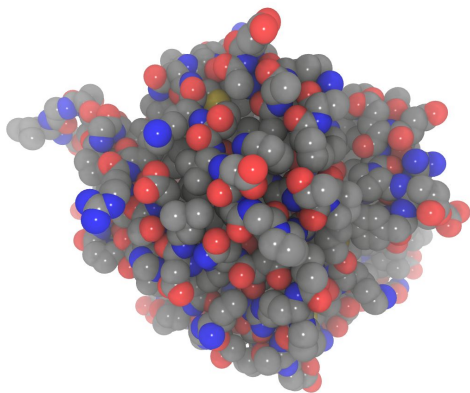


EEG signals - initial analysis

Martin Maňák
11.3.2025

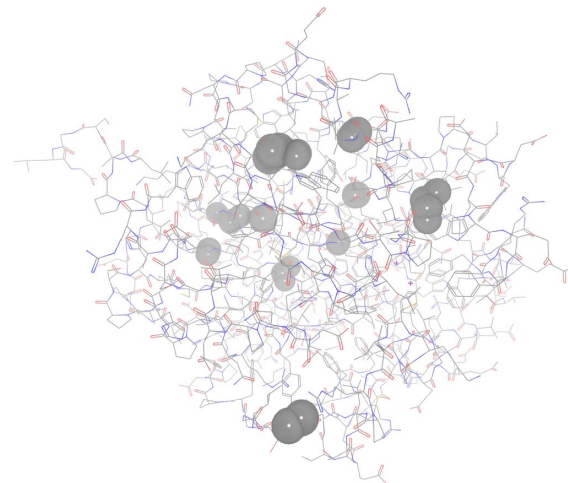
Compression of molecular structures



atoms represented as balls
scale: $1\text{\AA} = 10^{-9}\text{m}$



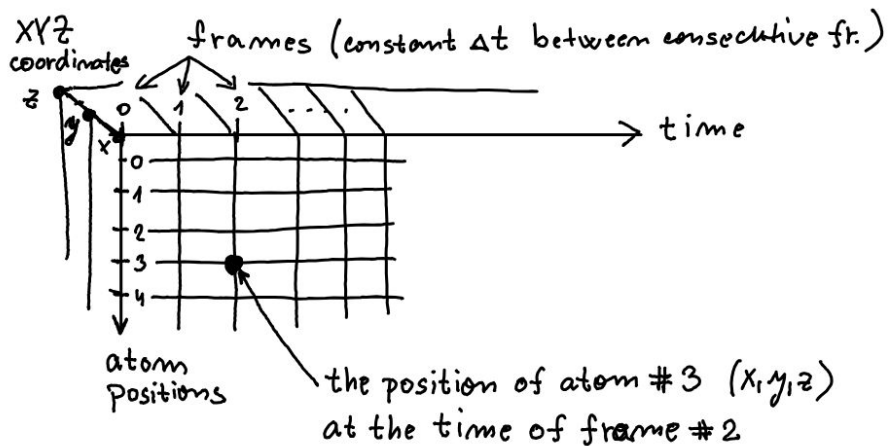
backbone (alpha-trace)



bonds between atoms

- Molecular dynamics: simulated motion of atoms.
- Constraints - bonds, forces, ...

Compression of molecular structures



N atoms, F frames, 32-bit float per coordinate

$\Rightarrow N \cdot F \cdot 3 \cdot 4$ BYTES to store the trajectory

(real datasets ~ 5 GB (or even more))

Compact representations of trajectories

- NetCDF files (.nc)
 - zlib (deflate)
- Gromacs trajectory files (.xtc)
 - float to integer coordinates
 - store with reduced precision
- High-res. traj. compression (HRTC)
 - traj \rightarrow piecewise lin. func.
 - (tolerance corridor)
 - 1 bit per sample (particle simul.)
- Predictive compression (PMC)
 - uses the structure of bonds
 - next-frame positions prediction
 - error control (e.g., $\text{diff} < 10^{-3}\text{\AA}$)



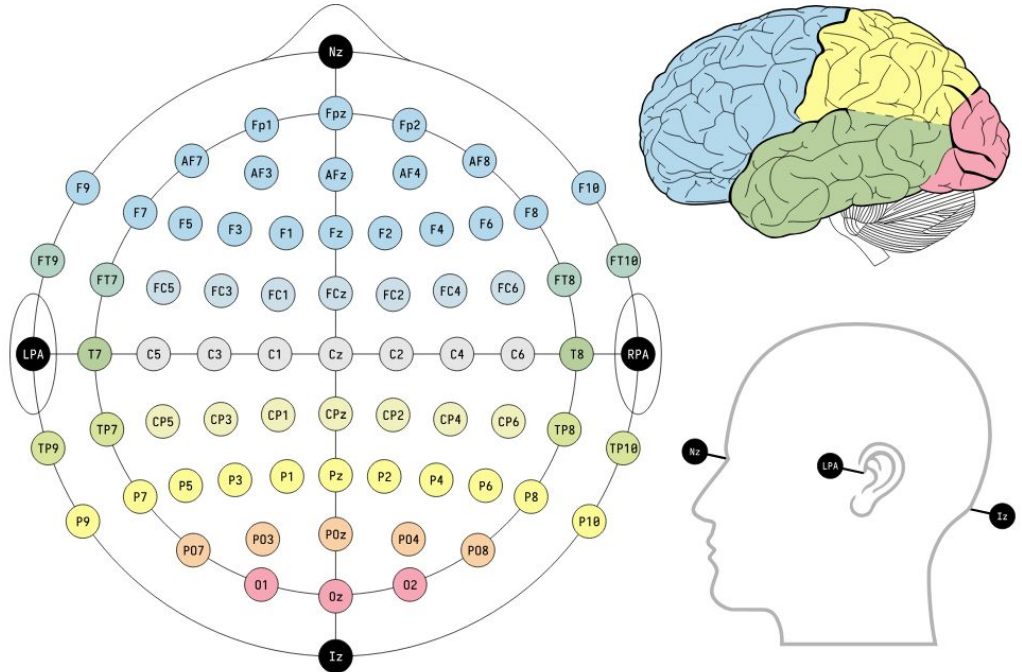
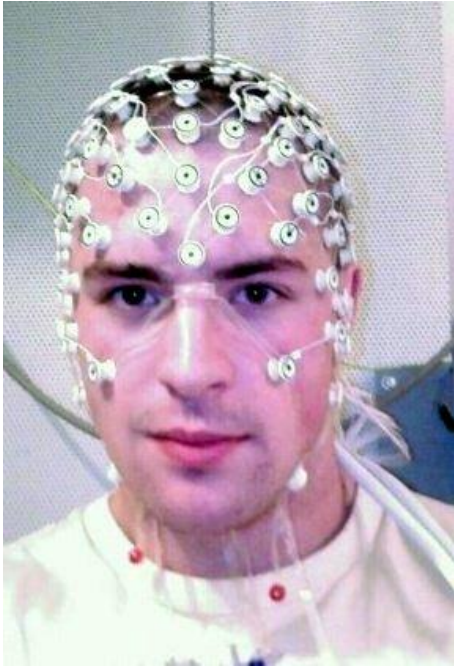
Applicability to digital EEG signals is limited

- different kind of data
- voltage measured at electrodes (up to 257)
- 1000 samples per second
- stored as 32-bit floats or 64-bit doubles
- reduction of precision is “problematic”
- no such connectivity as bonds in molecules



<https://commons.wikimedia.org/wiki/File:ElectroEncephalogram.png>

Data acquisition (wiki example)



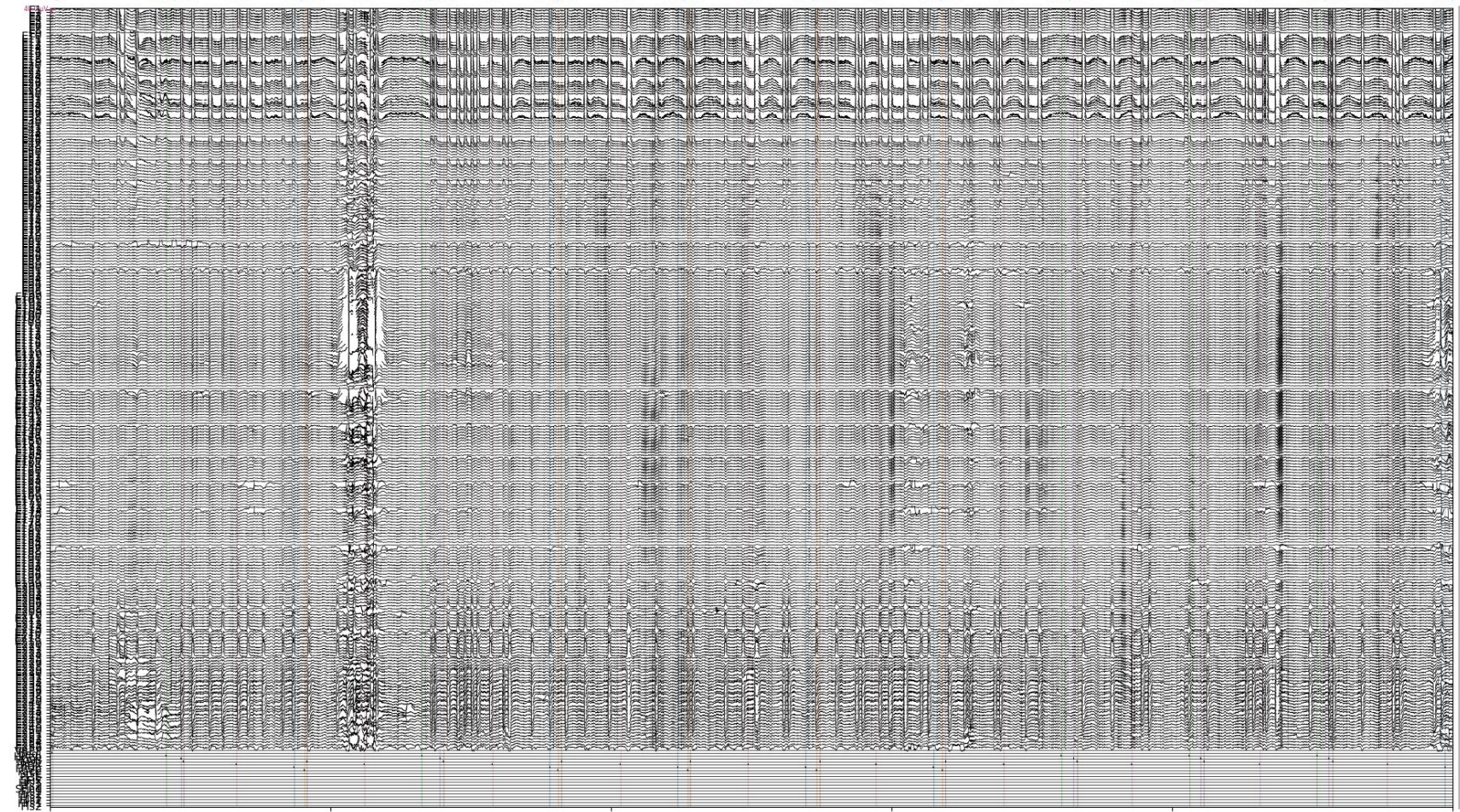


NUDZ Binocular rivalry dataset





1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525



Help

Government	Percentage
Current government	100%
Previous government	0%



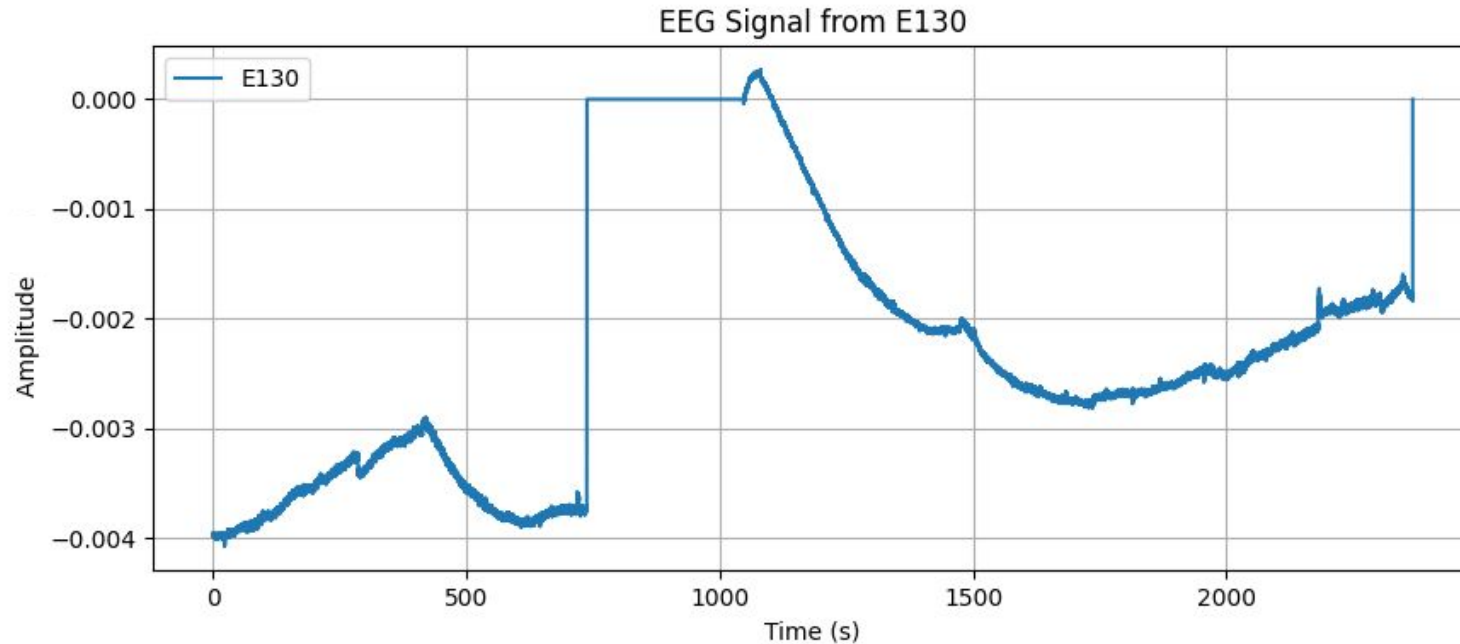
The size of EEG binary data (EGI MFF dataset)

- 257 EEG channels
- 2,3M samples/channel (1,000 samples/s, 40 minutes of recording)
- stored as 32-bit floats (probably), loaded as 64-bit doubles

- $257 * 2,367,324 * 4$ bytes
=>
2,1 GB file (signal1.bin)
+ some 1 MB of meta-data

- Q: Could we compress the data somehow? Some correlation is visually apparent...

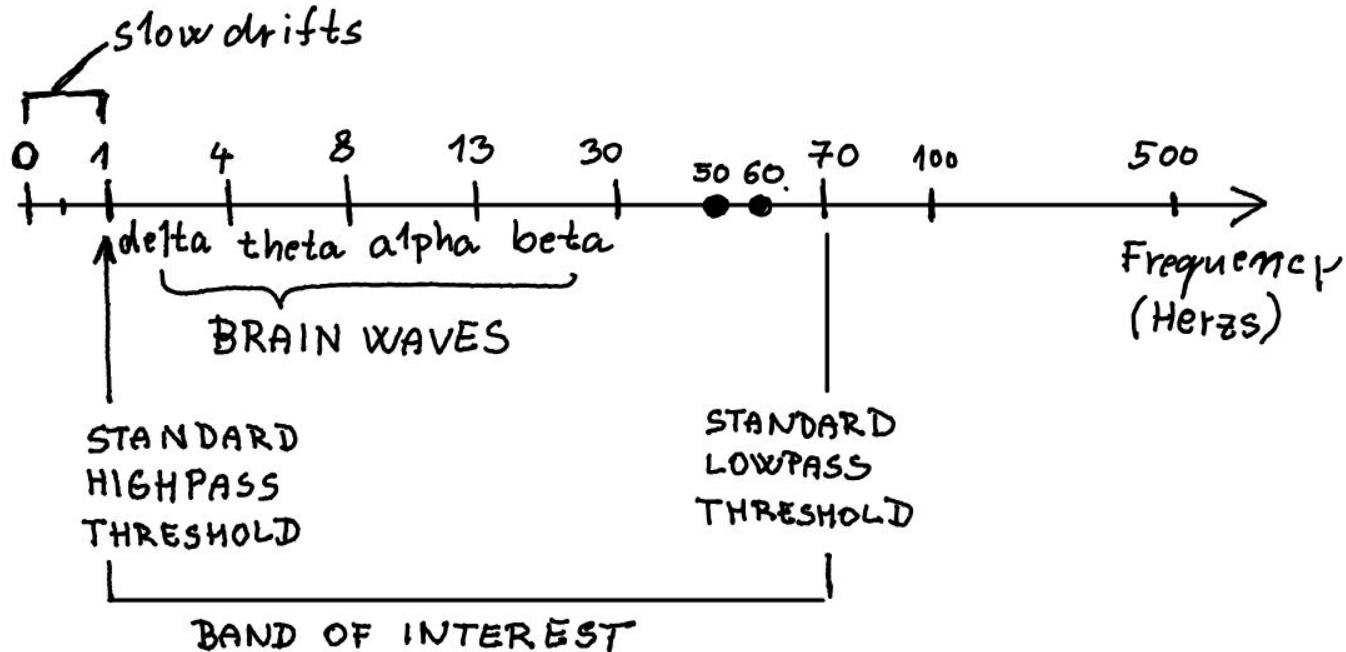
Actual RAW data for a single electrode (randomly selected one)



A spectral band (ai generated, DALL-E)

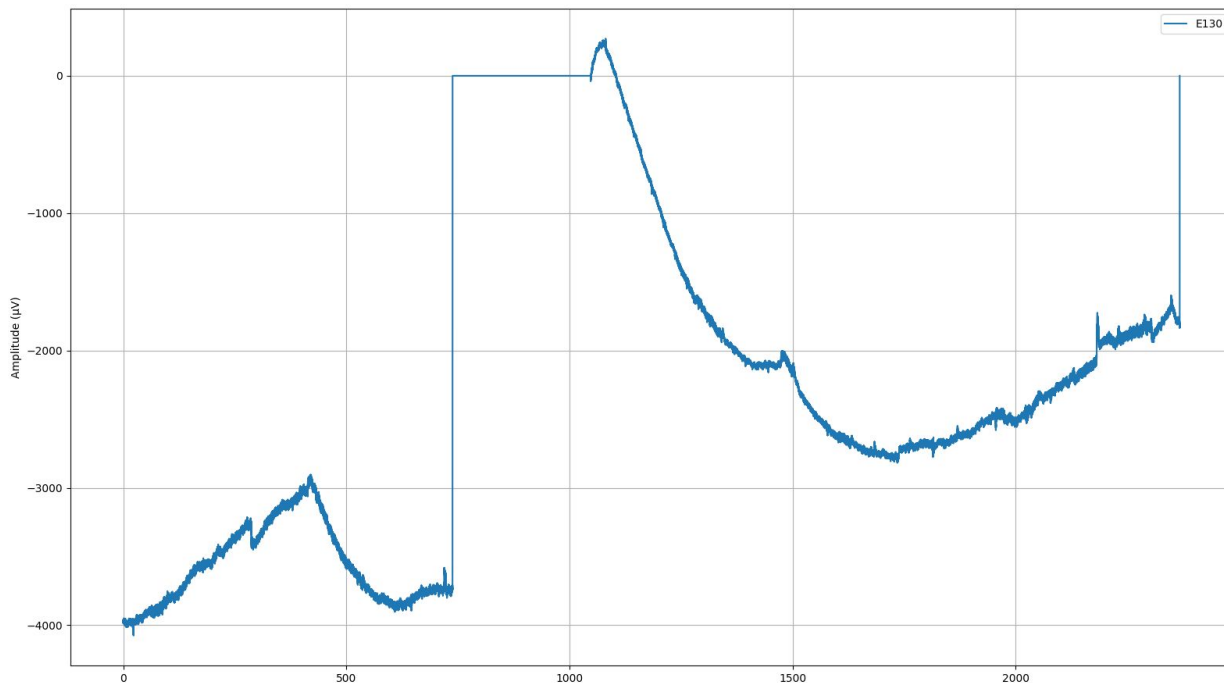


Spectral bands / EEG signals





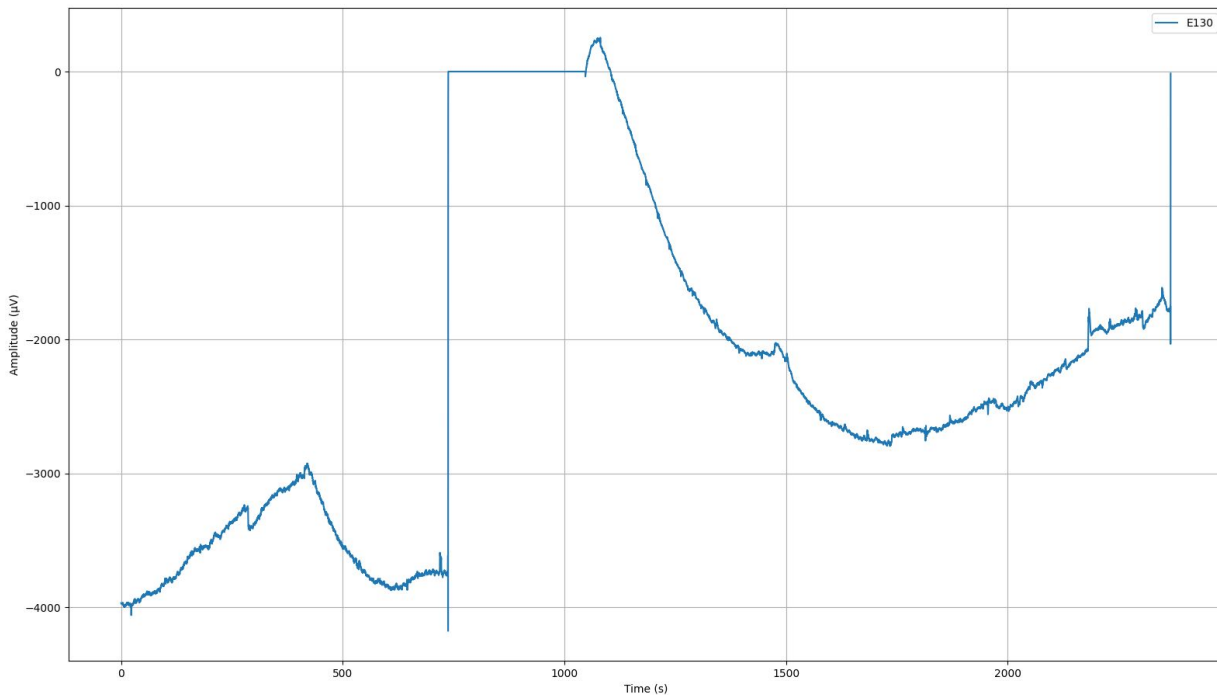
Channel E130 - RAW data



- Frequencies up to 500 Hz
- Nyquist:
 - need 1000 samples / s to represent the signal
 - means **2,1 GB** for all channels together



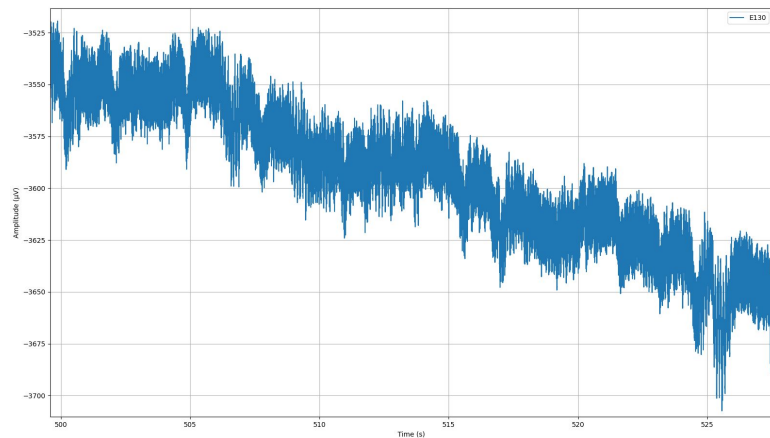
Channel E130 - lowpass 1 Hz



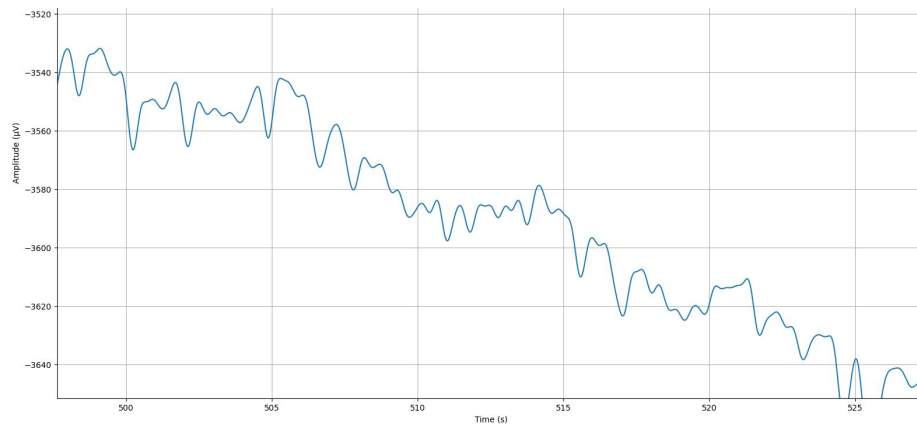
- Frequencies up to 1 Hz
- Nyquist
 - need 2 samples / s to represent the signal
- We have: 1000 samples/s
- Downsampling (after lowpass filtering)
 - keep every 500th sample (0.2 %)
 - means **4,2 MB** for all channels together
- Upsampling (to get it back)



Channel E130 - a closer look on details



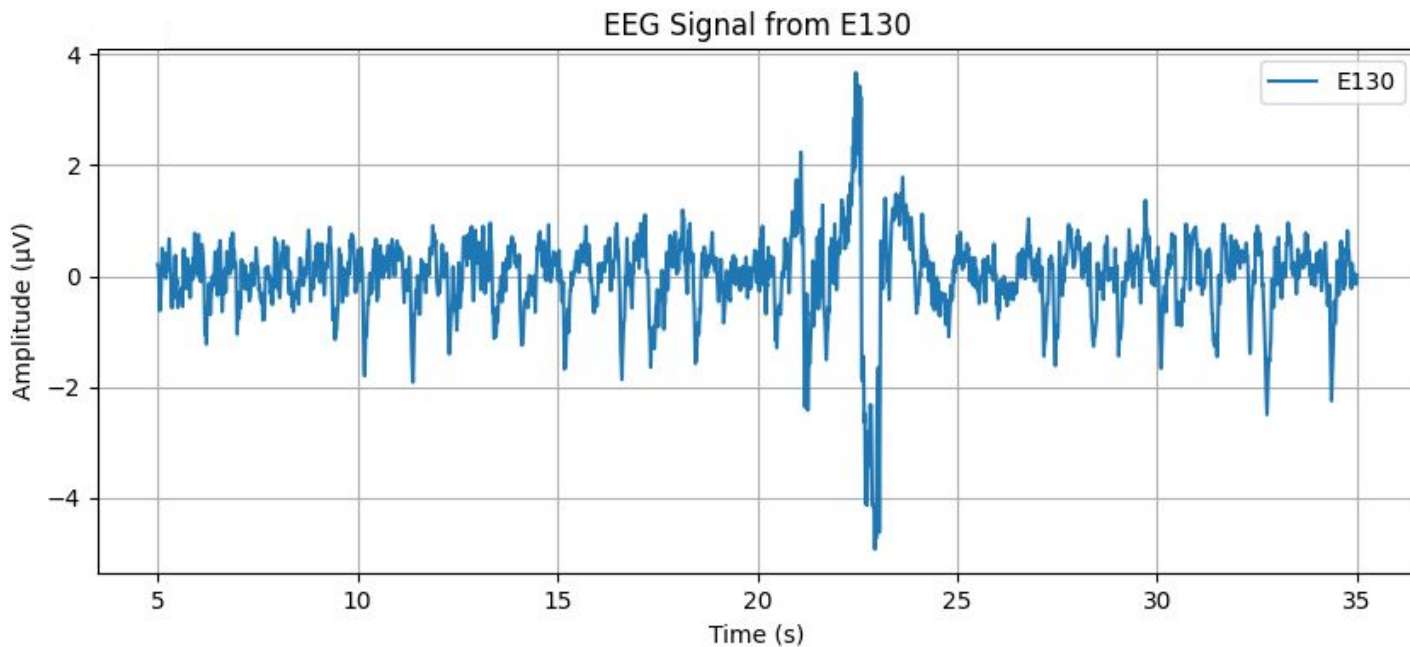
25 seconds, RAW data, 1000 samples/s



25 seconds, 1Hz lowpass (yet still 1000 samples/s)

Channel E130, bandpass 1-30Hz

The filter what a domain-expert might set up to analyze the signal (may further stretch the signal horizontally/vertically according to standards - sensitivity / reading speed)





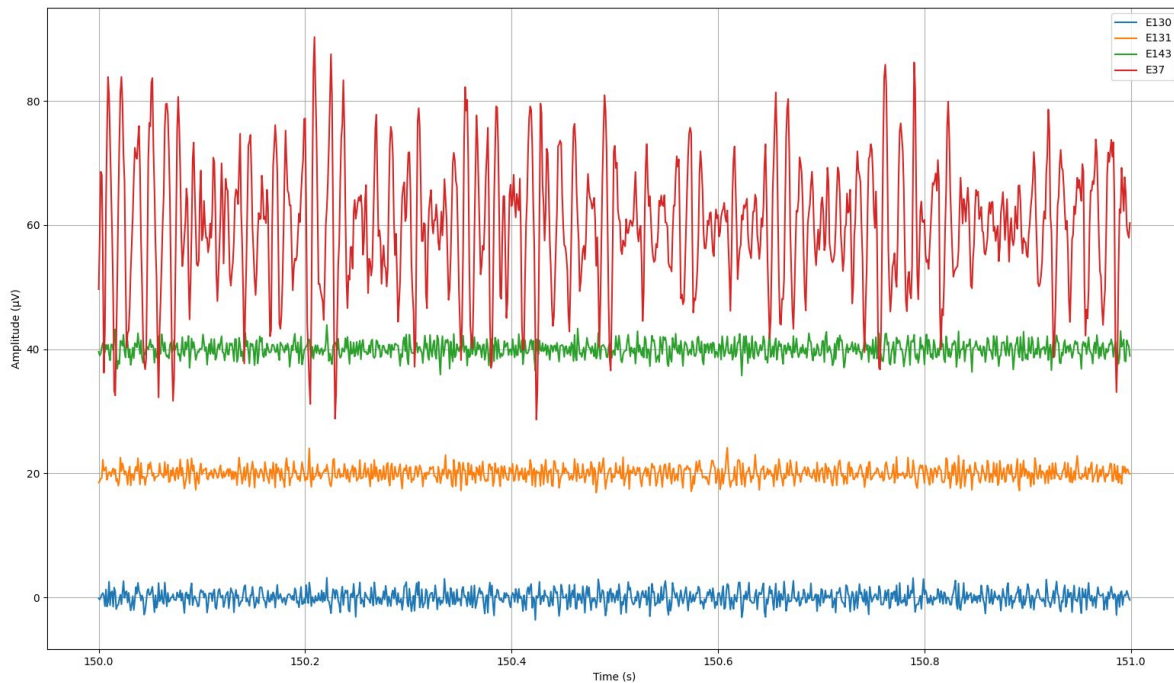
Spectral bands - some conclusions

worth to keep

- 0-1Hz (slow drifts)
 - domain-experts filter-out these frequencies explicitly
 - **0.2 %** of data (**4-5 MB** out of **2.1 GB**)
- 1-30-100 Hz (interesting brain waves, spikes, artifacts)
 - sampling rate 60-200 Hz
 - **5,8 - 19,8 %** of data (**126-420 MB** out of **2.1 GB** - if we are 32 bits/sample generous)
- 100-500 Hz (mostly noise)
 - also epileptic seizure spikes, some drug-effects, high temporal resolution, etc.
 - **94,2 - 80,2 %** of data (**1.68-1.974 GB** out of **2.1 GB**)



...mostly noise / residual content



70 Hz highpass filter

E130, E131, E143 are close to each other (on the head-mount). E37 is far.

“low” amplitude (a few uV)



Final thoughts

Is it worth to further compress the interesting 10% of data if we have to keep the remaining 90% of noise? (lossless compression)

- Probably not so much.

How to define information loss threshold?

- Perhaps as the ability to restore data to 1 uV resolution (basic unit)

How to compactly represent high frequency noise (the residual content)?

- An appropriate number-coding scheme could give us $1 + \log_2(\text{std_dev})$ bits / sample