

INDEX

A

ACB algorithm, Buynovsky's, 201–203

Adaptive coding

- binary arithmetic coding, 134–139
- cumulative distribution update
 - direct, 133–134
 - periodic, 140–142
 - tree-based, 139–140
- strategies for computing symbol distributions, 132

Adaptive entropy coder

- code selection, 317
- fundamental sequence encoding, 313–314
- low-entropy options
 - second-extension option, 316
 - zero-block option, 316–317
- no compression, 317
- split-sample option, 314–315

Adaptive Huffman coding

- algorithm *M*, 94
- brute force, 89–90
- Faller, Gallager, and Knuth (FGK) algorithm, 91–93
- splay tree algorithm, 93–94
- Vitter's algorithm Λ , 93

Additive codes, 72–73, 77

Algorithmic Information Theory, *see* Kolmogorov complexity

Algorithmic Prefix Complexity, 42–43, 48

Algorithm *M*, 94

Amplitude range, and segmentation, 259–260

Apostolico and Fraenkel codes, 67–69, 77

Approximate arithmetic, 115–118

Arithmetic coding

adaptive, 336

approximate arithmetic, 115–118

code values, 104–106

context-based, 397

correct decoding: conditions for, 118–120

decoding process, 110–111

dynamic sources, 112–113

encoder and decoder synchronized decisions, 113

encoding process, 106–109

entropy-coding methods, 102

intervals

- positive-length and disjoint, 118–119
- rescaling, 114–115

inverse operations, 120

model for Burrows–Wheeler compression, 177–179

nested subintervals, 119–120

notation, 103–104

optimality, 111–112

practical problems, 102–103

separation of coding and source modeling, 113–114

Arithmetic coding implementation, 120–147

- adaptive coding, 132–142
- coding with fixed-precision arithmetic, 121–132
- complexity analysis, 142–147
- integer-based, 147–150

Arithmetic encoder, 19–21

Arithmetic operations, speed of, 146–147

Asymptotic equipartition property, 26–34

ATMEL digital audio decoder, 432–433

Average filter, 380

B

Backup, remote, massive data sets, 280

Balanced pairs, *rsync* algorithm, 283–284

Bandwidth, in telemetry compression, 250–251

Bayesian model averaging, 218

Bidirectional dictionary methods, 161

Bilevel image data, 336–338

Binary arithmetic coding, 134–139

Binary polynomial representation, 56, 74–75

Binary sources, 10

Biocompress programs, 162

Biomolecular sequences, compression, 162–163

Bisection search, 143–145

Bit-plane coding, JPEG2000, 360–364

Bit rate

- for lossless audio signal, 256
- variable, peak vs. average, 264

Bits per character

- averages, 240
- entropy and, 228–229
- for new transforms, 242

Bitstream scaling, architecture for, 421

Blending

- model-based, 218–219
- performance-based, 218
- probability domain, 219

Blocking system, 17–18

Block moves, in delta compression, 273

Block size, *rsync* algorithm, 281–283

Block-sorting compression, *see*
Burrows–Wheeler compression
Bottom layer templates, 335–336
Bottom layer typical prediction,
334
Bounds
error probability, 34
lower-bounding Kolmogorov
complexity, 46–47
and redundancy and entropy,
16–17
Brute force adaptive Huffman, 89–90
Buffer carries, implementation with,
122–126
Burrows–Wheeler algorithm, 170
Burrows–Wheeler compression,
73–74, 77
improvements to, 180–181, 190
Move-to-Front step, 183–185
elimination, 187–189
permutation, 181–183
preprocessing, 181
statistical compressor, 185–187
and symbol-ranking, 200
Burrows–Wheeler transform
in file synchronization, 189
final coder, 190
forward, 170–171
implementation, 173–174
for lossless algorithms, 233
for PNG, 374
relationship to other algorithms,
180
reverse, 171
algorithms for, 172–173
Buzynsky's ACB algorithm,
201–203
Bzip2 compression algorithm, 234,
236–241

C

CALIC, 301
approach to modeling prediction
error, 212
edge detection by, 216
Canonical codes, Huffman, 84
Carries
buffer, implementation with,
122–126
propagation, 130–131
Cascaded coding model, 178
CCSDS, e_Rice algorithm
recommendation, 312
C-Cube video products, 427–428
Chain rule, 7
Channel coding, 251
Checkpointing, 278

Checksums, blockwise, of current
file, 280–281
Chip-sets, 438
Church–Turing thesis, 38–39
Codebook
computed for classes only, 87
Huffman, 80–83
Coded data format, 321
Coder/decoder pairs, Huffman,
94–96
Code selection function, adaptive
entropy coder, 317
Codestream, JPEG2000, 365–366
Code values
arithmetic coding, 104–111
scaling of, 115
Codewords
descendants, 14
end-of-line, 398
groups, in Golomb code,
62–64
ideal length, 16–17
in prefix code, 13–14
in tree, 15
variable-length, 12
Coding distribution, 19
Combined predictors, 217–219
model-based blending, 218–219
performance-based blending of
predictors, 218
probability domain blending, 219
Comma codes
adaptive entropy coder, 313–314
ternary, 70–71
Competitive optimality, 25–26
Complementary palindromes, 162
Complete tree, 93
Component transforms, JPEG2000,
354–355
Compress anything claim, 43–44
Compressed bitstream scaling,
architecture for, 421
Compressed domain image
processing, JPEG2000, 353
Compressed pointer macro scheme,
157
Compression
algorithms, 208
Burrows–Wheeler, *see*
Burrows–Wheeler
compression
data, *see* Data compression
dynamic Markov, 232
practical tips, 383–385
sequence, 20–21
symbol-ranking, 195–201
telemetry, *see* Telemetry
compression
tests, 385–388

text, *see* Text compression
Unicode, 293–294
Compression engine, PNG, 374–376
compress program, 165
Computation
issues, of Kolmogorov complexity,
44–47
sequential, of interval, 21–23
symbol distributions, strategies
for, 132
Computational complexity analysis
arithmetic operations, 146–147
cumulative distribution estimation,
145–146
interval renormalization, 142–143
symbol search, 143–145
Concrete Kolmogorov complexity,
40
Consultative Committee for Space
Data Systems, *see* CCSDS
Content progressive representations,
340
Context adaptive lossless image
compression algorithm, *see*
CALIC
Correcting one-pass algorithm, 278
Counting argument, 41
Cumulative distribution
of code values, 105–106
estimation, 145–146
updates
direct, 133–134
periodic, 140–142
tree-based, 139–140
Current context, 202
Current file, 270–271
blockwise checksums of, 280–281
encoding, 274
Cyclic redundancy code, 373

D

Database records, and file systems:
reconciling, 286–287
Data compression
audio, 255–267
different approaches, 35–36
lossless, CCSDS recommendation
for space applications,
311–326
Data compression: dictionary-based
benchmark programs and
standards, 165–166
biomolecular sequence
compression, 162–163
data structures
Karp–Rabin fingerprints,
164–165

- suffix trees, 163–164
 - tries and compact tries, 163
 - dictionary construction, 154–161
 - stages, 153
 - Data structure
 - and decoding control: JBIG2, 346–348
 - dictionary-based data compression, 163–165
 - JBIG, 336–338
 - telemetry frame or packet, 251
 - DCT, *see* Discrete Cosine Transform
 - Decoding
 - combined with updating and coding, 145
 - correct: conditions for, 118–120
 - Golomb code, 63–64
 - JBIG, 338–339
 - JPEG-LS, 309
 - postprocessor and adaptive entropy decoder for, 321–323
 - process, in arithmetic coding, 110–111
 - Decoding: JBIG2
 - generic refinement region, 342–343
 - generic region, 341–342
 - halftone region, 346
 - pattern dictionary, 345–346
 - rest region, 345
 - symbol dictionary, 343–345
 - Decoding control, and data structures: JBIG2, 346–348
 - Decompression program
 - custom-built, 41–42
 - lossless audio data compression, 264–265
 - Deflate, core of PNG's compression scheme, 374–376
 - Delta compression
 - applications, 271–273
 - block moves, 273
 - choosing reference files, 278–279
 - experimental results, 275–277
 - LZ77-based, 274–275
 - problem definition, 270–271
 - space-constrained, 277–278
 - Descriptive language, Kolmogorov complexity, 39–40
 - Design decisions, PNG, 372–374
 - Deterministic prediction, 333–334
 - Dictionary
 - for Golomb code, 62
 - methods for lossless algorithms, 232–233
 - pattern, decoding, 345–346
 - static vs. dynamic construction, 154–161
 - symbol, decoding, 343–345
 - Dictionary-based data compression
 - benchmark programs and standards, 165–166
 - compressing biomolecular sequences, 162–163
 - data structures, 163–165
 - dictionary construction
 - parsing issues, 155–157
 - semidynamic and dynamic, 157–161
 - static methods, 154–155
 - stages, 153
 - Differential compression, *see* Delta compression
 - Differential layer prediction, 332–334
 - Differential layer templates, 334
 - diff* utility, 272
 - Digit vector, combined with weight vector, 56–57
 - Discrete Cosine Transform
 - hardware, 416
 - video compression, 419–420
 - Distance coding, 233
 - Distance measures, *rsync* algorithm, 283
 - results, 285
 - Distribution scheme, efficient, 270
 - Dynamic dictionary methods, 157–161
 - bidirectional, 161
 - unidirectional, 158–161
 - Dynamic Markov compression, 232
 - Dynamic sources, arithmetic coding, 112–113
- E**
- e-commerce, security, 36
 - Edge detection, by CALIC, 216
 - Efficiency
 - entropy coding, 208
 - Huffman codes, 86
 - implementation of Huffman codecs, 94–96
 - memory-efficient algorithms, 95
 - output, and renormalization, 127–130
 - Rice codes, 61, 64
 - speed-efficient algorithms, 95–96
 - Elias algorithm, 19, 22–25
 - Elias gamma codes, 58–59, 77
 - Elias omega codes, 59
 - emacs* data set, 275–276
 - Encoding
 - current file, 274
 - fundamental sequence, 313–314
 - JPEG-LS, 302–309
 - prefix-free, 42–44
 - process, in arithmetic coding, 106–109
 - of ranks, 196
 - run-length color, 398
 - Encoding: JBIG
 - adaptive arithmetic coding, 336
 - bottom layer typical prediction, 334
 - differential layer prediction, 332–334
 - model templates, 334–336
 - resolution reduction, 330–331
 - Entropy
 - achieving, 23–24
 - adaptive entropy coder, 313–317
 - and bits per character, 228–229
 - conditional, 6
 - interpretation, 8–9
 - properties, 7–8
 - definition, 3–4
 - as information measure, 5–6
 - joint, 6
 - properties, 7
 - properties, 4–5
 - Entropy coding
 - in compression systems, 102
 - efficiency, 208
 - JPEG-LS, 307–309
 - lossless audio data compression, 262–263
 - Entropy rate, 11
 - Equivalence classes, Huffman code for, 87
 - e-Rice algorithm, 312–313
 - Error accumulation, 120
 - Error modeling, in lossless image compression, 212
 - Error probability
 - bounded, 34
 - and rate, 28
 - optimal balance between, 31–33
 - Escape character, 231
 - Even-Rodeh code, 59–60
 - Exclusion list, 198
 - Explicit detection, predictors based on, 215–216
 - Extended Huffman codes, 87–89
 - External pointer macro scheme, 157
- F**
- Facsimile compression
 - algorithms
 - context-based arithmetic coding, 397
 - modified Huffman, 393

Facsimile compression (*cont.*)
 modified READ, 393–397
 run-length color encoding, 398
 historical overview, 391–393
 standards
 ITU-T Group 3 (T.4), 398–399
 ITU-T Group 4 (T.6), 399
 JBIG and JBIG2 (T.82 and T.88), 399
 MRC–T.44, 399–402
 Facsimile transmission, 87
 JBIG compression for, 327–328
 Faller, Gallager, and Knuth (FGK) algorithm, 91–93
 Fibonacci codes
 Apostolico and Fraenkel codes, 67–69
 Fraenkel and Klein codes, 66
 higher-order Fibonacci representation, 66–67
 new order-3, 69–70
 use as universal codes, 65
 Zeckendorf representation, 66
 Fibonacci polynomial representation, 57
 Fictitious pixel, LNTP, 332–333
 File synchronization
 Burrows–Wheeler transform in, 189
 remote, *see* Remote file synchronization
 File systems, and database records: reconciling, 286–287
 Filters, PNG, 378–383
 Final coder, simplified, 190
 Fingerprints, Karp–Rabin, 164–165
 Fixed-length codes, for memoryless sources, 26–34
 Fixed-precision arithmetic, coding with, 121–132
 Flexible greedy parsing, 156
 Formatting, JBIG, 336–338
 Fraenkel and Klein codes, 66
 Frame telemetry, 248
 Freeze heuristics, 161
 Frequency decay method, 91
 Fujitsu, audio video products, 431–432
 Fundamental sequence encoding, 313–314

G

Gamma codes, Levenstein and Elias, 58–59
gcc data set, 275–276
GenCompress program, 162–163
 GIF compression standard, 166, 372–373, 385–386

Gödel's incompleteness theorem, 36–37
 Goldbach *G1* codes, 72, 77
 Golomb codes, 62–64
 limited-length, 307
 Gradient adaptive predictor, 216
 Grayscale image
 GIF and PNG, 374
 JBIG2, 339–340
 Greedy strategy, of parsing, 156–157, 164
gzip compression program, 165, 233–234, 241, 276

H

Halftone region, decoding, 346
 Hamming ball, 30–31
 Hardware implementation
 image compression, 415–417
 text compression, 407–415
 video compression, 417–442
 Hierarchical lossless image coding, 220–222
 Hierarchical prediction, in lossless image compression, 211–212
 Hilbert scans, 213–214
 Historical perspective
 facsimile compression, 391–393
 Kolmogorov complexity, 51
 PNG, 371–372
 symbol-ranking compressors, 197
 unidirectional dictionary methods, 158
 HTTP
 performance improvement, 272
 transfer, file synchronization for, 280
 Huffman codes, 49–50
 building of, 80–83
 canonical, 84, 230
 extended, 87–89
 length-constrained, 89
 modified, 87
 N-ary, 83–84
 performance, 84–86
 prefixed, 87
 Shannon–Fano coding, 80
 Huffman coding
 adaptive, 89–94
 statistical method for lossless algorithms, 229–232
 Huffman tree, 408–409
 Human genome, 162

I

IBM, MPEG-2 digital audio/video decoder, 436–437

Image compression
 bilevel, *see* JBIG; JBIG2
 highly scalable, *see* JPEG2000
 lossless, *see* Lossless image compression
 Image compression hardware, 415–417
 DCT, 416
 JPEG, 417
 wavelet architectures, 416–417
 Image compression standard, GIF, 166
 Image objects, 389
 Image stripes, 328, 330
 Implementations
 arithmetic coding
 adaptive coding, 132–142
 complexity analysis, 142–147
 with fixed-precision arithmetic, 121–132
 with buffer carries, 122–126
 Burrows–Wheeler compressor, 173–180
 data compression for space applications, 324–326
 and efficient output, 127–130
 with integer arithmetic, 126–127
 integer-based, 147–150
 Lempel–Ziv methods, 161
 numerical, lossless audio data compression, 263
 Implementations: hardware
 image compression, 415–417
 text compression, 407–415
 video compression, 417–442
 Incompressibility, in Kolmogorov complexity, 40–42
 Independent and identically distributed source, 10, 48, 102
 Independent updating, 145–146
 Inequality
 frequently used in information theory, 4–5
 Kraft–McMillan, 84–85
 Kraft's
 in competitive optimality, 26
 for prefix code, 14–17
 Information measure
 entropy as, 5–6
 universal, 50–51
 Information sources
 discrete, 9
 discrete stationary, 10–11
 with memory, variable-length codes for, 18–26
 memoryless
 fixed-length codes for, 26–34
 variable-length codes for, 11–18

memoryless or i.i.d., 10
 Information theory, background,
 228–229
 Initial letter preserving transform,
 241–243
 Innovation entropy, 11
 Instantaneous code, 13
 Integer arithmetic, implementation
 with, 126–127, 147–150
 Integer wavelet transform, 221–222
 Internet
 efficient web page storage,
 272–273
 fax standard, 402
 HTTP performance improvement,
 272
 MPEG-4 transportation on,
 422–423
 Intervals
 in arithmetic encoding process,
 106–109
 positive-length and disjoint,
 118–119
 rescaling, 114–115, *see also*
 Renormalization
 sequences represented by,
 19–24
 Intrapixel differencing, 389
 Invariance Theorem, 40, 45
 Inverse arithmetic operations, 120
 Inversion frequencies, 188
 Inversion ranks, 190
 Irreversible color transform,
 JPEG2000, 354–355
 ITU-T Recommendations
 Group 3 (T.4), 398–399
 Group 4 (T.6), 399

J

JBIG
 data structure and formatting,
 336–338
 decoding, 338–339
 encoding
 adaptive arithmetic coding, 336
 bottom layer typical prediction,
 334
 differential layer prediction,
 332–334
 model templates, 334–336
 resolution reduction, 330–332
 overview of encoding/decoding,
 327–330
 standard (T.82), 399
 JBIG2
 decoding control and data
 structures, 346–348

generic refinement region
 decoding, 342–343
 generic region decoding, 341–342
 halftone region decoding, 346
 overview, 339–341
 pattern dictionary decoding,
 345–346
 standard (T.88), 399
 symbol dictionary decoding,
 343–345
 text region decoding, 345
 JPEG2000
 features, 352–354
 compressed domain image
 processing/editing, 353
 progression, 353–354
 performance, 366–369
 JPEG2000 algorithm
 bit-plane coding, 360–364
 JPEG2000 codestream, 365–366
 packets and layers, 364–365
 quantization, 358–360
 tiles and component transforms,
 354–355
 wavelet transform, 355–358
 JPEG hardware, 417
 JPEG-LS
 entropy coding, 307–309
 modifications for multicomponent
 images, 304–306
 overview, 301–302
 prediction error correction, 306
 single-component images,
 303–304

K

Karp–Rabin fingerprints, 164–165
 Kolmogorov complexity
 basic definitions, 39–40
 computational issues, 44–47
 historical perspective, 51
 incompressibility, 40–42
 lower-bounding, 46–47
 notion of describing data, 38
 prefix-free encoding, 42–44
 relationship to Shannon
 Information Theory, 47–51
 resource-bounded, 45–46
 Kraft–McMillan inequality, 84–85
 Kraft's inequality
 in competitive optimality, 26
 for prefix code, 14–17

L

Laplacian pyramid structure,
 reduced, 221

Layers, JPEG2000, 364–365
 Learning-based switch predictors,
 216–217
 Least frequently used heuristic, 161
 Least recently used updating, in
 symbol-ranking compressor,
 200–201
 Least significant bits, 259–260
 Lempel–Ziv encoder, 412–415
 processing element, 414–415
 Lempel–Ziv-77 method, 159,
 232–233, 273, 407–408
 delta compressors based on,
 274–275
 LZU compressor, 294–296
 Unicode compressors, 294
 Lempel–Ziv-78 method, 160,
 232–233, 407–408
 Length-constrained Huffman codes,
 89
 Length-index preserving transform,
 236–240
 Letter index transform, 242–243
 Levenstein codes, 58–59
 Lexicographical ordering
 FGK algorithm, 91
 over source sequence, 21
 Lifting, in hierarchical lossless image
 coding, 221–222
 Linear first-order unit-delay
 predictor, 319
 Line interleaving, 305–306
 Line Not Typical, fictitious pixel,
 332–333
 Lossless audio data compression
 amplitude range and segmentation,
 259–260
 basic redundancy removal,
 257–258
 entropy coding, 262–263
 multiple-channel redundancy,
 260
 numerical implementation and
 portability, 263
 prediction, 260–262
 segmentation and
 resynchronization, 263–264
 software systems
 MLP, 266
 PCA, 266–267
 Shorten, 265
 speed and complexity, 264–265
 variable bit rate, 264
 Lossless compression algorithms
 dictionary methods, 232–233
 performance comparison,
 233–234
 statistical methods, 229–232
 transform-based methods, 233

Lossless image compression
 bilevel, *see* JBIG; JBIG2
 combined predictors, 217–219
 error modeling, 212
 hierarchical prediction, 211–212
 JPEG-LS
 decoding, 309
 encoding, 302–309
 overview, 301–302
 probability mass function
 prediction, 208
 pyramid coding scheme, 220–222
 scanning techniques, 212–214
 spatial prediction, 209–211
 switched predictors, 214–217
 Lossy audio compression, 256
 Low-entropy options, adaptive
 entropy coder, 316–317
 LSI Logic products, 428–430
 Lucent system-layer decoder, 428
 LZW algorithm, 160

M

Macro schemes, dynamic dictionary
 methods, 157
 Magnitude refinement pass, bit-plane
 coding, 363–364
 MARVLE architecture, 407
 elements of, 411–412
 memory mapping, 410
 Mask layer, facsimile, 401–402
 Maximum ambiguity, of current and
 reference files, 284–285
 Median edge detection, predictor
 based on, 216
 Memory-efficient algorithms, 95
 Memoryless sources, 10
 fixed-length codes for, 26–34
 variable-length codes for,
 11–18
 Memory mapping, 409–410
 Meridian Lossless Packing, 266
 Minimum description length
 principle, 39
 Mixed Raster Content (MRC)
 standard, 399–402
 MMR coding, JBIG2, 340–341
 Model templates, JBIG encoding,
 334–336
 Modem, compression standards,
 166
 Modified Huffman algorithm, 393
 Modified Huffman codes, 87
 Modified READ algorithm, 393–397
 Most significant bits, 259
 JPEG2000, 360–361
 Motorola, MPEG-2 video decoder,
 436

Move-to-Front recoding, 170,
 174–176, 185, 190
 Move-to-Front step
 Burrows–Wheeler, 183–185
 elimination, 187–189
 BWT, 233
 MOVIE project, 425
 MPEG-2
 digital audio/video decoder by
 IBM, 436–437
 video decoder by Motorola,
 436
 MPEG-4
 decoder by Sigma Designs, Inc.,
 438
 products by Toshiba, 437–438
 transportation on Internet,
 422–423
 video codec by Toshiba, 423–425
 MPEG 1 Layer 3, 256
 Multicomponent images,
 modifications for, 304–306
 Multiple-channel redundancy, 260
 Multiple-Image Network Graphics,
 388–390
 Multiplications, inexact, 115–118

N

N-ary Huffman codes, 83–84
 NEAR, JPEG-LS, 302, 306, 309
 NEC audio video codec, 430–431
 Nested loops, 389
 Nested subintervals, 119–120
 Noiseless source-coding theorem,
 Shannon's, 48, 50
 Noncomputability, 36–37
None filter, 380
 Non-prefix codes, subtle problems
 with, 43–44
 Notation, arithmetic coding,
 103–104
 Number index transform, 241–243

O

Occam's Razor, 39
 Omega code, Elias, 59
 One-step lookahead, greedy parsing
 with, 156–157, 164
 Optimality
 arithmetic coding, 111–112
 competitive, 25–26
 Optimal tree-based search,
 143–144
 Order-3 Fibonacci code, 69–70
 Original external pointer macro
 scheme, 157

Original pointer macro scheme, 157
 Output fragments, good and bad,
 185–186

P

Packetization algorithm, MPEG-4,
 422–423
 Packets, JPEG2000, 364–365
 Packet telemetry, 248
Paeth filter, 380–383
 Palindrome, complementary, 162
 Parsing
 bidirectional, 161
 issues in dictionary construction,
 155–157
 modifications, 160–161
 Patch, of minimal size, 270
 Pattern dictionary, decoding,
 345–346
 Peak bit rate, 264
 Peer-to-peer systems, 280
 Perceptual coding, lossy,
 255–256
 Perfect Clarity Audio, from Sonic
 Foundry, Inc., 266–267
 Performance
 HTTP, improvement, 272
 Huffman codes, 84–86
 JPEG2000, 366–369
 lossless compression algorithms,
 comparison, 233–234
 timing, degraded, 240–241
 Per letter entropy, 11
 Permutation, Burrows–Wheeler,
 173–174, 181–183
 PNG, *see* Portable Network Graphics
 Polynomial representations,
 56–57
 Portable Network Graphics
 compression engine, 374–376
 compression tests and
 comparisons, 385–388
 design decisions, 372–374
 filters, 378–383
 historical background, 371–372
 MNG, 388–390
 practical compression tips,
 383–385
 zlib format, 376
 zlib library, 376–378
 Prediction
 bottom layer typical, 334
 differential layer, 332–334
 lossless audio data compression,
 260–262
 process, symbol-ranking
 compressors, 198–199

- Prediction by Partial Match (PPM)
 - Burrows–Wheeler transform and, 180
 - constant-order compressor, 295
 - finite-context statistical compressors, 293
 - method C, 231
 - method D, 237–241
 - Prediction error correction, 306
 - Prediction error mapper, 320
 - Predictors
 - combined, 217–219
 - FIR and IIR, 262
 - linear and nonlinear, 211
 - Paeth, 380–383
 - preprocessor, 318–319
 - Shannon coder, 196
 - switched, 214–217
 - symbol, 198
 - using previous pixel in image, 210
 - Prefix codes
 - codewords in, 13–14
 - in Kolmogorov complexity, 42–44
 - Kraft’s inequality for, 14–16
 - sharing of prefixes, 81–83
 - Preprocessing
 - Burrows–Wheeler compression, 181
 - prediction error mapper, 320
 - predictor, 318–319
 - reference file, 274
 - reference sample, 319–320
 - Probabilities, *see also* Error
 - probability
 - consistent, 10
 - for memoryless sources, 27
 - shift invariant, 11
 - static and dynamic, 228–229
 - Probability distribution
 - combined, 217
 - dyadic, 24–26
 - given for free, 49–50
 - over source sequences, 23
 - Probability domain blending, of
 - subpredictors, 219
 - Probability mass function,
 - prediction, 208
 - Processing elements, Lempel–Ziv, 413–415
 - Programmability, hardware systems, 407
 - Programmable logic, 96
 - Programming language, *see* Descriptive language
 - Progressive-compatible sequential mode, JBIG, 328, 330
 - Progressivity dimensions, JPEG2000, 353–354
 - Pyramid coding scheme, 211–212
- Q**
- Quality progressive representations, 340
 - Quantization
 - JPEG2000, 358–360
 - and prediction error correction, 306
- R**
- Recoding, MTF, 170, 174–176, 185, 190
 - Reconstruction
 - and prediction error correction, 306
 - progressive, in JBIG transmission, 328
 - Recursively enumerable set, 45
 - Reduced-pyramid structure, 221
 - Redundancy
 - coding, acceptable, 22–23
 - individual and expected, 16
 - multiple-channel, 260
 - removal, 257–258
 - Reference bitmap, refined, 342–345
 - Reference file, delta compression, 270–271
 - choosing, 278–279
 - preprocessing, 274
 - Reference sample, preprocessor
 - predictor, 319–320
 - Remote file synchronization
 - applications, 279–280
 - balanced pairs, 283–284
 - database records and file systems:
 - reconciling, 286–287
 - distance measures, 283
 - results, 285
 - estimation of file distances, 286
 - fundamental results, 284–285
 - problem definition, 270–271
 - rsync* algorithm, 280–281
 - experimental results, 282–283
 - Renormalization, 121–132, 142–143
 - Rescaling, interval, 114–115, *see also* Renormalization
 - Resolution reduction, JBIG
 - encoding, 330–331
 - Resynchronization, segmentation
 - and, 263–264
 - Reverse binary tree, 408–409
 - Reversible color transform, JPEG2000, 354–355
 - Revision Control System, software package, 271–272
- S**
- Rice codes, 60–62, 75, *see also* e.Rice algorithms
 - rsync* algorithm, 280–281
 - experimental results, 282–283
 - Run-length coding, 179–180
 - Run-length color encoding, 398
 - Run-lengths, and Wheeler 1/2 code, 73–75
 - Run mode, JPEG-LS, 304
 - entropy coding, 307–309
- S**
- Sample interleaving, 306
 - Scalability, architecture, 406
 - Scanning techniques, in lossless
 - image compression, 212–214
 - Second-extension option, adaptive
 - entropy coder, 316
 - Segmentation
 - amplitude range and, 259–260
 - resynchronization and, 263–264
 - Self-delimiting codes, 55
 - Self-delimiting process, 43
 - Sequences
 - biomolecular, compression, 162–163
 - complex nonrandom, 37
 - empty, 9
 - representation by intervals, 19–24
 - of sources, infinite, 48–49
 - Settled symbols, 131
 - Shannon coder, 196
 - Shannon–Fano coding, 80
 - Shannon Information Theory
 - complex nonrandom sequences, 37
 - relationship to Kolmogorov complexity, 47–51
 - structured random strings, 37–38
 - Shorten* software, 265
 - Sigma Designs, Inc., MPEG-4
 - decoder, 438
 - Single-chip decoder, by Zoran, 435
 - Single-component images, 303–304
 - Single-pass methods, 158
 - Single-progression sequential mode, JBIG, 328
 - Software programs, lossless audio
 - data compression, 265–267
 - Sorting algorithm, improvements to, 182–183
 - Source coding system
 - fixed-length codes for memoryless sources, 26–34
 - variable-length codes for source symbols, 12–13
 - Source modeling, separation from
 - coding, 113–114
 - Source samples, blocks, 312–313

Source symbols
 blocks, 17–18
 variable-length codes for, 12–13

Space applications
 adaptive entropy coder, 313–317
 coded data format, 321
 e-Rice algorithm, 312–313
 implementation issues and
 applications, 324–326
 lossless decoder, 321–323
 preprocessing stage, 318–320

Space-constrained delta
 compression, 277–278

Spatial prediction, in lossless image
 compression, 209–211

Speed-efficient algorithms, 95–96

Splay tree algorithm, 93–94

Split-sample option, adaptive entropy
 coder, 314–315

Star (*) transformation, 234–236

Start–step–stop codes, 64–65

Static dictionary, 154–155

Stationary sources, discrete, 10–11

Statistical coding
 Burrows–Wheeler compressor,
 176–180
 for lossless algorithms, 229–232
 Statistical compressors, 185–187
 finite- and unbounded-context,
 293

Sticky MTF, 184–185

STMicroelectronics, video decoder,
 433–435

String-matching, PNG, 378

Strings
 algorithmic prefix complexity,
 42–43
 incompressible, 41–42
 run-length compressed, 73–75
 searching, fingerprints for,
 164–165
 vs. sources, 37
 structured random, 37–38

Stripes, document partitioned into,
 400–401

Structured coding model, 179

Sub filter, 380

Subpredictors, combining, 217–219

Suffix-complete dictionary, 156

Suffix trees
 as alternative to explicit sort,
 182–183
 construction, 163–164

Summation codes, 71–73

Switched predictors, 214–217
 explicit detection-based, 215–216
 learning-based, 216–217

Symbol dictionary, decoding,
 343–345

Symbol-ranking compression,
 195–201
 and Burrows–Wheeler
 compression, 200
 compressor history, 197
 fast, 200–201
 prediction flagging, 197–198
 Shannon coder, 196

Symbol search
 bisection search, 143
 on sorted symbols,
 144–145
 optimal tree-based, 143–144
 sequential search on sorted
 symbols, 143

Synchronization, *see also* Remote
 file synchronization
 encoder–decoder decisions,
 113
 file, BWT in, 189

T

Tandem repeats, 162

TANGRAM coprocessor, 420

Telemetry, description of,
 247–250

Telemetry compression
 existing, 252–253
 issues in, 250–251

Ternary comma codes, 70–71

Ternary polynomial representation,
 56–57

Test compressors, 294

Text compression
 ACB algorithm of Buynovsky,
 201–203
 classification of lossless
 compression algorithms,
 229–234
 hardware, 407–415
 Lempel–Ziv encoder, 412–415
 tree-based encoder, 408–412
 initial letter preserving transform,
 241–243
 length-index preserving transform,
 236–240
 letter index transform, 241–243
 number index transform,
 241–243
 star (*) transformation, 234–236
 timing performance
 measurements, 240–241

Text region, decoding, 345

Textual substitution, compression
 with, 153

TIFF-FAX standard, 402

Tile-parts, 366

Tiles, JPEG2000, 354, 364

Timing performance, degraded,
 240–241

Toshiba
 MPEG-4 products, 437–438
 MPEG-4 video codec, 423–425

Transform-based methods, for
 lossless algorithms, 233

Tree-based encoder, 408–412
 MARVLE architecture, 411–412
 memory mapping, 409–410

Tree-structured codebook, Huffman,
 81–83

Trie–reverse trie pairs, 164

Tries, and compact tries, 163

Tunstall code, 88–89

U

Unary codes, 57–58, 307

Uncertainty, resolved through
 information, 5–6

Unicode
 character codings, 291–293
 big-endian vs. little-endian,
 292
 UTF-8 coding, 292–293

compression
 comparison of compressors,
 295–296
 finite- and unbounded-context
 statistical compressors,
 293
 LZ77 compressors, 294
 file test suite, 294–295

Unidirectional dictionary methods
 construction modifications, 160
 dictionary maintenance in
 dynamic methods, 161
 historical overview, 158
 Lempel–Ziv-77, 159
 Lempel–Ziv-78, 160
 LZ77 variants, 159
 parsing modifications, 160–161

Uniquely decipherable code, 230

Uniquely decodable code, 13–14

Universal codes
 characteristics, 55–56
 comparison of representations,
 75–77
 Elias omega and Even–Rodeh
 codes, 59–60
 Fibonacci codes, 65–70
 Golomb codes, 62–64
 Levenstein and Elias gamma
 codes, 58–59
 polynomial representations,
 56–57

Rice codes, 60–62
 start–step–stop codes, 64–65
 summation codes, 71–73
 ternary comma codes, 70–71
 unary codes, 57–58
 Wheeler 1/2 code and run-lengths,
 73–75
Up filter, 380
 User files, synchronization, 279–280

V

Variable-length codes
 for memoryless sources, 11–18
 for sources with memory, 18–26
 Varn's algorithm, 97
vcdiff, compared to *zdelta* and *xdelta*,
 275–277
 Video compression hardware,
 417–442
 commercial video and audio
 products, 426–442

compressed bitstream scaling
 architecture, 421
 DCT, 419–420
 MOVIE project, 425
 MPEG-4 transportation on
 Internet, 422–423
 MPEG-4 video codec by Toshiba,
 423–425
 TANGRAM coprocessor, 420
 Video object plane size, 422–423
 Vitter's algorithm Λ , 93
 VLSI architectures, technology
 advances, 406–407

W

Wavelet architectures, 416–417
 Wavelet transform, JPEG2000,
 355–358
 Web page
 commercial, for vendors, 442
 efficient storage, 272–273
 Wheeler 1/2 code, 74–75

X

xdelta, compared to *zdelta* and *vcdiff*,
 275–277
 Xdelta File System, 272

Z

zdelta, compared to *xdelta* and *vcdiff*,
 275–277
 Zeckendorf representation, 66
 Zero-block option
 adaptive entropy coder,
 316–317
 coded data format, 323
 testing and, 324
 Zero-frequency leaf, 93
 Zip variants, of LZ77, 159
 Ziv–Lempel dictionary-based
 scheme, 89, 97, 232–233, *see*
 also Lempel–Ziv-77 method
 zlib, format and library, 376–378
 Zoran, single-chip decoder, 435