

YGNEQOG
VQ
ETARVQITCRJA NGEVWTG

DO YOU KNOW
WHAT YOU SHOULD THINK ABOUT MOST OF
YOUR TIME
????

**MOST OF YOUR TIME
YOU SHOULD THINK ABOUT
WHAT YOU SHOULD THINK ABOUT
MOST OF YOUR TIME
!!!!!!!!!!!!**

IV045, CODING THEORY, CRYPTOGRAPHY

and

CRYPTOGRAPHIC PROTOCOLS - 2020

Prof. Jozef Gruska
<http://www.fi.muni.cz/usr/gruska/crypto20>

Technické řešení této výukové pomůcky je spolufinancováno Evropským sociálním fondem a státním rozpočtem České republiky.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Old cryptography focused, until the end of 19th century, on the **art** of **designing and breaking secrecy codes**.

Modern cryptography has significantly enlarged its scope. It enlarged its scope to designs and rigorous analysis of any system that is a potential subject to malicious attacks and threats and to the design of system than can withstand such threats and attacks.

As a consequence many **new goals have been added to modern cryptography** and they will also be deal with in this lecture concerning contents.

Cryptography has also moved from an **engineering art** concentrating on **heuristic techniques** to both a **scientific and engineering discipline** concentrating on **rigorous and efficient techniques** and correctness proofs. All that will also be reflected in the style of this lecture.

- 1 Basics of coding theory
- 2 Linear codes
- 3 Cyclic, convolution and Turbo codes - list decoding
- 4 **Secret-key cryptosystems**
- 5 **Public-key cryptosystems, I. Key exchange, knapsack, RSA**
- 6 **Public-key cryptosystems, II. Other cryptosystems, security, PRG, hash functions**
- 7 **Digital signatures**
- 8 **Elliptic curves cryptography and factorization**
- 9 Identification, authentication, privacy, secret sharing and e-commerce
- 10 Protocols to do seemingly impossible and zero-knowledge protocols
- 11 Steganography and Watermarking
- 12 **Quantum cryptography**
- 13 **History and machines of cryptography**

BASIC INFORMATION - I.

- Usually, some important things will be said at the very beginning of each lecture.
- **Materials/slides of the lecture will be on <http://www.fi.muni.cz/usr/gruska/crypto20> and in IS, mostly 1-2days before scheduled lecture.**
- Videos of each lecture will appear, after postprocessing, one week later in the lecture materials in IS.
- **For each of the first 10 lectures there will be home exercises. They will be posted in my web page and in IS always on Tuesday before the lecture, at 18.00 .**
- At the lecture web page and in IS you also find instructions how to submit solutions of exercises and how they will be evaluated.

BASIC INFORMATION - II.

- There will be also nonobligatory exercise-tutorial sessions for this course. They will discuss subjects dealt with in the lecture in more details.
RNDr Matej Pivoluska PhD will charge tutorials.

Tutorials, will discuss, in details, some of key points or new examples related to lectures.
Tutorials (in English) in the form of videos will be inserted every week into the course study materials.

It ill be possible to ask questions related to tutorials (or lectures) via Google Hangout Meets (each Thursday in time 9.00-9.15 or at 10.00-10.15) or via discussion forums of tutorials or course.

Likely, the most efficient use of the lectures is to print materials of each lecture before the lecture and to make your comments into them during the lecture.

- Lecture's web page contains also access to so called **Appendix** -

Appendix contains few very basic and important facts from the number theory and abstract algebra that you should, but may not yet, know and you will need.

/bigskip

- **Read and learn Appendix carefully!**

- Whenever you find an error or a misprint in the lecture notes, let me know - extra points you get for that.

To your disposal there are also lecture notes called the "Exercises Book" that you can upload from the IS for the lecture IV054, through links "Ucební materialy – Exercise Book"

Exercises book (100 pages) contains selected exercises from the homeworks of the past lectures on Coding, Cryptography and Cryptography Protocols" with solutions. Exercise book is available

Lecture: Prof. Jozef Gruska DrSc

Tutorials: RNDr. Matej Pivoluska, PhD

Exercises creating and evaluating team:

RNDr. Lukáš Boháč, head

RNDR. Matej Pivoluska PhD

RNDR Luděk Matyska, doctorant

RNDr Libor Caha, doctorant

Bc Henrieta Michelova, one of the best of the 2019 course IV054

LITERATURE

- R. Hill: A first course in coding theory, Claredon Press, 1985
- V. Pless: Introduction to the theory of error-correcting codes, John Willey, 1998
- J. Gruska: Foundations of computing, Thomson International Computer Press, 1997
- J. Gruska: Quantum computing, McGraw-Hill, 1999
- A. Salomaa: Public-key cryptography, Springer, 1990
- D. R. Stinson: Cryptography: theory and practice, CRC Press, 1995
- W. Trappe, L. Washington: Introduction to cryptography with coding theory, 2006
- B. Schneier: Applied cryptography, John Willey and Sons, 1996
- S. Singh: The code book, Anchor Books, 1999
- D. Kahn: The codebreakers. Two story of secret writing. Macmillan, 1996 (An entertaining and informative history of cryptography.)
- Vaudenay: A classical introduction to cryptography, Springer, 2006
- J. Gruska: Coding, Cryptography and Cryptographic Protocols, lecture notes, <http://www.fi.muni.cz/usr/gruska/crypto17>
- J. Fridrich: Steganography in Digital Media, Cambridge University Press, 2010.
- J. Gruska and collective: Exercises and their solutions for IV054, 2015, FI, MU Brno; <http://www.fi.muni.c/xbohac/crypto/exercice-book.pdf>
- A. J. Menezes, P. C. van Oorschot, S. A. Vanstone: The Handbook of Applied Cryptography, 1996

HISTORY OF CRYPTOGRAPHY

The history of cryptography is the story of centuries-old battles between codemakers (cipharmakers) and codebreakers (cipherbreakers). It is an intellectual arms race that has had a dramatic impact on the course of history.

This ongoing battle between codemakers and codebreakers has inspired a whole series of remarkable scientific breakthroughs.

History is full of ciphers (cryptosystems). They have decided the outcomes of battles and led to the deaths of kings and queens.

Security of communication and data, as well as identity or privacy of users, are of the key importance for information society.

Cryptography, when broadly understood, is an important tool to achieve such goals.

STORIES

STORY I.

Mary - Queen of Scots - picture





Born: 1542

Crowned: 1543

Imprisoned: 1567

Trained - killed: 1586

- Mary was a catholic and in charge of the tron in England
- She was imprisoned by her sister in law, Elisabeth I, a protestant
- Mary was considered to be very beautiful and had many admirers.
- After spending 19 years in jail a group of her admirers established a communication with Mary with the goal to free Mary (and to put Mary on the tron in England).
- Main cryptographer of Elisabeth I, Sir Francis Walsingham, expected that and was able to decrypt special encrypted communication between Mary and her admirers.

Mary - cryptosystem she used

a b c d e f g h i k l m n o p q r s t u x y z
 0 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 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518

Zimmerman telegram - I. Story

On January 16, 1918, Arthur Zimmerman, German Foreign Affairs State Secretary,

sent, from Sweden, through US a special telegram to the Mexico government.

Telegraph suggested that Mexico should join the alliance with Germany in the case US would enter WWI against Germany, and should attack US.

This telegram was captured and decoded by British. They used the telegram to convince US president to declare war to Germany what very much influenced the outcome of the WWI.

Part I

Basics of the coding theory

Zimmerman's telegram II.

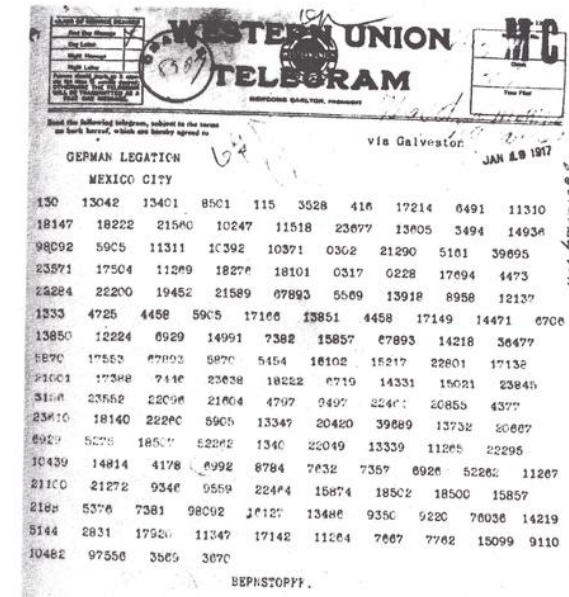


Figure 28 The Zimmermann telegram, as forwarded by von Bernstorff, the German Ambassador in Washington, to the British.

PROLOGUE - I.

PROLOGUE - I.

ROSETTA SPACECRAFT

- In 1993 in Europe **Rosetta spacecraft project** started.
- In 2004 Rosetta spacecraft was launched.
- In August 2015 Rosetta spacecraft got on the orbit of the comet 67P (one of 4000 known comets of the solar systems) and sent to earth a lot of photos of 67P.
- In spite of the fact that the comet 67P is 720 millions of kilometers from the earth and there is a lot of noise for signals on the way encoding of photos arrived in such a form that they could be decoded to get excellent photos of the comet.
- **All that was, to the large extent, due to the enormously high level coding theory already had in 1993.**
- **Since that time coding theory has made another enormous progress that has allowed, among other things, almost perfect mobile communication and transmission of music in time and space.**

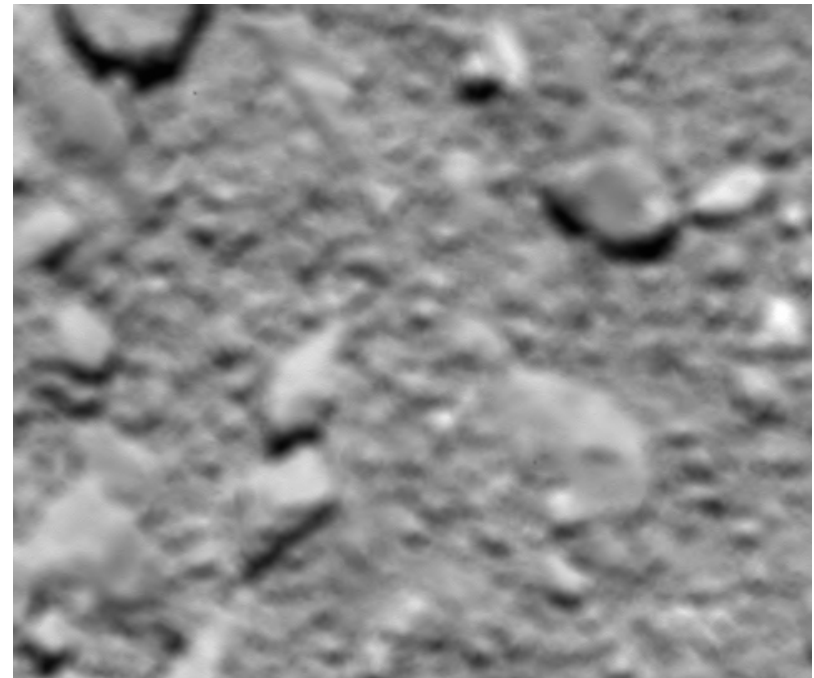
ROSETTA spacecraft



ROSETTA LANDING - VIEW from 21 km -29.9.2016



ROSETTA LANDING - VIEW from 51 m -29.9.2016



ABSTRACT

Coding theory - **theory of error correcting codes** - is one of the most interesting and applied part of informatics.

Goals of coding theory are to develop systems and methods that allow to detect/correct errors caused when information is transmitted through **noisy channels**.

All **real communication systems** that work with digitally represented data, as CD players, TV, fax machines, internet, satellites, mobiles, **require to use error correcting codes** because all **real channels** are, to some extent, **noisy** – due to various **interference/destruction caused by the environment**

- Coding theory problems are therefore among the very basic and most frequent problems of storage and transmission of information.
- Coding theory results allow to create reliable systems out of unreliable systems to store and/or to transmit information.
- Coding theory methods are often elegant applications of very basic concepts and methods of (abstract) algebra.

This first chapter presents and illustrates the very basic problems, concepts, methods and results of coding theory.

PROLOGUE - II.

INFORMATION

INFORMATION

is often an important and very valuable commodity.

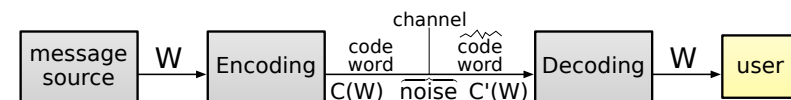
This lecture is about how to protect or even hide information

against noise or even unintended user,

using mainly classical, but also quantum tools.

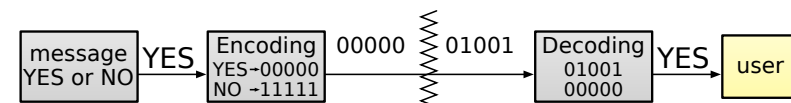
CODING - BASIC CONCEPTS

Error-correcting codes are used to correct messages when they are (erroneously) transmitted through **noisy channels**.



Error correcting framework

Example



A **code** C over an alphabet Σ is a nonempty subset of Σ^* ($C \subseteq \Sigma^*$).

A **q-nary** code is a code over an alphabet of q -symbols.

A **binary code** is a code over the alphabet $\{0, 1\}$.

Examples of codes

$C1 = \{00, 01, 10, 11\}$ $C2 = \{000, 010, 101, 100\}$

$C3 = \{00000, 01101, 10111, 11011\}$

CHANNEL

is any physical medium in which information is stored or through which information is transmitted.

(Telephone lines, optical fibres and also the atmosphere are examples of channels.)

NOISE

may be caused by sunspots, lighting, meteor showers, random radio disturbances, poor typing, poor hearing,

TRANSMISSION GOALS

- 1 Encoding of information should be very fast.
- 2 Very similar messages should be encoded very differently.
- 3 Transmission of encoded messages should be very easy.
- 4 Decoding of received messages should be very easy.
- 5 Correction of errors introduced in the channel should be reasonably easy.
- 6 As large as possible amount of information should be transferred reliably per a time unit.

BASIC METHOD OF FIGHTING ERRORS: REDUNDANCY!!!

Example: 0 is encoded as 00000 and 1 is encoded as 11111.

CHANNELS - MAIN TYPES

Discrete channels and **continuous channels** are main types of channels.

With an example of continuous channels we will deal in chapter 3. **Main model of the noise in discrete channels is:**

- **Shannon stochastic (probabilistic) noise model:**
 $Pr(y|x)$ (probability of the output y if the input is x) is known and the probability of too many errors is low.

DISCRETE CHANNELS - MATHEMATICAL VIEWS

Formally, a discrete Shannon stochastic channel is described by a triple $C = (\Sigma, \Omega, p)$, where

- Σ is an **input alphabet**
- Ω is an **output alphabet**
- Pr is a **probability distribution on $\Sigma \times \Omega$** and for each $i \in \Sigma, o \in \Omega$, $Pr(i, o)$ is the **probability that the output of the channel is o if the input is i .**

IMPORTANT CHANNELS

- **Binary symmetric channel** maps, with fixed probability p_0 , each binary input into opposite one. Hence, $Pr(0, 1) = Pr(1, 0) = p_0$ and $Pr(0, 0) = Pr(1, 1) = 1 - p_0$.
- **Binary erasure channel** maps, with fixed probability p_0 , binary inputs into $\{0, 1, e\}$, where e is so called the **erasure symbol**, and $Pr(0, 0) = Pr(1, 1) = p_0$, $Pr(0, e) = Pr(1, e) = 1 - p_0$.

BASIC CHANNEL CODING PROBLEMS

Summary: The task of a communication channel coding is to encode the information to be sent over the channel in such a way that even in the presence of some channel noise, several (a specific number of) errors can be detected and/or corrected.

Details of the techniques used to protect information against noise in practice are sometimes rather complicated, but basic principles are mostly easily understood.

The key idea is that in order to protect a message against a noise, we should encode the message by adding some **redundant information** to the message.

This should be done in such a way that even if the message is corrupted by a noise, there will be enough redundancy in the encoded message to recover – to decode the message completely.

The basic idea of so called **majority voting decoding/principle** or of **maximal likelihood decoding/principle**, when a code C is used, is

to decode a received message w'

by a codeword w that is the **closest** one to w' in the whole set of the codewords of the given code C .

EXAMPLE

In case: (a) the **encoding**

$$0 \rightarrow 000 \quad 1 \rightarrow 111,$$

is used,

(b) the **probability of the bit error** is $p < \frac{1}{2}$ and,

(c) the following **majority voting decoding**

$$000, 001, 010, 100 \rightarrow 000 \quad \text{and} \quad 111, 110, 101, 011 \rightarrow 111$$

is used,

then the probability of an erroneous decoding (for the case of 2 or 3 errors) is

$$3p^2(1-p) + p^3 = 3p^2 - 2p^3 < p$$

EXAMPLE: Coding of a path avoiding an enemy territory

Story Alice and Bob share an identical map (Fig. 1) shown in Fig.1. Only Alice knows the route through which Bob can reach her avoiding the enemy (gridded) territory. Alice wants to send Bob the information about the safe route he should take.

NNWNNWWSSWWNNNNWWN

Three ways to encode the safe route (by steps North, West, South, East) from Bob to Alice are:

1 $C1 = \{N = 00, W = 01, S = 11, E = 10\}$

In such a case **any error** in the code word

000001000001011111010100000000010100

would be a disaster.

2 $C2 = \{000, 011, 101, 110\}$

3 $C3 = \{00000, 01101, 10110, 11011\}$

A single error in encoding each of symbols N, W, S, E can be detected.

A single error in decoding each of symbols N, W, S, E can be corrected.

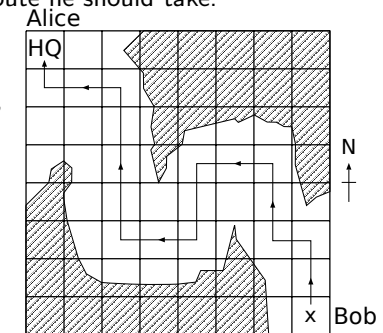


Fig. 1

BASIC TERMINOLOGY

Datawords - words of a message

Codewords - words of some code.

Block code - a code with all codewords of the same length.

Basic assumptions about channels

- 1 **Code length preservation.** Each output word of a channel it should have the same length as the corresponding input codeword.
- 2 **Independence of errors.** The probability of any one symbol being affected by an error in transmissions is the same.

Basic strategy for decoding

For decoding we use the so-called **maximal likelihood principle**, or **nearest neighbor decoding strategy**, or **majority voting decoding strategy** which says that

the receiver should decode a received word w'

as

the codeword w that is the **closest one** to w' .

HAMMING DISTANCE

The intuitive concept of "**closeness**" of two words is well formalized through **Hamming distance** $h(x, y)$ of words x, y . For two words x, y

$h(x, y)$ = the number of symbols in which the words x and y differ.

Example: $h(10101, 01100) = 3$, $h(\text{fourth}, \text{eighth}) = 4$

Properties of Hamming distance

- 1 $h(x, y) = 0 \Leftrightarrow x = y$
- 2 $h(x, y) = h(y, x)$
- 3 $h(x, z) \leq h(x, y) + h(y, z)$ **triangle inequality**

An important parameter of codes C is their **minimal distance**.

$$h(C) = \min\{h(x, y) \mid x, y \in C, x \neq y\},$$

Therefore, $h(C)$ is the smallest number of errors that can change one codeword into another.

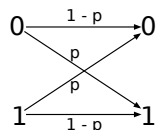
Basic error correcting theorem

- 1 A code C can detect up to s errors if $h(C) \geq s + 1$.
- 2 A code C can correct up to t errors if $h(C) \geq 2t + 1$.

Proof (1) Trivial. (2) Suppose $h(C) \geq 2t + 1$. Let a codeword x is transmitted and a word y is received such that $h(x, y) \leq t$. If $x' \neq x$ is any codeword, then $h(y, x') \geq t + 1$ because otherwise $h(y, x') < t + 1$ and therefore $h(x, x') \leq h(x, y) + h(y, x') < 2t + 1$ what contradicts the assumption $h(C) \geq 2t + 1$.

BINARY SYMMETRIC CHANNEL

Consider a transition of binary symbols such that each symbol has probability of error $p < \frac{1}{2}$.



Binary symmetric channel

If n symbols are transmitted, then the probability of t errors is

$$p^t(1-p)^{n-t} \binom{n}{t}$$

In the case of binary symmetric channels, the "nearest neighbour decoding strategy" is also "maximum likelihood decoding strategy".

POWER of PARITY BITS

Example Let all 2^{11} of binary words of length 11 be codewords and let the probability of a bit error be $p = 10^{-8}$.

Let bits be transmitted at the rate 10^7 bits per second.

The probability that a word is transmitted incorrectly is approximately

$$11p(1-p)^{10} \approx \frac{11}{10^8}.$$

Therefore $\frac{11}{10^8} \cdot \frac{10^7}{11} = 0.1$ of words per second are transmitted incorrectly.

Therefore, one wrong word is transmitted every 10 seconds, 360 erroneous words every hour and 8640 words every day without being detected!

Let now one parity bit be added.

Any single error can be detected!!!

The probability of at least two errors is:

$$1 - (1-p)^{12} - 12(1-p)^{11}p \approx \binom{12}{2}(1-p)^{10}p^2 \approx \frac{66}{10^{16}}$$

Therefore, approximately $\frac{66}{10^{16}} \cdot \frac{10^7}{12} \approx 5.5 \cdot 10^{-9}$ words per second are transmitted with an undetectable error.

Corollary One undetected error occurs only once every 2000 days! ($2000 \approx \frac{10^9}{5.5 \times 86400}$).

TWO-DIMENSIONAL PARITY CODE

This is a generalization of the previous (called also as one-dimensional) parity code. The **two-dimensional parity code** arranges first the to be transmitted message into a two-dimensional array and then to each row (column) of the array parity bits are attached.

Example Binary string

10001011000100101111

is represented and encoded as follows

1	0	0	0	1	→	1	0	0	0	1	0
0	1	1	0	0		0	1	1	0	0	0
0	1	0	0	1		0	1	0	0	1	0
0	1	1	1	1		0	1	1	1	1	0
						1	1	0	1	1	0

Question How much better is two-dimensional encoding than one-dimensional encoding?

NOTATIONS and EXAMPLES

Notation: An (n, M, d) -code C is a code such that

- n - is the **length** of codewords.
- M - is the **number** of codewords.
- d - is the **minimum distance** in C .

Example:

$C_1 = \{00, 01, 10, 11\}$ is a $(2, 4, 1)$ -code.

$C_2 = \{000, 011, 101, 110\}$ is a $(3, 4, 2)$ -code.

$C_3 = \{00000, 01101, 10110, 11011\}$ is a $(5, 4, 3)$ -code.

Comment: A **good** (n, M, d) -code has small n , large M and also large d .

EXAMPLES from DEEP SPACE TRAVELS

Examples (Transmission of photographs from the deep space)

- In 1965-69 **Mariner 4-5** probes took the first photographs of another planet - 22 photos. Each photo was divided into 200×200 elementary squares - pixels. Each pixel was assigned 6 bits representing 64 levels of brightness. and so called **Hadamard code** was used.

Transmission rate was 8.3 bits per second.

- In 1970-72 **Mariners 6-8** took such photographs that each picture was broken into 700×832 squares. So called Reed-Muller $(32, 64, 16)$ code was used.

Transmission rate was 16200 bits per second. (Much better quality pictures could be received)

HADAMARD CODE

In Mariner 5, 6-bit pixels were encoded using 32-bit long Hadamard code that could correct up to 7 errors.

Hadamard code has 64 codewords. 32 of them are represented by the 32×32 matrix $H = \{h_{ij}\}$, where $0 \leq i, j \leq 31$ and

$$h_{ij} = (-1)^{a_0 b_0 + a_1 b_1 + \dots + a_4 b_4}$$

where i and j have binary representations

$$i = a_4 a_3 a_2 a_1 a_0, j = b_4 b_3 b_2 b_1 b_0$$

The remaining 32 codewords are represented by the matrix $-H$.
Decoding was quite simple.

CODES RATES

For q -nary (n, M, d) -code C we define the **code rate**, or **information rate**, R_C , by

$$R_C = \frac{\lg_q M}{n}.$$

The code rate represents the ratio of the number of needed input data symbols to the number of transmitted code symbols.

If a q -nary code has code rate R , then we say that it transmits R q -symbols per a channel use - or R is a number of bits per a channel use (bpc) - in the case of binary alphabet.

Code rate (6/32 for Hadamard code), is an important parameter for real implementations, because it shows what fraction of the communication bandwidth is being used to transmit actual data.

The ISBN-code I

Each book till 1.1.2007 had **International Standard Book Number** which was a 10-digit codeword produced by the publisher with the following structure:

l	p	m	w	
language	publisher	number	weighted check sum	$= x_1 \dots x_{10}$
0	07	709503	0	

such that $\sum_{i=1}^{10} (11 - i)x_i \equiv 0 \pmod{11}$

The publisher has to put $x_{10} = X$ if x_{10} is to be 10.

The ISBN code was designed to detect: (a) any single error (b) any double error created by a transposition

Single error detection

Let $X = x_1 \dots x_{10}$ be a correct code and let

$$Y = x_1 \dots x_{j-1} y_j x_{j+1} \dots x_{10} \text{ with } y_j = x_j + a, a \neq 0$$

In such a case:

$$\sum_{i=1}^{10} (11 - i)y_i = \sum_{i=1}^{10} (11 - i)x_i + (11 - j)a \neq 0 \pmod{11}$$

The ISBN-code II

Transposition detection

Let x_j and x_k be exchanged.

$$\sum_{i=1}^{10} (11 - i)y_i = \sum_{i=1}^{10} (11 - i)x_i + (k - j)x_j + (j - k)x_k = (k - j)(x_j - x_k) \neq 0 \pmod{11}$$

if $k \neq j$ and $x_j \neq x_k$.

New ISBN code

Starting 1.1.2007 instead of 10-digit ISBN code a 13-digit ISBN code is being used.

New ISBN number can be obtained from the old one by preceding the old code with three digits 978.

For details about 13-digit ISBN see

https://en.wikipedia.org/wiki/International_Standard_Book_Number

EQUIVALENCE of CODES

Definition Two q -ary codes are called **equivalent** if one can be obtained from the other by a combination of operations of the following type:

- (a) a permutation of the positions of the code.
- (b) a permutation of symbols appearing in a fixed position.

Question: Let a code be displayed as an $M \times n$ matrix. To what correspond operations (a) and (b)?

Claim: Distances between codewords are unchanged by operations (a), (b).
Consequently, equivalent codes have the same parameters (n, M, d) (and correct the same number of errors).

Examples of equivalent codes

$$(1) \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

$$(2) \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

Lemma Any q -ary (n, M, d) -code over an alphabet $\{0, 1, \dots, q-1\}$ is equivalent to an (n, M, d) -code which contains the all-zero codeword $00 \dots 0$.

Proof Trivial.

THE MAIN CODING THEORY PROBLEM

A good (n, M, d) -code should have a small n , large M and large d .

The main coding theory problem is to optimize one of the parameters n, M, d for given values of the other two.

Notation: $A_q(n, d)$ is the largest M such that there is an q -ary (n, M, d) -code.

Theorem

- (a) $A_q(n, 1) = q^n$;
- (b) $A_q(n, n) = q$.

Proof

- (a) First claim is obvious;
- (b) Let C be an q -ary (n, M, n) -code. Any two distinct codewords of C have to differ in all n positions. Hence symbols in any fixed position of M codewords have to be different. Therefore $\Rightarrow A_q(n, n) \leq q$. Since the q -ary repetition code is (n, q, n) -code, we get $A_q(n, n) \geq q$.

EXAMPLE

Example Proof that $A_2(5, 3) = 4$.

- (a) Code C_3 , page (??), is a $(5, 4, 3)$ -code, hence $A_2(5, 3) \geq 4$.
- (b) Let C be a $(5, M, 3)$ -code with $M = 5$.
 - By previous lemma we can assume that $00000 \in C$.
 - C has to contain at most one codeword with at least four 1's. (otherwise $d(x, y) \leq 2$ for two such codewords x, y)
 - Since $00000 \in C$, there can be no codeword in C with at most one or two 1's.
 - Since $d = 3$, C cannot contain three codewords with three 1's.
 - Since $M \geq 4$, there have to be in C two codewords with three 1's. (say 11100, 00111), the only possible codeword with four or five 1's is then 11011.

DESIGN of ONE CODE from ANOTHER ONE

Theorem Suppose d is odd. Then a binary (n, M, d) -code exists iff a binary $(n+1, M, d+1)$ -code exists.

Proof Only if case: Let C be a binary (n, M, d) code. Let

$$C' = \{x_1 \dots x_n x_{n+1} \mid x_1 \dots x_n \in C, x_{n+1} = (\sum_{i=1}^n x_i) \bmod 2\}$$

Since parity of all codewords in C' is even, $d(x', y')$ is even for all

$$x', y' \in C'.$$

Hence $d(C')$ is even. Since $d \leq d(C') \leq d+1$ and d is odd,

$$d(C') = d+1.$$

Hence C' is an $(n+1, M, d+1)$ -code.

If case: Let D be an $(n+1, M, d+1)$ -code. Choose code words x, y of D such that $d(x, y) = d+1$.

Find a position in which x, y differ and delete this position from all codewords of D . Resulting code is an (n, M, d) -code.

A COROLLARY

Corollary:

If d is odd, then $A_2(n, d) = A_2(n + 1, d + 1)$.

If d is even, then $A_2(n, d) = A_2(n - 1, d - 1)$.

Example

$$A_2(5, 3) = 4 \Rightarrow A_2(6, 4) = 4$$

$$(5, 4, 3)\text{-code} \Rightarrow (6, 4, 4)\text{-code}$$

$$\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{array} \text{ by adding check.}$$

A SPHERE and its VOLUME

Notation F_q^n - is a set of all words of length n over the alphabet $\{0, 1, 2, \dots, q - 1\}$

Definition For any codeword $u \in F_q^n$ and any integer $r \geq 0$ the **sphere of radius r and centre u** is denoted by

$$S(u, r) = \{v \in F_q^n \mid h(u, v) \leq r\}.$$

Theorem A sphere of radius r in F_q^n , $0 \leq r \leq n$ contains

$$\binom{n}{0} + \binom{n}{1}(q - 1) + \binom{n}{2}(q - 1)^2 + \dots + \binom{n}{r}(q - 1)^r$$

words.

Proof Let u be a fixed word in F_q^n . The number of words that differ from u in m positions is

$$\binom{n}{m}(q - 1)^m.$$

GENERAL UPPER BOUNDS on CODE PARAMETERS

Theorem (The sphere-packing (or Hamming) bound)

If C is a q -nary $(n, M, 2t + 1)$ -code, then

$$M \left\{ \binom{n}{0} + \binom{n}{1}(q - 1) + \dots + \binom{n}{t}(q - 1)^t \right\} \leq q^n \quad (1)$$

Proof Since minimal distance of the code C is $2t + 1$, any two spheres of radius t centred on distinct codewords have no codeword in common. Hence the total number of words in M spheres of radius t centred on M codewords is given by the left side in (1). This number has to be less or equal to q^n .

A code which achieves the sphere-packing bound from (1), i.e. such a code that equality holds in (1), is called a **perfect code**.

Singleton bound: If C is an q -ary (n, M, d) code, then

$$M \leq q^{n-d+1}$$

A GENERAL UPPER BOUND on $A_q(n, d)$

Example An $(7, M, 3)$ -code is perfect if

$$M \left(\binom{7}{0} + \binom{7}{1} \right) = 2^7$$

i.e. $M = 16$

An example of such a code:

$C_4 = \{0000000, 1111111, 1000101, 1100010, 0110001, 1011000, 0101100, 0010110, 0001011, 0111010, 0011101, 1001110, 0100111, 1010011, 1101001, 1110100\}$

Table of $A_2(n, d)$ from 1981

n	$d = 3$	$d = 5$	$d = 7$
5	4	2	-
6	8	2	-
7	16	2	2
8	20	4	2
9	40	6	2
10	72-79	12	2
11	144-158	24	4
12	256	32	4
13	512	64	8
14	1024	128	16
15	2048	256	32
16	2560-3276	256-340	36-37

For current best results see <http://www.codetables.de>

The following lower bound for $A_q(n, d)$ is known as [Gilbert-Varshamov bound](#):

Theorem Given $d \leq n$, there exists a q -ary (n, M, d) -code with

$$M \geq \frac{q^n}{\sum_{j=0}^{d-1} \binom{n}{j} (q-1)^j}$$

and therefore

$$A_q(n, d) \geq \frac{q^n}{\sum_{j=0}^{d-1} \binom{n}{j} (q-1)^j}$$

Error detection is much more modest aim than error correction.

Error detection is suitable in the cases that channel is so good that probability of an error is small and if an error is detected, the receiver can ask the sender to renew the transmission.

For example, two main requirements for many telegraphy codes used to be:

- Any two codewords had to have distance at least 2;
- No codeword could be obtained from another codeword by transposition of two adjacent letters.

PICTURES of SATURN TAKEN by VOYAGER

Pictures of Saturn taken by Voyager, in 1980, had 800×800 pixels with 8 levels of brightness.

Since pictures were in color, each picture was transmitted three times; each time through different color filter. The full color picture was represented by

$$3 \times 800 \times 800 \times 8 = 13360000 \text{ bits.}$$

To transmit pictures Voyager used the so called **Golay code** G_{24} .

GENERAL CODING PROBLEM

Important problems of information theory are how to define formally such concepts as information and how to store or transmit information efficiently.

Let X be a [random variable](#) (source) which takes any value x with probability $p(x)$. The [entropy](#) of X is defined by

$$S(X) = - \sum_x p(x) \lg p(x)$$

and it is considered to be the [information content](#) of X .

In a special case, of a binary variable X which takes on the value 1 with probability p and the value 0 with probability $1 - p$, then the information content of X is:

$$S(X) = H(p) = -p \lg p - (1 - p) \lg (1 - p)^1$$

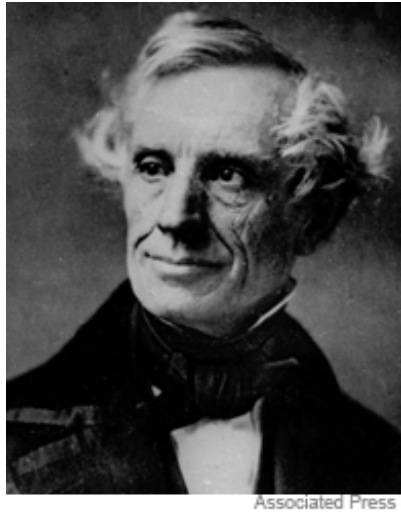
Problem: What is the minimal number of bits needed to transmit n values of X ?

Basic idea: Encode more (less) probable outputs of X by shorter (longer) binary words.

Example ([Morse code - 1838](#))

a .-	b -...	c -.-.	d -..	e .	f ..-	g -.
h	i ..	j .—	k -.-	l .-..	m —	n -.
o —	p .-	q -.-	r .-	s ...	t -	u ..-
v ...-	w .-	x -.-	y -.-	z --.		

¹Notation \lg (\ln) [\log] will be used for binary, natural and decimal logarithms.



Associated Press

Shannon's noiseless coding theorem says that in order to transmit n values of X , we need, and it is sufficient, to use $nS(X)$ bits.

More exactly, we cannot do better than the bound $nS(X)$ says, and we can reach the bound $nS(X)$ as close as desirable.

Example: Let a source X produce the value 1 with probability $p = \frac{1}{4}$ and the value 0 with probability $1 - p = \frac{3}{4}$

Assume we want to encode blocks of the outputs of X of length 4.

By Shannon's theorem we need $4H(\frac{1}{4}) = 3.245$ bits per blocks (in average)

A simple and practical method known as **Huffman code** requires in this case 3.273 bits per a 4-bit message.

mess.	code	mess.	code	mess.	code	mess.	code
0000	10	0100	010	1000	011	1100	11101
0001	000	0101	11001	1001	11011	1101	111110
0010	001	0110	11010	1010	11100	1110	111101
0011	11000	0111	1111000	1011	111111	1111	1111001

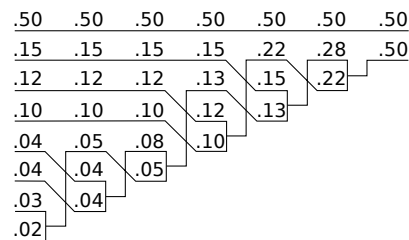
Observe that this is a **prefix code** - no codeword is a prefix of another codeword.

DESIGN of HUFFMAN CODE II

Given a sequence of n objects, x_1, \dots, x_n with probabilities $p_1 \geq \dots \geq p_n$.

Stage 1 - shrinking of the sequence.

- Replace x_{n-1}, x_n with a new object y_{n-1} with probability $p_{n-1} + p_n$ and rearrange sequence so one has again non-increasing probabilities.
- Keep doing the above step till the sequence shrinks to two objects.



Stage 2 - extending the code - Apply again and again the following method.

If $C = \{c_1, \dots, c_r\}$ is a prefix optimal code for a source S_r , then $C' = \{c'_1, \dots, c'_{r+1}\}$ is an optimal code for S_{r+1} , where

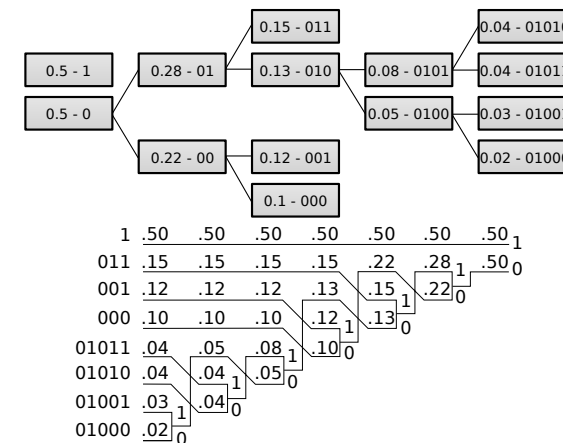
$$\begin{aligned} c'_i &= c_i & 1 \leq i \leq r-1 \\ c'_r &= c_r 1 \\ c'_{r+1} &= c_r 0. \end{aligned}$$

DESIGN of HUFFMAN CODE II

Stage 2 Apply again and again the following method:

If $C = \{c_1, \dots, c_r\}$ is a prefix optimal code for a source S_r , then $C' = \{c'_1, \dots, c'_{r+1}\}$ is an optimal code for S_{r+1} , where

$$\begin{aligned} c'_i &= c_i & 1 \leq i \leq r-1 \\ c'_r &= c_r 1 \\ c'_{r+1} &= c_r 0. \end{aligned}$$



A BIT OF HISTORY I

The subject of error-correcting codes arose originally as a response to practical problems in the reliable communication of digitally encoded information.

The discipline was initiated in the paper

Claude Shannon: A mathematical theory of communication, Bell Syst.Tech. Journal V27, 1948, 379-423, 623-656

Shannon's paper started the scientific discipline **information theory** and **error-correcting codes** are its part.

Originally, information theory was a part of electrical engineering. Nowadays, it is an important part of mathematics and also of informatics.

ENTROPY - basics - I.

The concept of **ENTROPY** is one of the most basic and important in modern science, especially in physics, mathematics and information theory.

So called **physical entropy** is a measure of the unavailable energy in a closed thermodynamics system (that is usually considered to be a measure of the system's disorder).

Entropy of an object is a measure of the amount of energy in the object which is unable to do some work.

Entropy is also a measure of the number of possible arrangements of the atoms a system can have.

Information entropy

So called **information entropy** is a measure of uncertainty and randomness.

Example If we have a process (a random variable) X producing values 0 and 1, both with probability $\frac{1}{2}$, then we are completely uncertain what will be the next value produced by the process.

On the other side, if we have a process (random variable) Y producing value 0 with probability $\frac{1}{4}$ and value 1 with probability $\frac{3}{4}$, then we are more certain that the next value of the process will be 1 than 0.

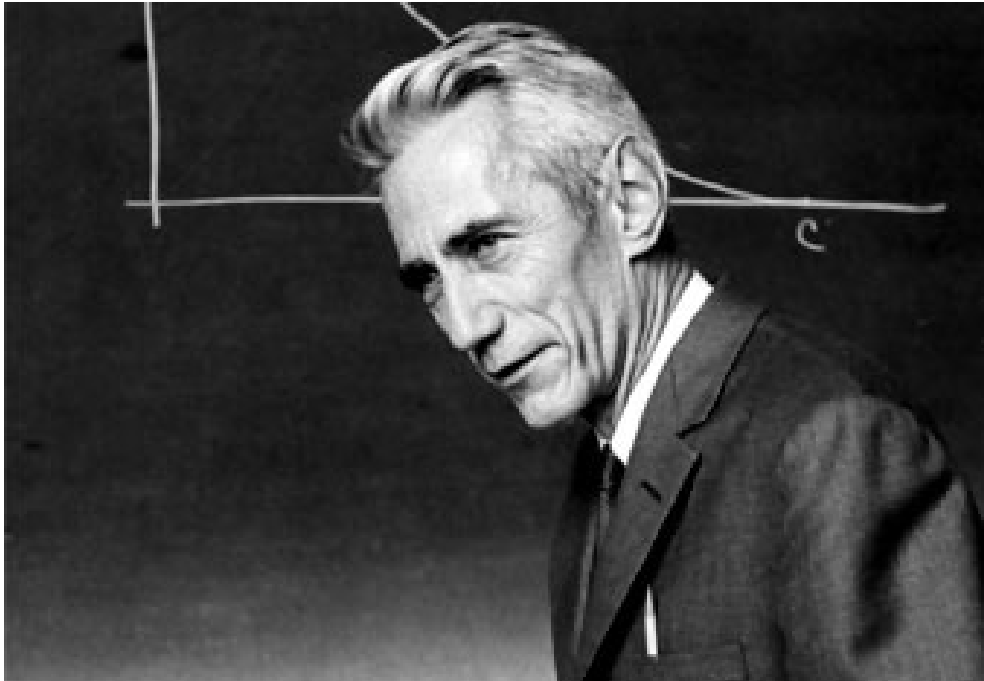
History Rudolf Clausius coined the term **entropy** in 1865.

A BIT OF HISTORY II

SHANNON's VIEW

In the introduction to his seminal paper "A mathematical theory of communication" Shannon wrote:

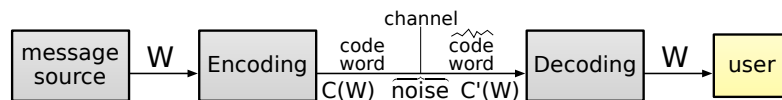
The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.



APPENDIX

HARD VERSUS SOFT DECODING I

At the beginning of this chapter the process **encoding-channel transmission-decoding** was illustrated as follows:

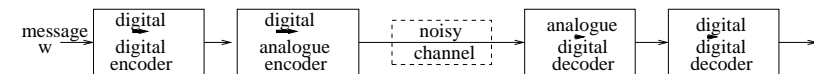


In that process a binary message is at first encoded into a binary codeword, then transmitted through a noisy channel, and, finally, the decoder receives, for decoding, a potentially erroneous binary message and makes an error correction.

This is a simplified view of the whole process. **In practice the whole process looks quite differently.**

HARD versus SOFT DECODING II

Here is a more realistic view of the whole **encoding-transmission-decoding** process:



that is

- a binary message is at first transferred to a binary codeword;
- the binary codeword is then transferred to an analogue signal;
- the analogue signal is then transmitted through a noisy channel
- the received analogous signal is then transferred to a binary form that can be used for decoding and, finally
- decoding takes place.

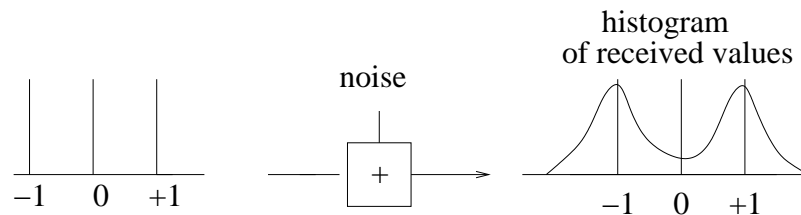
In case the analogous noisy signal is transferred before decoding to the binary signal we talk about a **hard decoding**;

In case the output of analogous-digital decoding is a pair (p_b, b) where p_b is the probability that the output is the bit b (or a weight of such a binary output (often given by a number from an interval $(-V_{max}, V_{max})$), we talk about a **soft decoding**.

HARD versus SOFT DECODING III

In order to deal with such a more general model of transmission and soft decoding, it is common to use, instead of the binary symbols 0 and 1 so-called **antipodal binary symbols** +1 and -1 that are represented electronically by voltage +1 and -1.

A transmission channel with analogue antipodal signals can then be depicted as follows.



A very important case in practise, especially for space communication, is so-called **additive white Gaussian noise (AWGN)** and the channel with such a noise is called **Gaussian channel**.

HARD versus SOFT DECODING - COMMENTS

When the signal received by the decoder comes from a device capable of producing estimations of an analogue nature on the binary transmitted data the error correction capability of the decoder can greatly be improved.

Since the decoder has in such a case an information about the reliability of data received, decoding on the basis of finding the codeword with minimal **Hamming distance** does not have to be optimal and the optimal decoding may depend on the type of noise involved.

For example, in an important practical case of the Gaussian white noise one search at the minimal likelihood decoding for a codeword with minimal **Euclidean distance**.

BASIC FAMILIES of CODES

Two basic families of codes are

Block codes called also as **algebraic codes** that are appropriate to encode blocks of data of the same length and independent one from the other. Their encoders have often a huge number of internal states and decoding algorithms are based on techniques specific for each code.

Stream codes called also as **convolution codes** that are used to protect continuous flows of data. Their encoders often have only small number of internal states and then decoders can use a complete representation of states using so called *trellises*, iterative approaches via several simple decoders and an exchange of information of probabilistic nature.

Hard decoding is used mainly for block codes and soft one for stream codes. However, distinctions between these two families of codes are tending to blur.

NOTATIONAL COMMENT

The term **code** is often used also to denote **a specific encoding algorithm that transfers any dataword, say of the size k , into a codeword, say of the size n** . The set of all such codewords then forms the code in the original sense.

For the same code there can be many encoding algorithms that map the same set of datawords into different codewords.

STORY of MORSE TELEGRAPH - I.

- In 1825 William Sturgeon discovered electromagnet and showed that using electricity one can make to ring a ring that was far away.
- The first telegraph designed Charles Wheate Stone and demonstrated it at the distance 2.4 km.
- Samuel Morse made a significant improvement by designing a telegraph that could not only send information, but using a magnet at other end it could also write the transmitted symbol on a paper.
- Morse was a portrait painter whose hobby were electrical machines.
- Morse and his assistant Alfred Vailem invented "Morse alphabet" around 1842.
- After US Congress approved 30,000 \$ on 3.3.1943 for building a telegraph connection between Washington and Baltimore, the line was built fast, and already on 24.3.1943 the first telegraph message was sent: "What hat God wrought" - "Čo Boh vykonal".
- The era of Morse telegraph ended on 26.1.2006 when the main telegraph company in US, Western Union, announced cancelation of all telegraph services.

STORY of MORSE TELEGRAPH - II.

In his telegraphs Moore used the following two-character audio alphabet

- **TIT** or **dot** — a short tone lasting four hundredths of second;
- **TAT** or **dash** — a long tone lasting twelve hundredths of second.

Morse could called these tones as 0 and 1

The binary elements 0 and 1 were first called **bits** by J. W. Tuckley in 1943.