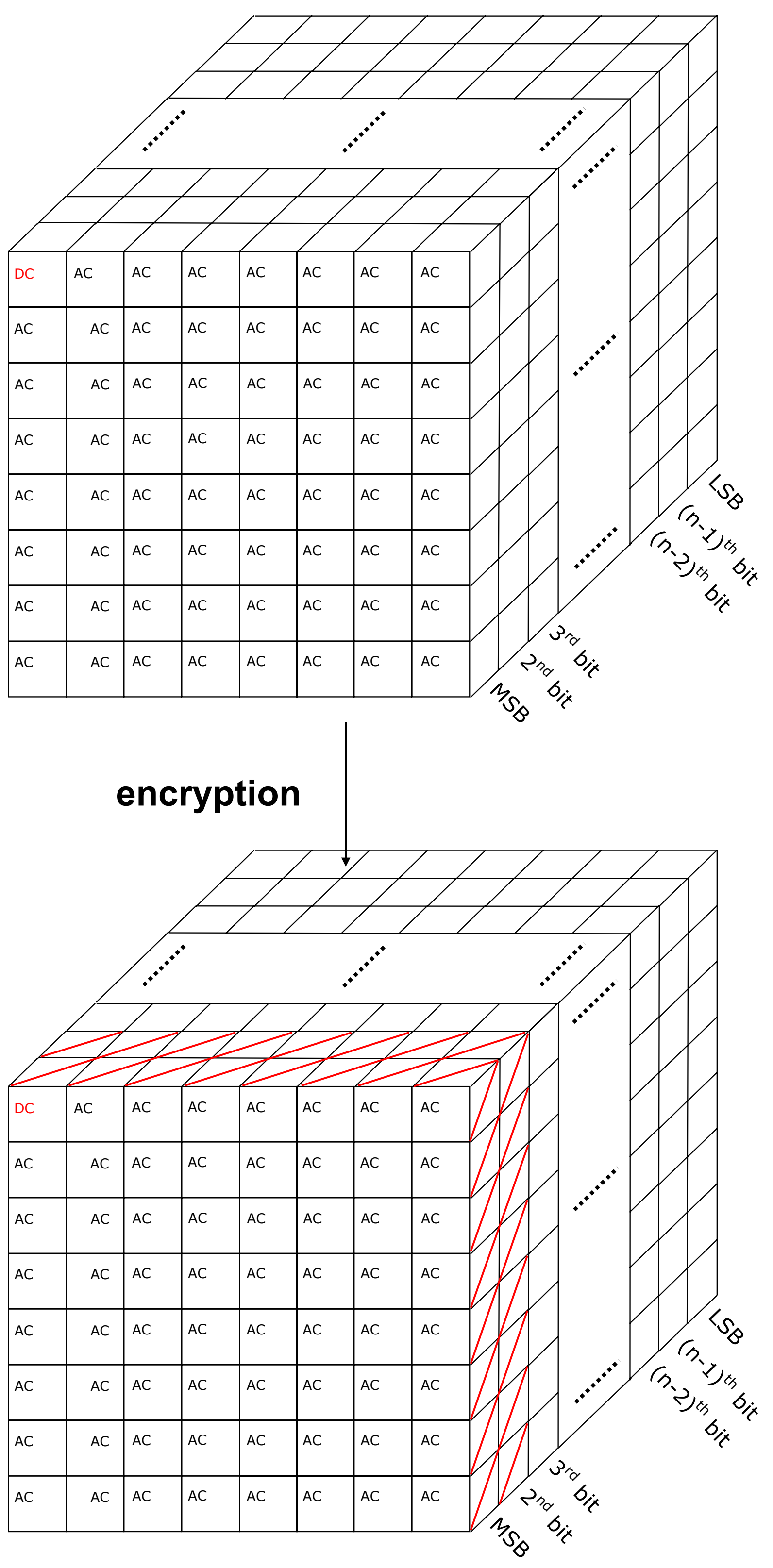
DC coefficients set to mean

PSNR [dB]

Lena, 9.3% enc.	14.6
Lena, 29.7% enc.	14.5

Successive approximation

JPEG 8x8 pixel block



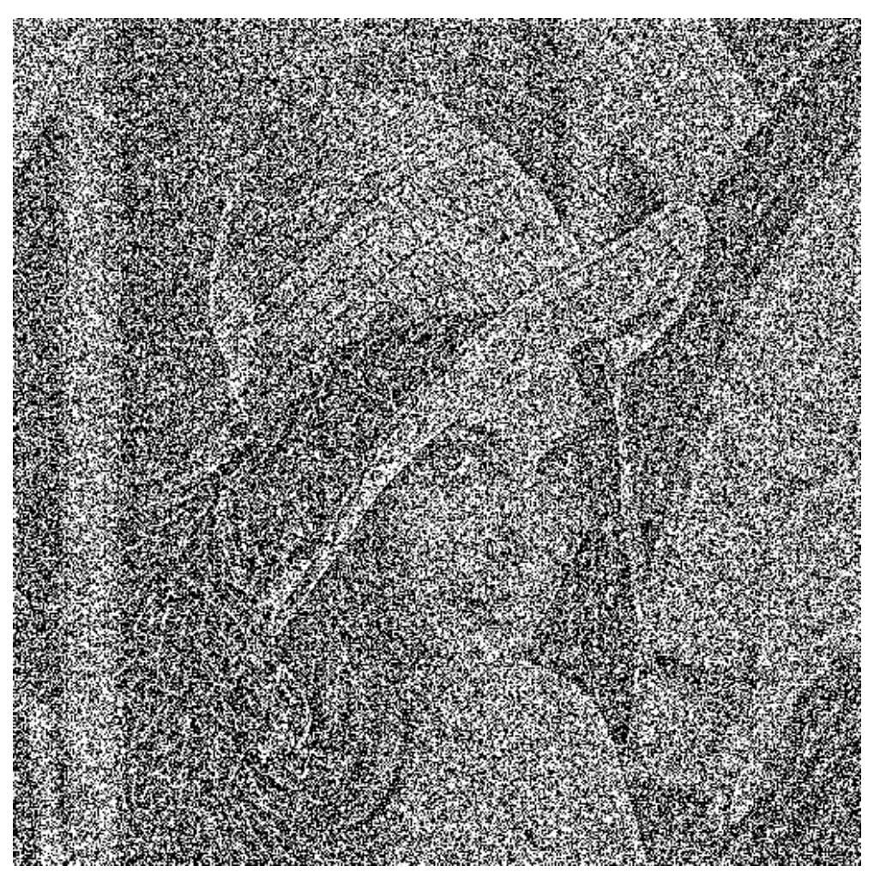
The most significant bits of all coefficients are organized in the first scan, the second scan contains the next bit corresponding to the binary representation of the coefficients, and so on. Since quantization is highly related to reducing the bit depth of coefficents, this mode behaves similarly to SNR scalability.

Example

Original Image: Lena.jpg

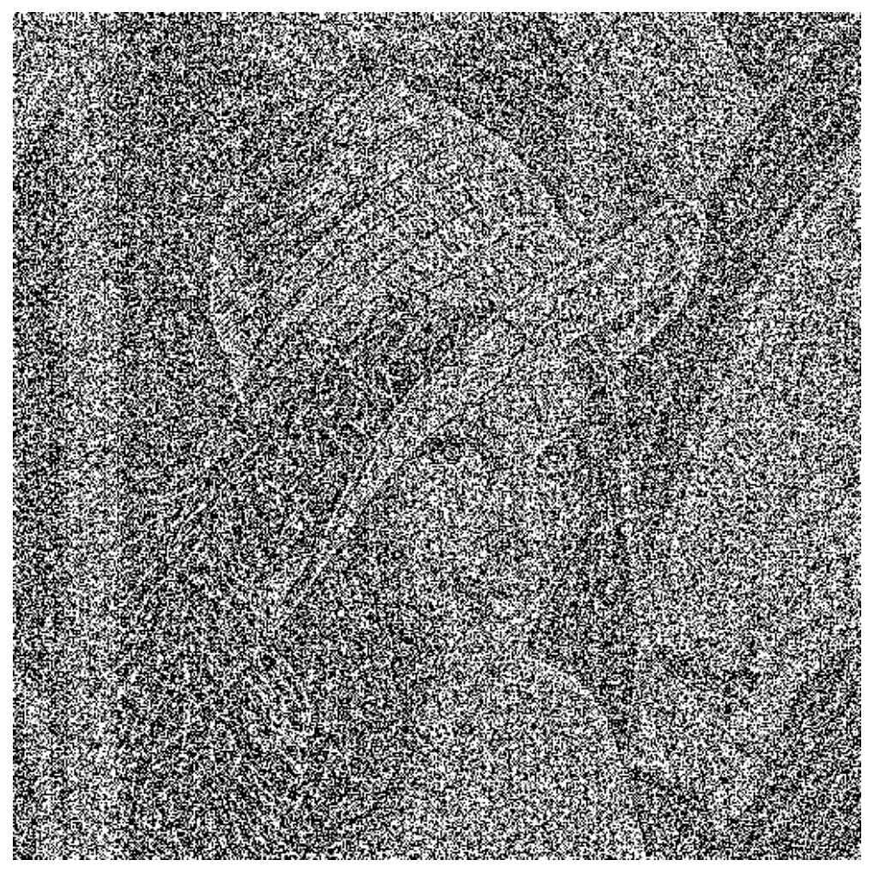
JPEG Baseline coded

10 %



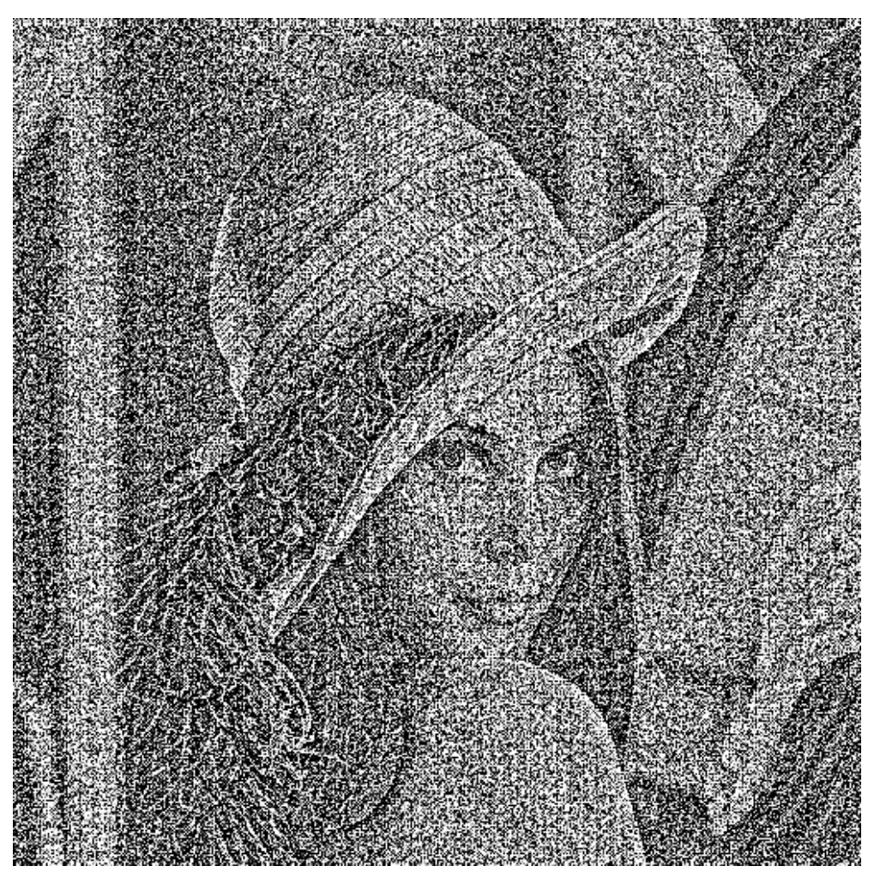
MSB of all coefficients

30 %

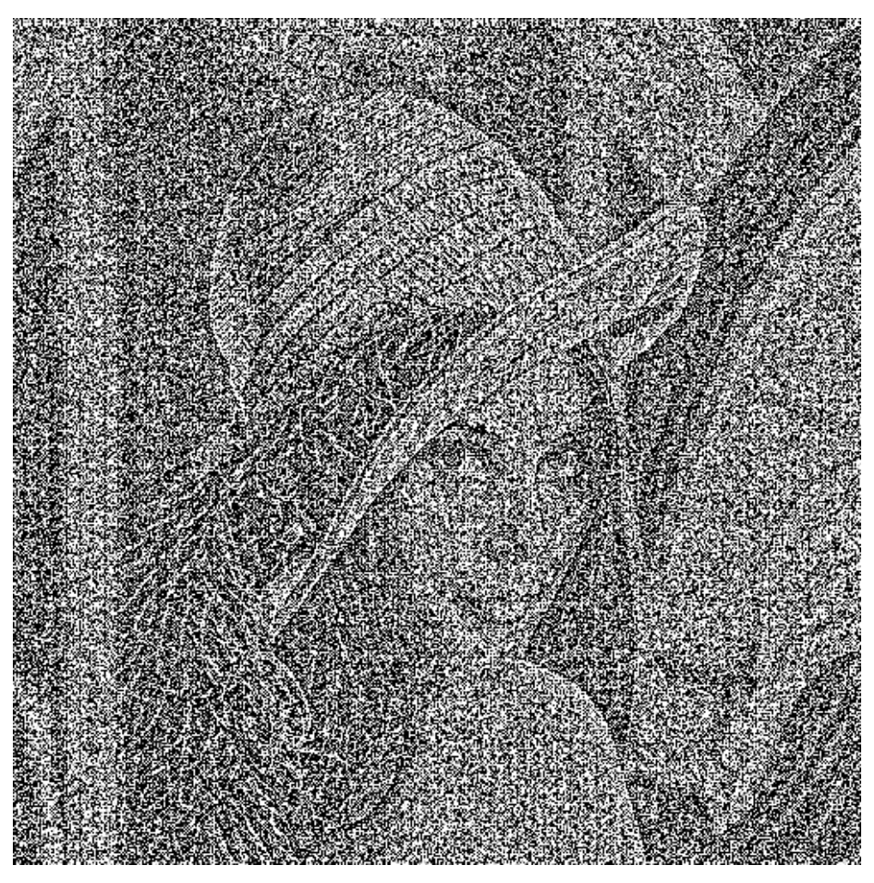


MSB + 2 Bits of all coefficients

noise cancelling: replacement attack



Encrypted Bits set to mean value

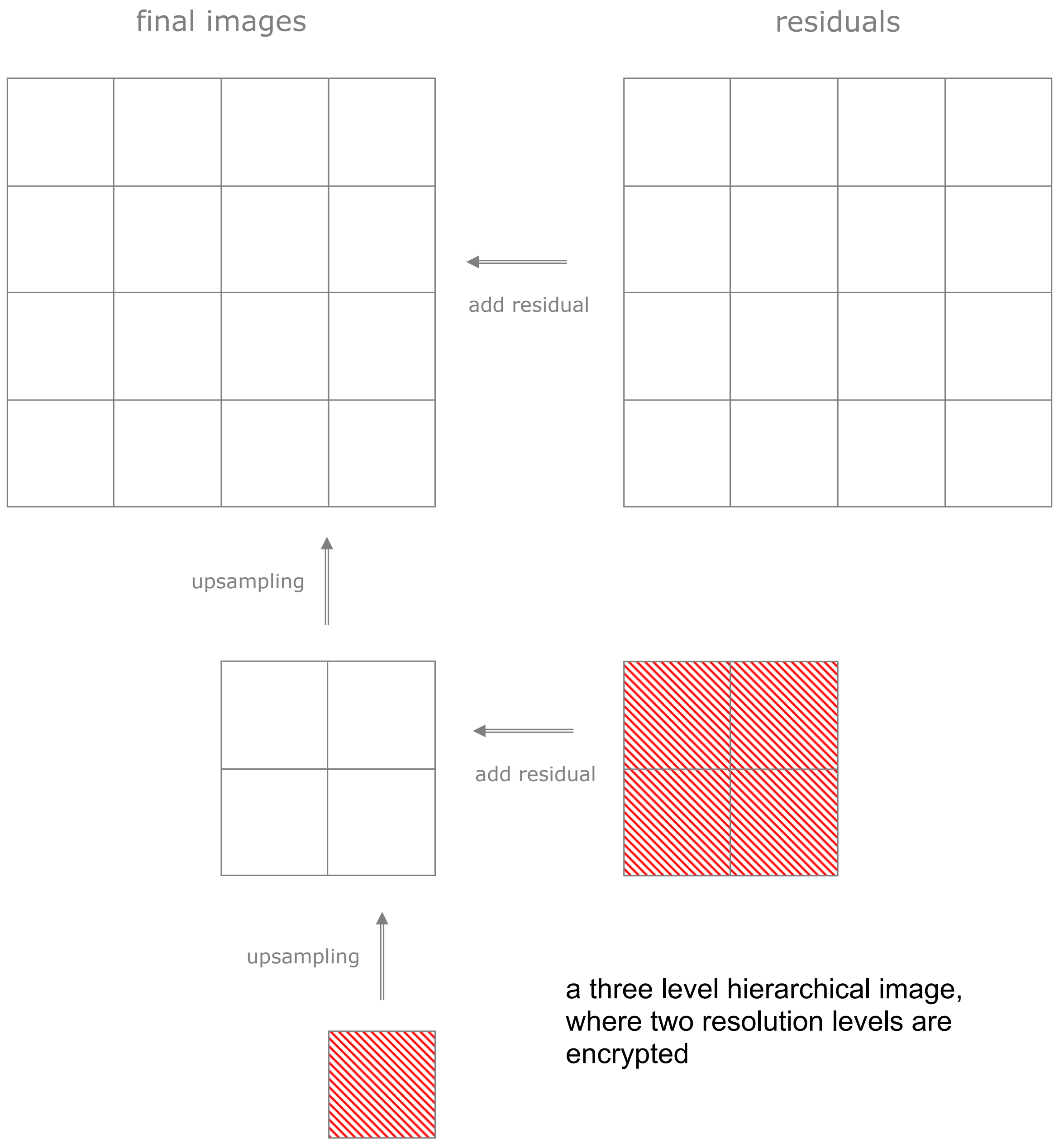


Encrypted Bits set to mean

PSNR [dB]	
Lena, 10% enc.	7.0
Lena, 30% enc.	6.2

Hierarchical progressive

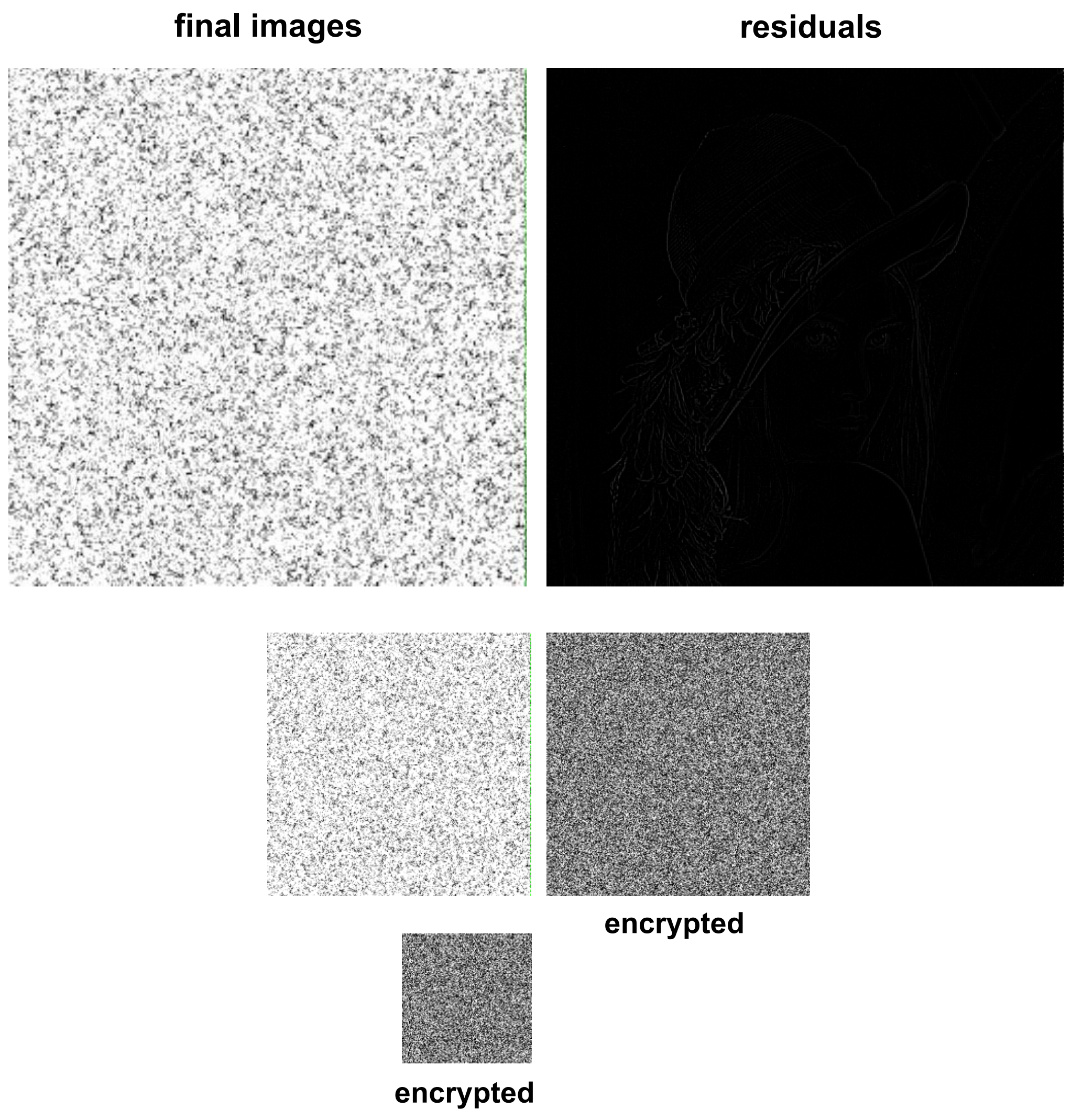
three level hierarchical pyramid

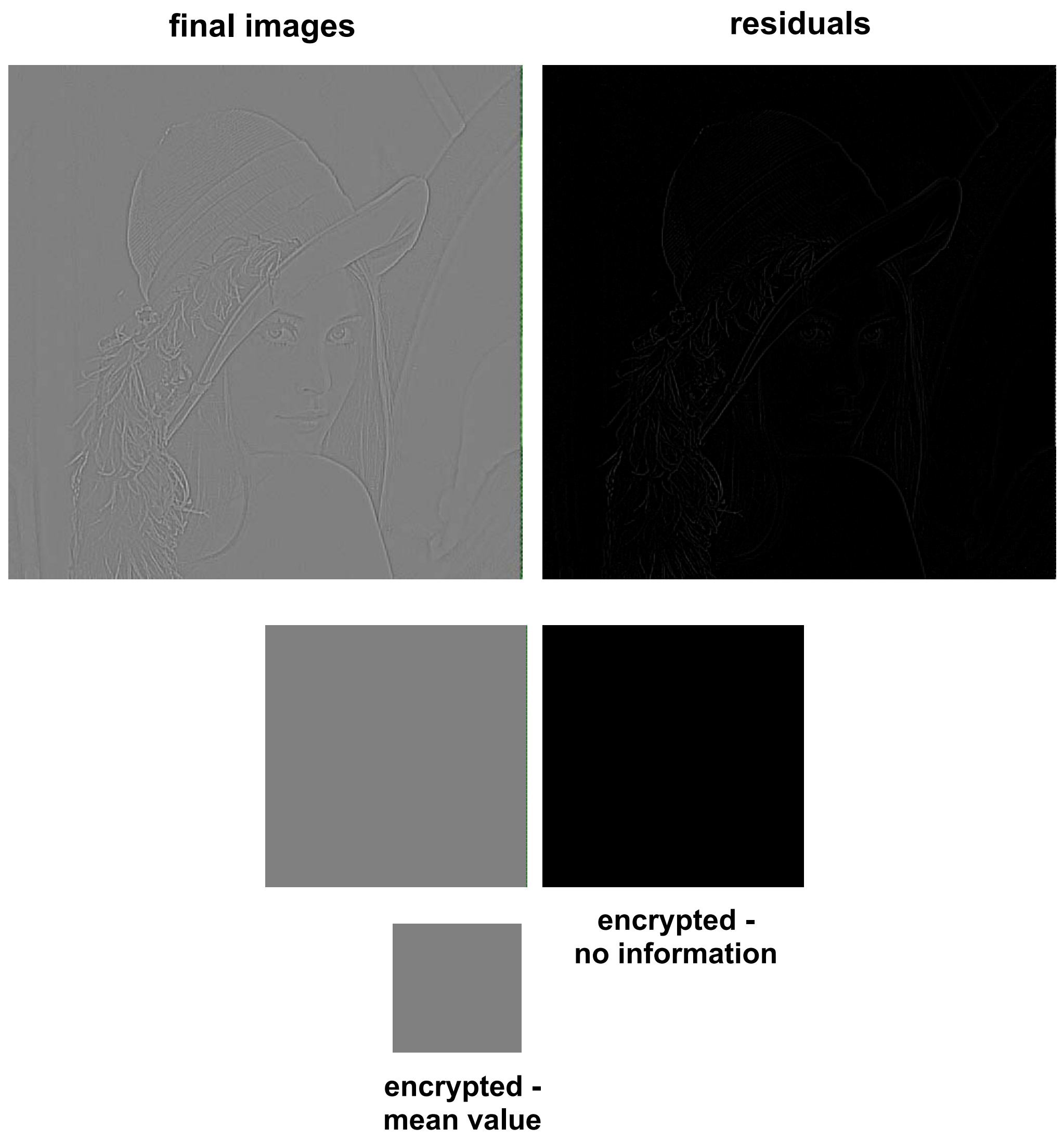


a three level hierarchical image, where two resolution levels are encrypted

An image pyramid is constructed by repeated weighted averaging and downsampling. The lowest resolution approximation is stored as JPEG (i.e. the first scan), reconstructed, bilinearly upsampled, and the difference to the next resolution level is computed and stored as JPEG with different quantization strategy (similar to P and B frames in MPEG). This is repeated until the top level of the pyramid is reached. This mode corresponds well to MPEG-2 resolution scalability.

three level pyramid where 2 levels (31,25 %) were





Example

Original Image: Lena.jpg

JPEG Baseline coded

PSNR [dB]	
Lena, 8.3%* enc.	14.8
Lena, 31.25% enc.	14.7

* six level pyramid with lowest resolution and the first three levels encrypted