USA TERRITORIES ONLY

CargoBit

This is the addendum to the CargoBit paper of February 2023 explaining the difference between current modern day data compression, and the original Psitron M-100 Supercomputer Subspace Conversion back in 1983. Dell is used because of their guaranteed performance of non pattern sensitive data, and a leader in 2023, so far. Dell does NOT appear to break 'A Mathematical Theory of Communication' By C. E. Shannon - 1948.

Included is a copy of StorageReview (www.storagereview.com) article written in January 30, 2023. Also included are pages of the paper COURSAYRE Processor Element stored on www.scribd.com that has been available on the Internet since 2013.

The original approach, long before 56K modems existed was to use the existing 9,600 bps modem, and send the data in Subspace Conversion mode 6 times faster. Digital Signal Processing (DSP) was used for the interface, however this was misinterpreted by almost everyone to be the critical component, ignoring the data compression completely. Subspace Conversion began in 1980 to break 'A Mathematical Theory of Communication' By C. E. Shannon - 1948.

This is a comparison of where I was in 1983, and where the world is in 2023 only, and NOT a practical application in 2023. However the technology that was practical in 1983 was more than capable of a 6 times Subspace Conversion achieving lossless data compression for all input data sets, and is not pattern sensitive. This also proves the security of this Trade Secret.

The 30.72 TB SSD NVME disk drive is used as a practical example. One 30.72TB disk will hold 3 to 4 times the amount of data with Dell in 2023.

One 30.72TB SSD NVME disk drive will hold 6 times the amount of data with the Psitron M-100 Supercomputer Subspace Conversion in 1983.

I have been working independent Research and Development since 1980, and I have made great progress since 1983 in data storage.

New Dell PowerMax Transforms Cybersecurity, Data Reduction, and Intelligent Automation

written by Brian Beeler | January 30, 2023

Dell PowerMax has been a constant in the primary storage space for decades. Mission-critical applications depend on the security and reliability of PowerMax. In fact, 95 of the Fortune 100 use PowerMax in their businesses.

This past summer, Dell launched the next generation of PowerMax and PowerMaxOS 10, with over 200 new features, including critical new cybersecurity and advanced automation. From the mainframe perspective, primary updates include the industry's first data reduction guarantee for mainframe and a series of cybersecurity and resiliency benefits.

The tremendous growth in data retention has strained primary storage systems. Dell PowerMax has addressed this phenomenon by offering the industry's first data reduction for mainframe while advancing data reduction rates for open systems storage. PowerMax is the only solution offering a 3:1 data reduction guarantee and an improved 4:1 guarantee for open systems (non-mainframe) data.

Another significant PowerMax enhancement focuses on cybersecurity, and Dell has taken advantage of CloudIQ's alerting capabilities to detect attacks and minimize risk and exposure proactively. PowerMax offers up to 65 million secure snapshots to enable rapid recovery from a cyber attack. The next generation of PowerMax also benefits multi-cloud deployments by providing secure data mobility between cloud and on-premises data centers.

Dell

Compression reduces the size of the data.
Deduplication stores data as a single instance.
Pattern detection includes a non-zero allocation function that excludes strings of consecutive zeros stored as partof compressed data.
Compression, dedupe, and pattern matching use hardware assistance to reduce overhead.
Machine Learning identifies the data stored on disk that is accessed repeatedly and keeps it unreduced.
Using a function called compaction, data is stored strategically to minimize wasted space and reduce the need fordefrag functions.
Activity Based Reduction (ABR) reduces processing resources.

2023 = 3:1 data reduction guarantee

2023 = Improved 4:1 guarantee for open systems [non-mainframe] data

PSITRON M-100 Supercomputer

1983 = 6:1 Subspace Conversion guarantee

Quick History

After a successful demonstration of storing 100 million bytes of binary storage onto an audio cassette tape in 1982-1983, we began to raise funding for that product. It was determined that the market for digital data transmission over phone lines called modems, (modulator demodulator) would be a better application. The digital signal processing of the cassette storage system did greatly improve the performance over the very electrical noisy, and expensive long distance phone calls.

Long distance phone calls were very, very expensive, often with very poor acoustic quality, and a very limited frequency range. The fastest modem available at the time for computer telecommunications about 1983 was 9,600 bits per second over a leased line. I invented a modem that would transmit 6 times 9,600 bits per second, or similar to what is now called the 56K modem. The current 56K modem was developed a full fourteen years or so later using digital signal processing only, and involves a very different design compared to what I was doing in 1983-1984. This was possible 14 years later because of the many improvements over the Public Switch Telephone Network (PSTN) greatly improving the signal to noise ratio. I do not believe that the current 56K modem would have worked on the 1983 PSTN signal to noise ratio, especially over long distance.

Because of the very poor analog network signal quality of the PSTN at the time, my digital signal processing would only improve the bits per second less than I wanted. A very extensive digital computer architecture was designed not only for the digital signal processing, it was also required for the lossless digital compression system. I did design a supercomputer architecture in 1984 called the —100 for 100 Million 64 bit Floating Point Operations Per second (MFLOPS). This supercomputer architecture was then used in a reduced optimized physical package for the 6 X 9600 bits per second modem.

When I designed my first supercomputer architecture in 1984, the CRAY was in the neighborhood of about 100 Million Floating Point Operations per second. A few Large Main Frames were running about 8 Million Instructions per second at the time. The physical dimensions of the 6 X 9600 bits per second modem were that of a very large microwave oven costing about $6,000.00 each with a minimum of two units required at two different locations. (Please see the attached copy) That was economically practical to large banks at the time because of the very expense cost of long distance phone calls.

An agreement with the Union Planters Corporation was made on July 4, 1983 to begin develop the Ultra High Speed Modem (UHSM). A condominium was to be provided just down the street from the Union Planters Bank, and the very expensive Printed Circuit Board Computer Aided Design machine located in the condominium, where access by Union Planters Bank was possible. (Please see attached copy of letter received). The reasons for this agreement failing are well documented concerning previously agreed to rights concerning the equipment with ownership, and have nothing to do with the UHSM technology itself in any way, shape, or form, and other breached previous agreements. (Please see attached copy of letters) We then put together another group of people to Research & Develop the UHSM.

Manufacturing in the United States of America was not a foreign concept at that time, and we proceeded to purchase a required printed circuit board Computer Aided Design machine. The best unit available was Scicards out of New York. The current versions were about $500,000.00 with $250,000.00 in VAX 11780 hardware requiring a very large room, with about $250,000.00 in software. We went to a trade show, and the salesman said yes, they just came out with the ability to work with ECL on their new workstation that reduced cost to about $150,000.00. We ordered one and waited a few months for the workstation hardware to materialize. We went to the school at the business location in New York, and received a stack of manuals from the floor to almost the ceiling. My business partners from Colorado snow country did not mix well with the 3 piece suits of East Coast New York. After a few days of study by myself after school in the day and dinner at night, it did not appear that the software could work with ECL. I started to ask the teacher some difficult questions about working with ECL, and we were immediately sent to the head office. We were informed that they had called the FBI on us, because they thought we were secret agents from another country. I explained who I was and what I was doing, and they continued to ask me if I worked for Motorola over and over. I found out that they could work with ECL symbols and automatically drop load resistors in the schematic, however there was no intelligence as to transmission line theory whatsoever. I asked them while looking at their many, many programmers working away, if they could solve this problem for us. The formal response was that "You do not represent market share." We ended being escorted out of the building by the corporate attorney, and with a quick "Mexican stand off", received all of our money back from them.

We called Bell Laboratories and asked them what they used when working with ECL, and we were told that they take this one million dollar machine, and connect it with that two million dollar machine, adding a couple more million dollar machines in the process.

In my travels, I had come across the Zuken Corporation out of Japan. I knew that at that time every printed circuit board coming out of Japan, including SONY consumer electronics was using Zuken software. We flew out to San Hose, California and spoke with the representative. He informed me that they did not have the ability to work with ECL at that time, however they were willing to create solutions for me that would solve our problem of transmission line theory, surface area calculations, and microwave transitions. (Please see attached copy of fax sent to Japan)

We purchased the Zuken PCB CAD. (Please see attached copy of invoice)

I was using Fairchild Emitter Coupled Logic semiconductors that were military allocated at the time (see attached copy of first order). The received units were delivered to the wrong address by UPS, boxes were broken open as if they were thrown on the ground arriving on a Wells Cargo Stage Coach across a rocky desert, and the UPS man assured us that they treat each package as if it were $100.00 diamond ring. That delivery was something like $86,000.00 dollars scattered to the wind. We immediately forced the factory to use Federal Express from then onward, with no problems after that. It was a fortuitous sign of things to come, unfortunately.

Binary Digit (Bit) to Decimal

64 Bit binary digit to decimal =

18446744073709551616

640 Bit binary digit to decimal =

4562440617622195218641171605700291324893228507248559930579192517899275167208677386505912811317371399778642309573594407310688704721375437998252661319722214188251994674360264950082874192246603776

4096 Bit binary digit to decimal =

1044388881413152506691752710716624382579964249047383780384233483283953907971557456848826811934997558340890106714439262837987573438185793607263236087851365277945956976543709998340361590134383718314428070011855946226376318839397712745672334684344586617496807908705803704071284048740118609114467977783598029006686938976881787785946905630190260940599579453432823469303026696443059025015972399867714215541693835559885291486318237914434496734087811872639496475100189041349008417061675093668333850551032972088269550769983616369411933015213796825837188091833656751221318492846368125550225998300412344784862595674492194617023806505913245610825731835380087608622102834270197698202313169017678006675195485079921636419370285375124784014907159135459982790513399611551794271106831134090584272884279791554849782954323534517065223269061394905987693002122963395687782878948440616007412945674919823050571642377154816321380631045902916136926708342856440730447899971901781465763473223850267253059899795996090799469201774624817718449867455659250178329070473119433165550807568221846571746373296884912819520317457002440926616910874148385078411929804522981857338977648103126085903001302413467189726673216491511131602920781738033436090243804708340403154190336

65,536 Bit binary digit to decimal = (Following 6 Pages)

2003529930406846464979072351560255750447825475569751419265016973710894059556311453089506130880933348101038234342907263181822949382118812668869506364761547029165041871916351587966347219442930927982084309104855990570159318959639524863372367203002916969592156108764948889254090805911457037675208500206671563702366126359747144807111774815880914135742720967190151836282560618091458852699826141425030123391108273603843767876449043205960379124490905707560314035076162562476031863793126484703743782954975613770981604614413308692118102485959152380195331030292162800160568670105651646750568038741529463842244845292537361442533614373729088303794601274724958414864915930647252015155693922628180691650796381064132275307267143998158508811292628901134237782705567421080070065283963322155077831214288551675554073345107213112427399562982719769150054883905223804357045848197956393157853510018992000024141963706813559840464039472194016069517690156119726982337890017641517190051133466306898140219383481435426387306539552969691388024158161859561100640362119796101859534802787167200122604642492385111393400464351623867567078745259464670903886547743483217897012764455529409092021959585751622973333576159552394885297579954028471943529913543763705986928913757153740001986394332464890052543106629669165243419174691389632476560289415199775477703138064781342309596190960654591300890188887588084733625956065444888501447335706058817090162108499714529568344061979690565469813631162053579369791403236328496233046421066136200220175787851857409162050489711781820400187282939943446186224328009837323764931814789848119452713007440220765680910376203999203492023906626264491909167985461515778839060397720759279378852241294301017458086862263369284725851403039615558564330385450688652213114813638408384778263790459607186876728509763471271988890680478243230394718650525660978150729861141430305816927924971409161059417185352275887504477592218301158780701975535722241400019548102005661773589781499532325208589753463547007786690406429016763808161740550405117670093673202804549339027992491867306539931640720492238474815280619166900933805732120816350707634351669869625020969023162859350071874190579161241536897514808261904847946571736601005892476655445840838334790544144817684255327207315586349347605137419779525190365032198020108764738368682531025183377533908861426184800374008082238104076468878471647552945326947661700424461063311238021134588694532200116564076327023074292426051582811070387018345324567635625951430032037432740780879056283663406965030844225855967039271869461158513793386475699748568670079823960604393478850861649260304945061743412365828352144806726676841807083754862211408236579802961200027441324438432402331257403545019352428776430880232850855886089962774458164680857875115807014743763867976955049991643998284357290415378143438847303484261903388841494031366139854257635577105335580206622185577060082551288893332226436281984838613239570676191409638533832374343758830859233722284644287996245605476932428998432652677378373173288063210753211238680604674708428051166488709084770291208161104912555598322366244868556651402684641209694982590565519216188104341226838996283071654868525536914850299539675503954938371853405900096187489473992880432496373165753803673586710175783994818471798498246948060532081996066183434012476096639519778021441199752546704080608499344178256285092726523709898650

15394621930046073645079262129759176982938923670151709920915315678144397912484 7
5706237804600009918293321306880570046591458387208088016887445835557926258465 12
4763087148566313528934166117490617526671492672176128330845273936469244582892 57
1388877839056300482483799839692029222215486145902373478222682521639957440801 72
7144146179559226175083889020074169926238300282286249284182671243405751424188 56
9994272331606998712986882771820617214453142574944015066139463169197629181506 57
9745526236191224848063890033669074365989226349564114665503062965960199720636 20
2603521917776740668777463549375318899587866282125469797102065747232721372918 14
4666659421872003474508942830911535189271114287108376159222380276605327823351 66
1555149369375778466670145717971901227117812780450240026384758788339396817962 95
0690798817121690686929538248529830023476068454114178139110648560236549754227 49
7231007615131870024053910510913817843721791422528587432098524957878034683703 33
7818421444017138688124249984418618129271198533315382567321870421530631197748 53
5214670955334626336610864667332292409879849256691109516143618601548909740241 91
3509623043612196128165950518666022030715613684732364660868905014263913906515 06
3908199378852318365059897299125404479443425166774299659811849233151555272883 27
4028352688442408752811283289980625912673699546247341543333500147231430612750 39
0307397135252069338173843322950701049061867539433130784798015655130384758155 68
5236218010419650255596181934986315913233036096461905990236112681196023441843 36
3334594927631946101716652913823717182394299216272538461776065694542297877071 38
3198817036964588689811863210976900355735884624464835706291453052757101278872 02
7965364479724025405448132748391794128826423835171949197209797145936887537198 72
9130831738033911016128547415377377715951728084111627597186384924222802373441 92
5469991983672192131287035585307966942713416391033882754318613643490100943197 40
9047331014476299861725424423355612237435715825933382804986243892498222780715 95
1762757847109475119033482241412025182688713728193104253478196128440176479531 50
5057110722974314569915223451643121848657575786528197564843508958384722923534 55
9464521215831657751471298708225909292655638836651120681943836904116252668710 04
4560243704200663709001941185557160472044643696932850060046928140507119069261 39
3993902735534545567470314903886022024639948260501762431969305640666366626090 20
7048887438898907498152865444381862917382901051820869936382661868303915273264 58
1286782806601337500096593364625146091723180312930347877421234679118454791311 10
9897794648216922505629399956793483801699157439700537542134485874586856047286 75
1065423341893839099110586465595113646061055156838541217459801807133163612573 07
9611168343863767667307354583494789788316330129240800836356825939157113130978 03
0516441716682518346573675934198084958947940983292500086389778563494693212473 42
6103062713745077286156922596628573857905533240641849018451328284632709269753 83
0867308409142247659474439973348130810986399417379789657010687026734161967196 59
1599588537834822988270125605842365589539690306474965584147981310997157542043 25
6395776070485100881578291408250777385597901291294073094627859445058594122731 9
4812753225152324801503466519048228961406646890305102510916237770448486230229 48
8966711380555607956620732449373374027836767300203011615227008921843515652121 37
9215748206859356920790214502277133099987729459596952817044582181956080965811 70
2798062669891205061560742325686842271306295009864421853470810407128917646906 55
0836129916694778023822502789667843489199409657361704586786242554006942516693 97
9292624714524945408858422726153755260071904336329196375777502176005195800693 84

13 of NN January 07, 2013 Run-002

763578958687848953687212289855780682651819270363209948015587445557517531273647
142129553649408438558661520801211507907506855334448925869328385965301327204697
069457154695935365857178889486233329246520273585318853337094845540333656535698
817258252891805663548836374379334841184558016833182767683464629199560551347003
914787680864032262961664156066750815371064672310846196424753749055374480531822
600271021640098058449752602303564003808347205314994117296573678506642140084269
649710324191918212121320693976914392336837470922826773870813223668008692470349
158684099115309831541206356612318750430546753698323082796645741762080659317726
568584168183796610614496343254411170694170022265781735835125982108076910196105
222926387974504901925431190062056190657745241619191318753398404934397682331029
846589331837301580959252282920682086223033258528011926649631444131644277300323
779227471233069641714994553226103547514563129066885434542686978844774298177749
371011761465162418361668025481529633530849084994300676365480610294009469375060
984558855804397048591444958444507997849704558355068540874516331646411808312307
970438984919050658758642581073842242059119194167418249045270028826398305795005
734171148703118714283418449915345670291528010448514517605530697144176136858238
410278765932466268997841831962031226242117739147720800488357833356920453393595
325456489702855858973550575123512953654050284208102278524877660357424636667314
868027948605244578267362623085297826505711462484659591421027812278894144816399
497388188462276824485162205181707672216986326570165431691974265123004175732990
447353767253684579275436541282655358185804684006936771860502007054724754840080
553042495185449526724726134731817474218007857469346544713603697588411802940803
961674694628854067917213860122541950381970453841726800639882065632879283958270
851091995883944829777564715202613287108952616341770715164289948795356485455355
314875497813400996485449863582484769059003311696130376612792346432312970662841
130742704620203201336835038542536031363676357521260470742531120923340283748294
945310472741896928727557202761527226828337674139342565265328306846999759709775
000556088993268502504921288406827413988163154045649035077587168007405568572402
175868543905322813377070741583075626962831695568742406052772648585305061135638
485196591896864959633556821697543762143077866593473045016482243296489127070989
807667662567151726906205881554966638257382927418208227896068448822298339481667
098403902428351430681376725346012600726926296946867275079434619043999661897961
192875051944235640264430327173734159128149605616835398818856948404534231142461
355992527233006488162746672352375123431189344211888508507935816384899448754475
633168921386967557430273795378526254232902488104718193903722066689470220425883
689584093999845356094886994683385257967516188215941098162491874181336472696512
398067756194791255795744647142786862405375057610420426714936608498023827468057
598259133100691994190465190653117190892607794911921794640735512963386452303567
334558803331319708036545718479155043265489955970586288828686660661802188224860
214499997312216413817065348017551043840662441282280361664890425737764095632648
282525840766904560843949032529052633753231650908768133661424239830953080654966
187938194912003391948949406513239881664208008839555494223709673484007264270570
116508907519615537018626479745638118785617545711340047381076276301495330973517
418065547911266093803431137853253288353335202493436597912934128485497094682632
907583019307266533778255931433111096384805394085928398890779621047984791968687
653998747709591278872747587443980677982496827827220092644994455938041460877064

14 of NN January 07, 2013 Run-002

194181044075826980568803894965461658798390466058764534181028990719429302177451
997610449504319684150345551404482092893337865736305283061999007774872692299860
827905317169187657886090894181705799340489021844155979109267686279659758395248
392673488363474565168701616624064242424122896111801061568234253939218005248345
472377921991122859591419187749179382334001007812832650671028178139602912091472
010094787875255126337288422235386949006792766451163475810119387531965724212147
603828477477457170457861041738574791130190858387789015233434301300528279703858
035981518292960030568261209195094373732545417105638388704752895056396102984364
136093564163258940813798151169333861979733982167076100460798009601602482309694
304380695662012321365014054958625061528258803302290838581247846931572032323360
189946943764772672187937682643182838260356452069946863021604887452842436359355
862233350623594500289055858161127534178375045593612613085264082805121387317749
020024955273873458595640516083058305377073253397155262044470542957353836111367
752316997274029294167420442324811387507563131907827218886405337469421384216992
886294047963530515056078812636620649723125757901959887304119562622734372890051
656111109411174527796548279047125058199907749806382155937688554649882293898540
829132512907647838632249478101675349169348928810420301561028338614382737816094
634133538357834076531432141715065587754782025245478065730134227747061674424196
895261316427410469547462148375628829977180418678508454696561915090869587425118
443583730659095146098045124740941137389992782249298336779601101538709612974970
556630163730720275073475992294379239382442742118615823616131788639255309511718
842129850830723825972914414225157940388301135908333165185823496722125962181250
705811375949552502274727467436988713192667076929919908446716122873885845758462
272657333075373557282395161696417519867501268174542932373829414382481437713986
190671665757294580780482055951188168718807521297183263644215533678775127476694
079011705750981957508456356521738954417987507452385445520013357203333237989507
439390531291821225525983379090946363020218535384885482506289771561696386071238
277172562131346054940177041358173193176337013633225281912754719144345092071184
883836681817426334294961187009150304916533946476371776643912079834749462739782
217150209067019030246976215127852195614207080646163137323651785397629209202550
028896201297014137964003805573494926907353514596120867479654773369295877362863
566014376796403843079686413856344780132826128458918489852804804884418082163942
397401436290348166545811445436646003249061876303950235640204453074821024136689
519664422133920075747912868380517515063466256939193774028351207566626082989049
187728783385217852279204577184696585527879044756219266399200840930207567392536
373562839082981757790215320210640961737328359849406665214119818381088451545977
289516457213189779790749194101314836854463961690460703010759681893374121757598
816512700076126278916951040631585763753478742007022205107089125761236165802680
681585849985263146587808661680073326467683020639169720306489440562819540619068
524200305346315662189132730906968735318164109451428803660599522024824888671155
442910472192913424834643870536850864874909917881267056566538719104972182004237
149274016446094345984539253670613221061653308566202118896823400575267548610147
699368873820958455221157192347968688816085363161586288015039594941852948922707
441082820716930338781808493620401825522227101098565344481720747075601924591559
943107294957819787859057894005254012286751714251118435643718405356302418122547
326609330271039796809106493927272268303541046763259135527968383770501985523462

15 of NN January 07, 2013 Run-002

122285841055711992173171796980433931770775075562705604783177984444763756025463
703336924711422081551997369137197516324130274871219986340454824852457011855334
267526471597831073124566342980522145549415625272402891533335434934121786203700
726031527987077187249123449447714790952073476138542548531155277330103034247683
586549609372232400715451812973269208105842409055772564580368146223449318970813
889714329983134761779967971245378231070373915147387869211918756670031932128189
680332269659445928621060743882741691946516226763254066507088107103039417886056
489376981673415902592519461182364294565266937220315550470021359884629275801252
771542201662995486313032491231102962792372389976641680349714122652793190763632
613681414551637665655983978848938173308266877990196288693229659737995193162118
721545528739417024366988559388879331674453336311954151840408828381519342123412
282003095031334105070476015998798547252919066522247931971544033179483683737322
082188577334162385644138070054191353024594391350255453188645479625226025176292
837433046510236105758351455073944333961021622967546141578112719700173861149427
950141125328062125477581051297208846526315809480663368767014731073354071771087
661593585681409821296773075919738297344144525668877085532457088895832099382343
210271822411476373279135756861542125284965790333509315277692550584564401055219
264450531207375628774499816364633283581614033017581396735942732769044892036188
038675495575180689005853292720149392350052584514670698262854825788326739873522
045722823929020714482221988558710289699193587307427781515975762076402395124386
020203259659625021257834995771008562638611823381331850901468657706401067627861
758377277289589274603940393033727187385053691295712671506689668849388088514294
360996201296675907922508227531381284985152690293170026313632894209579757795932
763553116206675348865131732387243874806351331451264488996758982881292548007642
518658649024111112730135719718138160258317850693224400799865663537154408845486
639318170839573578079905973083909488180406093595919090747396090441015051632174
968141210076571917748376735575100073361692238653742907945780320004233745280756
615304292901449578062963413838355178359976470885134900485697369796523869584599
459559209070905895689145114141268450546211794502661175016692826025095077077821
195043261738322356243760177679936279609936897519139496503335850715541843645685
261667424368892037103749532842592713161053783498074073915863381796765842525803
673720646935124865223848134166380806150570482905989069645193644001859712042572
3007316410009916987524260377362177763430621616744888493081092990100951797454156
425120482208671458684925513244426677712786372821133153622430109182439124338021
4046242223349153555951689081628848798998827363044537243217428021575577796702166
631704796972817248339284101564227450727177926939992974030807277039501358154514
249404902653610582540937311465310494338248437971860693721444460082679800247122
948940576185389220342560830269705287662137737359439422411470707407290272546130
735854174569141944648762435768239706570318416846754073346634629367398362000404
140071405427763248013274220268539369886978760700959004868465062677136307097982
100655728510130660101078063374334477307347865388174268123074376606664331277535
646657860371519292276844045827328324380821284121877613204246046490080105473142
674926082692215563740548624171703102791999694264562095561981645454766204502241
144940474934983220680719135276798674781345820385957041346617793722853494003163
159954409368408957253343870298671782977037333280680176463950209002394193149911
500910527682111951099906316615031158558283558260717941005252858361136996130344

279017381178741206128818206202326384986151565645123004779296756361834576810504
334176954306753804111392855379252924134733948105053202570872818630729115891133
594201476187266429156403637192760230628384065042544174233546454998705531872688
792642410214736369862546374715974435494344389973005174252511087735788639094681
209667342815258591992485764048805507132981429935991146323991911395992675257635
900744657281019180584180734222773472139772321823177171691640010882611254909336
118678057572239101818616854910850088527227437421208652485237245624869766224538
481929867112945294551549703058591930719849710541418163696897613112674402700964
866754593456705993699546450055892162804797636568613331656390739570327203438917
541526750091501119885687270884819553167693168127289214303137681801644547736751
835349785792427646335416243360112596025210950161226411034608346564823559793427
405686884922445874549377675212032470380303549115754483129527589193989368087632
768543876955769488142284431199859570072752139317683783177033913042306095899913
731468456901042209516196707050642025673387344611565527617599272715187766001023
894476053978951694570880272873622512107622409181006670088347473760515628553394
356584375627124124445765166306408593950794755092046393224520253546363444479175
566172596218719927918657549085785295001284022903506151493731010700944615101161
371242376142672254173205595920278212932572594714641722497732131638184532655527
960427054187149623658525245864893325414506264233788565146467060429856478196846
159366328895429978072254226479040061601975197500746054515006029180663827149701
611098795133663377137843441619405312144529185518013657555866761501937302969193
207612000925506508158327550849934076879725236998702356793102680413674571895664
143185267905471716996299036301554564509004480278905570196832831363071899769915
316667920895876857229060091547291963638167359667395997571032601557192023734858
052112811745861006515259888384311451189488055212914577569914657753004138471712
4577965048175856395072895337539755822087777506072339445587895905719156736

JIM MCDANIEL

MICROSYSTEMS
ALTAIR COMPUTER CENTER

PERSONAL COMPUTING '76
CONSUMER TRADE FAIR
Atlantic City, New Jersey
EXHIBITOR

MICROSYSTEMS

Jim McDaniel

MICROSYSTEMS
COMPUTER CORPORATION

Jim McDaniel
HARDWARE SPECIALIST

6605A BACKLICK ROAD
SPRINGFIELD, VIRGINIA 22150

SHOWROOM
(703) 569-1110
COMMERCIAL
(703) 569-2523

Union Planters Corporation

July 6, 1983

Mr. James L. McDaniel
1-C Buck Hill-Deerfield Plantation
Myrtle Beach, South Carolina 29577

Dear Mr. McDaniel:

Enclosed is a check for $2,000 which represents an advance to you relative to the project to develop an improved communications device. Rick Ireland will be in touch with you to set up the Hewlett-Packard demonstrations.

I look forward to seeing you in Memphis within the next two weeks.

Sincerely,

Wm M Matthews

William M. Matthews, Jr.
Chairman of the Board

WMMjr:nq
Enclosure

P. O. BOX 387 / MEMPHIS, TENNESSEE 38147

LAW OFFICES
MOLLOY & JOHNSON, P. C.
307 MAPLE AVENUE WEST (SUITE E)
VIENNA, VIRGINIA 22180
703 281-2323

ROBERT T. MOLLOY
RICHARD M. JOHNSON
GARY D. STANFIELD
JAMES L. BORING
W. THOMAS PARROTT, III
BENJAMIN R. DUNCAN

COUNSEL
ROBERT E. SIMPSON

July 14, 1983

Mr. William Matthews, Chairman
Union Planters National Bank
P. O. Box 387
Memphis, Tennessee 38147

Re: Letter of Intent - Jimmy McDaniel

Dear Mr. Matthews:

This is the letter of intent about which we spoke by phone. I have spoken with Jim and Jimmy McDaniel about the specifics of what we discussed, and they have approved this letter.

A corporation would be formed, with Jimmy McDaniel owning fifty percent of the stock, and you owning the other fifty. Capitalization would be minimal. Jimmy would be employed by the corporation to work on a project for development of the modem/high speed data concentrator, which we have referred to as the first level. The corporation would obtain the proprietary rights in the product developed. You would provide a condominium in which Jimmy would live and work, with the thought that the space would be sufficient for this project. We have estimated 2000 square feet would be appropriate initially.

You would provide funds in an amount which is reasonable for the proper and efficient development of this project in the form of loans to the corporation. These funds will be used to pay for operating expenses, as well as necessary equipment and supplies. Although we did not discuss this point, since you would have an equity position, we would think that these loans should be without interest. Because of the debt-equity rules, you may face dividend treatment at the repayment of these loans out of the corporation's earnings and profits, but my initial reaction is that this remains the most practical way to capitalize the corporation and fund its operations, given the situations of the parties. If your tax counsel has another approach, we would be glad to discuss it.

Jimmy would be provided loans advanced by you against future earnings in the amount of $2,000 per month for living expenses. This would be repaid on a reasonable schedule out of his salary when the corporation is in a position to provide him a salary out of earnings. I would think that the only reasonable approach to

this, given Jimmy's present financial situation, is to make such advances non-refundable in the unlikely event that future earnings should be insufficient to allow a pay back.

Jimmy would also be provided a Volkswagen GTl Rabbit automobile for his use during his employment by the corporation.

Production and marketing of products developed out of this project would be accomplished through the corporation, whether directly or by contracting out or other appropriate agreements with third parties.

The two other projects, or levels, are the development of a high-speed computer and the development of a data compression system to enhance memory capacity. The corporation would have a right of first refusal for participation in those projects in return for providing reasonable funding for development of those projects and reasonable compensation to Jimmy McDaniel for assigning his rights in those projects to the corporation.

Since I talked with you, and as this matter has gotten closer to fruition, the McDaniels have realized that it would be financially difficult for them to fund a move of their household to Memphis. While this is a new matter, I would think it not outside the spirit of this agreement for them to be provided the funds for this. We do not have figures yet, but can get a quote from a moving company to confirm the amount needed. If you wish to speak with the McDaniels directly about this, please do so.

I think this covers the major points involved. This, of course, is a letter of intent only, and is subject to the negotiation and preparation of written agreements setting forth more fully and finally the particulars of this transaction.

It was nice to talk with you by phone. Best wishes.

Sincerely,

W. Thomas Parrott, III

cc: Jimmy McDaniel
Jim McDaniel

LAW OFFICES
MOLLOY & JOHNSON, P. C.
307 MAPLE AVENUE WEST (SUITE E)
VIENNA, VIRGINIA 22180
703 281-2323

ROBERT T. MOLLOY
RICHARD M. JOHNSON
GARY D. STANFIELD
JAMES L. BORING
W. THOMAS PARROTT, III
BENJAMIN R. DUNCAN

COUNSEL
ROBERT E. SIMPSON

August 31, 1983

Mr. William Matthews, Chairman
Union Planters National Bank
P. O. Box 387
Memphis, Tennessee 38147

Re: McDaniel

Dear Mr. Matthews:

This will confirm my phone conversation of August 26 with Mr. Ireland. In that conversation he stated your present position as follows:

1. Jimmy would have a 30% equity position in the product developed. Mr. Ireland indicated that this may be negotiable.

2. You would pay for Jimmy's move to Memphis, but would not pay for the move of the remainder of the family to Memphis.

3. The Bank would want to have use of the modem for a period of six months before placing it on the market for other users. However, at the time the Bank begins using the product, it will pay a fair market price for the product, with benefits accruing to Jimmy.

4. Jimmy would not be provided a car initially, but would be provided a car out of the earnings of the corporation, at such time as there are earnings.

5. Jimmy would have the use of the condominium we have discussed previously.

6. Jimmy would be given a salary of $1,000 per month in addition to the use of the condominium, in lieu of the former proposal for him to have $2,000 per month against future earnings.

7. Jimmy would initially become an employee of Propsoft and would be loaned by Propsoft to the corporation to work on the project.

I have discussed this new proposal in detail with the McDaniels. They feel that an agreement was reached in Myrtle Beach, and it was substantially in accordance with the terms of my letter to you of July 14. They feel that they dealt with you

in good faith, and that their good faith to you was very clear as far as they were concerned at all times. As you realize, the new proposal is substantially at variance with the prior agreement reached in Myrtle Beach.

In response to your new proposal, the McDaniels will accept your provision for moving only Jimmy to Memphis. Otherwise, they will not accept material terms of the agreement that are different from the agreement as outlined in my letter of July 14. Based on what I have come to understand about this project, I am sure that the McDaniels will be able to find an agreement with another investor or institution in line with the agreement they previously reached with you.

Sincerely,

W. Thomas Parrott, III

cc: Mr. James McDaniel
Mr. James L. McDaniel ✓

FEDERAL EXPRESS

PLEASE COMPLETE ALL INFORMATION IN THE 5 BLOCKS OUTLINED IN ORANGE
SEE BACK OF FORM SET FOR COMPLETE PREPARATION INSTRUCTIONS.

AIRBILL NUMBER 703213490

YOUR FEDERAL EXPRESS ACCOUNT NUMBER: 380 0308 9

DATE: 7-6-83

FROM (Your Name): Wm. M. Matthews, Jr.
COMPANY: Union Planters Corporation
DEPARTMENT/FLOOR NO.: 2nd
STREET ADDRESS: 67 Madison Avenue
CITY: Memphis
STATE: TN
ZIP (ACCURATE ZIP CODE REQUIRED FOR CORRECT INVOICING): 38103

AIRBILL NO. 703213490

TO (Recipient's Name): James L. McDaniel
If Hold For Pick-Up or Saturday Delivery, Recipient's Phone Number:
COMPANY:
DEPARTMENT/FLOOR NO.:
STREET ADDRESS (P.O. BOX NUMBERS ARE NOT DELIVERABLE): 1-C Buck Hill Deerfield Plantation
CITY: Myrtle Beach
STATE: S.C.
ZIP (ACCURATE ZIP CODE REQUIRED FOR OVERNIGHT DELIVERY): 29577

IN TENDERING THIS SHIPMENT, SHIPPER AGREES THAT F.E.C. SHALL NOT BE LIABLE FOR SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING FROM CARRIAGE HEREOF. F.E.C. DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, WITH RESPECT TO THIS SHIPMENT. THIS IS A NON-NEGOTIABLE AIRBILL SUBJECT TO CONDITIONS OF CONTRACT SET FORTH ON REVERSE OF SHIPPER'S COPY. UNLESS YOU DECLARE A HIGHER VALUE, THE LIABILITY OF FEDERAL EXPRESS CORPORATION IS LIMITED TO $100.00. FEDERAL EXPRESS DOES NOT CARRY CARGO LIABILITY INSURANCE.

YOUR NOTES/REFERENCE NUMBERS (FIRST 12 CHARACTERS WILL ALSO APPEAR ON INVOICE)

PAYMENT: [X] Bill Shipper ☐ Bill Recipient's F.E.C. Acct. ☐ Bill 3rd Party F.E.C. Acct. ☐ Bill Credit Card ☐ Cash In Advance
Account Number/Credit Card Number ____

SERVICES — CHECK ONLY ONE BOX

PRIORITY 1
1 ☐ (OVERNIGHT PACKAGES) (Up to 70 LBS.)
COURIER PAK
2 ☐ OVERNIGHT ENVELOPE (Up to 2 LBS.)
3 ☐ OVERNIGHT BOX (Up to 5 LBS.)
4 ☐ OVERNIGHT TUBE (Up to 5 LBS.)
STANDARD AIR
5 ☐ DELIVERY 2ND BUSINESS DAY FOLLOWING PICK UP (Up to 70 LBS.)

OVERNIGHT LETTER
6 [X] (Up to 2 OZ.)
7 ☐
8 ☐
9 ☐

"OVERNIGHT" IS NEXT BUSINESS DAY (MONDAY THROUGH FRIDAY); TWO DAYS FROM ALASKA/HAWAII. SATURDAY DELIVERY AVAILABLE IN CONTINENTAL U.S. SEE "SPECIAL HANDLING."

DELIVERY AND SPECIAL HANDLING — CHECK SERVICES REQUIRED

1 ☐ HOLD FOR PICK UP AT FOLLOWING FEDERAL EXPRESS LOCATION SHOWN IN SERVICE GUIDE. RECIPIENT'S PHONE NUMBER REQUIRED
2 [X] DELIVER
3 ☐ SATURDAY SERVICE REQUIRED See Reverse (Extra charge applies for delivery.)
4 ☐ RESTRICTED ARTICLES SERVICE (P-1 and Standard Air Packages only, extra charge)
5 ☐ SSS (Signature Security Service required, extra charge applies)
6 ☐ DRY ICE ____ LBS.
7 ☐ OTHER SPECIAL SERVICE ____
8 ☐ ____
9 ☐

PACKAGES	WEIGHT	DECLARED VALUE	O/S
TOTAL	TOTAL	TOTAL	

RECEIVED AT SHIPPER'S DOOR ☐ REGULAR STOP ☐ ON-CALL STOP ☐ F.E.C. LOC.
Federal Express Corporation Employee No.
DATE/TIME For Federal Express Use

EMP. NO. | DATE
☐ CASH RECEIVED | AGT/PRO
☐ RETURN SHIPMENT
☐ THIRD PARTY | AGT/PRO
☐ CHG. TO DEL. ☐ CHG. TO HOLD
STREET ADDRESS
CITY | STATE | ZIP
RECEIVED BY: (Signature) X
DATE/TIME RECEIVED | F.E.C. EMPLOYEE NUMBER

FEDERAL EXPRESS USE
FREIGHT CHARGES
DECLARED VALUE CHARGE
ADVANCE ORIGIN
ADVANCE DESTINATION
OTHER
TOTAL CHARGES

PART# 2041730750
FEC-S-0750 D/O/W
REVISION DATE 7-82 GBF
PRINTED USA

RECIPIENT COPY (AFFIXED TO PACKAGE, GIVEN TO RECIPIENT AT DELIVERY)

ULTRA HIGH SPEED MODEM FEATURES

I. Six times faster than current high speed 9,600 bit per second modems.

2. Two wire half duplex. Four wire full duplex.

3. Designed for unconditioned voice grade lines.

4. Built in data compression and data cipher.

5. Digital signal processing.

6. Digital signal reconstruction.

7. Greater than 30 million instruction per second arithmetic processor.

8. Large scale integrated circuits and large scale programmable logic array utilization.

9. Look down emulation of any modem designed for voice grade lines.

I0. Multipurpose front-end programmable interface units.

II. Designed with Computer Aided Design machine.

I2. Multiplexed in any format defined by end user up to 256 divisions.

I. The ultra high speed in this modem is only possible because of the radical departure from traditional methods, that do not apply to the combination of unique concepts utilized in this modem design.

If traditional accepted methods were utilized in this modem design, the modem would only be slightly faster than 9,600 bits per second, less expensive, and very much less desirable.

This modem is the result of three years independent research into random data compression, super high speed computers, and low cost high density digital recording for tape drives. The answers supplied from the research have been combined to form a new code structure for use in transmission and storage of raw binary data. The modem market has been chosen for implementation of the new code structure because the performance increase is incredibly dramatic, compared to any transmission rate available for normal voice grade lines. Although the increase of data storage on a standard audio cassette tape is dramatic, there are other means of mass data storage. There is currently no way to transmit six times ninety six hundred bits per second on a voice grade line.

2. The term "two wire" will be used in this paper to describe the two electrical wires utilized to make one normal voice grade line. The term "four wire" will be used to describe the four electrical wires used to make two normal voice grade lines. Half duplex describes transmission in two directions, but only one way at a time. Full duplex describes transmission in both directions simultaneously.

The modem appears to the end user as two seperate devices contained in one physical structure. This means one full duplex configuration is possible and two half duplex configurations are possible. A dual half duplex unit will transmit twelve times faster than ninety six hundred bit per second modems. All combinations are end user defined.

3. This modem is designed for use on normal unconditioned voice grade lines. The advantages are low cost, easy availability, universal compatability, and portability, because no special conditioning is required. The complexity of this project is necessary to offset the limitations contained in a normal voice grade line. The complexity is the major reason this project has never been approached before.

4. The data compression is a fundamental part of the new code structure. This compression system is pattern sensitive, however the six times ninety six hundred bit per second data rate, is a worst case figure. Therefore the compression system is not pattern sensitive to random data at the modems full rated speed.

The data cipher is a fundamental part of the new code structure and becomes beneficial to the end user in the form of security. This cipher can be influenced by the end user to provide selective compatability.

5. The majority of all present day modem circuits utilize analog signal processing, which is a simple way of doing simple things. Analog signal processing has been the only practical possibility of designing a modem circuit. This fact has been eliminated with the introduction of high speed, low cost, digital computers.

Digital signal processing provides a relative simple way of doing difficult analysis. The flexibility of analog processing is extremely limited, and there is a large performance and cost trade-off involved. Flexibility of digital signal processing is only limited to the speed cf the computer used. Also one digital signal processing circuit can replace many analog processing circuits.

This flexibility will allow the modem to compensate for the many problems involved in transmitting digital data over a normal voice grade line. Although digital signal processing is essential to high speed communications, it is not a current state of the art process. The history of known digital signal processing began with I7th and I8th century mathematics.

6. Digital signal reconstruction is included to provide more reliability. The modem transmission structure is programmed to transmit blocks of bits. If the checking algorithm determines that an error has occured inside of a block of bits, digital signal reconstruction is engaged. Error correction that is used in digital memory systems is not practical in high speed modems, because if an error is detected, it will be because of signal distortion or complete signal destruction. The bad signal will not typically transmit the original bit structure close enough to do corrections. The only possible thing a typical high speed traditional modem can do is order a retransmission of the bad block. This is very time consuming and a complete negative effort.

This modem not only stores the bits transmitted inside of a block, but it also stores the complete transmitted block signal. Therefore when an error is detected, the modem will continue to process the continuous transmission, and not have to stop and order a retransmission. The computer is fast enough to go back to the bad block signal and attempt reconstruction. Digital signal reconstruction is a mathematic process that counter acts normal line disturbances. If the signal was totally destroyed, such as a direct lightning strike on the lines, retransmission of that individual bad block will be ordered. Up to ten bad block signals can be stored and operated on before a disruption in service is necessary. This is all practical because the computer is millions of times faster than a normal voice grade line.

7. The high speed of this computer is necessary to deal with the many simultaneous events occuring during normal operations. The majority of calculations will be processing the mathematics that are necessary to achieve the ultra high speed of this modem. This is why an arithmetic processor is a better choice than a general purpose central processor design.

The speed and power of the computer in this project is much greater than many general purpose computers costing more than three hundred thousand dollars. The low cost of this projects computer is possible because it is dedicated to this modem, and not for general purpose use.

8. The complexity of this project is minimized with large scale integrated circuits and large scale programmable logic arrays. Many major functions can be achieved with the combination of a few large scale integrated circuits. Where it is economical, a large printed circuit board full of integrated circuits can be reduced to one large scale programmable logic array. This is a major consideration because the physical size of this modem is close to the size of a typical microwave food processor.

9. Because the speed, power, and flexibility of the arithmetic processor in this modem is so great, the end user has the option of simulating any modem designed for normal voice grade lines. This is desirable because no current traditional modem has this capability, and the end user does not have to buy ten different modems to be compatable with ten different modems.

10. The multipurpose front-end programmable interface units are external devices that allow our modem to enhance and adapt to many different environments. Some units will be simple computer interface devices, and other units will be complicated time shared division distribution and simulation structures. These will appear as future options that enhance the basic modem unit.

11. The realization of this project is completely impractical without the centralized, top down use of a Computer Aided Design machine. The hundreds of thousands of electrical interconnections. This means that with a manual design, a minor change means a major disaster. With a Computer Aided Design machine, all interactions are handled automatically and the major disaster becomes a minor change. The Computer Aided Design machine will reduce the design cycle by years.

12. The computer in this modem can divide the bit structure up to 256 divisions. This will allow computer sites with one hundred and ninety two remote keypunch terminals to transmit at three hundred bits per second each, full duplex over two normal voice grade lines. Any division rate is possible up to six times 9,600 bits per second.

I2. With the proper interface option, terminals, digital voice processors and data links can be connected directly to this modem and a large computer will not be necessary to handle the multiplexing.

One possibility is to use digitized voice units such as the VADIC 5, TSP-I00, or ICS TELEMUX and our modem in a full duplex structure to achieve twenty four individual conversations over two normal voice grade lines. The possibility of this option is only limited to the end users desire.

November 29, 1983

Mr. Jim McDaniels
Psitron Corporation
Unit 1C Buck Hill
Deerfield Plantation
Myrtle Beach, SC 29577

Dear Mr. McDaniels:

On behalf of Scientific Calculations, Inc., I want to thank you for the recent Design Station order from Psitron Corporation.

May I assure you of our continued support and cooperation. We look forward to a long and mutually profitable partnership between our two companies.

Sincerely,

Neil K. Clayton

Neil Clayton
Sales Representative

NC/jjm

SCIENTIFIC CALCULATIONS, INC. 7635 MAIN STREET, FISHERS, NEW YORK 14453 (716) 924-9303 TWX 510-254-8546 TELEX 97-8316

March 15, 1984

Mr. Jim McDaniels
Psitron Corporation
Deerfield Plantation
P.O. Box 2084
Myrtle Beach, South Carolina 29577

Re: Cancellation of Design Station Order

Dear Mr. McDaniels:

In light of our discussion on March 14, 1984, in which it was determined that the Design Station SCDS/1 and SCICARDS® and SCHEMACTIVE® Programs were not compatible with the high speed ECL Logic which your application requires, we have decided to cancel SC License Quotation 175-DSL and SC Equipment Quotation 176-DSE, both dated October 25, 1983.

Upon receipt by SC of a copy of this letter signed by an officer of Psitron, SC shall (i) release Psitron from its obligation to pay cancellation charges pursuant to Section 9 of SC's Standard Terms and Conditions of Sale; (ii) terminate the SC Standard Domestic Software License Agreement--REV. 1.1 DS, dated December 29, 1983; (iii) terminate the SC Maintenance Agreement for Design Stations, accepted by SC on February 8, 1984; and (iv) terminate the Operating System License Agreement dated February 8, 1984. Please indicate your acceptance of such cancellations and terminations by signing and returning the enclosed copy of this letter in the enclosed self-addressed envelope. In addition, please sign and return the enclosed Licensee Statement of Return as required by Section 13.2 (d) of the SC Standard Software License Agreement.

Please be advised that Psitron's obligations of confidentiality as set forth in the Standard Software License Agreement and the Operating System License Agreement shall survive the termination of such Agreements.

SCIENTIFIC CALCULATIONS, INC. 7635 MAIN STREET, FISHERS, NEW YORK 14453 (716) 924-9303 TWX 510-254-8546 TELEX 97-8316

Thank you very much for your cooperation in this matter.

Very truly yours,

SCIENTIFIC CALCULATIONS, INC.

Mark W. Leunig

Mark W. Leunig
Assistant General Counsel

ACCEPTED THIS 25 DAY
OF MARCH, 1984.

BY: James T. McDaniel

TITLE: C.E.O.

MWL/bt

cc: J. Montieth Estes, Executive Vice President
David M. Jacobstein, Vice President-Finance and General Counsel
Neil Clayton, Regional Sales Manager
Steve Testa, Design Station Director of Sales

LICENSEE

STATEMENT OF RETURN

On behalf of PSITRON Corporation (the "Company"), I, James L McDaniel, affirm that the Company has returned to SC Management control all SC Software source or object code whether contained on or within magnetic tape, cards, disk packs, and/or listings, together with all SC Software documentation and other SC Proprietary Information. There are no outstanding copies of SC Software source or object code or other SC Proprietary Information in the Company's possession, except as follows:

Signed James L McDaniel

Title C.E.O.

Date 3/25/84

Signed for Scientific Calculations, Inc.

Title

Date

Ed Koller
Box 2202
Vail, Co.
81658

E. J. KOLLER
EMBASSY APARTMENTS • 475 SOUTH PERKINS
MEMPHIS, TENN. • 38117

Jimmy -

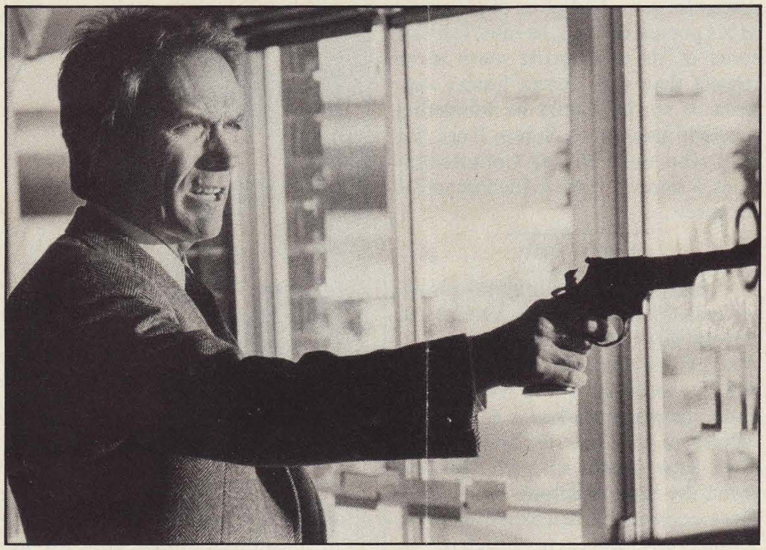
Thought you migit like this picture - "you don't make me feel too welcome now!"

Just you are preparing a new avenue to pursue.

See you soon,

Eddie

"You don't make me feel too welcome, now!"



Eastwood in "Sudden Impact": Warner's film library may be worth $500 million.

d
n-
an
ra-
ter
half
st a
ar-
its
ts
w
re
led
are
ridge
eepen

"They've
...'s pro-

electronics colos... makes sense to
at least ...

done well. It holds franchises in seven
...ities and parts of New York
...be interactive cable opera-

Eddi Koller
Box 2202
Vail, Co. 81658

VAIL, CO
MAR 19
PM
1984
81657

USA 20c

Jimmy McDaniels
Deerfield Plantation
P.O. Box 2084
Myrtle Beach, So. Carolina
29577


VAX 11/780

- separate displays for graphics and control functions
- enhanced color raster graphics
- convenient ergonometric design
- SCICARDS™ system's ease of operation

. . . all in all, it's the PCB designer's dream machine.

SCICARDS™ PROGRAM

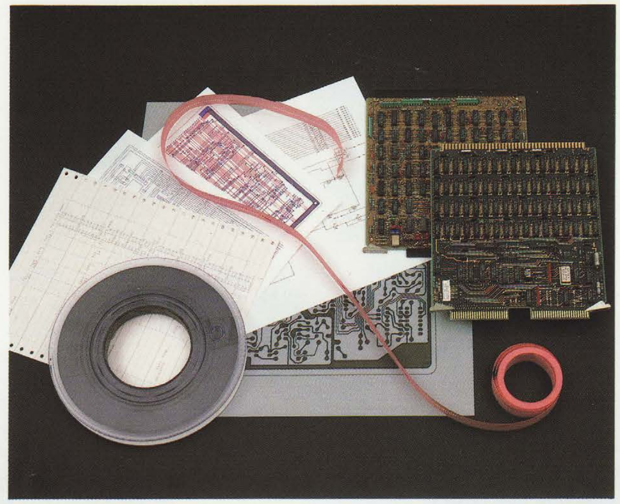
Make PCB design time much more productive

- Auto-Plus design slashes time and costs, speeds turnaround, cuts product lead time
- Speeds handling of engineering changes and redesigns
- Generates accurate fabrication artwork, N/C tapes and documentation
- Easily handles large digital, analog or mixed designs, dense multilayer boards
- Gives designer unmatched interactive control of powerful automatic functions
- Full-color graphic display speeds design, aids accuracy and efficiency
- Expandable turnkey computer system supports one or a number of PCB designers
- Aids standardization of PCB design and production
- Provides easy-to-use man/machine interface via interactive graphic display and designer-oriented macro commands
- Affords archival storage of comprehensive, accurate design data in machine-retrievable form

SC
SCIENTIFIC CALCULATIONS

ZUKEN SYSTEM-2000

SYSTEM 2000 is the most comprehensive printed circuit board and hybrid integrated circuit design system in the world today. The 2000 provides a complete software environment for schematic capture, digital design, analog design, mixed digital and analog design, hybrid design, automatic routing, automatic placement, automatic insertion, numerically controlled drill machines, and automatic test machines. SYSTEM 2000 is based around a Hewlett Packard A900 central processing unit with high-resolution color and monochrome displays, digitizer tables and tablets, and the complete spectrum of hard-copy and plot output devices.
Welcome to Zuken's world.



ZUKEN AMERICA

ZUKEN was founded in 1976 by five engineers from Mutoh, the largest drafting equipment vendor in Japan. This strong background in output and manufacturing technology directed them to create the SYSTEM 2000 exclusively for printed circuit board and hybrid integrated circuit design. By avoiding the limitations of traditional vector based CAD systems Zuken engineers based the 2000 on true curve data. This unique approach allowed users to enter designs exactly as the photo-plotter would draw them, with true curves, tear drop shapes, necked-up and necked-down traces. In addition to the fully functional artwork system Zuken maintained complete electrical connectivity capability, parts location and placement information, and complete table driven interface capability to virtually any NC machine, ATE machine or AW (artwork generation) machine. Today we offer the most comprehensive computer aided manufacturing and computer aided design system in the world.

ZUKEN
AMERICA

ZUKEN
AMERICA INCORPORATED
2010 North First Street Suite 205
San Jose, California 95131
(408) 947-2070

10 NOV 1984

NOV-9'84

CALIF

PSITRON CORPORATION
801 Main Street
Conway, South Carolina 29526

ATTN: Mr. James L. McDaniel

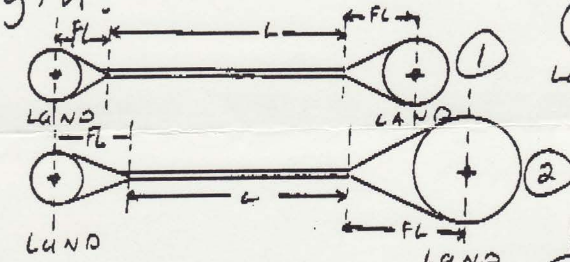
To: TOM WADE (ZA) 4/8/84

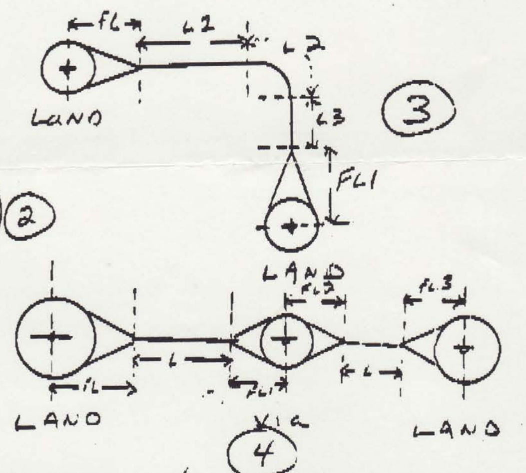
From: Chris Menze (ZJ)

Dear Tom;

Please inform Jim of the following results of my discussion with Mr WADA, Mr. SAITOH and Mr. Katsube.

1. Length:





FL = Fillet lenght
L = Length
(No Fillet angle restriction)

all total lengths can be measured and displayed

Data Base contains "FL" (LAND center to trace start)

IF Feed throughs (vias) are ignored (Figure 4) the Length is STANDARD ("FL" + "L" + "FL") IF NOT CUSTOMER CAN provide via Length (Depth) (Via count/Trace Available) though it should be nominal for calculation.

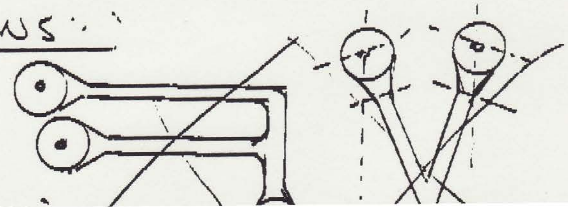
Ⓐ Length will be displayed on screen, later on Text node placed on "N" layer. (POST NCGA)

Ⓑ Length can be TOTAL or segment Length

Suggested restrictions:

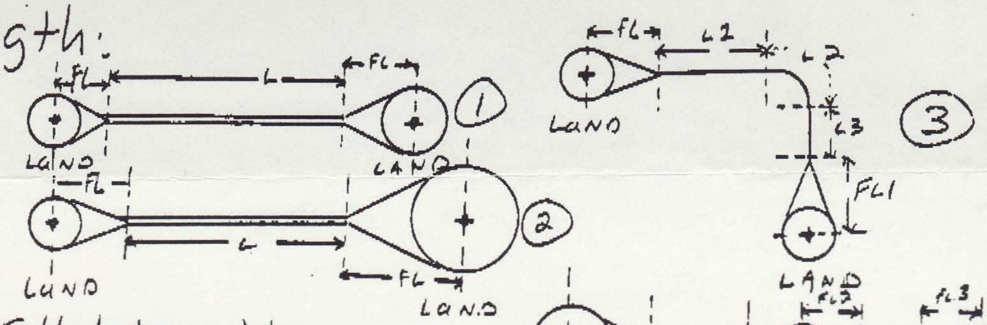
Ⓐ NO T CONNECTIONS

Ⓑ NO Y CONNECTIONS

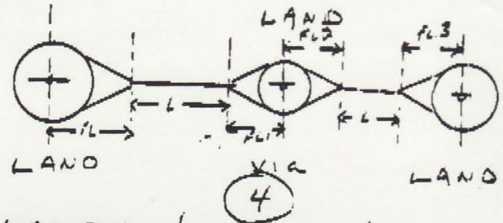


MR. SAITOH and MR. Katsube.

1. Length:



FL = Fillet lenght
L = Length
(No Fillet angle restriction)



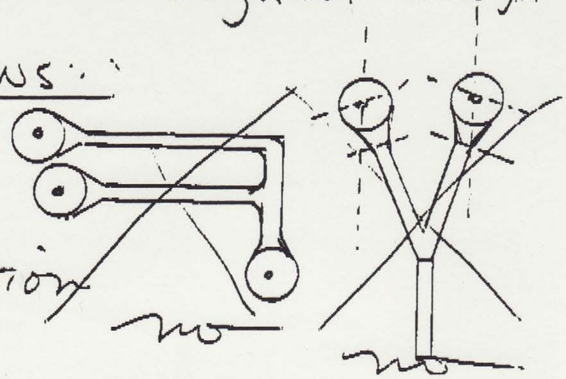
all total lengths can be measured and displayed
Data Base contains "FL" (LAND center to trace start)
IF Feed throughs (vias) are ignored (Figure 4)
the Length is standard ("FL" + "L" + "FL") IF NOT
customer can provide via Length (Depth) (Via count/trace available)
though it should be nominal for calculation.

Ⓐ Length will be displayed on screen, later on Text node placed on "N" layer. (POST NCGA)

Ⓑ Length can be total or segment Length

Suggested restrictions:

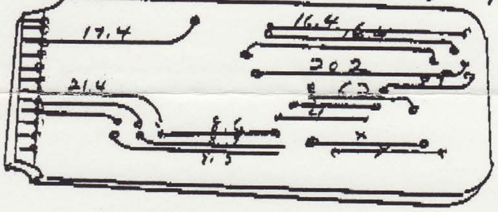
Ⓐ NO T connections
Ⓑ NO Y connections
Due to contention



Page 2

(2) Area

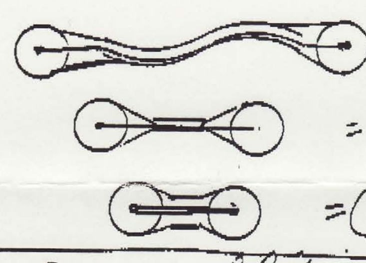
Area can be calculated and annotated to each individual trace/net through a Batch Program. The Board will return with each net area calculated as shown. (Text placement is not currently justified intelligently, some strings may overlap) (need to be adjusted)



After changes and adjustments Batch Area must be rerun.

(3) Speed Traps

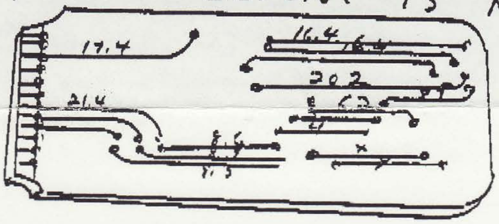
Speed traps may be calculated for value by using area (batch) and Length (interactive)



(TOM:) Please fax for additional long term information (1) Boolean OR to fine line pattern to.

Regards

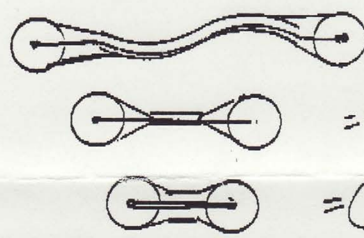
each net area calculated as shown.
(TEXT PLACEMENT is NOT CURRENTLY JUSTIFIED INTELLEGENTLY, SOME STRINGS MAY OVERLAP) (Need to be adjusted)



after changes and adjustments Batch area MUST BE RERUN.

③ Speed Traps

Speed traps may be calucated for value by using area (batch) and Length (INTERACTIVE)



=(10) imped
=(5) imped
=(1) imped

(TOM:) Please Fax For additional Long term INFORMATION

① Boolean OR to Fine Line pattern

$$A = 1/2 \sum_{i=1}^{N} (x_i - x_{i+1}) * (y_i + y_{i+1})$$

(STD area)

(2) interactive Formula

Regards
Chin

SYSTEM QUOTATION
PSITRON

April 9, 1984

QUANTITY	MODEL #	DESCRIPTION	LIST PRICE
1	2199D	A900 SYSTEM	$33,400
1	-006	CABLE TO SYSTEM CONSOLE	
1	-014	DELETE STD MEMORY	-6,100
1	-061	1600BPI MEDIA	
1	12220C	3MB ECC MEMORY	16,000
1	12222D	MEMORY CONNECTOR	160
1	7914ST	DISC/TAPE CONSISTING OF: 1-7914R DISC 1-7974A MAG TAPE	26,495
1	-800	ADD 800 BPI OPTION	2,500
1	2621B	CRT TERMINAL	1,295
1	-100	EXTENDED KEYBOARD	150
1	-050	ADD BUILT-IN 120 CPS THERMAL PRINTER	1,210
1	13242Y	2621B CABLE	75
1	2932A	GENERAL PURPOSE PRINTER	2,495
1	92214P	PRINTER STAND	275
1	92219G	PRINTER CABLE	60
1	12042B	PSI CARD	2,030
1	-002	EDGE CONNECTOR	-305
2	12005B	ASYNCH. I/F CARD	850
2	-003	RS-232 CABLE	850
1	12009A	HPIB I/F CARD	1,100
1	-001	4 METER CABLE	35
------			-------
1	ZA 5010	MONOCHROME WORKSTATION CONSISTING OF: 1-TEK 4016-1 GRAPHICS TERM. 1-OPTION 004 HIGH SPEED INTERFACE 1-CALCOMP 2000 DIGITIZING TABLET 1 HIGH SPEED LEVEL SHIFTER (RECEIVER)	35,000
1	ZA 5011	TEK HARD COPY UNIT	4,550
1	CCP 1077	CALCOMP 1077 PLOTTER	24,500
------			-------
		HARDWARE SUBTOTAL	$145,775
------			-------

ZUKEN SOFTWARE

1	Z2000	BASIC PACKAGE	$ 55,000
1	Z2010	ANALOG PACKAGE	25,000
1	Z2020	DIGITAL AUTOROUTE/AUTOPLACE	47,500
1	Z2030	SCHEMATIC CREATION	18,500
1	Z2100	CAM DATA GENERATOR	8,500
1	Z2200	PAINTING FUNCTION	4,500
1	Z2500	DESIGN RULE CHECKING	12,500
------			-------
		SOFTWARE SUBTOTAL	$171,500
------			-------
		TOTAL SYSTEM PRICE	$317,275
------			-------

ZUKEN INTERNATIONAL SOFTWARE AGREEMENT

Agreement No. ZA1003.0
Purchase Order No. Ltr 5/10/84
Date October 29, 1984

ZUKEN AMERICA INC.
2010 North First Street
San Jose, CA 95131

This agreement is between "LICENSOR", ZUKEN AMERICA INC. and "LICENSEE",

PSITRON CORPORATION

801 Main Street, Conway, South Carolina, 29526.

The Licensee agrees to acquire the right to use the ZUKEN SYSTEM 2000 CAD/CAM system software in accordance with the following terms. This proprietary software system is the property of Zuken Inc., 3-1-1, Shinyokohama, Kihoku-ky, Yokohama 222, JAPAN.

1. LICENSE TO USE LIMITATIONS

A. Licensee is hereby granted the right to use the software products listed in Table I on one computer system identified in Table I. No title to or ownership of the software or any of its parts is transmitted to the Licensee.

B. Licensee shall return these software products to ZUKEN AMERICA INC. within thirty (30) days after termination of this agreement or upon default by the licensee without the prior written consent of ZUKEN AMERICA INC.

C. The Licensee's right to use shall at all times be subject to the use restrictions and copyright restrictions described in this agreement. Any subsequent software upgrades or releases shall be included in this agreement, as addendum to Table I.

2. DUPLICATION RESTRICTIONS AND COPYRIGHT

A. The software products listed in Table I may not be copied by Licensee except for archive purposes, to replace a defective copy, or for program error verification.

B. These software products may not be copied into any media; such as magnetic tape, paper tape, disc memory cartridges, Read Only Memories, etc., for any other purpose. This authorization to duplicate the software products shall not be constructed to grant the Licensee the right to use the software in any manner other than that prescribed by ZUKEN AMERICA INC., unless otherwise approved in writing by ZUKEN AMERICA INC.

C. The Licensee agrees to label each copy of the orginally supplied software with the following copyright notice:

Copyright ZUKEN AMERICA INC., 198_, Copy by permission of ZUKEN AMERICA INC.

3. USE

The Licensee agrees to limit the use of these software products and their derivations to "ZUKEN CAD/CAM SYSTEM 2000", and refrain from making these software products available to third parties without prior written approval of ZUKEN AMERICA INC.

4. LICENSES TO U.S. GOVERNMENT

This paragraph is applicable if the Licensee is the U.S. Government, or an agency, or other entity of the U.S. Government.

A. The software products listed in Table I are furnished with RESTRICTED RIGHTS. Use, duplication or disclosures by the Government is subject to restrictions as set forth in paragraph (b) (3) (B) of the RIGHTS IN TECHNICAL DATA AND COMPUTER SOFTWARE clause in DAR-7-104.9(a).

B. The U.S. Government agrees that any software products licensed hereunder which do not have appropriate RESTRICTED RIGHTS legends applied thereto shall be deemed to be provided only with RESTRICTED RIGHTS.

5. GOVERNMENT PRIME CONTRACTOR LICENSES

This paragraph is applicable only if the listed software products are being supplied to the U.S. Government under a prime contract, or to a contractor operating under a U.S. Government contract. The following additional terms and conditions apply. Orders that deviate are not acceptable.

A. The purchase order for the software product license must contain the Standard Department of Defense, "Rights in Technical Data and Computer Software", Clause DAR 7-104.9(a), or the equivalent clause for other agencies.

B. The following paragraph must be included in the purchase order:

The above-identified software is furnished hereunder with RESTRICTED RIGHTS. Use, duplication, or disclosure by the Government is subject to restrictions as set forth in paragraph (b) (3) (B) RIGHTS IN TECHNICAL DATA AND COMPUTER SOFTWARE clause in DAR 7-104.9(a).

C. Licensee agrees to notify ZUKEN AMERICA INC. in writing, within ten (10) days, in the event that the Government challenges ZUKEN AMERICA's Restricted Rights.

D. Licensee agrees to use its best efforts to ensure that the RESTRICTED RIGHTS legend for commercial computer products is applied to any software products received hereunder for government use.

6. TERMINATION

A. This agreement terminates when the Licensee has stopped using these software products any more, or when the Licensee has breached any provision in this agreement and has not cured the breach or has not taken actions to cure the breach within 30 days of receipt of written notice of termination.

Agreement No. ZA1003.0

Purchase Order No. Ltr 5/10/84

Date October 29, 1984

TABLE I

The Hewlett-Packard HP900 or HP700 (Series A) for which the below listed software is to be licensed is identified as:

System Serial Number:

Mainframe Serial Number:

2409A00164, 2201A00725

Storage Device:

Model No. HP7914ST

Serial No. 2430A05988,2434E00931

List of Software Supplied:

Z2000	BASIC DESIGN SOFTWARE?FIRMWARE
Z2010	ANALOG DESIGN SUBSYSTEM
Z2020	DIGITAL AUTOROUTE/AUTOPLACE
Z2030	SCHEMATIC CREATION
Z2100	CAM DATA GENERATOR
Z2200	PAINTING FUNCTION
Z2500	DESIGN RULE CHECKING
Z2510	GRAPHIC CALCULATION FOR PICTURE FILE DATA

Agreement No. ZA1003.0

Purchase Order No. Ltr. 5/10/84

Date October 29, 1984

For CUSTOMER: [illegible] PSITRON CORPORATION

Oct 29 1984
(Date)

For ZUKEN AMERICA INC.

Tom Wade

10-31-84
(Date)

ORDER NO. 1850

PURCHASE ORDER

Page 1 of 2

PSITRON CORPORATION
801 Main Street
Conway, South Carolina 29526

TO	Fairchild Camera and Instrument Corporation	SHIP TO	~~PSITRON CORPORATION~~
ADDRESS	5970-C Six Forks Road	ADDRESS	~~801 Main Street~~
CITY	Raleigh, North Carolina 27609	CITY	~~Conway, SC 29526~~

REQ. NO.	FOR	DATE REQUIRED	TERMS	HOW SHIP	DATE
111	M-100	ASAP	N-30	UPS	7/1/84

	QUANTITY ORDERED	QUANTITY RECEIVED	PLEASE SUPPLY ITEMS LISTED BELOW	PRICE		UNIT
1	400		F100101DC	2	63	EA.
2	400		F100102DC	2	63	EA.
3	400		F100107DC	3	44	EA.
4	800		F100113DC	5	13	EA.
5	800		F100114DC	4	00	EA.
6	400		F100117DC	2	98	EA.
7	400		F100118DC	2	94	EA.
8	800		F100122DC	3	00	EA.
9	800		F100123DC	3	63	EA.
10	800		F100124DC	5	19	EA.
11	800		F100125DC	5	06	EA.
12	400		F100130DC	5	25	EA.
13	400		F100131DC	4	81	EA.
14	400		F100136DC	10	36	EA.
15	200		F100141DC	5	69	EA.
16	1600		F100150DC	5	25	EA.
17	1600		F100151DC	5	56	EA.
18	400		F100155DC	5	44	EA.
19	400		F100156DC	5	23	EA.
20	200		F100158DC	8	34	EA.
21	200		F100160DC	5	61	EA.
22	200		F100163DC	4	81	EA.

INVOICES PASSED FOR PAYMENT

PLEASE SEND COPIES OF YOUR INVOICE WITH ORIGINAL BILL OF LADING

PURCHASING AGENT

1H 146 Rediform

OFFICE COPY

PURCHASE ORDER

PSITRON CORPORATION
801 Main Street
Conway, South Carolina 29526

ORDER NO. 1802

TO Fairchild Camera and Instrument Corporation
ADDRESS 5970-C Six Forks Road
CITY Raleigh, North Carolina 27609

SHIP TO PSITRON CORPORATION
ADDRESS 801 Main Street
CITY Conway, SC 29526

REQ. NO.	FOR	DATE REQUIRED	TERMS	HOW SHIP	DATE
111	M-100	ASAP	N-30	UPS	7/1/84

	QUANTITY ORDERED	RECEIVED	PLEASE SUPPLY ITEMS LISTED BELOW	PRICE		UNIT
1	200		F100164DC	4	88	EA.
2	200		F100165DC	5	19	EA.
3	400		F100166DC	5	00	EA.
4	200		F100170DC	4	75	EA.
5	200		F100171DC	5	58	EA.
6	600		F100179DC	6	38	EA.
7	1600		F100180DC	7	13	EA.
8	800		F100181DC	12	00	EA.
9	700		F100182DC	6	00	EA.
10	600		F100183DC	6	00	EA.
11	800		F100194DC	5	38	EA.
12						
13						
14			Total This PO	104298	00	
15						
16						
17						
18						
19						
20						
21						
22						

INVOICES PASSED FOR PAYMENT

PLEASE SEND COPIES OF YOUR INVOICE WITH ORIGINAL BILL OF LADING

James L. McDaniel
PURCHASING AGENT

1H 146 Rediform

OFFICE COPY

New Dell PowerMax Transforms Cybersecurity, Data Reduction, and Intelligent Automation

written by Brian Beeler | January 30, 2023

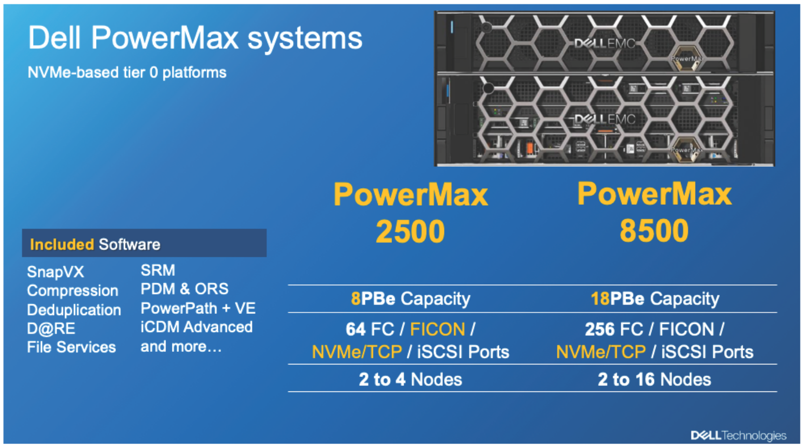
Dell PowerMax has been a constant in the primary storage space for decades. Mission-critical applications depend on the security and reliability of PowerMax. In fact, 95 of the Fortune 100 use PowerMax in their businesses.

This past summer, Dell launched the next generation of PowerMax and PowerMaxOS 10, with over 200 new features, including critical new cybersecurity and advanced automation. From the mainframe perspective, primary updates include the industry's first data reduction guarantee for mainframe and a series of cybersecurity and resiliency benefits.

The tremendous growth in data retention has strained primary storage systems. Dell PowerMax has addressed this phenomenon by offering the industry's first data reduction for mainframe while advancing data reduction rates for open systems storage. PowerMax is the only solution offering a 3:1 data reduction guarantee and an improved 4:1 guarantee for open systems (non-mainframe) data.

Another significant PowerMax enhancement focuses on cybersecurity, and Dell has taken advantage of CloudIQ's alerting capabilities to detect attacks and minimize risk and exposure proactively. PowerMax offers up to 65 million secure snapshots to enable rapid recovery from a cyber attack. The next generation of PowerMax also benefits multi-cloud deployments by providing secure data mobility between cloud and on-premises data centers.

In addition to these significant software updates, Dell launched two new PowerMax units: PowerMax 2500 and PowerMax 8500. The new PowerMax systems offer inclusive software to simplify ordering. The 2500 is a powerhouse in a small footprint and scales to 8PBe of effective capacity (2PB physical). That is 7x more capacity in half the rack space than previous models. The 8500 delivers 2x faster performance, 50 percent lower response times, and scales to 18PBe of effective capacity (4.5PB physical). For mainframe, both units support IBM zHyperLink read for applications that demand the fastest connectivity via synchronous IO.



Dell PowerMax is unique in that it delivers efficient workload consolidation, consolidating mainframe and mainstream (open systems) enterprise (heterogeneous environments) and IBM i data types, and file in the same unit. Several technical elements are required to enable this, but the fact that one physical array can handle multiple data sets is important.

Owing to the fact PowerMaxOS code base can be implemented across different data types, the mainframe use cases stand to benefit from the overall investment Dell continues to make in PowerMax.

High-Performance and Increased Capacity in the New PowerMax 2500 and 8500

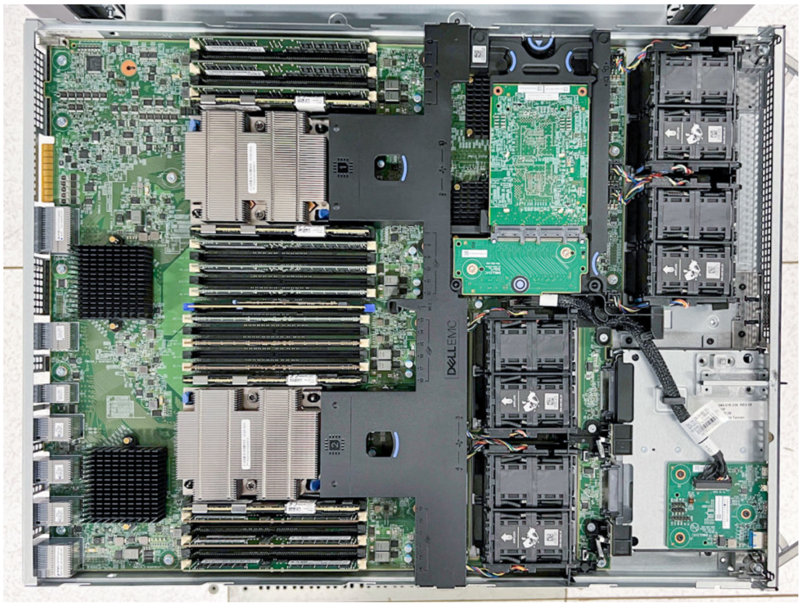
Dell has significantly enhanced the new PowerMax 2500 and 8500 storage arrays. Based on NVMe dynamic fabric technology, PowerMax eliminates traditional storage boundaries in performance, scalability, capacity, and security. Next-generation cloud-based applications and traditional workloads benefit from this new architecture.

The PowerMax 2500 is a compact, high-performance storage server, storing up to 7x more capacity (8 PBe) in half the rack space compared with the previous PowerMax 2000. Along with the high-efficiency design, the 2500 supports data services for heterogeneous environments, block, file, and virtual environments.

The PowerMax 8500 delivers outstanding performance at scale for demanding mixed workloads that require predictable performance and always-on availability while delivering up to 18 PBe capacity. The performance of the PowerMax 8500 is 2x faster, with 50 percent lower response times than the PowerMax 8000. Like the PowerMax 2500, the 8500 can easily consolidate open systems, mainframe, IBM i, file, and virtualized storage to simplify operations, significantly reducing TCO and increasing ROI.

The PowerMax 2500 and 8500 are the most energy-efficient enterprise-level storage platforms Dell has ever produced, delivering over 5x the effective capacity per watt consumed (PBe/watt) over the previous PowerMax generation.

The second-generation PowerMax models are more efficient, cramming 14x more capacity per rack unit, delivering 80 percent power savings per effective terabyte, doubling the performance, and cutting latency by 50 percent. With those numbers, PowerMax can significantly reduce TCO and increase ROI.



For a complete listing of PowerMax Specifications, download the PowerMax spec sheet.

More Efficient Modular Storage Hardware Platform

The next-generation arrays are built on modular building blocks called "nodes," similar to the PowerBrick from the previous-generation PowerMax. Nodes contain the primary compute elements (CPU and memory) of the second-generation systems. Each second-generation PowerMax system has at least two nodes or a "node pair." Each node has dual Intel Xeon Scalable processors, 24 DDR4 DIMM slots, and two 64-lane PCIe switches for front-end connections, among other advanced features.

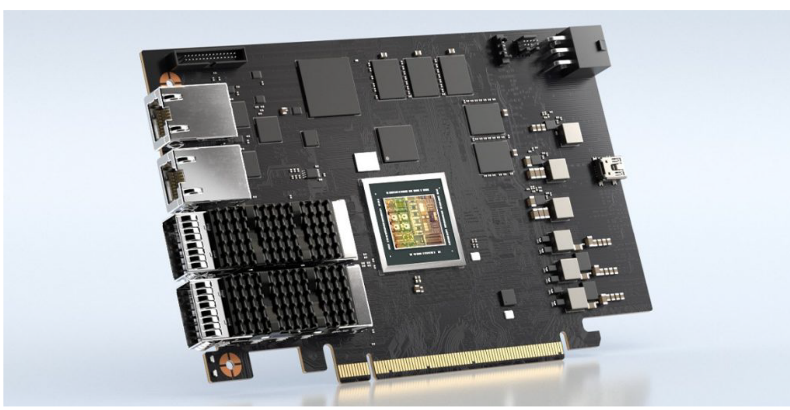
Dynamic Media Enclosure

The storage component for the PowerMax 2500 and 8500 is called the Dynamic Media Enclosure (DME). Each DME has 48 top-loading slots for 2.5" U.2-based NVMe flash drives that are side loaded into the DME slots. The PowerMax 2500 can scale up to two node pairs and two DMEs, while the PowerMax 8500 can scale up to eight node pairs and up to eight DMEs.

The DMEs are more than just high-density drive enclosures. The DMEs are “smart” fabric-attached units that include dual link controller cards (LCCs) for high availability. Each LCC includes NVIDIA’s BlueField DPU. The core of the PowerMax architecture is built around NVMe-oF with NVIDIA Quantum InfiniBand as the transport layer. The BlueField DPUs are key to making this possible.

Enhanced Scale and Performance with NVIDIA BlueField DPU

NVIDIA DPUs come standard with the new PowerMax 8500 and enhance the system’s overall scale and performance. Having the compute nodes and DMEs connected on the dynamic NVMe/InfiniBand fabric allows compute and media to scale independently. This means if an application is compute-intensive, customers can remedy the situation by adding more compute nodes. If capacity is an issue, add more DMEs and NVMe flash drives.



The dynamic fabric architecture of the PowerMax means that any node on the fabric has access to any data drive in the system, no matter which DME it is physically located in. This is a significant performance benefit since I/O from any host connected to any node in the array can be routed to any drive with efficiency and low latency. The DPU-enhanced architecture streamlines access, removing extra “hops” through adjacent nodes.

BlueField DPUs support secure boot, adding to the security of the PowerMax platform. PowerMax is designed for Zero Trust security architectures, with end-to-end security features in place to safeguard customer data.

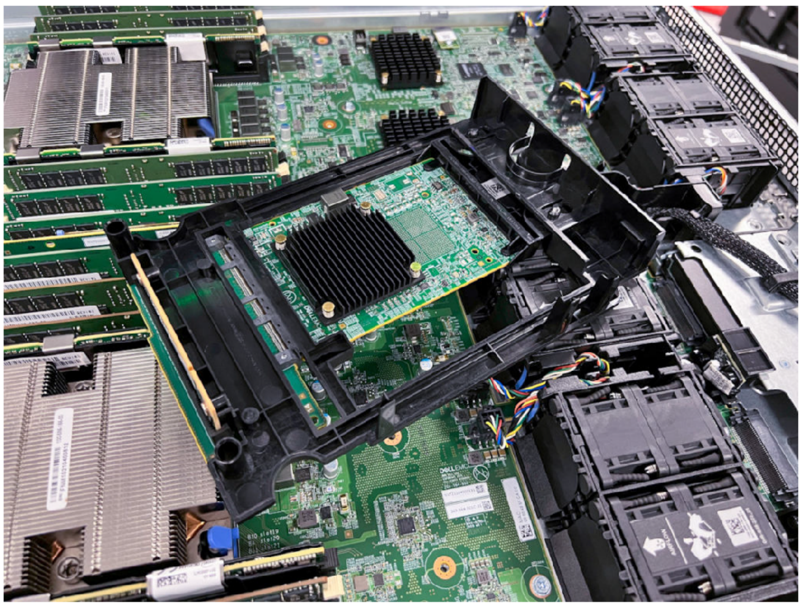
For more details, check out Dell's Blog on NVIDIA BlueField DPU Technology.

PowerMax Data Reduction Enhances Data Storage Efficiency

PowerMax storage platform is the first in the industry to offer complete mainframe data reduction. In fact, Dell guarantees a 3:1 data reduction ratio for mainframe and 4:1 for open systems on the 8500 and 2500.

How does Dell hit these lofty data reduction goals?

A key element (missing from the mainframe environment until this release) is the ability to boost overall system efficiency using data-reduction techniques.



By combining inline compression, inline deduplication (for non-mainframe data), pattern detection, efficient data placement, and machine learning, the system can write more host data than the overall available physical capacity and continue delivering the performance expected from an enterprise storage system.

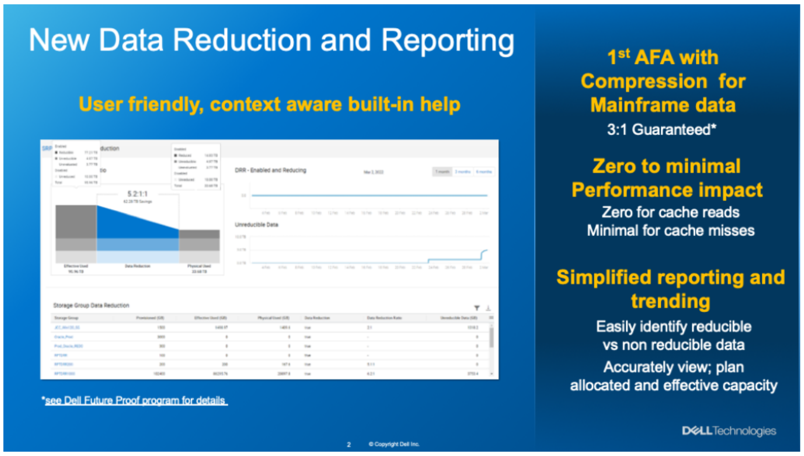
PowerMax looks at two specific resource areas for enhancing data storage efficiency: physical capacity and effective capacity.

- Physical Capacity is the total amount of physical capacity available after applying RAID protection.
- Effective Capacity is the amount of data the host can write with data reduction enabled. The effective capacity for the 2500 is 8PBe. The effective capacity for the 8500 is 18PBe.

All these features are great but implementing them independently would fail to deliver the required efficiency or performance. So, let's look at the individual components of each feature.

- Compression reduces the size of the data.
- Deduplication stores data as a single instance.
- Pattern detection includes a non-zero allocation function that excludes strings of consecutive zeros stored as part of compressed data.

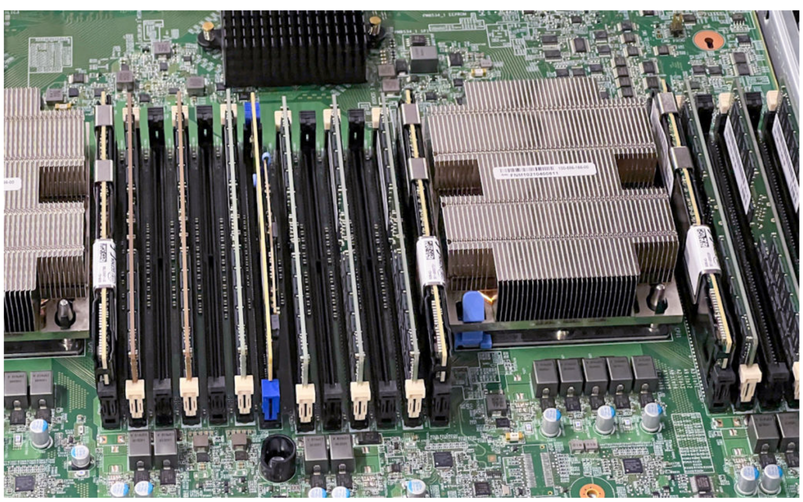
- Compression, dedupe, and pattern matching use hardware assistance to reduce overhead.
- Machine Learning identifies the data stored on disk that is accessed repeatedly and keeps it unreduced.
- Using a function called compaction, data is stored strategically to minimize wasted space and reduce the need for defrag functions.
- Activity Based Reduction (ABR) reduces processing resources.



To get a detailed report of how the system is managing resources, PowerMax uses Unisphere for PowerMax. Information regarding capacity usage, data reduction, and system resources is displayed in the capacity dashboard, with Unisphere for PowerMax providing multiple ways to display capacity usage.

PowerMax is the First Dell Storage Platform to Use Persistent Memory

The second-generation PowerMax system is the first Dell Technologies storage platform to use Persistent Memory (PMEM) DIMMs (Dual Inline Memory Modules). PowerMax uses PMEM to store system metadata, improve data vaulting efficiency, lower TCO, and reduce overall system footprint.



The primary value proposition for using PMEM in the new PowerMax is it enables an overall lower cost of system ownership because workload density and capacity can be increased using a smaller overall system footprint.

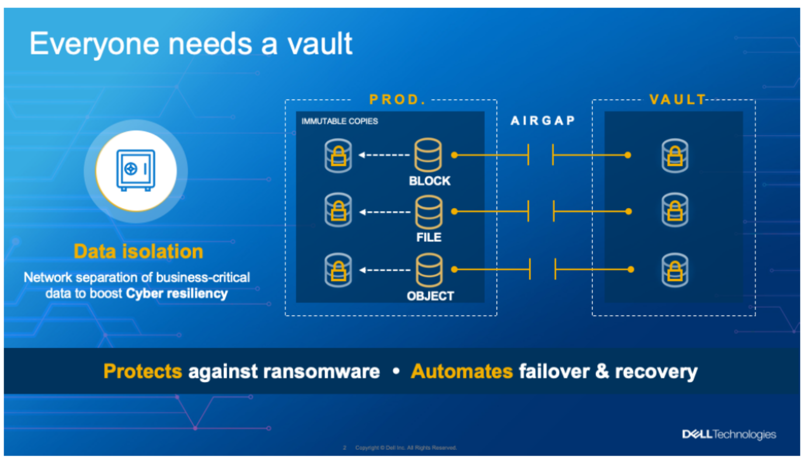
Flexible RAID Reduces Overhead

Disaggregation of storage and compute has allowed Dell to implement a new RAID distribution scheme called Flexible RAID Technology. Flexible RAID gives customers more granularity and configuration options, reduced RAID overhead, and higher availability.

Flexible RAID provides all compute nodes in the system with active-active access to storage resources that are distributed across all DMEs. The technology reduces RAID overhead, allowing higher system capacity while using fewer drives. For example, with Flexible RAID, 1 TB can be rebuilt in less than 10 minutes.

The Importance of Cybersecurity for Storage

Cyber security remains a hot topic, given the continued damaging cyber attacks over the past few years. Cyber attacks are expensive, brutal to prevent, and can ruin the reputation of any organization. Ransomware attacks are on the increase from all directions. IT administrators are typically held to account when an attack occurs, so they take an aggressive, proactive position in preventing cyber-attacks.



Cyber Recovery (CR) vaults and immutable copies of data provide the mandatory "fail-safe" protection to protect and preserve an organization's business strategy. Although the concept of Disaster Recovery (DR) remains a requirement, both DR and CR must coexist to recover from a cyber attack.

Addressable but immutable (in other words, indestructible) copies of current production data serve as the first line of defense in the recovery of production data after an attack. Immutable copies are non-addressable, essentially "invisible" mirror images of production data that have a defined "time to expire" date and time. Dell's immutable copies are "space-efficient" in terms of the amount of storage capacity required because they are pointer-based.

A Cyber vault is mostly disconnected and non-addressable (optionally air-gapped) by any network and inaccessible by even production servers. Think of the cyber vault data as the "plan B" for the immutable copies of production data. Air-gapping or decoupling the vault from your network is what makes the vault inaccessible to cyber criminals and is a means to "double down" on ensuring the company will have uncorrupted data to restore the business after an attack or ransom event.

Dell has taken a comprehensive approach to create data preservation and resilience regarding data protection on PowerMax. Unified controls, features, and functions related to cyber protection have been implemented across the entire product line, not simply limited to PowerMax. Dell has defined security controls based on industry standards referred to as the Dell secure development life cycle (SDL) which includes analysis activities and prescriptive proactive controls. There is also the Dell Product Security Incident Response Team (PSIRT), a chartered team responsible for coordinating the response and disclosure of all reported product vulnerabilities.

PowerMax Cyber Security for Heterogeneous Environments

With PowerMax, users benefit from a large number of cyber data protection and resiliency enhancements for storing both open systems and mainframe data, such as:

- Hardware Root of Trust (HWRoT) cryptographically affirms the integrity of BIOS and BMC firmware. HWRoT is based on one-time programmable, read-only public keys provisioned by Dell in the factory to protect against malware tampering.
- Secure Boot represents an industry-wide standard for security in the preboot environment. Secure boot verifies the image to be booted is precisely the image expected.
- Snapshot Policies protect applications automatically and require little or no maintenance. PowerMax supports up to 1,024 snapshots per device and 65 million snapshots per storage array. Users have the option of creating secure snaps by setting a retention period on snapshots. A secure snap cannot be terminated during the retention period. Once the retention time is reached, the snapshot is automatically terminated.
- Data at Rest Encryption (D@RE) provides hardware-based, on-array, backend encryption for PowerMax and VMAX All Flash systems. Back-end encryption protects information from unauthorized access when drives are removed from the system.
- CloudIQ combines proactive monitoring, machine learning, and predictive analytics to identify risks and anomalies in the storage environment. The cybersecurity component of CloudIQ continuously compares the configuration of the PowerMax array based on a set of customer-selected, security-related evaluation tests. If a deviation is detected, CloudIQ will notify the customer and provide remediation to correct the issue.
- Unisphere is a web-based application that enables customers to configure, administer, monitor, and troubleshoot PowerMax.

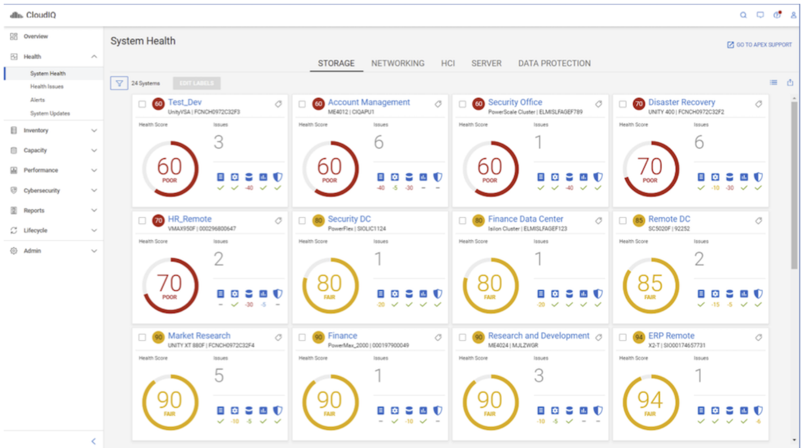
Dell also designed some unique mainframe capabilities into the PowerMax for improved cyber resiliency. Although not the first thing to come to mind when thinking of cybersecurity, mainframes play an important role for customers. Mainframes are viewed as the System of Record (SOR) for organizations. The platform delivers unmatched high-speed transaction processing, and mainframes offer Reliability, Availability, and Survivability (RAS).

Additional mainframe capabilities include:

- zDP, Data Protector for z Systems, provides a granular level of protection for mainframe data which can create up to 1024 point-in-time copies of an entire z/OS environment (tens of thousands of devices across multiple PowerMax arrays) as frequently as every five minutes. This ensures that a processing error, intentional or due to human error, or a malicious attack, such as ransomware, does not impact the IT environment. The optional zDP two-actor security feature, based on z/OS security and enabled by PowerMax, ensures two people are required to alter the operational setting of the zDP solution.
- Geographically Dispersed Disaster Restart (GDDR) is a mainframe software product that automates storage failover and business-recovery procedures by reacting to events in the IT environment. Dell recently added a new component to enhance cyber resiliency called zCPA, Cyber Protection Automation for z Systems to automate the preservation of data in cyber vault arrays on a periodic basis.

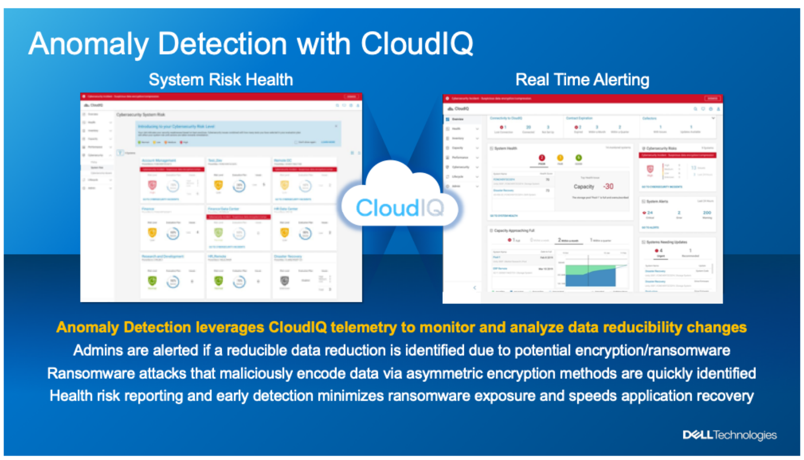
Proactively Monitoring PowerMax with CloudIQ

PowerMax includes CloudIQ, a cloud-based application that provides a detailed investigation of PowerMax arrays, including integration with VMware. Customers are provided with an independent, secure portal that allows them to register their PowerMax arrays and monitor storage from a single console. The secure portal ensures customers are restricted to those devices in their environment. CloudIQ is a powerful application that leverages machine learning and also tracks system health through pattern recognition and advanced analytics.



CloudIQ constantly compares the configuration of the PowerMax system to a set of user-selected, security-related evaluation tests. Upon identifying a deviation between the actual and desired configuration setting, CloudIQ proactively notifies users of the violation and provides remediation steps to correct the issue.

CloudIQ can be set to provide proactive notifications to the user in the event of an infrastructure security risk. The Security Advisories section of the Cybersecurity feature in CloudIQ notifies users of relevant Dell and VMware Security Advisories. Users quickly see a summary of vulnerabilities specific to their systems and versions, along with links to remediation details.



With CloudIQ cybersecurity, users can define legal configurations for PowerMax, monitor the system, and receive alerts if the array is out of compliance. CloudIQ can also track data patterns and detect anomalies, including changes to data reduction rates, to determine whether ransomware or malware may have infected the system. When suspicious anomalies are detected, CloudIQ alerts IT management to take corrective action.

Intelligent Automation

PowerMax systems are designed with intelligent automation in mind. They support advanced AIOps, DevOps, and containers to streamline operations and eliminate redundancy, so IT practitioners can focus on strategic initiatives. PowerMax brings autonomous storage to life with built-in machine learning that uses predictive analytics and pattern recognition to maximize performance with no management overhead. Automated storage provisioning for open systems workloads is accomplished using a simple REST API, saving considerable time and effort. And PowerMaxOS 10 provides the industry's first software-defined NVMe/TCP utility for storage resource automation, resulting in 44% less

time to set up NVMe/TCP resources. NVMe/TCP helps lower deployment costs, reduces SAN design complexity, and allows for building a highly scalable PowerMax storage environment for mission-critical workloads.

Automating storage administration has become increasingly important as infrastructures scale to meet increasing demands. Automation needs to be well thought through and designed in a way that can scale across organizations, processes, and hybrid cloud infrastructure. Dell Technologies offers a range of solutions to integrate with automation tools using the PowerMax REST API that are becoming industry standards.

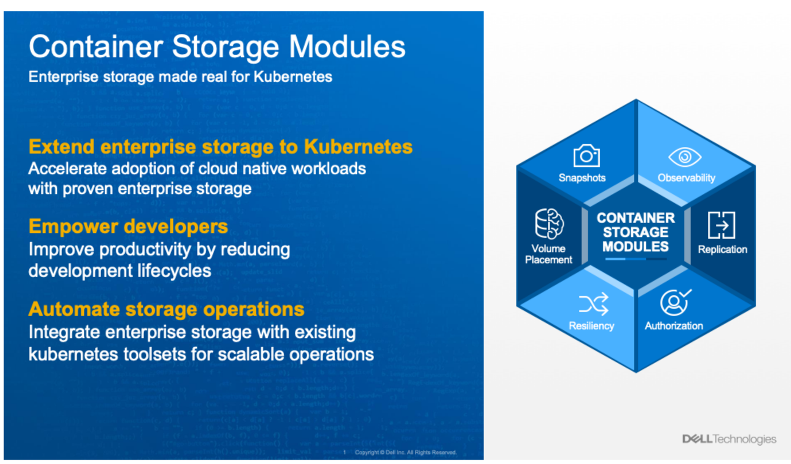
With PowerMax, customers can use a REST API that provides access to the storage system to build automation scripts and playbooks using tools such as Ansible. The PowerMax platform is the first storage platform in the industry that can perform provisioning and other administrative tasks for open-system block, file, and mainframe workloads using a single unified, comprehensive REST API toolkit.

Multi-Array Workload Optimization

Multi-array workload planner analyzes the storage infrastructure across multiple PowerMax/VMAX arrays and recommends the best place to host workloads for optimal performance and resource utilization. Built-in data movement technology provides seamless data mobility across PowerMax and VMAX arrays using array-based orchestration and replication services to automatically discover, configure, and migrate data online. CloudIQ Health Check gives administrators faster time to insight; with all the information needed to take quick action and efficiently manage their storage environment. It enables proactive monitoring and predictive analytics to deliver alerts, aggregated PowerMax health scores, and provide proactive assistance with actionable insights and recommended remediation – all from the cloud and from your mobile devices, free of charge.

DevOps Automation and Containers

PowerMax customers can seamlessly consume storage infrastructure as code in a variety of development and automation environments using powerful APIs, SDKs, plugins for VMware automation tools like vRO and vRA, and modules for the most popular configuration management tools like Ansible. PowerMax supports a major shift in software development by being the first major enterprise storage solution to implement the Container Storage Interface (CSI) driver standard to enable containerized storage workloads to optimize productivity.



Dell Container Storage Modules (CSM) build on top of the Container Storage Interface (CSI) foundation to deliver unique, powerful storage and enterprise capabilities. CSM makes enterprise storage real for Kubernetes with simple, consistent integration and automation for DevOps and IT across storage and cloud-native stateful apps.

CSM accelerates cloud-native workload adoption with enterprise storage and enables a high-performing and resilient storage foundation for Kubernetes. CSM delivers a full stack of enterprise capabilities such as replication, authorization, failure recovery, and management. The inclusion of these capabilities accelerates deployment testing, which results in a faster application deployment lifecycle. Get involved on GitHub's CSI Driver for Dell PowerMax for even more detail.

Storage admins take advantage of these benefits for PowerMax, automating storage operations with existing Kubernetes toolsets for scalable operations and delivering an integrated experience bridging the gap between Kubernetes admins/developers and traditional IT admins.

Another core automation feature is optimized workload placement. In a storage infrastructure with multiple PowerMax arrays, the systems send information to Unisphere regarding storage usage and utilization. Whether a user provisions storage manually through Unisphere or using scripts, they can allow the system to determine which PowerMax storage array is best suited to support the new workload.

Efficient Workload Consolidation

PowerMax is designed to consolidate mixed workloads while delivering consistently high performance. PowerMax's scale-up and scale-out architecture are ideal for relational databases, real-time analytics, demanding transaction processing workloads, and big data applications that require uncompromising uptime and extremely low latency.

As previously described, the new PowerMax is the only platform in the industry where customers can natively run mainframe, open systems block, and file workloads on the same system without using gateways or third-party solutions. The ability to run mainframe workloads natively along with OS block and file is available on the PowerMax 2500 and 8500 and the PowerMax 8000.

The PowerMax 2500 and 8500 also feature a completely redesigned 64-bit, fully embedded NAS file platform for SMB and NFS workloads. This new file platform can provide four data movers (virtual machines acting as file servers) on the PowerMax 2500 and eight data movers on the PowerMax 8500. Each data mover can provide up to 512 TB of usable capacity, which can be carved up into 64 TB file systems.

Mainframe and File Enhancements

Dell also made significant mainframe and file enhancements.

File enhancements include:

- Redesigned 64-bit containerized microservice architecture
- Support for up to four file servers (PowerMax 2500); support for up to eight file servers (PowerMax 8500)
- Active/Active high-availability architecture scalable across 2-16 nodes
- Single Global Namespace file access scalable across all nodes
- 64 TB SMB and NFS file system support
- Data service integration for SRDF/S and SRDF/A, TimeFinder Snap, service levels, data reduction, D@RE, and non-disruptive upgrades
- Single I/O module for file, iSCSI, and NVMe/TCP

Mainframe enhancements include:

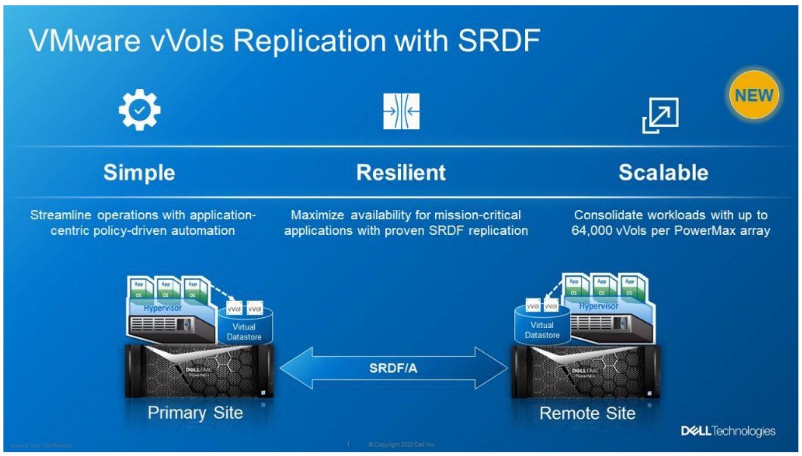
- Mainframe workloads can run on the entry class PowerMax 2500
- The PowerMax 2500 and 8500 are the first storage platform to offer complete mainframe data reduction, guaranteeing a 3:1 data reduction ratio

- End-to-end 32Gb FICON support using the same I/O module as 32Gb Fibre Channel
- Hardware support for 8Gb zHyperLink reads using a PCIe connection to IBM z Systems hosts

Dell PowerMax Tight Integration with VMware

Dell has years of experience working with VMware and support is tightly integrated into many Dell servers, storage systems, and HCI appliances. PowerMax is no different and provides seamless VMware integration, delivering the highest scalability for VMware vVols deployments and high availability. With impressive linear scale-out, PowerMax increases storage capacity and performance while achieving 100 percent uptime.

PowerMax has consistently delivered innovative capabilities to support VMware environments running mission-critical workloads. Examples include offering PowerMax as a principal storage platform for VMware Cloud Foundation (VCF), providing customers with high-performance storage for their hybrid cloud deployments. In addition, Dell Technologies and VMware have partnered to provide integrations for vVols, vRealize Orchestrator, and vRealize Automation to further extend high-end storage access to mission-critical workloads.



PowerMax has been architected to consolidate virtualized workloads, resulting in increased performance. vSphere deployments benefit from integrated data reduction, automatic data placement, and machine learning that streamline storage operations. PowerMax is also qualified for deployment in a Virtual Infrastructure domain through VCF.

The latest iteration of PowerMax brings greater simplicity, scalability, and data resiliency for VMware deployments by integrating PowerMax SRDF/A replication with VMware vSphere vVols and VMware Site Recovery Manager (SRM). Automating VM movement between sites ensures maximum availability for mission-critical applications. Further, as customers migrate from a hardware-centric storage approach to an application-centric method for managing storage with VMware vVols, Dell PowerMax runs virtual volumes at scale (64,000 vVols) with the highest levels of resiliency.

PowerMax integration with VMware management makes it easy for VMware administrators to manage storage and improve efficiency. Critically, with deep VMware support, this means PowerMax has a unique ability to help organizations consolidate mainframe, block, file, and virtualized data on a single mission-critical storage array.

VMware vRealize Orchestrator (vRO) is a process automation tool that performs automated management and operational tasks across VMware and third-party applications. With vRO, automated routines are created for workflows using simple drag-and-drop. Dell offers vRO plug-ins for PowerMax, bringing a deeper range of storage functionality in the form of programmable blocks that can be dropped into a workflow process map. Specifically, the functionality includes storage provisioning, scheduled and on-demand snapshots, VMware-integrated storage operations, and more.

VMware vRealize Automation (vRA) makes PowerMax vRO workflow automation recipes into an anything-as-a-service catalog for the entire IT ecosystem. Workflows automated in vRO can also be used in the self-serve catalog.

Final Thoughts

The Dell PowerMax storage platform features two new hardware models, the 2500 and 8500. PowerMaxOS 10 has been released with over 200 new features and functions, making PowerMax the standout storage platform for mainframes. A key feature of the new platform is the ability to support provisioning and administrative tasks for open-system block, file, and mainframe workloads using a single REST API toolkit.

Dell has incorporated a comprehensive set of cyber security features to prevent malicious attacks and secure corporate data. The addition of HWRoT, Secure Boot, MFA, Secure access controls, and detection and response using CloudIQ make this new release a storage system fortress. And PowerMaxOS 10 provides the industry's first software-defined NVMe/TCP utility for storage resource automation, resulting in 44 percent less setup time for NVMe/TCP resources.

Mainframes offer large organizations reliable, secure and high-performance storage capable of handling mission-critical applications. Dell PowerMax not only addresses these needs but adds additional arrows to the quiver by offering guaranteed 3:1 data reduction, additional security, and data visibility features for mainframe workloads. Support for heterogeneous environments is unique in this space. Given the extremely deep data services and the ability to handle literally any enterprise workload in a single cluster, Dell PowerMax clearly has differentiated itself in the mainframe world as arguably the most versatile solution on the market.

This next-generation PowerMax delivers performance, extreme scalability, low latency, and high availability. Customers benefit from the ability to modernize without disruption with data-in-place Anytime Upgrade and Dell's Future-proof program.

Additional resources:

- Dell Paper on PowerMax Data Reduction
- Dell PowerMax website
- Dell Mainframe Solutions website

This report is sponsored by Dell Technologies. All views and opinions expressed in this report are based on our unbiased view of the product(s) under consideration.

Engage with StorageReview

Newsletter | YouTube | Podcast iTunes /Spotify | Instagram | Twitter | TikTok | RSS Feed

BRIAN BEELER

Brian is the owner and President of StorageReview.com.

previous post
Lenovo ThinkPad X13s Snapdragon Review

next post
Dell Latitude 9330 2-in-1 Review

Facebook Twitter Instagram Linkedin Youtube Email Spotify Red

Reprinted with corrections from *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities. For example, adding one relay to a group doubles the number of possible states of the relays. It adds 1 to the base 2 logarithm of this number. Doubling the time roughly squares the number of possible messages, or doubles the logarithm, etc.

2. It is nearer to our intuitive feeling as to the proper measure. This is closely related to (1) since we intuitively measures entities by linear comparison with common standards. One feels, for example, that two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information.

3. It is mathematically more suitable. Many of the limiting operations are simple in terms of the logarithm but would require clumsy restatement in terms of the number of possibilities.

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits*, a word suggested by J. W. Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2^N and $\log_2 2^N = N$. If the base 10 is used the units may be called decimal digits. Since

$$\begin{aligned} \log_2 M &= \log_{10} M / \log_{10} 2 \\ &= 3.32 \log_{10} M, \end{aligned}$$

[1]Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal,* April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A.I.E.E. Trans.,* v. 47, April 1928, p. 617.

[2]Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal,* July 1928, p. 535.

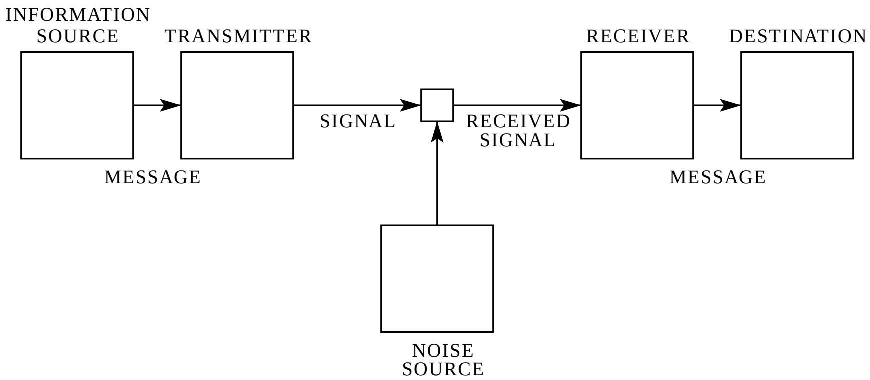


Fig. 1 — Schematic diagram of a general communication system.

a decimal digit is about $3\frac{1}{3}$ bits. A digit wheel on a desk computing machine has ten stable positions and therefore has a storage capacity of one decimal digit. In analytical work where integration and differentiation are involved the base e is sometimes useful. The resulting units of information will be called natural units. Change from the base a to base b merely requires multiplication by $\log_b a$.

By a communication system we will mean a system of the type indicated schematically in Fig. 1. It consists of essentially five parts:

1. An *information source* which produces a message or sequence of messages to be communicated to the receiving terminal. The message may be of various types: (a) A sequence of letters as in a telegraph of teletype system; (b) A single function of time $f(t)$ as in radio or telephony; (c) A function of time and other variables as in black and white television — here the message may be thought of as a function $f(x,y,t)$ of two space coordinates and time, the light intensity at point (x,y) and time t on a pickup tube plate; (d) Two or more functions of time, say $f(t)$, $g(t)$, $h(t)$ — this is the case in "three-dimensional" sound transmission or if the system is intended to service several individual channels in multiplex; (e) Several functions of several variables — in color television the message consists of three functions $f(x,y,t)$, $g(x,y,t)$, $h(x,y,t)$ defined in a three-dimensional continuum — we may also think of these three functions as components of a vector field defined in the region — similarly, several black and white television sources would produce "messages" consisting of a number of functions of three variables; (f) Various combinations also occur, for example in television with an associated audio channel.

2. A *transmitter* which operates on the message in some way to produce a signal suitable for transmission over the channel. In telephony this operation consists merely of changing sound pressure into a proportional electrical current. In telegraphy we have an encoding operation which produces a sequence of dots, dashes and spaces on the channel corresponding to the message. In a multiplex PCM system the different speech functions must be sampled, compressed, quantized and encoded, and finally interleaved properly to construct the signal. Vocoder systems, television and frequency modulation are other examples of complex operations applied to the message to obtain the signal.

3. The *channel* is merely the medium used to transmit the signal from transmitter to receiver. It may be a pair of wires, a coaxial cable, a band of radio frequencies, a beam of light, etc.

4. The *receiver* ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal.

5. The *destination* is the person (or thing) for whom the message is intended.

We wish to consider certain general problems involving communication systems. To do this it is first necessary to represent the various elements involved as mathematical entities, suitably idealized from their

physical counterparts. We may roughly classify communication systems into three main categories: discrete, continuous and mixed. By a discrete system we will mean one in which both the message and the signal are a sequence of discrete symbols. A typical case is telegraphy where the message is a sequence of letters and the signal a sequence of dots, dashes and spaces. A continuous system is one in which the message and signal are both treated as continuous functions, e.g., radio or television. A mixed system is one in which both discrete and continuous variables appear, e.g., PCM transmission of speech.

We first consider the discrete case. This case has applications not only in communication theory, but also in the theory of computing machines, the design of telephone exchanges and other fields. In addition the discrete case forms a foundation for the continuous and mixed cases which will be treated in the second half of the paper.

PART I: DISCRETE NOISELESS SYSTEMS

1. The Discrete Noiseless Channel

Teletype and telegraphy are two simple examples of a discrete channel for transmitting information. Generally, a discrete channel will mean a system whereby a sequence of choices from a finite set of elementary symbols $S_1, \dots, S_n$ can be transmitted from one point to another. Each of the symbols S_i is assumed to have a certain duration in time t_i seconds (not necessarily the same for different S_i, for example the dots and dashes in telegraphy). It is not required that all possible sequences of the S_i be capable of transmission on the system; certain sequences only may be allowed. These will be possible signals for the channel. Thus in telegraphy suppose the symbols are: (1) A dot, consisting of line closure for a unit of time and then line open for a unit of time; (2) A dash, consisting of three time units of closure and one unit open; (3) A letter space consisting of, say, three units of line open; (4) A word space of six units of line open. We might place the restriction on allowable sequences that no spaces follow each other (for if two letter spaces are adjacent, it is identical with a word space). The question we now consider is how one can measure the capacity of such a channel to transmit information.

In the teletype case where all symbols are of the same duration, and any sequence of the 32 symbols is allowed the answer is easy. Each symbol represents five bits of information. If the system transmits n symbols per second it is natural to say that the channel has a capacity of $5n$ bits per second. This does not mean that the teletype channel will always be transmitting information at this rate — this is the maximum possible rate and whether or not the actual rate reaches this maximum depends on the source of information which feeds the channel, as will appear later.

In the more general case with different lengths of symbols and constraints on the allowed sequences, we make the following definition:
Definition: The capacity C of a discrete channel is given by

$$C = \operatorname*{Lim}_{T\to\infty} \frac{\log N(T)}{T}$$

where $N(T)$ is the number of allowed signals of duration T.

It is easily seen that in the teletype case this reduces to the previous result. It can be shown that the limit in question will exist as a finite number in most cases of interest. Suppose all sequences of the symbols $S_1, \dots, S_n$ are allowed and these symbols have durations $t_1, \dots, t_n$. What is the channel capacity? If $N(t)$ represents the number of sequences of duration t we have

$$N(t) = N(t-t_1) + N(t-t_2) + \cdots + N(t-t_n).$$

The total number is equal to the sum of the numbers of sequences ending in $S_1, S_2, \dots, S_n$ and these are $N(t-t_1), N(t-t_2), \dots, N(t-t_n)$, respectively. According to a well-known result in finite differences, $N(t)$ is then asymptotic for large t to X_0^t where X_0 is the largest real solution of the characteristic equation:

$$X^{-t_1} + X^{-t_2} + \cdots + X^{-t_n} = 1$$

and therefore

$$C = \log X_0.$$

In case there are restrictions on allowed sequences we may still often obtain a difference equation of this type and find C from the characteristic equation. In the telegraphy case mentioned above

$$N(t) = N(t-2) + N(t-4) + N(t-5) + N(t-7) + N(t-8) + N(t-10)$$

as we see by counting sequences of symbols according to the last or next to the last symbol occurring. Hence C is $-\log \mu_0$ where μ_0 is the positive root of $1 = \mu^2 + \mu^4 + \mu^5 + \mu^7 + \mu^8 + \mu^{10}$. Solving this we find $C = 0.539$.

A very general type of restriction which may be placed on allowed sequences is the following: We imagine a number of possible states $a_1, a_2, \ldots, a_m$. For each state only certain symbols from the set $S_1, \ldots, S_n$ can be transmitted (different subsets for the different states). When one of these has been transmitted the state changes to a new state depending both on the old state and the particular symbol transmitted. The telegraph case is a simple example of this. There are two states depending on whether or not a space was the last symbol transmitted. If so, then only a dot or a dash can be sent next and the state always changes. If not, any symbol can be transmitted and the state changes if a space is sent, otherwise it remains the same. The conditions can be indicated in a linear graph as shown in Fig. 2. The junction points correspond to the

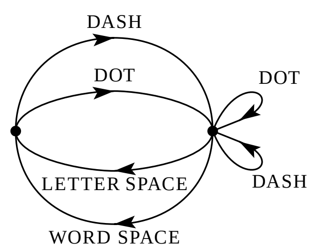


Fig. 2—Graphical representation of the constraints on telegraph symbols.

states and the lines indicate the symbols possible in a state and the resulting state. In Appendix 1 it is shown that if the conditions on allowed sequences can be described in this form C will exist and can be calculated in accordance with the following result:

Theorem 1: Let $b_{ij}^{(s)}$ be the duration of the s^{th} symbol which is allowable in state i and leads to state j. Then the channel capacity C is equal to $\log W$ where W is the largest real root of the determinant equation:

$$\left| \sum_s W^{-b_{ij}^{(s)}} - \delta_{ij} \right| = 0$$

where $\delta_{ij} = 1$ if $i = j$ and is zero otherwise.

For example, in the telegraph case (Fig. 2) the determinant is:

$$\begin{vmatrix} -1 & (W^{-2} + W^{-4}) \\ (W^{-3} + W^{-6}) & (W^{-2} + W^{-4} - 1) \end{vmatrix} = 0.$$

On expansion this leads to the equation given above for this case.

2. The Discrete Source of Information

We have seen that under very general conditions the logarithm of the number of possible signals in a discrete channel increases linearly with time. The capacity to transmit information can be specified by giving this rate of increase, the number of bits per second required to specify the particular signal used.

We now consider the information source. How is an information source to be described mathematically, and how much information in bits per second is produced in a given source? The main point at issue is the effect of statistical knowledge about the source in reducing the required capacity of the channel, by the use

of proper encoding of the information. In telegraphy, for example, the messages to be transmitted consist of sequences of letters. These sequences, however, are not completely random. In general, they form sentences and have the statistical structure of, say, English. The letter E occurs more frequently than Q, the sequence TH more frequently than XP, etc. The existence of this structure allows one to make a saving in time (or channel capacity) by properly encoding the message sequences into signal sequences. This is already done to a limited extent in telegraphy by using the shortest channel symbol, a dot, for the most common English letter E; while the infrequent letters, Q, X, Z are represented by longer sequences of dots and dashes. This idea is carried still further in certain commercial codes where common words and phrases are represented by four- or five-letter code groups with a considerable saving in average time. The standardized greeting and anniversary telegrams now in use extend this to the point of encoding a sentence or two into a relatively short sequence of numbers.

We can think of a discrete source as generating the message, symbol by symbol. It will choose successive symbols according to certain probabilities depending, in general, on preceding choices as well as the particular symbols in question. A physical system, or a mathematical model of a system which produces such a sequence of symbols governed by a set of probabilities, is known as a stochastic process.[3] We may consider a discrete source, therefore, to be represented by a stochastic process. Conversely, any stochastic process which produces a discrete sequence of symbols chosen from a finite set may be considered a discrete source. This will include such cases as:

1. Natural written languages such as English, German, Chinese.

2. Continuous information sources that have been rendered discrete by some quantizing process. For example, the quantized speech from a PCM transmitter, or a quantized television signal.

3. Mathematical cases where we merely define abstractly a stochastic process which generates a sequence of symbols. The following are examples of this last type of source.

   (A) Suppose we have five letters A, B, C, D, E which are chosen each with probability .2, successive choices being independent. This would lead to a sequence of which the following is a typical example.

   B D C B C E C C C A D C B D D A A E C E E A
   A B B D A E E C A C E E B A E E C B C E A D.

   This was constructed with the use of a table of random numbers.[4]

   (B) Using the same five letters let the probabilities be .4, .1, .2, .2, .1, respectively, with successive choices independent. A typical message from this source is then:

   A A A C D C B D C E A A D A D A C E D A
   E A D C A B E D A D D C E C A A A A A D.

   (C) A more complicated structure is obtained if successive symbols are not chosen independently but their probabilities depend on preceding letters. In the simplest case of this type a choice depends only on the preceding letter and not on ones before that. The statistical structure can then be described by a set of transition probabilities $p_i(j)$, the probability that letter i is followed by letter j. The indices i and j range over all the possible symbols. A second equivalent way of specifying the structure is to give the "digram" probabilities $p(i,j)$, i.e., the relative frequency of the digram $i\,j$. The letter frequencies $p(i)$, (the probability of letter i), the transition probabilities

[3] See, for example, S. Chandrasekhar, "Stochastic Problems in Physics and Astronomy," *Reviews of Modern Physics*, v. 15, No. 1, January 1943, p. 1.

[4] Kendall and Smith, *Tables of Random Sampling Numbers*, Cambridge, 1939.

$p_i(j)$ and the digram probabilities $p(i,j)$ are related by the following formulas:

$$p(i) = \sum_j p(i,j) = \sum_j p(j,i) = \sum_j p(j) p_j(i)$$

$$p(i,j) = p(i) p_i(j)$$

$$\sum_j p_i(j) = \sum_i p(i) = \sum_{i,j} p(i,j) = 1.$$

As a specific example suppose there are three letters A, B, C with the probability tables:

$p_i(j)$		j: A	B	C
	A	0	$\frac{4}{5}$	$\frac{1}{5}$
i	B	$\frac{1}{2}$	$\frac{1}{2}$	0
	C	$\frac{1}{2}$	$\frac{2}{5}$	$\frac{1}{10}$

i	$p(i)$
A	$\frac{9}{27}$
B	$\frac{16}{27}$
C	$\frac{2}{27}$

$p(i,j)$		j: A	B	C
	A	0	$\frac{4}{15}$	$\frac{1}{15}$
i	B	$\frac{8}{27}$	$\frac{8}{27}$	0
	C	$\frac{1}{27}$	$\frac{4}{135}$	$\frac{1}{135}$

A typical message from this source is the following:

A B B A B A B A B A B A B A B B B A B B B B B A B A B A B A B A B B B A C A C A B B A B B B B A B B A B A C B B B A B A.

The next increase in complexity would involve trigram frequencies but no more. The choice of a letter would depend on the preceding two letters but not on the message before that point. A set of trigram frequencies $p(i,j,k)$ or equivalently a set of transition probabilities $p_{ij}(k)$ would be required. Continuing in this way one obtains successively more complicated stochastic processes. In the general n-gram case a set of n-gram probabilities $p(i_1,i_2,\dots,i_n)$ or of transition probabilities $p_{i_1,i_2,\dots,i_{n-1}}(i_n)$ is required to specify the statistical structure.

(D) Stochastic processes can also be defined which produce a text consisting of a sequence of "words." Suppose there are five letters A, B, C, D, E and 16 "words" in the language with associated probabilities:

.10 A	.16 BEBE	.11 CABED	.04 DEB
.04 ADEB	.04 BED	.05 CEED	.15 DEED
.05 ADEE	.02 BEED	.08 DAB	.01 EAB
.01 BADD	.05 CA	.04 DAD	.05 EE

Suppose successive "words" are chosen independently and are separated by a space. A typical message might be:

DAB EE A BEBE DEED DEB ADEE ADEE EE DEB BEBE BEBE BEBE ADEE BED DEED DEED CEED ADEE A DEED DEED BEBE CABED BEBE BED DAB DEED ADEB.

If all the words are of finite length this process is equivalent to one of the preceding type, but the description may be simpler in terms of the word structure and probabilities. We may also generalize here and introduce transition probabilities between words, etc.

These artificial languages are useful in constructing simple problems and examples to illustrate various possibilities. We can also approximate to a natural language by means of a series of simple artificial languages. The zero-order approximation is obtained by choosing all letters with the same probability and independently. The first-order approximation is obtained by choosing successive letters independently but each letter having the same probability that it has in the natural language.[5] Thus, in the first-order approximation to English, E is chosen with probability .12 (its frequency in normal English) and W with probability .02, but there is no influence between adjacent letters and no tendency to form the preferred

[5]Letter, digram and trigram frequencies are given in *Secret and Urgent* by Fletcher Pratt, Blue Ribbon Books, 1939. Word frequencies are tabulated in *Relative Frequency of English Speech Sounds,* G. Dewey, Harvard University Press, 1923.

digrams such as TH, ED, etc. In the second-order approximation, digram structure is introduced. After a letter is chosen, the next one is chosen in accordance with the frequencies with which the various letters follow the first one. This requires a table of digram frequencies $p_i(j)$. In the third-order approximation, trigram structure is introduced. Each letter is chosen with probabilities which depend on the preceding two letters.

3. The Series of Approximations to English

To give a visual idea of how this series of processes approaches a language, typical sequences in the approximations to English have been constructed and are given below. In all cases we have assumed a 27-symbol "alphabet," the 26 letters and a space.

1. Zero-order approximation (symbols independent and equiprobable).

   XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

2. First-order approximation (symbols independent but with frequencies of English text).

   OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

3. Second-order approximation (digram structure as in English).

   ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

4. Third-order approximation (trigram structure as in English).

   IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

5. First-order word approximation. Rather than continue with tetragram, . . . , *n*-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.

   REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.

   THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

The resemblance to ordinary English text increases quite noticeably at each of the above steps. Note that these samples have reasonably good structure out to about twice the range that is taken into account in their construction. Thus in (3) the statistical process insures reasonable text for two-letter sequences, but four-letter sequences from the sample can usually be fitted into good sentences. In (6) sequences of four or more words can easily be placed in sentences without unusual or strained constructions. The particular sequence of ten words "attack on an English writer that the character of this" is not at all unreasonable. It appears then that a sufficiently complex stochastic process will give a satisfactory representation of a discrete source.

The first two samples were constructed by the use of a book of random numbers in conjunction with (for example 2) a table of letter frequencies. This method might have been continued for (3), (4) and (5), since digram, trigram and word frequency tables are available, but a simpler equivalent method was used.

To construct (3) for example, one opens a book at random and selects a letter at random on the page. This letter is recorded. The book is then opened to another page and one reads until this letter is encountered. The succeeding letter is then recorded. Turning to another page this second letter is searched for and the succeeding letter recorded, etc. A similar process was used for (4), (5) and (6). It would be interesting if further approximations could be constructed, but the labor involved becomes enormous at the next stage.

4. Graphical Representation of a Markoff Process

Stochastic processes of the type described above are known mathematically as discrete Markoff processes and have been extensively studied in the literature.[6] The general case can be described as follows: There exist a finite number of possible "states" of a system; $S_1, S_2, \ldots, S_n$. In addition there is a set of transition probabilities; $p_i(j)$ the probability that if the system is in state S_i it will next go to state S_j. To make this Markoff process into an information source we need only assume that a letter is produced for each transition from one state to another. The states will correspond to the "residue of influence" from preceding letters.

The situation can be represented graphically as shown in Figs. 3, 4 and 5. The "states" are the junction

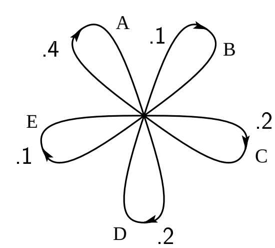


Fig. 3—A graph corresponding to the source in example B.

points in the graph and the probabilities and letters produced for a transition are given beside the corresponding line. Figure 3 is for the example B in Section 2, while Fig. 4 corresponds to the example C. In Fig. 3

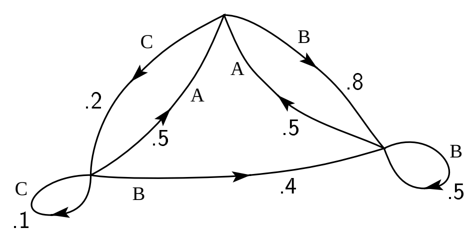


Fig. 4—A graph corresponding to the source in example C.

there is only one state since successive letters are independent. In Fig. 4 there are as many states as letters. If a trigram example were constructed there would be at most n^2 states corresponding to the possible pairs of letters preceding the one being chosen. Figure 5 is a graph for the case of word structure in example D. Here S corresponds to the "space" symbol.

5. Ergodic and Mixed Sources

As we have indicated above a discrete source for our purposes can be considered to be represented by a Markoff process. Among the possible discrete Markoff processes there is a group with special properties of significance in communication theory. This special class consists of the "ergodic" processes and we shall call the corresponding sources ergodic sources. Although a rigorous definition of an ergodic process is somewhat involved, the general idea is simple. In an ergodic process every sequence produced by the process

[6] For a detailed treatment see M. Fréchet, *Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles*. Paris, Gauthier-Villars, 1938.

is the same in statistical properties. Thus the letter frequencies, digram frequencies, etc., obtained from particular sequences, will, as the lengths of the sequences increase, approach definite limits independent of the particular sequence. Actually this is not true of every sequence but the set for which it is false has probability zero. Roughly the ergodic property means statistical homogeneity.

All the examples of artificial languages given above are ergodic. This property is related to the structure of the corresponding graph. If the graph has the following two properties[7] the corresponding process will be ergodic:

1. The graph does not consist of two isolated parts A and B such that it is impossible to go from junction points in part A to junction points in part B along lines of the graph in the direction of arrows and also impossible to go from junctions in part B to junctions in part A.
2. A closed series of lines in the graph with all arrows on the lines pointing in the same orientation will be called a "circuit." The "length" of a circuit is the number of lines in it. Thus in Fig. 5 series BEBES is a circuit of length 5. The second property required is that the greatest common divisor of the lengths of all circuits in the graph be one.

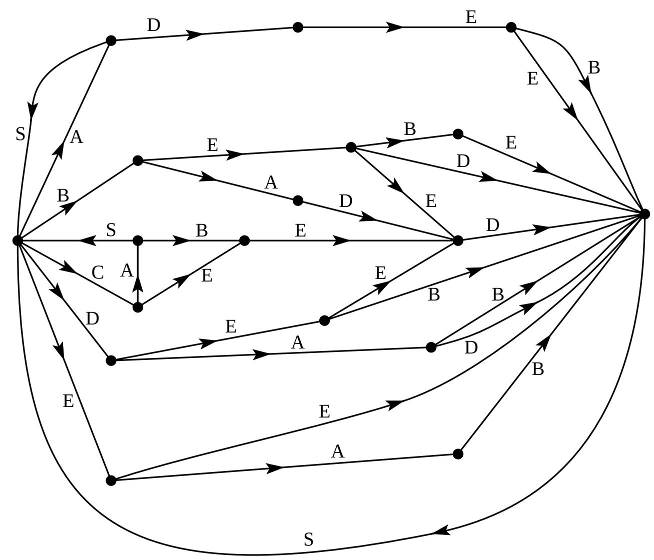


Fig. 5—A graph corresponding to the source in example D.

If the first condition is satisfied but the second one violated by having the greatest common divisor equal to $d > 1$, the sequences have a certain type of periodic structure. The various sequences fall into d different classes which are statistically the same apart from a shift of the origin (i.e., which letter in the sequence is called letter 1). By a shift of from 0 up to $d-1$ any sequence can be made statistically equivalent to any other. A simple example with $d = 2$ is the following: There are three possible letters a, b, c. Letter a is followed with either b or c with probabilities $\frac{1}{3}$ and $\frac{2}{3}$ respectively. Either b or c is always followed by letter a. Thus a typical sequence is

$$a\,b\,a\,c\,a\,c\,a\,c\,a\,b\,a\,c\,a\,b\,a\,b\,a\,c\,a\,c.$$

This type of situation is not of much importance for our work.

If the first condition is violated the graph may be separated into a set of subgraphs each of which satisfies the first condition. We will assume that the second condition is also satisfied for each subgraph. We have in this case what may be called a "mixed" source made up of a number of pure components. The components correspond to the various subgraphs. If $L_1, L_2, L_3, \ldots$ are the component sources we may write

$$L = p_1 L_1 + p_2 L_2 + p_3 L_3 + \cdots$$

[7]These are restatements in terms of the graph of conditions given in Fréchet.

where p_i is the probability of the component source L_i.

Physically the situation represented is this: There are several different sources L_1, L_2, $L_3, \ldots$ which are each of homogeneous statistical structure (i.e., they are ergodic). We do not know *a priori* which is to be used, but once the sequence starts in a given pure component L_i, it continues indefinitely according to the statistical structure of that component.

As an example one may take two of the processes defined above and assume $p_1 = .2$ and $p_2 = .8$. A sequence from the mixed source

$$L = .2L_1 + .8L_2$$

would be obtained by choosing first L_1 or L_2 with probabilities .2 and .8 and after this choice generating a sequence from whichever was chosen.

Except when the contrary is stated we shall assume a source to be ergodic. This assumption enables one to identify averages along a sequence with averages over the ensemble of possible sequences (the probability of a discrepancy being zero). For example the relative frequency of the letter A in a particular infinite sequence will be, with probability one, equal to its relative frequency in the ensemble of sequences.

If P_i is the probability of state i and $p_i(j)$ the transition probability to state j, then for the process to be stationary it is clear that the P_i must satisfy equilibrium conditions:

$$P_j = \sum_i P_i p_i(j).$$

In the ergodic case it can be shown that with any starting conditions the probabilities $P_j(N)$ of being in state j after N symbols, approach the equilibrium values as $N \to \infty$.

6. Choice, Uncertainty and Entropy

We have represented a discrete information source as a Markoff process. Can we define a quantity which will measure, in some sense, how much information is "produced" by such a process, or better, at what rate information is produced?

Suppose we have a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?

If there is such a measure, say $H(p_1, p_2, \ldots, p_n)$, it is reasonable to require of it the following properties:

1. H should be continuous in the p_i.
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n. With equally likely events there is more choice, or uncertainty, when there are more possible events.
3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Fig. 6. At the left we have three

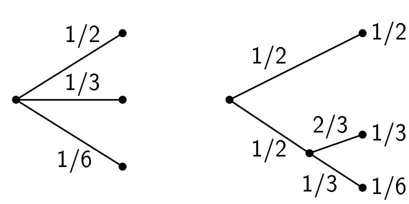


Fig. 6—Decomposition of a choice from three possibilities.

possibilities $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$. On the right we first choose between two possibilities each with probability $\frac{1}{2}$, and if the second occurs make another choice with probabilities $\frac{2}{3}$, $\frac{1}{3}$. The final results have the same probabilities as before. We require, in this special case, that

$$H(\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{6}) = H(\tfrac{1}{2}, \tfrac{1}{2}) + \tfrac{1}{2}H(\tfrac{2}{3}, \tfrac{1}{3}).$$

The coefficient $\frac{1}{2}$ is because this second choice only occurs half the time.

In Appendix 2, the following result is established:

Theorem 2: The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant.

This theorem, and the assumptions required for its proof, are in no way necessary for the present theory. It is given chiefly to lend a certain plausibility to some of our later definitions. The real justification of these definitions, however, will reside in their implications.

Quantities of the form $H = -\sum p_i \log p_i$ (the constant K merely amounts to a choice of a unit of measure) play a central role in information theory as measures of information, choice and uncertainty. The form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics[8] where p_i is the probability of a system being in cell i of its phase space. H is then, for example, the H in Boltzmann's famous H theorem. We shall call $H = -\sum p_i \log p_i$ the entropy of the set of probabilities $p_1, \ldots, p_n$. If x is a chance variable we will write $H(x)$ for its entropy; thus x is not an argument of a function but a label for a number, to differentiate it from $H(y)$ say, the entropy of the chance variable y.

The entropy in the case of two possibilities with probabilities p and $q = 1 - p$, namely

$$H = -(p \log p + q \log q)$$

is plotted in Fig. 7 as a function of p.

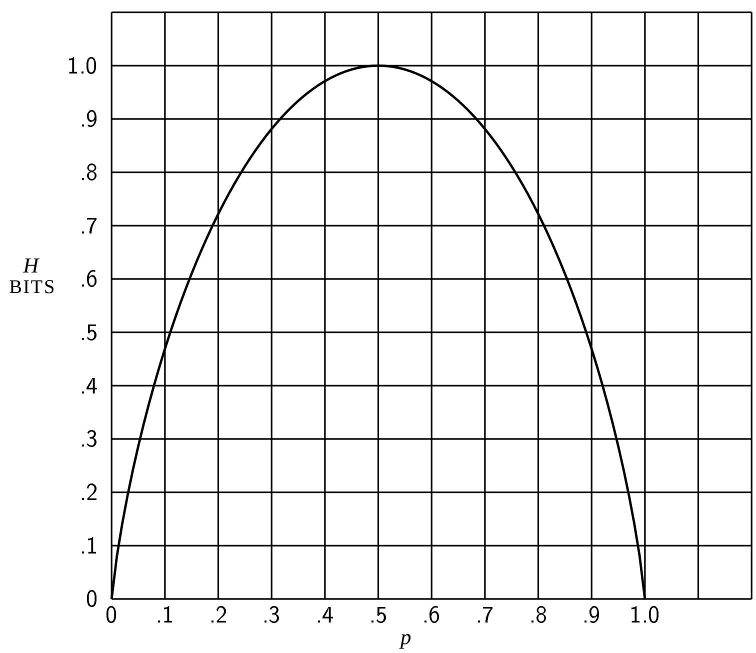


Fig. 7 — Entropy in the case of two possibilities with probabilities p and $(1 - p)$.

The quantity H has a number of interesting properties which further substantiate it as a reasonable measure of choice or information.

1. $H = 0$ if and only if all the p_i but one are zero, this one having the value unity. Thus only when we are certain of the outcome does H vanish. Otherwise H is positive.

2. For a given n, H is a maximum and equal to $\log n$ when all the p_i are equal (i.e., $\frac{1}{n}$). This is also intuitively the most uncertain situation.

[8]See, for example, R. C. Tolman, *Principles of Statistical Mechanics,* Oxford, Clarendon, 1938.

3. Suppose there are two events, x and y, in question with m possibilities for the first and n for the second. Let $p(i,j)$ be the probability of the joint occurrence of i for the first and j for the second. The entropy of the joint event is

$$H(x,y) = -\sum_{i,j} p(i,j) \log p(i,j)$$

while

$$H(x) = -\sum_{i,j} p(i,j) \log \sum_{j} p(i,j)$$
$$H(y) = -\sum_{i,j} p(i,j) \log \sum_{i} p(i,j).$$

It is easily shown that

$$H(x,y) \leq H(x) + H(y)$$

with equality only if the events are independent (i.e., $p(i,j) = p(i)p(j)$). The uncertainty of a joint event is less than or equal to the sum of the individual uncertainties.

4. Any change toward equalization of the probabilities $p_1, p_2, \ldots, p_n$ increases H. Thus if $p_1 < p_2$ and we increase p_1, decreasing p_2 an equal amount so that p_1 and p_2 are more nearly equal, then H increases. More generally, if we perform any "averaging" operation on the p_i of the form

$$p'_i = \sum_j a_{ij} p_j$$

where $\sum_i a_{ij} = \sum_j a_{ij} = 1$, and all $a_{ij} \geq 0$, then H increases (except in the special case where this transformation amounts to no more than a permutation of the p_j with H of course remaining the same).

5. Suppose there are two chance events x and y as in 3, not necessarily independent. For any particular value i that x can assume there is a conditional probability $p_i(j)$ that y has the value j. This is given by

$$p_i(j) = \frac{p(i,j)}{\sum_j p(i,j)}.$$

We define the *conditional entropy* of y, $H_x(y)$ as the average of the entropy of y for each value of x, weighted according to the probability of getting that particular x. That is

$$H_x(y) = -\sum_{i,j} p(i,j) \log p_i(j).$$

This quantity measures how uncertain we are of y on the average when we know x. Substituting the value of $p_i(j)$ we obtain

$$\begin{aligned} H_x(y) &= -\sum_{i,j} p(i,j) \log p(i,j) + \sum_{i,j} p(i,j) \log \sum_j p(i,j) \\ &= H(x,y) - H(x) \end{aligned}$$

or

$$H(x,y) = H(x) + H_x(y).$$

The uncertainty (or entropy) of the joint event x,y is the uncertainty of x plus the uncertainty of y when x is known.

6. From 3 and 5 we have

$$H(x) + H(y) \geq H(x,y) = H(x) + H_x(y).$$

Hence

$$H(y) \geq H_x(y).$$

The uncertainty of y is never increased by knowledge of x. It will be decreased unless x and y are independent events, in which case it is not changed.

7. The Entropy of an Information Source

Consider a discrete source of the finite state type considered above. For each possible state i there will be a set of probabilities $p_i(j)$ of producing the various possible symbols j. Thus there is an entropy H_i for each state. The entropy of the source will be defined as the average of these H_i weighted in accordance with the probability of occurrence of the states in question:

$$\begin{aligned} H &= \sum_i P_i H_i \\ &= -\sum_{i,j} P_i p_i(j) \log p_i(j)\,. \end{aligned}$$

This is the entropy of the source per symbol of text. If the Markoff process is proceeding at a definite time rate there is also an entropy per second

$$H' = \sum_i f_i H_i$$

where f_i is the average frequency (occurrences per second) of state i. Clearly

$$H' = mH$$

where m is the average number of symbols produced per second. H or H' measures the amount of information generated by the source per symbol or per second. If the logarithmic base is 2, they will represent bits per symbol or per second.

If successive symbols are independent then H is simply $-\sum p_i \log p_i$ where p_i is the probability of symbol i. Suppose in this case we consider a long message of N symbols. It will contain with high probability about $p_1 N$ occurrences of the first symbol, $p_2 N$ occurrences of the second, etc. Hence the probability of this particular message will be roughly

$$p = p_1^{p_1 N} p_2^{p_2 N} \cdots p_n^{p_n N}$$

or

$$\begin{aligned} \log p &\doteq N \sum_i p_i \log p_i \\ \log p &\doteq -NH \\ H &\doteq \frac{\log 1/p}{N}. \end{aligned}$$

H is thus approximately the logarithm of the reciprocal probability of a typical long sequence divided by the number of symbols in the sequence. The same result holds for any source. Stated more precisely we have (see Appendix 3):

Theorem 3: Given any $\epsilon > 0$ and $\delta > 0$, we can find an N_0 such that the sequences of any length $N \geq N_0$ fall into two classes:

1. *A set whose total probability is less than ϵ.*
2. *The remainder, all of whose members have probabilities satisfying the inequality*

$$\left| \frac{\log p^{-1}}{N} - H \right| < \delta.$$

In other words we are almost certain to have $\frac{\log p^{-1}}{N}$ very close to H when N is large.

A closely related result deals with the number of sequences of various probabilities. Consider again the sequences of length N and let them be arranged in order of decreasing probability. We define $n(q)$ to be the number we must take from this set starting with the most probable one in order to accumulate a total probability q for those taken.

Theorem 4:

$$\lim_{N\to\infty} \frac{\log n(q)}{N} = H$$

when q does not equal 0 or 1.

We may interpret $\log n(q)$ as the number of bits required to specify the sequence when we consider only the most probable sequences with a total probability q. Then $\frac{\log n(q)}{N}$ is the number of bits per symbol for the specification. The theorem says that for large N this will be independent of q and equal to H. The rate of growth of the logarithm of the number of reasonably probable sequences is given by H, regardless of our interpretation of "reasonably probable." Due to these results, which are proved in Appendix 3, it is possible for most purposes to treat the long sequences as though there were just 2^{HN} of them, each with a probability 2^{-HN}.

The next two theorems show that H and H' can be determined by limiting operations directly from the statistics of the message sequences, without reference to the states and transition probabilities between states.

Theorem 5: Let $p(B_i)$ be the probability of a sequence B_i of symbols from the source. Let

$$G_N = -\frac{1}{N}\sum_i p(B_i)\log p(B_i)$$

where the sum is over all sequences B_i containing N symbols. Then G_N is a monotonic decreasing function of N and

$$\lim_{N\to\infty} G_N = H.$$

Theorem 6: Let $p(B_i,S_j)$ be the probability of sequence B_i followed by symbol S_j and $p_{B_i}(S_j) = p(B_i,S_j)/p(B_i)$ be the conditional probability of S_j after B_i. Let

$$F_N = -\sum_{i,j} p(B_i,S_j)\log p_{B_i}(S_j)$$

where the sum is over all blocks B_i of $N-1$ symbols and over all symbols S_j. Then F_N is a monotonic decreasing function of N,

$$F_N = NG_N - (N-1)G_{N-1},$$

$$G_N = \frac{1}{N}\sum_{n=1}^{N} F_n,$$

$$F_N \le G_N,$$

and $\lim_{N\to\infty} F_N = H$.

These results are derived in Appendix 3. They show that a series of approximations to H can be obtained by considering only the statistical structure of the sequences extending over $1,2,\ldots,N$ symbols. F_N is the better approximation. In fact F_N is the entropy of the N^{th} order approximation to the source of the type discussed above. If there are no statistical influences extending over more than N symbols, that is if the conditional probability of the next symbol knowing the preceding $(N-1)$ is not changed by a knowledge of any before that, then $F_N = H$. F_N of course is the conditional entropy of the next symbol when the $(N-1)$ preceding ones are known, while G_N is the entropy per symbol of blocks of N symbols.

The ratio of the entropy of a source to the maximum value it could have while still restricted to the same symbols will be called its *relative entropy*. This is the maximum compression possible when we encode into the same alphabet. One minus the relative entropy is the *redundancy*. The redundancy of ordinary English, not considering statistical structure over greater distances than about eight letters, is roughly 50%. This means that when we write English half of what we write is determined by the structure of the language and half is chosen freely. The figure 50% was found by several independent methods which all gave results in

this neighborhood. One is by calculation of the entropy of the approximations to English. A second method is to delete a certain fraction of the letters from a sample of English text and then let someone attempt to restore them. If they can be restored when 50% are deleted the redundancy must be greater than 50%. A third method depends on certain known results in cryptography.

Two extremes of redundancy in English prose are represented by Basic English and by James Joyce's book "Finnegans Wake". The Basic English vocabulary is limited to 850 words and the redundancy is very high. This is reflected in the expansion that occurs when a passage is translated into Basic English. Joyce on the other hand enlarges the vocabulary and is alleged to achieve a compression of semantic content.

The redundancy of a language is related to the existence of crossword puzzles. If the redundancy is zero any sequence of letters is a reasonable text in the language and any two-dimensional array of letters forms a crossword puzzle. If the redundancy is too high the language imposes too many constraints for large crossword puzzles to be possible. A more detailed analysis shows that if we assume the constraints imposed by the language are of a rather chaotic and random nature, large crossword puzzles are just possible when the redundancy is 50%. If the redundancy is 33%, three-dimensional crossword puzzles should be possible, etc.

8. Representation of the Encoding and Decoding Operations

We have yet to represent mathematically the operations performed by the transmitter and receiver in encoding and decoding the information. Either of these will be called a discrete transducer. The input to the transducer is a sequence of input symbols and its output a sequence of output symbols. The transducer may have an internal memory so that its output depends not only on the present input symbol but also on the past history. We assume that the internal memory is finite, i.e., there exist a finite number m of possible states of the transducer and that its output is a function of the present state and the present input symbol. The next state will be a second function of these two quantities. Thus a transducer can be described by two functions:

$$\begin{aligned} y_n &= f(x_n, \alpha_n) \\ \alpha_{n+1} &= g(x_n, \alpha_n) \end{aligned}$$

where

x_n is the n^{th} input symbol,

α_n is the state of the transducer when the n^{th} input symbol is introduced,

y_n is the output symbol (or sequence of output symbols) produced when x_n is introduced if the state is α_n.

If the output symbols of one transducer can be identified with the input symbols of a second, they can be connected in tandem and the result is also a transducer. If there exists a second transducer which operates on the output of the first and recovers the original input, the first transducer will be called non-singular and the second will be called its inverse.

Theorem 7: The output of a finite state transducer driven by a finite state statistical source is a finite state statistical source, with entropy (per unit time) less than or equal to that of the input. If the transducer is non-singular they are equal.

Let α represent the state of the source, which produces a sequence of symbols x_i; and let β be the state of the transducer, which produces, in its output, blocks of symbols y_j. The combined system can be represented by the "product state space" of pairs (α, β). Two points in the space (α_1, β_1) and (α_2, β_2), are connected by a line if α_1 can produce an x which changes β_1 to β_2, and this line is given the probability of that x in this case. The line is labeled with the block of y_j symbols produced by the transducer. The entropy of the output can be calculated as the weighted sum over the states. If we sum first on β each resulting term is less than or equal to the corresponding term for α, hence the entropy is not increased. If the transducer is non-singular let its output be connected to the inverse transducer. If H'_1, H'_2 and H'_3 are the output entropies of the source, the first and second transducers respectively, then $H'_1 \geq H'_2 \geq H'_3 = H'_1$ and therefore $H'_1 = H'_2$.

Suppose we have a system of constraints on possible sequences of the type which can be represented by a linear graph as in Fig. 2. If probabilities $p_{ij}^{(s)}$ were assigned to the various lines connecting state i to state j this would become a source. There is one particular assignment which maximizes the resulting entropy (see Appendix 4).

Theorem 8: Let the system of constraints considered as a channel have a capacity $C = \log W$. If we assign

$$p_{ij}^{(s)} = \frac{B_j}{B_i} W^{-\ell_{ij}^{(s)}}$$

where $\ell_{ij}^{(s)}$ is the duration of the s^{th} symbol leading from state i to state j and the B_i satisfy

$$B_i = \sum_{s,j} B_j W^{-\ell_{ij}^{(s)}}$$

then H is maximized and equal to C.

By proper assignment of the transition probabilities the entropy of symbols on a channel can be maximized at the channel capacity.

9. The Fundamental Theorem for a Noiseless Channel

We will now justify our interpretation of H as the rate of generating information by proving that H determines the channel capacity required with most efficient coding.

Theorem 9: Let a source have entropy H (bits per symbol) and a channel have a capacity C (bits per second). Then it is possible to encode the output of the source in such a way as to transmit at the average rate $\frac{C}{H} - \epsilon$ symbols per second over the channel where ϵ is arbitrarily small. It is not possible to transmit at an average rate greater than $\frac{C}{H}$.

The converse part of the theorem, that $\frac{C}{H}$ cannot be exceeded, may be proved by noting that the entropy of the channel input per second is equal to that of the source, since the transmitter must be non-singular, and also this entropy cannot exceed the channel capacity. Hence $H' \le C$ and the number of symbols per second $= H'/H \le C/H$.

The first part of the theorem will be proved in two different ways. The first method is to consider the set of all sequences of N symbols produced by the source. For N large we can divide these into two groups, one containing less than $2^{(H+\eta)N}$ members and the second containing less than 2^{RN} members (where R is the logarithm of the number of different symbols) and having a total probability less than μ. As N increases η and μ approach zero. The number of signals of duration T in the channel is greater than $2^{(C-\theta)T}$ with θ small when T is large. if we choose

$$T = \left(\frac{H}{C} + \lambda\right) N$$

then there will be a sufficient number of sequences of channel symbols for the high probability group when N and T are sufficiently large (however small λ) and also some additional ones. The high probability group is coded in an arbitrary one-to-one way into this set. The remaining sequences are represented by larger sequences, starting and ending with one of the sequences not used for the high probability group. This special sequence acts as a start and stop signal for a different code. In between a sufficient time is allowed to give enough different sequences for all the low probability messages. This will require

$$T_1 = \left(\frac{R}{C} + \varphi\right) N$$

where φ is small. The mean rate of transmission in message symbols per second will then be greater than

$$\left[(1-\delta)\frac{T}{N} + \delta\frac{T_1}{N}\right]^{-1} = \left[(1-\delta)\left(\frac{H}{C} + \lambda\right) + \delta\left(\frac{R}{C} + \varphi\right)\right]^{-1}.$$

As N increases δ, λ and φ approach zero and the rate approaches $\dfrac{C}{H}$.

Another method of performing this coding and thereby proving the theorem can be described as follows: Arrange the messages of length N in order of decreasing probability and suppose their probabilities are $p_1 \geq p_2 \geq p_3 \cdots \geq p_n$. Let $P_s = \sum_1^{s-1} p_i$; that is P_s is the cumulative probability up to, but not including, p_s. We first encode into a binary system. The binary code for message s is obtained by expanding P_s as a binary number. The expansion is carried out to m_s places, where m_s is the integer satisfying:

$$\log_2 \frac{1}{p_s} \leq m_s < 1 + \log_2 \frac{1}{p_s}.$$

Thus the messages of high probability are represented by short codes and those of low probability by long codes. From these inequalities we have

$$\frac{1}{2^{m_s}} \leq p_s < \frac{1}{2^{m_s - 1}}.$$

The code for P_s will differ from all succeeding ones in one or more of its m_s places, since all the remaining P_i are at least $\frac{1}{2^{m_s}}$ larger and their binary expansions therefore differ in the first m_s places. Consequently all the codes are different and it is possible to recover the message from its code. If the channel sequences are not already sequences of binary digits, they can be ascribed binary numbers in an arbitrary fashion and the binary code thus translated into signals suitable for the channel.

The average number H' of binary digits used per symbol of original message is easily estimated. We have

$$H' = \frac{1}{N} \sum m_s p_s.$$

But,

$$\frac{1}{N} \sum \Big(\log_2 \frac{1}{p_s}\Big) p_s \leq \frac{1}{N} \sum m_s p_s < \frac{1}{N} \sum \Big(1 + \log_2 \frac{1}{p_s}\Big) p_s$$

and therefore,

$$G_N \leq H' < G_N + \frac{1}{N}$$

As N increases G_N approaches H, the entropy of the source and H' approaches H.

We see from this that the inefficiency in coding, when only a finite delay of N symbols is used, need not be greater than $\frac{1}{N}$ plus the difference between the true entropy H and the entropy G_N calculated for sequences of length N. The per cent excess time needed over the ideal is therefore less than

$$\frac{G_N}{H} + \frac{1}{HN} - 1.$$

This method of encoding is substantially the same as one found independently by R. M. Fano.[9] His method is to arrange the messages of length N in order of decreasing probability. Divide this series into two groups of as nearly equal probability as possible. If the message is in the first group its first binary digit will be 0, otherwise 1. The groups are similarly divided into subsets of nearly equal probability and the particular subset determines the second binary digit. This process is continued until each subset contains only one message. It is easily seen that apart from minor differences (generally in the last digit) this amounts to the same thing as the arithmetic process described above.

10. Discussion and Examples

In order to obtain the maximum power transfer from a generator to a load, a transformer must in general be introduced so that the generator as seen from the load has the load resistance. The situation here is roughly analogous. The transducer which does the encoding should match the source to the channel in a statistical sense. The source as seen from the channel through the transducer should have the same statistical structure

[9] Technical Report No. 65, The Research Laboratory of Electronics, M.I.T., March 17, 1949.

as the source which maximizes the entropy in the channel. The content of Theorem 9 is that, although an exact match is not in general possible, we can approximate it as closely as desired. The ratio of the actual rate of transmission to the capacity C may be called the efficiency of the coding system. This is of course equal to the ratio of the actual entropy of the channel symbols to the maximum possible entropy.

In general, ideal or nearly ideal encoding requires a long delay in the transmitter and receiver. In the noiseless case which we have been considering, the main function of this delay is to allow reasonably good matching of probabilities to corresponding lengths of sequences. With a good code the logarithm of the reciprocal probability of a long message must be proportional to the duration of the corresponding signal, in fact

$$\left|\frac{\log p^{-1}}{T} - C\right|$$

must be small for all but a small fraction of the long messages.

If a source can produce only one particular message its entropy is zero, and no channel is required. For example, a computing machine set up to calculate the successive digits of π produces a definite sequence with no chance element. No channel is required to "transmit" this to another point. One could construct a second machine to compute the same sequence at the point. However, this may be impractical. In such a case we can choose to ignore some or all of the statistical knowledge we have of the source. We might consider the digits of π to be a random sequence in that we construct a system capable of sending any sequence of digits. In a similar way we may choose to use some of our statistical knowledge of English in constructing a code, but not all of it. In such a case we consider the source with the maximum entropy subject to the statistical conditions we wish to retain. The entropy of this source determines the channel capacity which is necessary and sufficient. In the π example the only information retained is that all the digits are chosen from the set $0,1,\ldots,9$. In the case of English one might wish to use the statistical saving possible due to letter frequencies, but nothing else. The maximum entropy source is then the first approximation to English and its entropy determines the required channel capacity.

As a simple example of some of these results consider a source which produces a sequence of letters chosen from among A, B, C, D with probabilities $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{8}$, successive symbols being chosen independently. We have

$$\begin{aligned} H &= -\left(\tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{4}\log\tfrac{1}{4} + \tfrac{2}{8}\log\tfrac{1}{8}\right) \\ &= \tfrac{7}{4} \text{ bits per symbol.} \end{aligned}$$

Thus we can approximate a coding system to encode messages from this source into binary digits with an average of $\frac{7}{4}$ binary digit per symbol. In this case we can actually achieve the limiting value by the following code (obtained by the method of the second proof of Theorem 9):

A	0
B	10
C	110
D	111

The average number of binary digits used in encoding a sequence of N symbols will be

$$N\left(\tfrac{1}{2}\times 1 + \tfrac{1}{4}\times 2 + \frac{2}{8}\times 3\right) = \tfrac{7}{4}N.$$

It is easily seen that the binary digits 0, 1 have probabilities $\frac{1}{2}$, $\frac{1}{2}$ so the H for the coded sequences is one bit per symbol. Since, on the average, we have $\frac{7}{4}$ binary symbols per original letter, the entropies on a time basis are the same. The maximum possible entropy for the original set is $\log 4 = 2$, occurring when A, B, C, D have probabilities $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$. Hence the relative entropy is $\frac{7}{8}$. We can translate the binary sequences into the original set of symbols on a two-to-one basis by the following table:

00	A'
01	B'
10	C'
11	D'

This double process then encodes the original message into the same symbols but with an average compression ratio $\frac{7}{8}$.

As a second example consider a source which produces a sequence of A's and B's with probability p for A and q for B. If $p \ll q$ we have

$$\begin{aligned} H &= -\log p^p (1-p)^{1-p} \\ &= -p \log p (1-p)^{(1-p)/p} \\ &\doteq p \log \frac{e}{p}. \end{aligned}$$

In such a case one can construct a fairly good coding of the message on a 0, 1 channel by sending a special sequence, say 0000, for the infrequent symbol A and then a sequence indicating the *number* of B's following it. This could be indicated by the binary representation with all numbers containing the special sequence deleted. All numbers up to 16 are represented as usual; 16 is represented by the next binary number after 16 which does not contain four zeros, namely $17 = 10001$, etc.

It can be shown that as $p \to 0$ the coding approaches ideal provided the length of the special sequence is properly adjusted.

PART II: THE DISCRETE CHANNEL WITH NOISE

11. Representation of a Noisy Discrete Channel

We now consider the case where the signal is perturbed by noise during transmission or at one or the other of the terminals. This means that the received signal is not necessarily the same as that sent out by the transmitter. Two cases may be distinguished. If a particular transmitted signal always produces the same received signal, i.e., the received signal is a definite function of the transmitted signal, then the effect may be called distortion. If this function has an inverse — no two transmitted signals producing the same received signal — distortion may be corrected, at least in principle, by merely performing the inverse functional operation on the received signal.

The case of interest here is that in which the signal does not always undergo the same change in transmission. In this case we may assume the received signal E to be a function of the transmitted signal S and a second variable, the noise N.

$$E = f(S,N)$$

The noise is considered to be a chance variable just as the message was above. In general it may be represented by a suitable stochastic process. The most general type of noisy discrete channel we shall consider is a generalization of the finite state noise-free channel described previously. We assume a finite number of states and a set of probabilities

$$p_{\alpha,i}(\beta, j).$$

This is the probability, if the channel is in state α and symbol i is transmitted, that symbol j will be received and the channel left in state β. Thus α and β range over the possible states, i over the possible transmitted signals and j over the possible received signals. In the case where successive symbols are independently perturbed by the noise there is only one state, and the channel is described by the set of transition probabilities $p_i(j)$, the probability of transmitted symbol i being received as j.

If a noisy channel is fed by a source there are two statistical processes at work: the source and the noise. Thus there are a number of entropies that can be calculated. First there is the entropy $H(x)$ of the source or of the input to the channel (these will be equal if the transmitter is non-singular). The entropy of the output of the channel, i.e., the received signal, will be denoted by $H(y)$. In the noiseless case $H(y) = H(x)$. The joint entropy of input and output will be $H(xy)$. Finally there are two conditional entropies $H_x(y)$ and $H_y(x)$, the entropy of the output when the input is known and conversely. Among these quantities we have the relations

$$H(x,y) = H(x) + H_x(y) = H(y) + H_y(x).$$

All of these entropies can be measured on a per-second or a per-symbol basis.

12. EQUIVOCATION AND CHANNEL CAPACITY

If the channel is noisy it is not in general possible to reconstruct the original message or the transmitted signal with *certainty* by any operation on the received signal E. There are, however, ways of transmitting the information which are optimal in combating noise. This is the problem which we now consider.

Suppose there are two possible symbols 0 and 1, and we are transmitting at a rate of 1000 symbols per second with probabilities $p_0 = p_1 = \frac{1}{2}$. Thus our source is producing information at the rate of 1000 bits per second. During transmission the noise introduces errors so that, on the average, 1 in 100 is received incorrectly (a 0 as 1, or 1 as 0). What is the rate of transmission of information? Certainly less than 1000 bits per second since about 1% of the received symbols are incorrect. Our first impulse might be to say the rate is 990 bits per second, merely subtracting the expected number of errors. This is not satisfactory since it fails to take into account the recipient's lack of knowledge of where the errors occur. We may carry it to an extreme case and suppose the noise so great that the received symbols are entirely independent of the transmitted symbols. The probability of receiving 1 is $\frac{1}{2}$ whatever was transmitted and similarly for 0. Then about half of the received symbols are correct due to chance alone, and we would be giving the system credit for transmitting 500 bits per second while actually no information is being transmitted at all. Equally "good" transmission would be obtained by dispensing with the channel entirely and flipping a coin at the receiving point.

Evidently the proper correction to apply to the amount of information transmitted is the amount of this information which is missing in the received signal, or alternatively the uncertainty when we have received a signal of what was actually sent. From our previous discussion of entropy as a measure of uncertainty it seems reasonable to use the conditional entropy of the message, knowing the received signal, as a measure of this missing information. This is indeed the proper definition, as we shall see later. Following this idea the rate of actual transmission, R, would be obtained by subtracting from the rate of production (i.e., the entropy of the source) the average rate of conditional entropy.

$$R = H(x) - H_y(x)$$

The conditional entropy $H_y(x)$ will, for convenience, be called the equivocation. It measures the average ambiguity of the received signal.

In the example considered above, if a 0 is received the *a posteriori* probability that a 0 was transmitted is .99, and that a 1 was transmitted is .01. These figures are reversed if a 1 is received. Hence

$$\begin{aligned} H_y(x) &= -[.99 \log .99 + 0.01 \log 0.01] \\ &= .081 \text{ bits/symbol} \end{aligned}$$

or 81 bits per second. We may say that the system is transmitting at a rate $1000 - 81 = 919$ bits per second. In the extreme case where a 0 is equally likely to be received as a 0 or 1 and similarly for 1, the *a posteriori* probabilities are $\frac{1}{2}$, $\frac{1}{2}$ and

$$\begin{aligned} H_y(x) &= -\left[\tfrac{1}{2} \log \tfrac{1}{2} + \tfrac{1}{2} \log \tfrac{1}{2}\right] \\ &= 1 \text{ bit per symbol} \end{aligned}$$

or 1000 bits per second. The rate of transmission is then 0 as it should be.

The following theorem gives a direct intuitive interpretation of the equivocation and also serves to justify it as the unique appropriate measure. We consider a communication system and an observer (or auxiliary device) who can see both what is sent and what is recovered (with errors due to noise). This observer notes the errors in the recovered message and transmits data to the receiving point over a "correction channel" to enable the receiver to correct the errors. The situation is indicated schematically in Fig. 8.

Theorem 10: If the correction channel has a capacity equal to $H_y(x)$ it is possible to so encode the correction data as to send it over this channel and correct all but an arbitrarily small fraction ϵ of the errors. This is not possible if the channel capacity is less than $H_y(x)$.

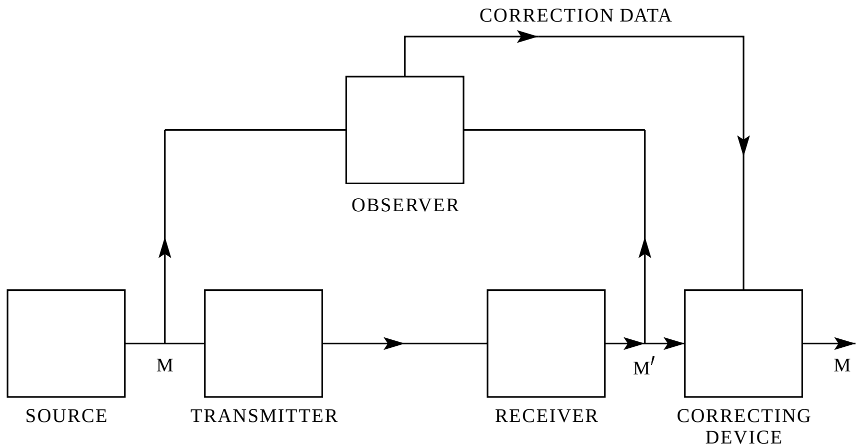


Fig. 8—Schematic diagram of a correction system.

Roughly then, $H_y(x)$ is the amount of additional information that must be supplied per second at the receiving point to correct the received message.

To prove the first part, consider long sequences of received message M' and corresponding original message M. There will be logarithmically $TH_y(x)$ of the M's which could reasonably have produced each M'. Thus we have $TH_y(x)$ binary digits to send each T seconds. This can be done with ϵ frequency of errors on a channel of capacity $H_y(x)$.

The second part can be proved by noting, first, that for any discrete chance variables x, y, z

$$H_y(x,z) \geq H_y(x).$$

The left-hand side can be expanded to give

$$H_y(z) + H_{yz}(x) \geq H_y(x)$$
$$H_{yz}(x) \geq H_y(x) - H_y(z) \geq H_y(x) - H(z).$$

If we identify x as the output of the source, y as the received signal and z as the signal sent over the correction channel, then the right-hand side is the equivocation less the rate of transmission over the correction channel. If the capacity of this channel is less than the equivocation the right-hand side will be greater than zero and $H_{yz}(x) > 0$. But this is the uncertainty of what was sent, knowing both the received signal and the correction signal. If this is greater than zero the frequency of errors cannot be arbitrarily small.

Example:

> Suppose the errors occur at random in a sequence of binary digits: probability p that a digit is wrong and $q = 1 - p$ that it is right. These errors can be corrected if their position is known. Thus the correction channel need only send information as to these positions. This amounts to transmitting from a source which produces binary digits with probability p for 1 (incorrect) and q for 0 (correct). This requires a channel of capacity
>
> $$-[p \log p + q \log q]$$
>
> which is the equivocation of the original system.

The rate of transmission R can be written in two other forms due to the identities noted above. We have

$$\begin{aligned} R &= H(x) - H_y(x) \\ &= H(y) - H_x(y) \\ &= H(x) + H(y) - H(x,y). \end{aligned}$$

The first defining expression has already been interpreted as the amount of information sent less the uncertainty of what was sent. The second measures the amount received less the part of this which is due to noise. The third is the sum of the two amounts less the joint entropy and therefore in a sense is the number of bits per second common to the two. Thus all three expressions have a certain intuitive significance.

The capacity C of a noisy channel should be the maximum possible rate of transmission, i.e., the rate when the source is properly matched to the channel. We therefore define the channel capacity by

$$C = \mathrm{Max}\big(H(x) - H_y(x)\big)$$

where the maximum is with respect to all possible information sources used as input to the channel. If the channel is noiseless, $H_y(x) = 0$. The definition is then equivalent to that already given for a noiseless channel since the maximum entropy for the channel is its capacity.

13. The Fundamental Theorem for a Discrete Channel with Noise

It may seem surprising that we should define a definite capacity C for a noisy channel since we can never send certain information in such a case. It is clear, however, that by sending the information in a redundant form the probability of errors can be reduced. For example, by repeating the message many times and by a statistical study of the different received versions of the message the probability of errors could be made very small. One would expect, however, that to make this probability of errors approach zero, the redundancy of the encoding must increase indefinitely, and the rate of transmission therefore approach zero. This is by no means true. If it were, there would not be a very well defined capacity, but only a capacity for a given frequency of errors, or a given equivocation; the capacity going down as the error requirements are made more stringent. Actually the capacity C defined above has a very definite significance. It is possible to send information at the rate C through the channel *with as small a frequency of errors or equivocation as desired* by proper encoding. This statement is not true for any rate greater than C. If an attempt is made to transmit at a higher rate than C, say $C + R_1$, then there will necessarily be an equivocation equal to or greater than the excess R_1. Nature takes payment by requiring just that much uncertainty, so that we are not actually getting any more than C through correctly.

The situation is indicated in Fig. 9. The rate of information into the channel is plotted horizontally and the equivocation vertically. Any point above the heavy line in the shaded region can be attained and those below cannot. The points on the line cannot in general be attained, but there will usually be two points on the line that can.

These results are the main justification for the definition of C and will now be proved.

Theorem 11: Let a discrete channel have the capacity C and a discrete source the entropy per second H. If $H \leq C$ there exists a coding system such that the output of the source can be transmitted over the channel with an arbitrarily small frequency of errors (or an arbitrarily small equivocation). If $H > C$ it is possible to encode the source so that the equivocation is less than $H - C + \epsilon$ where ϵ is arbitrarily small. There is no method of encoding which gives an equivocation less than $H - C$.

The method of proving the first part of this theorem is not by exhibiting a coding method having the desired properties, but by showing that such a code must exist in a certain group of codes. In fact we will

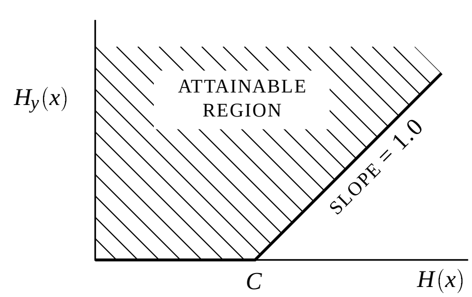


Fig. 9—The equivocation possible for a given input entropy to a channel.

average the frequency of errors over this group and show that this average can be made less than ϵ. If the average of a set of numbers is less than ϵ there must exist at least one in the set which is less than ϵ. This will establish the desired result.

The capacity C of a noisy channel has been defined as

$$C = \text{Max}\big(H(x) - H_y(x)\big)$$

where x is the input and y the output. The maximization is over all sources which might be used as input to the channel.

Let S_0 be a source which achieves the maximum capacity C. If this maximum is not actually achieved by any source let S_0 be a source which approximates to giving the maximum rate. Suppose S_0 is used as input to the channel. We consider the possible transmitted and received sequences of a long duration T. The following will be true:

1. The transmitted sequences fall into two classes, a high probability group with about $2^{TH(x)}$ members and the remaining sequences of small total probability.

2. Similarly the received sequences have a high probability set of about $2^{TH(y)}$ members and a low probability set of remaining sequences.

3. Each high probability output could be produced by about $2^{TH_y(x)}$ inputs. The probability of all other cases has a small total probability.

All the ϵ's and δ's implied by the words "small" and "about" in these statements approach zero as we allow T to increase and S_0 to approach the maximizing source.

The situation is summarized in Fig. 10 where the input sequences are points on the left and output sequences points on the right. The fan of cross lines represents the range of possible causes for a typical output.

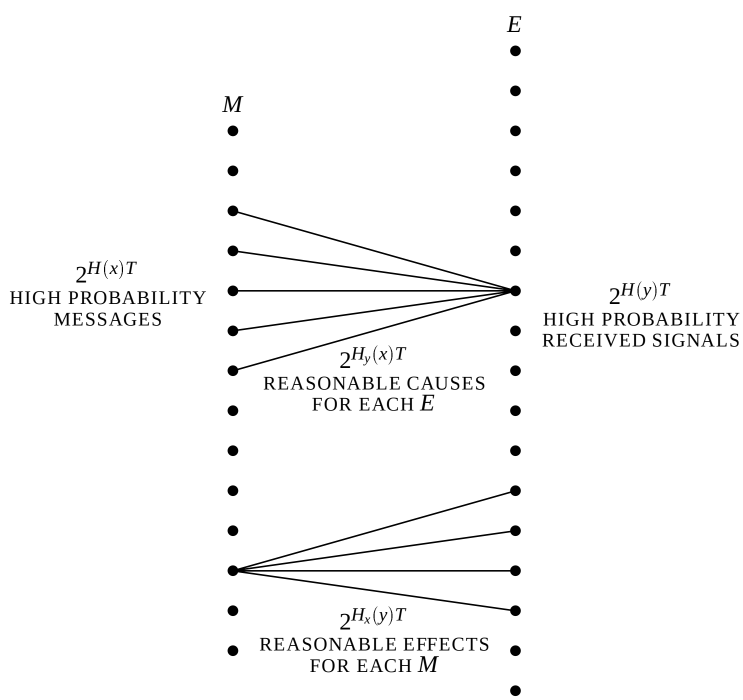


Fig. 10—Schematic representation of the relations between inputs and outputs in a channel.

Now suppose we have another source producing information at rate R with $R < C$. In the period T this source will have 2^{TR} high probability messages. We wish to associate these with a selection of the possible channel inputs in such a way as to get a small frequency of errors. We will set up this association in all

possible ways (using, however, only the high probability group of inputs as determined by the source S_0) and average the frequency of errors for this large class of possible coding systems. This is the same as calculating the frequency of errors for a random association of the messages and channel inputs of duration T. Suppose a particular output y_1 is observed. What is the probability of more than one message in the set of possible causes of y_1? There are 2^{TR} messages distributed at random in $2^{TH(x)}$ points. The probability of a particular point being a message is thus

$$2^{T(R-H(x))}.$$

The probability that none of the points in the fan is a message (apart from the actual originating message) is

$$P = \left[1 - 2^{T(R-H(x))}\right]^{2^{TH_y(x)}}.$$

Now $R < H(x) - H_y(x)$ so $R - H(x) = -H_y(x) - \eta$ with η positive. Consequently

$$P = \left[1 - 2^{-TH_y(x)-T\eta}\right]^{2^{TH_y(x)}}$$

approaches (as $T \to \infty$)

$$1 - 2^{-T\eta}.$$

Hence the probability of an error approaches zero and the first part of the theorem is proved.

The second part of the theorem is easily shown by noting that we could merely send C bits per second from the source, completely neglecting the remainder of the information generated. At the receiver the neglected part gives an equivocation $H(x) - C$ and the part transmitted need only add ϵ. This limit can also be attained in many other ways, as will be shown when we consider the continuous case.

The last statement of the theorem is a simple consequence of our definition of C. Suppose we can encode a source with $H(x) = C + a$ in such a way as to obtain an equivocation $H_y(x) = a - \epsilon$ with ϵ positive. Then $R = H(x) = C + a$ and

$$H(x) - H_y(x) = C + \epsilon$$

with ϵ positive. This contradicts the definition of C as the maximum of $H(x) - H_y(x)$.

Actually more has been proved than was stated in the theorem. If the average of a set of numbers is within ϵ of of their maximum, a fraction of at most $\sqrt{\epsilon}$ can be more than $\sqrt{\epsilon}$ below the maximum. Since ϵ is arbitrarily small we can say that almost all the systems are arbitrarily close to the ideal.

14. Discussion

The demonstration of Theorem 11, while not a pure existence proof, has some of the deficiencies of such proofs. An attempt to obtain a good approximation to ideal coding by following the method of the proof is generally impractical. In fact, apart from some rather trivial cases and certain limiting situations, no explicit description of a series of approximation to the ideal has been found. Probably this is no accident but is related to the difficulty of giving an explicit construction for a good approximation to a random sequence.

An approximation to the ideal would have the property that if the signal is altered in a reasonable way by the noise, the original can still be recovered. In other words the alteration will not in general bring it closer to another reasonable signal than the original. This is accomplished at the cost of a certain amount of redundancy in the coding. The redundancy must be introduced in the proper way to combat the particular noise structure involved. However, any redundancy in the source will usually help if it is utilized at the receiving point. In particular, if the source already has a certain redundancy and no attempt is made to eliminate it in matching to the channel, this redundancy will help combat noise. For example, in a noiseless telegraph channel one could save about 50% in time by proper encoding of the messages. This is not done and most of the redundancy of English remains in the channel symbols. This has the advantage, however, of allowing considerable noise in the channel. A sizable fraction of the letters can be received incorrectly and still reconstructed by the context. In fact this is probably not a bad approximation to the ideal in many cases, since the statistical structure of English is rather involved and the reasonable English sequences are not too far (in the sense required for the theorem) from a random selection.

As in the noiseless case a delay is generally required to approach the ideal encoding. It now has the additional function of allowing a large sample of noise to affect the signal before any judgment is made at the receiving point as to the original message. Increasing the sample size always sharpens the possible statistical assertions.

The content of Theorem 11 and its proof can be formulated in a somewhat different way which exhibits the connection with the noiseless case more clearly. Consider the possible signals of duration T and suppose a subset of them is selected to be used. Let those in the subset all be used with equal probability, and suppose the receiver is constructed to select, as the original signal, the most probable cause from the subset, when a perturbed signal is received. We define $N(T,q)$ to be the maximum number of signals we can choose for the subset such that the probability of an incorrect interpretation is less than or equal to q.

Theorem 12: $\displaystyle\lim_{T\to\infty} \frac{\log N(T,q)}{T} = C$, *where C is the channel capacity, provided that q does not equal 0 or 1.*

In other words, no matter how we set out limits of reliability, we can distinguish reliably in time T enough messages to correspond to about CT bits, when T is sufficiently large. Theorem 12 can be compared with the definition of the capacity of a noiseless channel given in Section 1.

15. Example of a Discrete Channel and its Capacity

A simple example of a discrete channel is indicated in Fig. 11. There are three possible symbols. The first is never affected by noise. The second and third each have probability p of coming through undisturbed, and q of being changed into the other of the pair. We have (letting $\alpha = -[p\log p + q\log q]$ and P and Q be the

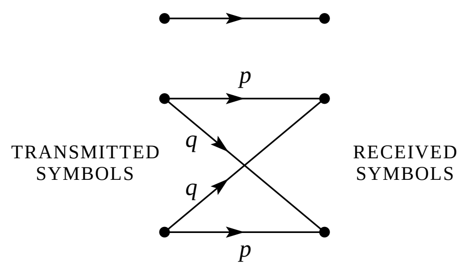


Fig. 11—Example of a discrete channel.

probabilities of using the first and second symbols)

$$H(x) = -P\log P - 2Q\log Q$$
$$H_y(x) = 2Q\alpha.$$

We wish to choose P and Q in such a way as to maximize $H(x) - H_y(x)$, subject to the constraint $P + 2Q = 1$. Hence we consider

$$U = -P\log P - 2Q\log Q - 2Q\alpha + \lambda(P + 2Q)$$

$$\frac{\partial U}{\partial P} = -1 - \log P + \lambda = 0$$
$$\frac{\partial U}{\partial Q} = -2 - 2\log Q - 2\alpha + 2\lambda = 0.$$

Eliminating λ

$$\log P = \log Q + \alpha$$
$$P = Qe^{\alpha} = Q\beta$$

$$P = \frac{\beta}{\beta + 2} \qquad\qquad Q = \frac{1}{\beta + 2}.$$

The channel capacity is then

$$C = \log \frac{\beta + 2}{\beta}.$$

Note how this checks the obvious values in the cases $p = 1$ and $p = \frac{1}{2}$. In the first, $\beta = 1$ and $C = \log 3$, which is correct since the channel is then noiseless with three possible symbols. If $p = \frac{1}{2}$, $\beta = 2$ and $C = \log 2$. Here the second and third symbols cannot be distinguished at all and act together like one symbol. The first symbol is used with probability $P = \frac{1}{2}$ and the second and third together with probability $\frac{1}{2}$. This may be distributed between them in any desired way and still achieve the maximum capacity.

For intermediate values of p the channel capacity will lie between $\log 2$ and $\log 3$. The distinction between the second and third symbols conveys some information but not as much as in the noiseless case. The first symbol is used somewhat more frequently than the other two because of its freedom from noise.

16. The Channel Capacity in Certain Special Cases

If the noise affects successive channel symbols independently it can be described by a set of transition probabilities p_{ij}. This is the probability, if symbol i is sent, that j will be received. The maximum channel rate is then given by the maximum of

$$-\sum_{i,j} P_i p_{ij} \log \sum_i P_i p_{ij} + \sum_{i,j} P_i p_{ij} \log p_{ij}$$

where we vary the P_i subject to $\sum P_i = 1$. This leads by the method of Lagrange to the equations,

$$\sum_j p_{sj} \log \frac{p_{sj}}{\sum_i P_i p_{ij}} = \mu \qquad s = 1, 2, \ldots.$$

Multiplying by P_s and summing on s shows that $\mu = C$. Let the inverse of p_{sj} (if it exists) be h_{st} so that $\sum_s h_{st} p_{sj} = \delta_{tj}$. Then:

$$\sum_{s,j} h_{st} p_{sj} \log p_{sj} - \log \sum_i P_i p_{it} = C \sum_s h_{st}.$$

Hence:

$$\sum_i P_i p_{it} = \exp\Big[-C \sum_s h_{st} + \sum_{s,j} h_{st} p_{sj} \log p_{sj}\Big]$$

or,

$$P_i = \sum_t h_{it} \exp\Big[-C \sum_s h_{st} + \sum_{s,j} h_{st} p_{sj} \log p_{sj}\Big].$$

This is the system of equations for determining the maximizing values of P_i, with C to be determined so that $\sum P_i = 1$. When this is done C will be the channel capacity, and the P_i the proper probabilities for the channel symbols to achieve this capacity.

If each input symbol has the same set of probabilities on the lines emerging from it, and the same is true of each output symbol, the capacity can be easily calculated. Examples are shown in Fig. 12. In such a case $H_x(y)$ is independent of the distribution of probabilities on the input symbols, and is given by $-\sum p_i \log p_i$ where the p_i are the values of the transition probabilities from any input symbol. The channel capacity is

$$\text{Max}\big[H(y) - H_x(y)\big] = \text{Max}\, H(y) + \sum p_i \log p_i.$$

The maximum of $H(y)$ is clearly $\log m$ where m is the number of output symbols, since it is possible to make them all equally probable by making the input symbols equally probable. The channel capacity is therefore

$$C = \log m + \sum p_i \log p_i.$$

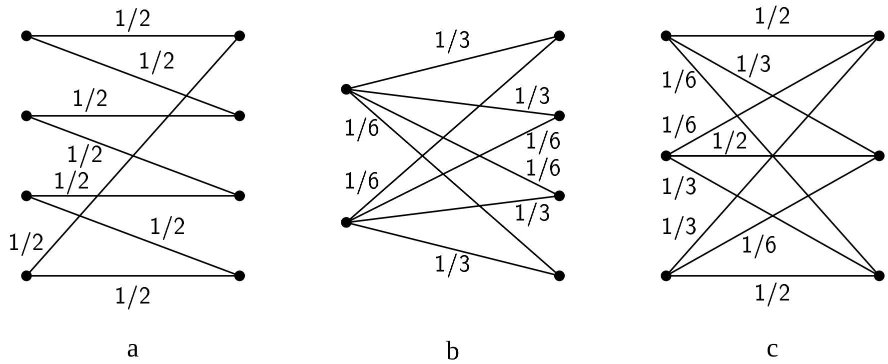


Fig. 12—Examples of discrete channels with the same transition probabilities for each input and for each output.

In Fig. 12a it would be

$$C = \log 4 - \log 2 = \log 2.$$

This could be achieved by using only the 1st and 3d symbols. In Fig. 12b

$$\begin{aligned} C &= \log 4 - \tfrac{2}{3}\log 3 - \tfrac{1}{3}\log 6 \\ &= \log 4 - \log 3 - \tfrac{1}{3}\log 2 \\ &= \log \tfrac{1}{3} 2^{\frac{5}{3}}. \end{aligned}$$

In Fig. 12c we have

$$\begin{aligned} C &= \log 3 - \tfrac{1}{2}\log 2 - \tfrac{1}{3}\log 3 - \tfrac{1}{6}\log 6 \\ &= \log \frac{3}{2^{\frac{1}{2}} 3^{\frac{1}{3}} 6^{\frac{1}{6}}}. \end{aligned}$$

Suppose the symbols fall into several groups such that the noise never causes a symbol in one group to be mistaken for a symbol in another group. Let the capacity for the nth group be C_n (in bits per second) when we use only the symbols in this group. Then it is easily shown that, for best use of the entire set, the total probability P_n of all symbols in the nth group should be

$$P_n = \frac{2^{C_n}}{\sum 2^{C_n}}.$$

Within a group the probability is distributed just as it would be if these were the only symbols being used. The channel capacity is

$$C = \log \sum 2^{C_n}.$$

17. An Example of Efficient Coding

The following example, although somewhat unrealistic, is a case in which exact matching to a noisy channel is possible. There are two channel symbols, 0 and 1, and the noise affects them in blocks of seven symbols. A block of seven is either transmitted without error, or exactly one symbol of the seven is incorrect. These eight possibilities are equally likely. We have

$$\begin{aligned} C &= \text{Max}\big[H(y) - H_x(y)\big] \\ &= \tfrac{1}{7}\big[7 + \tfrac{8}{8}\log \tfrac{1}{8}\big] \\ &= \tfrac{4}{7} \text{ bits/symbol}. \end{aligned}$$

An efficient code, allowing complete correction of errors and transmitting at the rate C, is the following (found by a method due to R. Hamming):

Let a block of seven symbols be $X_1, X_2, \ldots, X_7$. Of these X_3, X_5, X_6 and X_7 are message symbols and chosen arbitrarily by the source. The other three are redundant and calculated as follows:

X_4	is chosen to make	$\alpha = X_4 + X_5 + X_6 + X_7$	even
X_2	" " " "	$\beta = X_2 + X_3 + X_6 + X_7$	"
X_1	" " " "	$\gamma = X_1 + X_3 + X_5 + X_7$	"

When a block of seven is received α, β and γ are calculated and if even called zero, if odd called one. The binary number $\alpha\beta\gamma$ then gives the subscript of the X_i that is incorrect (if 0 there was no error).

APPENDIX 1

THE GROWTH OF THE NUMBER OF BLOCKS OF SYMBOLS WITH A FINITE STATE CONDITION

Let $N_i(L)$ be the number of blocks of symbols of length L ending in state i. Then we have

$$N_j(L) = \sum_{i,s} N_i\bigl(L - b_{ij}^{(s)}\bigr)$$

where $b_{ij}^1, b_{ij}^2, \ldots, b_{ij}^m$ are the length of the symbols which may be chosen in state i and lead to state j. These are linear difference equations and the behavior as $L \to \infty$ must be of the type

$$N_j = A_j W^L.$$

Substituting in the difference equation

$$A_j W^L = \sum_{i,s} A_i W^{L - b_{ij}^{(s)}}$$

or

$$A_j = \sum_{i,s} A_i W^{-b_{ij}^{(s)}}$$

$$\sum_i \Bigl(\sum_s W^{-b_{ij}^{(s)}} - \delta_{ij}\Bigr) A_i = 0.$$

For this to be possible the determinant

$$D(W) = |a_{ij}| = \Bigl|\sum_s W^{-b_{ij}^{(s)}} - \delta_{ij}\Bigr|$$

must vanish and this determines W, which is, of course, the largest real root of $D = 0$.

The quantity C is then given by

$$C = \mathop{\mathrm{Lim}}_{L \to \infty} \frac{\log \sum A_j W^L}{L} = \log W$$

and we also note that the same growth properties result if we require that all blocks start in the same (arbitrarily chosen) state.

APPENDIX 2

DERIVATION OF $H = -\sum p_i \log p_i$

Let $H\Bigl(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\Bigr) = A(n)$. From condition (3) we can decompose a choice from s^m equally likely possibilities into a series of m choices from s equally likely possibilities and obtain

$$A(s^m) = mA(s).$$

Similarly

$$A(t^n) = nA(t).$$

We can choose n arbitrarily large and find an m to satisfy

$$s^m \le t^n < s^{(m+1)}.$$

Thus, taking logarithms and dividing by $n \log s$,

$$\frac{m}{n} \le \frac{\log t}{\log s} \le \frac{m}{n} + \frac{1}{n} \quad \text{or} \quad \left| \frac{m}{n} - \frac{\log t}{\log s} \right| < \epsilon$$

where ϵ is arbitrarily small. Now from the monotonic property of $A(n)$,

$$A(s^m) \le A(t^n) \le A(s^{m+1})$$
$$mA(s) \le nA(t) \le (m+1)A(s).$$

Hence, dividing by $nA(s)$,

$$\frac{m}{n} \le \frac{A(t)}{A(s)} \le \frac{m}{n} + \frac{1}{n} \quad \text{or} \quad \left| \frac{m}{n} - \frac{A(t)}{A(s)} \right| < \epsilon$$

$$\left| \frac{A(t)}{A(s)} - \frac{\log t}{\log s} \right| < 2\epsilon \qquad A(t) = K \log t$$

where K must be positive to satisfy (2).

Now suppose we have a choice from n possibilities with commeasurable probabilities $p_i = \dfrac{n_i}{\sum n_i}$ where the n_i are integers. We can break down a choice from $\sum n_i$ possibilities into a choice from n possibilities with probabilities $p_1, \dots, p_n$ and then, if the ith was chosen, a choice from n_i with equal probabilities. Using condition (3) again, we equate the total choice from $\sum n_i$ as computed by two methods

$$K \log \sum n_i = H(p_1, \dots, p_n) + K \sum p_i \log n_i.$$

Hence

$$H = K\left[\sum p_i \log \sum n_i - \sum p_i \log n_i\right]$$
$$= -K \sum p_i \log \frac{n_i}{\sum n_i} = -K \sum p_i \log p_i.$$

If the p_i are incommeasurable, they may be approximated by rationals and the same expression must hold by our continuity assumption. Thus the expression holds in general. The choice of coefficient K is a matter of convenience and amounts to the choice of a unit of measure.

APPENDIX 3

THEOREMS ON ERGODIC SOURCES

If it is possible to go from any state with $P > 0$ to any other along a path of probability $p > 0$, the system is ergodic and the strong law of large numbers can be applied. Thus the number of times a given path p_{ij} in the network is traversed in a long sequence of length N is about proportional to the probability of being at i, say P_i, and then choosing this path, $P_i p_{ij} N$. If N is large enough the probability of percentage error $\pm\delta$ in this is less than ϵ so that for all but a set of small probability the actual numbers lie within the limits

$$(P_i p_{ij} \pm \delta)N.$$

Hence nearly all sequences have a probability p given by

$$p = \prod p_{ij}^{(P_i p_{ij} \pm \delta)N}$$

and $\frac{\log p}{N}$ is limited by

$$\frac{\log p}{N} = \sum (P_i p_{ij} \pm \delta) \log p_{ij}$$

or

$$\left| \frac{\log p}{N} - \sum P_i p_{ij} \log p_{ij} \right| < \eta.$$

This proves Theorem 3.

Theorem 4 follows immediately from this on calculating upper and lower bounds for $n(q)$ based on the possible range of values of p in Theorem 3.

In the mixed (not ergodic) case if

$$L = \sum p_i L_i$$

and the entropies of the components are $H_1 \geq H_2 \geq \cdots \geq H_n$ we have the

Theorem: $\underset{N\to\infty}{\operatorname{Lim}} \frac{\log n(q)}{N} = \varphi(q)$ *is a decreasing step function,*

$$\varphi(q) = H_s \quad \textit{in the interval} \quad \sum_1^{s-1} \alpha_i < q < \sum_1^{s} \alpha_i.$$

To prove Theorems 5 and 6 first note that F_N is monotonic decreasing because increasing N adds a subscript to a conditional entropy. A simple substitution for $p_{B_i}(S_j)$ in the definition of F_N shows that

$$F_N = NG_N - (N-1)G_{N-1}$$

and summing this for all N gives $G_N = \frac{1}{N}\sum F_n$. Hence $G_N \geq F_N$ and G_N monotonic decreasing. Also they must approach the same limit. By using Theorem 3 we see that $\underset{N\to\infty}{\operatorname{Lim}} G_N = H$.

APPENDIX 4

Maximizing the Rate for a System of Constraints

Suppose we have a set of constraints on sequences of symbols that is of the finite state type and can be represented therefore by a linear graph. Let $\ell_{ij}^{(s)}$ be the lengths of the various symbols that can occur in passing from state i to state j. What distribution of probabilities P_i for the different states and $p_{ij}^{(s)}$ for choosing symbol s in state i and going to state j maximizes the rate of generating information under these constraints? The constraints define a discrete channel and the maximum rate must be less than or equal to the capacity C of this channel, since if all blocks of large length were equally likely, this rate would result, and if possible this would be best. We will show that this rate can be achieved by proper choice of the P_i and $p_{ij}^{(s)}$.

The rate in question is

$$\frac{-\sum P_i p_{ij}^{(s)} \log p_{ij}^{(s)}}{\sum P_i p_{ij}^{(s)} \ell_{ij}^{(s)}} = \frac{N}{M}.$$

Let $\ell_{ij} = \sum_s \ell_{ij}^{(s)}$. Evidently for a maximum $p_{ij}^{(s)} = k \exp \ell_{ij}^{(s)}$. The constraints on maximization are $\sum P_i = 1$, $\sum_j p_{ij} = 1$, $\sum P_i(p_{ij} - \delta_{ij}) = 0$. Hence we maximize

$$U = \frac{-\sum P_i p_{ij} \log p_{ij}}{\sum P_i p_{ij} \ell_{ij}} + \lambda \sum_i P_i + \sum \mu_i p_{ij} + \sum \eta_j P_i (p_{ij} - \delta_{ij})$$

$$\frac{\partial U}{\partial p_{ij}} = -\frac{MP_i(1 + \log p_{ij}) + NP_i \ell_{ij}}{M^2} + \lambda + \mu_i + \eta_i P_i = 0.$$

Solving for p_{ij}

$$p_{ij} = A_i B_j D^{-\ell_{ij}}.$$

Since

$$\sum_j p_{ij} = 1, \quad A_i^{-1} = \sum_j B_j D^{-\ell_{ij}}$$

$$p_{ij} = \frac{B_j D^{-\ell_{ij}}}{\sum_s B_s D^{-\ell_{is}}}.$$

The correct value of D is the capacity C and the B_j are solutions of

$$B_i = \sum B_j C^{-\ell_{ij}}$$

for then

$$p_{ij} = \frac{B_j}{B_i} C^{-\ell_{ij}}$$

$$\sum P_i \frac{B_j}{B_i} C^{-\ell_{ij}} = P_j$$

or

$$\sum \frac{P_i}{B_i} C^{-\ell_{ij}} = \frac{P_j}{B_j}.$$

So that if λ_i satisfy

$$\sum \gamma_i C^{-\ell_{ij}} = \gamma_j$$

$$P_i = B_i \gamma_i.$$

Both the sets of equations for B_i and γ_i can be satisfied since C is such that

$$|C^{-\ell_{ij}} - \delta_{ij}| = 0.$$

In this case the rate is

$$-\frac{\sum P_i p_{ij} \log \frac{B_j}{B_i} C^{-\ell_{ij}}}{\sum P_i p_{ij} \ell_{ij}} = C - \frac{\sum P_i p_{ij} \log \frac{B_j}{B_i}}{\sum P_i p_{ij} \ell_{ij}}$$

but

$$\sum P_i p_{ij} (\log B_j - \log B_i) = \sum_j P_j \log B_j - \sum P_i \log B_i = 0$$

Hence the rate is C and as this could never be exceeded this is the maximum, justifying the assumed solution.

PART III: MATHEMATICAL PRELIMINARIES

In this final installment of the paper we consider the case where the signals or the messages or both are continuously variable, in contrast with the discrete nature assumed heretofore. To a considerable extent the continuous case can be obtained through a limiting process from the discrete case by dividing the continuum of messages and signals into a large but finite number of small regions and calculating the various parameters involved on a discrete basis. As the size of the regions is decreased these parameters in general approach as limits the proper values for the continuous case. There are, however, a few new effects that appear and also a general change of emphasis in the direction of specialization of the general results to particular cases.

We will not attempt, in the continuous case, to obtain our results with the greatest generality, or with the extreme rigor of pure mathematics, since this would involve a great deal of abstract measure theory and would obscure the main thread of the analysis. A preliminary study, however, indicates that the theory can be formulated in a completely axiomatic and rigorous manner which includes both the continuous and discrete cases and many others. The occasional liberties taken with limiting processes in the present analysis can be justified in all cases of practical interest.

18. Sets and Ensembles of Functions

We shall have to deal in the continuous case with sets of functions and ensembles of functions. A set of functions, as the name implies, is merely a class or collection of functions, generally of one variable, time. It can be specified by giving an explicit representation of the various functions in the set, or implicitly by giving a property which functions in the set possess and others do not. Some examples are:

1. The set of functions:

   $$f_\theta(t) = \sin(t+\theta).$$

   Each particular value of θ determines a particular function in the set.

2. The set of all functions of time containing no frequencies over W cycles per second.

3. The set of all functions limited in band to W and in amplitude to A.

4. The set of all English speech signals as functions of time.

An *ensemble* of functions is a set of functions together with a probability measure whereby we may determine the probability of a function in the set having certain properties.[1] For example with the set,

$$f_\theta(t) = \sin(t+\theta),$$

we may give a probability distribution for θ, $P(\theta)$. The set then becomes an ensemble.

Some further examples of ensembles of functions are:

1. A finite set of functions $f_k(t)$ $(k = 1,2,\dots,n)$ with the probability of f_k being p_k.

2. A finite dimensional family of functions

   $$f(\alpha_1,\alpha_2,\dots,\alpha_n;t)$$

   with a probability distribution on the parameters α_i:

   $$p(\alpha_1,\dots,\alpha_n).$$

   For example we could consider the ensemble defined by

   $$f(a_1,\dots,a_n,\theta_1,\dots,\theta_n;t) = \sum_{i=1}^{n} a_i \sin i(\omega t+\theta_i)$$

   with the amplitudes a_i distributed normally and independently, and the phases θ_i distributed uniformly (from 0 to 2π) and independently.

[1] In mathematical terminology the functions belong to a measure space whose total measure is unity.

3. The ensemble

$$f(a_i,t)=\sum_{n=-\infty}^{+\infty} a_n \frac{\sin \pi(2Wt-n)}{\pi(2Wt-n)}$$

with the a_i normal and independent all with the same standard deviation $\sqrt{N}$. This is a representation of "white" noise, band limited to the band from 0 to W cycles per second and with average power N.[2]

4. Let points be distributed on the t axis according to a Poisson distribution. At each selected point the function $f(t)$ is placed and the different functions added, giving the ensemble

$$\sum_{k=-\infty}^{\infty} f(t+t_k)$$

where the t_k are the points of the Poisson distribution. This ensemble can be considered as a type of impulse or shot noise where all the impulses are identical.

5. The set of English speech functions with the probability measure given by the frequency of occurrence in ordinary use.

An ensemble of functions $f_\alpha(t)$ is *stationary* if the same ensemble results when all functions are shifted any fixed amount in time. The ensemble

$$f_\theta(t)=\sin(t+\theta)$$

is stationary if θ is distributed uniformly from 0 to 2π. If we shift each function by t_1 we obtain

$$\begin{aligned} f_\theta(t+t_1) &= \sin(t+t_1+\theta) \\ &= \sin(t+\varphi) \end{aligned}$$

with φ distributed uniformly from 0 to 2π. Each function has changed but the ensemble as a whole is invariant under the translation. The other examples given above are also stationary.

An ensemble is *ergodic* if it is stationary, and there is no subset of the functions in the set with a probability different from 0 and 1 which is stationary. The ensemble

$$\sin(t+\theta)$$

is ergodic. No subset of these functions of probability $\neq 0,1$ is transformed into itself under all time translations. On the other hand the ensemble

$$a\sin(t+\theta)$$

with a distributed normally and θ uniform is stationary but not ergodic. The subset of these functions with a between 0 and 1 for example is stationary.

Of the examples given, 3 and 4 are ergodic, and 5 may perhaps be considered so. If an ensemble is ergodic we may say roughly that each function in the set is typical of the ensemble. More precisely it is known that with an ergodic ensemble an average of any statistic over the ensemble is equal (with probability 1) to an average over the time translations of a particular function of the set.[3] Roughly speaking, each function can be expected, as time progresses, to go through, with the proper frequency, all the convolutions of any of the functions in the set.

[2]This representation can be used as a definition of band limited white noise. It has certain advantages in that it involves fewer limiting operations than do definitions that have been used in the past. The name "white noise," already firmly entrenched in the literature, is perhaps somewhat unfortunate. In optics white light means either any continuous spectrum as contrasted with a point spectrum, or a spectrum which is flat with *wavelength* (which is not the same as a spectrum flat with frequency).

[3]This is the famous ergodic theorem or rather one aspect of this theorem which was proved in somewhat different formulations by Birkoff, von Neumann, and Koopman, and subsequently generalized by Wiener, Hopf, Hurewicz and others. The literature on ergodic theory is quite extensive and the reader is referred to the papers of these writers for precise and general formulations; e.g., E. Hopf, "Ergodentheorie," *Ergebnisse der Mathematik und ihrer Grenzgebiete,* v. 5; "On Causality Statistics and Probability," *Journal of Mathematics and Physics,* v. XIII, No. 1, 1934; N. Wiener, "The Ergodic Theorem," *Duke Mathematical Journal,* v. 5, 1939.

Just as we may perform various operations on numbers or functions to obtain new numbers or functions, we can perform operations on ensembles to obtain new ensembles. Suppose, for example, we have an ensemble of functions $f_\alpha(t)$ and an operator T which gives for each function $f_\alpha(t)$ a resulting function $g_\alpha(t)$:

$$g_\alpha(t) = T f_\alpha(t).$$

Probability measure is defined for the set $g_\alpha(t)$ by means of that for the set $f_\alpha(t)$. The probability of a certain subset of the $g_\alpha(t)$ functions is equal to that of the subset of the $f_\alpha(t)$ functions which produce members of the given subset of g functions under the operation T. Physically this corresponds to passing the ensemble through some device, for example, a filter, a rectifier or a modulator. The output functions of the device form the ensemble $g_\alpha(t)$.

A device or operator T will be called invariant if shifting the input merely shifts the output, i.e., if

$$g_\alpha(t) = T f_\alpha(t)$$

implies

$$g_\alpha(t+t_1) = T f_\alpha(t+t_1)$$

for all $f_\alpha(t)$ and all t_1. It is easily shown (see Appendix 5 that if T is invariant and the input ensemble is stationary then the output ensemble is stationary. Likewise if the input is ergodic the output will also be ergodic.

A filter or a rectifier is invariant under all time translations. The operation of modulation is not since the carrier phase gives a certain time structure. However, modulation is invariant under all translations which are multiples of the period of the carrier.

Wiener has pointed out the intimate relation between the invariance of physical devices under time translations and Fourier theory.[4] He has shown, in fact, that if a device is linear as well as invariant Fourier analysis is then the appropriate mathematical tool for dealing with the problem.

An ensemble of functions is the appropriate mathematical representation of the messages produced by a continuous source (for example, speech), of the signals produced by a transmitter, and of the perturbing noise. Communication theory is properly concerned, as has been emphasized by Wiener, not with operations on particular functions, but with operations on ensembles of functions. A communication system is designed not for a particular speech function and still less for a sine wave, but for the ensemble of speech functions.

19. Band Limited Ensembles of Functions

If a function of time $f(t)$ is limited to the band from 0 to W cycles per second it is completely determined by giving its ordinates at a series of discrete points spaced $\frac{1}{2W}$ seconds apart in the manner indicated by the following result.[5]

Theorem 13: Let $f(t)$ contain no frequencies over W. Then

$$f(t) = \sum_{-\infty}^{\infty} X_n \frac{\sin \pi(2Wt-n)}{\pi(2Wt-n)}$$

where

$$X_n = f\left(\frac{n}{2W}\right).$$

[4] Communication theory is heavily indebted to Wiener for much of its basic philosophy and theory. His classic NDRC report, *The Interpolation, Extrapolation and Smoothing of Stationary Time Series* (Wiley, 1949), contains the first clear-cut formulation of communication theory as a statistical problem, the study of operations on time series. This work, although chiefly concerned with the linear prediction and filtering problem, is an important collateral reference in connection with the present paper. We may also refer here to Wiener's *Cybernetics* (Wiley, 1948), dealing with the general problems of communication and control.

[5] For a proof of this theorem and further discussion see the author's paper "Communication in the Presence of Noise" published in the *Proceedings of the Institute of Radio Engineers*, v. 37, No. 1, Jan., 1949, pp. 10–21.

In this expansion $f(t)$ is represented as a sum of orthogonal functions. The coefficients X_n of the various terms can be considered as coordinates in an infinite dimensional "function space." In this space each function corresponds to precisely one point and each point to one function.

A function can be considered to be substantially limited to a time T if all the ordinates X_n outside this interval of time are zero. In this case all but $2TW$ of the coordinates will be zero. Thus functions limited to a band W and duration T correspond to points in a space of $2TW$ dimensions.

A subset of the functions of band W and duration T corresponds to a region in this space. For example, the functions whose total energy is less than or equal to E correspond to points in a $2TW$ dimensional sphere with radius $r = \sqrt{2WE}$.

An *ensemble* of functions of limited duration and band will be represented by a probability distribution $p(x_1, \ldots, x_n)$ in the corresponding n dimensional space. If the ensemble is not limited in time we can consider the $2TW$ coordinates in a given interval T to represent substantially the part of the function in the interval T and the probability distribution $p(x_1, \ldots, x_n)$ to give the statistical structure of the ensemble for intervals of that duration.

20. Entropy of a Continuous Distribution

The entropy of a discrete set of probabilities $p_1, \ldots, p_n$ has been defined as:

$$H = -\sum p_i \log p_i.$$

In an analogous manner we define the entropy of a continuous distribution with the density distribution function $p(x)$ by:

$$H = -\int_{-\infty}^{\infty} p(x) \log p(x)\, dx.$$

With an n dimensional distribution $p(x_1, \ldots, x_n)$ we have

$$H = -\int \cdots \int p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

If we have two arguments x and y (which may themselves be multidimensional) the joint and conditional entropies of $p(x,y)$ are given by

$$H(x,y) = -\iint p(x,y) \log p(x,y)\, dx\, dy$$

and

$$H_x(y) = -\iint p(x,y) \log \frac{p(x,y)}{p(x)}\, dx\, dy$$
$$H_y(x) = -\iint p(x,y) \log \frac{p(x,y)}{p(y)}\, dx\, dy$$

where

$$p(x) = \int p(x,y)\, dy$$
$$p(y) = \int p(x,y)\, dx.$$

The entropies of continuous distributions have most (but not all) of the properties of the discrete case. In particular we have the following:

1. If x is limited to a certain volume v in its space, then $H(x)$ is a maximum and equal to $\log v$ when $p(x)$ is constant $(1/v)$ in the volume.

2. With any two variables x, y we have

$$H(x,y) \le H(x) + H(y)$$

with equality if (and only if) x and y are independent, i.e., $p(x,y) = p(x)p(y)$ (apart possibly from a set of points of probability zero).

3. Consider a generalized averaging operation of the following type:

$$p'(y) = \int a(x,y)p(x)\,dx$$

with

$$\int a(x,y)\,dx = \int a(x,y)\,dy = 1, \qquad a(x,y) \ge 0.$$

Then the entropy of the averaged distribution $p'(y)$ is equal to or greater than that of the original distribution $p(x)$.

4. We have

$$H(x,y) = H(x) + H_x(y) = H(y) + H_y(x)$$

and

$$H_x(y) \le H(y).$$

5. Let $p(x)$ be a one-dimensional distribution. The form of $p(x)$ giving a maximum entropy subject to the condition that the standard deviation of x be fixed at σ is Gaussian. To show this we must maximize

$$H(x) = -\int p(x)\log p(x)\,dx$$

with

$$\sigma^2 = \int p(x)x^2\,dx \quad \text{and} \quad 1 = \int p(x)\,dx$$

as constraints. This requires, by the calculus of variations, maximizing

$$\int \left[-p(x)\log p(x) + \lambda p(x)x^2 + \mu p(x)\right] dx.$$

The condition for this is

$$-1 - \log p(x) + \lambda x^2 + \mu = 0$$

and consequently (adjusting the constants to satisfy the constraints)

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x^2/2\sigma^2)}.$$

Similarly in n dimensions, suppose the second order moments of $p(x_1,\ldots,x_n)$ are fixed at A_{ij}:

$$A_{ij} = \int\cdots\int x_i x_j p(x_1,\ldots,x_n)\,dx_1\cdots dx_n.$$

Then the maximum entropy occurs (by a similar calculation) when $p(x_1,\ldots,x_n)$ is the n dimensional Gaussian distribution with the second order moments A_{ij}.

6. The entropy of a one-dimensional Gaussian distribution whose standard deviation is σ is given by

$$H(x) = \log \sqrt{2\pi e}\sigma.$$

This is calculated as follows:

$$\begin{aligned}
p(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(x^2/2\sigma^2)} \\
-\log p(x) &= \log \sqrt{2\pi}\sigma + \frac{x^2}{2\sigma^2} \\
H(x) &= -\int p(x) \log p(x)\, dx \\
&= \int p(x) \log \sqrt{2\pi}\sigma\, dx + \int p(x) \frac{x^2}{2\sigma^2}\, dx \\
&= \log \sqrt{2\pi}\sigma + \frac{\sigma^2}{2\sigma^2} \\
&= \log \sqrt{2\pi}\sigma + \log \sqrt{e} \\
&= \log \sqrt{2\pi e}\sigma.
\end{aligned}$$

Similarly the n dimensional Gaussian distribution with associated quadratic form a_{ij} is given by

$$p(x_1, \ldots, x_n) = \frac{|a_{ij}|^{\frac{1}{2}}}{(2\pi)^{n/2}} \exp\Big(-\tfrac{1}{2} \sum a_{ij} x_i x_j\Big)$$

and the entropy can be calculated as

$$H = \log (2\pi e)^{n/2} |a_{ij}|^{-\frac{1}{2}}$$

where $|a_{ij}|$ is the determinant whose elements are a_{ij}.

7. If x is limited to a half line ($p(x) = 0$ for $x \leq 0$) and the first moment of x is fixed at a:

$$a = \int_0^\infty p(x) x\, dx,$$

then the maximum entropy occurs when

$$p(x) = \frac{1}{a} e^{-(x/a)}$$

and is equal to $\log ea$.

8. There is one important difference between the continuous and discrete entropies. In the discrete case the entropy measures in an *absolute* way the randomness of the chance variable. In the continuous case the measurement is *relative to the coordinate system.* If we change coordinates the entropy will in general change. In fact if we change to coordinates $y_1 \cdots y_n$ the new entropy is given by

$$H(y) = \int \cdots \int p(x_1, \ldots, x_n) J\Big(\frac{x}{y}\Big) \log p(x_1, \ldots, x_n) J\Big(\frac{x}{y}\Big)\, dy_1 \cdots dy_n$$

where $J\big(\frac{x}{y}\big)$ is the Jacobian of the coordinate transformation. On expanding the logarithm and changing the variables to $x_1 \cdots x_n$, we obtain:

$$H(y) = H(x) - \int \cdots \int p(x_1, \ldots, x_n) \log J\Big(\frac{x}{y}\Big)\, dx_1 \ldots dx_n.$$

Thus the new entropy is the old entropy less the expected logarithm of the Jacobian. In the continuous case the entropy can be considered a measure of randomness *relative to an assumed standard*, namely the coordinate system chosen with each small volume element $dx_1 \cdots dx_n$ given equal weight. When we change the coordinate system the entropy in the new system measures the randomness when equal volume elements $dy_1 \cdots dy_n$ in the new system are given equal weight.

In spite of this dependence on the coordinate system the entropy concept is as important in the continuous case as the discrete case. This is due to the fact that the derived concepts of information rate and channel capacity depend on the *difference* of two entropies and this difference *does not* depend on the coordinate frame, each of the two terms being changed by the same amount.

The entropy of a continuous distribution can be negative. The scale of measurements sets an arbitrary zero corresponding to a uniform distribution over a unit volume. A distribution which is more confined than this has less entropy and will be negative. The rates and capacities will, however, always be non-negative.

9. A particular case of changing coordinates is the linear transformation

$$y_j = \sum_i a_{ij} x_i.$$

In this case the Jacobian is simply the determinant $|a_{ij}|^{-1}$ and

$$H(y) = H(x) + \log |a_{ij}|.$$

In the case of a rotation of coordinates (or any measure preserving transformation) $J = 1$ and $H(y) = H(x)$.

21. Entropy of an Ensemble of Functions

Consider an ergodic ensemble of functions limited to a certain band of width W cycles per second. Let

$$p(x_1, \ldots, x_n)$$

be the density distribution function for amplitudes $x_1, \ldots, x_n$ at n successive sample points. We define the entropy of the ensemble per degree of freedom by

$$H' = -\lim_{n \to \infty} \frac{1}{n} \int \cdots \int p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

We may also define an entropy H per second by dividing, not by n, but by the time T in seconds for n samples. Since $n = 2TW$, $H = 2WH'$.

With white thermal noise p is Gaussian and we have

$$H' = \log \sqrt{2\pi e N},$$
$$H = W \log 2\pi e N.$$

For a given average power N, white noise has the maximum possible entropy. This follows from the maximizing properties of the Gaussian distribution noted above.

The entropy for a continuous stochastic process has many properties analogous to that for discrete processes. In the discrete case the entropy was related to the logarithm of the *probability* of long sequences, and to the *number* of reasonably probable sequences of long length. In the continuous case it is related in a similar fashion to the logarithm of the *probability density* for a long series of samples, and the *volume* of reasonably high probability in the function space.

More precisely, if we assume $p(x_1, \ldots, x_n)$ continuous in all the x_i for all n, then for sufficiently large n

$$\left| \frac{\log p}{n} - H' \right| < \epsilon$$

for all choices of $(x_1,\ldots,x_n)$ apart from a set whose total probability is less than δ, with δ and ϵ arbitrarily small. This follows form the ergodic property if we divide the space into a large number of small cells.

The relation of H to volume can be stated as follows: Under the same assumptions consider the n dimensional space corresponding to $p(x_1,\ldots,x_n)$. Let $V_n(q)$ be the smallest volume in this space which includes in its interior a total probability q. Then

$$\operatorname*{Lim}_{n\to\infty}\frac{\log V_n(q)}{n}=H'$$

provided q does not equal 0 or 1.

These results show that for large n there is a rather well-defined volume (at least in the logarithmic sense) of high probability, and that within this volume the probability density is relatively uniform (again in the logarithmic sense).

In the white noise case the distribution function is given by

$$p(x_1,\ldots,x_n)=\frac{1}{(2\pi N)^{n/2}}\exp-\frac{1}{2N}\sum x_i^2.$$

Since this depends only on $\sum x_i^2$ the surfaces of equal probability density are spheres and the entire distribution has spherical symmetry. The region of high probability is a sphere of radius $\sqrt{nN}$. As $n\to\infty$ the probability of being outside a sphere of radius $\sqrt{n(N+\epsilon)}$ approaches zero and $\frac{1}{n}$ times the logarithm of the volume of the sphere approaches $\log\sqrt{2\pi eN}$.

In the continuous case it is convenient to work not with the entropy H of an ensemble but with a derived quantity which we will call the entropy power. This is defined as the power in a white noise limited to the same band as the original ensemble and having the same entropy. In other words if H' is the entropy of an ensemble its entropy power is

$$N_1=\frac{1}{2\pi e}\exp 2H'.$$

In the geometrical picture this amounts to measuring the high probability volume by the squared radius of a sphere having the same volume. Since white noise has the maximum entropy for a given power, the entropy power of any noise is less than or equal to its actual power.

22. Entropy Loss in Linear Filters

Theorem 14: If an ensemble having an entropy H_1 per degree of freedom in band W is passed through a filter with characteristic $Y(f)$ the output ensemble has an entropy

$$H_2=H_1+\frac{1}{W}\int_W \log|Y(f)|^2\,df.$$

The operation of the filter is essentially a linear transformation of coordinates. If we think of the different frequency components as the original coordinate system, the new frequency components are merely the old ones multiplied by factors. The coordinate transformation matrix is thus essentially diagonalized in terms of these coordinates. The Jacobian of the transformation is (for n sine and n cosine components)

$$J=\prod_{i=1}^{n}|Y(f_i)|^2$$

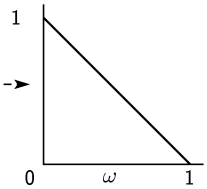
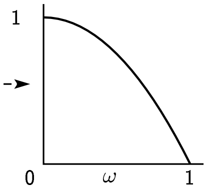
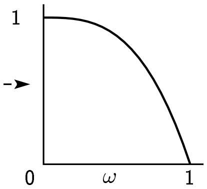
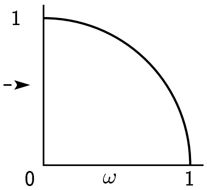
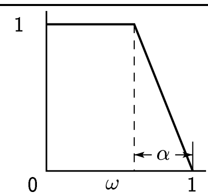
where the f_i are equally spaced through the band W. This becomes in the limit

$$\exp\frac{1}{W}\int_W \log|Y(f)|^2\,df.$$

Since J is constant its average value is the same quantity and applying the theorem on the change of entropy with a change of coordinates, the result follows. We may also phrase it in terms of the entropy power. Thus if the entropy power of the first ensemble is N_1 that of the second is

$$N_1\exp\frac{1}{W}\int_W \log|Y(f)|^2\,df.$$

TABLE I

GAIN	ENTROPY POWER FACTOR	ENTROPY POWER GAIN IN DECIBELS	IMPULSE RESPONSE
$1-\omega$ 	$\frac{1}{e^2}$	-8.69	$\frac{\sin^2(t/2)}{t^2/2}$
$1-\omega^2$ 	$\left(\frac{2}{e}\right)^4$	-5.33	$2\left[\frac{\sin t}{t^3}-\frac{\cos t}{t^2}\right]$
$1-\omega^3$ 	0.411	-3.87	$6\left[\frac{\cos t-1}{t^4}-\frac{\cos t}{2t^2}+\frac{\sin t}{t^3}\right]$
$\sqrt{1-\omega^2}$ 	$\left(\frac{2}{e}\right)^2$	-2.67	$\frac{\pi}{2}\frac{J_1(t)}{t}$
	$\frac{1}{e^{2\alpha}}$	-8.69α	$\frac{1}{\alpha t^2}\left[\cos(1-\alpha)t-\cos t\right]$

The final entropy power is the initial entropy power multiplied by the geometric mean gain of the filter. If the gain is measured in *db*, then the output entropy power will be increased by the arithmetic mean *db* gain over W.

In Table I the entropy power loss has been calculated (and also expressed in *db*) for a number of ideal gain characteristics. The impulsive responses of these filters are also given for $W = 2\pi$, with phase assumed to be 0.

The entropy loss for many other cases can be obtained from these results. For example the entropy power factor $1/e^2$ for the first case also applies to any gain characteristic obtain from $1-\omega$ by a measure preserving transformation of the ω axis. In particular a linearly increasing gain $G(\omega) = \omega$, or a "saw tooth" characteristic between 0 and 1 have the same entropy loss. The reciprocal gain has the reciprocal factor. Thus $1/\omega$ has the factor e^2. Raising the gain to any power raises the factor to this power.

23. Entropy of a Sum of Two Ensembles

If we have two ensembles of functions $f_\alpha(t)$ and $g_\beta(t)$ we can form a new ensemble by "addition." Suppose the first ensemble has the probability density function $p(x_1, \dots, x_n)$ and the second $q(x_1, \dots, x_n)$. Then the

density function for the sum is given by the convolution:

$$r(x_1,\ldots,x_n) = \int\cdots\int p(y_1,\ldots,y_n)q(x_1-y_1,\ldots,x_n-y_n)\,dy_1\cdots dy_n.$$

Physically this corresponds to adding the noises or signals represented by the original ensembles of functions.

The following result is derived in Appendix 6.

Theorem 15: Let the average power of two ensembles be N_1 *and* N_2 *and let their entropy powers be* $\overline{N}_1$ *and* $\overline{N}_2$*. Then the entropy power of the sum,* $\overline{N}_3$*, is bounded by*

$$\overline{N}_1+\overline{N}_2 \le \overline{N}_3 \le N_1+N_2.$$

White Gaussian noise has the peculiar property that it can absorb any other noise or signal ensemble which may be added to it with a resultant entropy power approximately equal to the sum of the white noise power and the signal power (measured from the average signal value, which is normally zero), provided the signal power is small, in a certain sense, compared to noise.

Consider the function space associated with these ensembles having n dimensions. The white noise corresponds to the spherical Gaussian distribution in this space. The signal ensemble corresponds to another probability distribution, not necessarily Gaussian or spherical. Let the second moments of this distribution about its center of gravity be a_{ij}. That is, if $p(x_1,\ldots,x_n)$ is the density distribution function

$$a_{ij} = \int\cdots\int p(x_i-\alpha_i)(x_j-\alpha_j)\,dx_1\cdots dx_n$$

where the α_i are the coordinates of the center of gravity. Now a_{ij} is a positive definite quadratic form, and we can rotate our coordinate system to align it with the principal directions of this form. a_{ij} is then reduced to diagonal form b_{ii}. We require that each b_{ii} be small compared to N, the squared radius of the spherical distribution.

In this case the convolution of the noise and signal produce approximately a Gaussian distribution whose corresponding quadratic form is

$$N+b_{ii}.$$

The entropy power of this distribution is

$$\left[\prod(N+b_{ii})\right]^{1/n}$$

or approximately

$$\begin{aligned} &= \left[(N)^n+\sum b_{ii}(N)^{n-1}\right]^{1/n} \\ &\doteq N+\frac{1}{n}\sum b_{ii}. \end{aligned}$$

The last term is the signal power, while the first is the noise power.

PART IV: THE CONTINUOUS CHANNEL

24. The Capacity of a Continuous Channel

In a continuous channel the input or transmitted signals will be continuous functions of time $f(t)$ belonging to a certain set, and the output or received signals will be perturbed versions of these. We will consider only the case where both transmitted and received signals are limited to a certain band W. They can then be specified, for a time T, by $2TW$ numbers, and their statistical structure by finite dimensional distribution functions. Thus the statistics of the transmitted signal will be determined by

$$P(x_1,\ldots,x_n) = P(x)$$

and those of the noise by the conditional probability distribution

$$P_{x_1,\ldots,x_n}(y_1,\ldots,y_n) = P_x(y).$$

The rate of transmission of information for a continuous channel is defined in a way analogous to that for a discrete channel, namely

$$R = H(x) - H_y(x)$$

where $H(x)$ is the entropy of the input and $H_y(x)$ the equivocation. The channel capacity C is defined as the maximum of R when we vary the input over all possible ensembles. This means that in a finite dimensional approximation we must vary $P(x) = P(x_1,\ldots,x_n)$ and maximize

$$-\int P(x)\log P(x)\,dx + \iint P(x,y)\log\frac{P(x,y)}{P(y)}\,dx\,dy.$$

This can be written

$$\iint P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\,dx\,dy$$

using the fact that $\iint P(x,y)\log P(x)\,dx\,dy = \int P(x)\log P(x)\,dx$. The channel capacity is thus expressed as follows:

$$C = \operatorname*{Lim}_{T\to\infty}\operatorname*{Max}_{P(x)}\frac{1}{T}\iint P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\,dx\,dy.$$

It is obvious in this form that R and C are independent of the coordinate system since the numerator and denominator in $\log\frac{P(x,y)}{P(x)P(y)}$ will be multiplied by the same factors when x and y are transformed in any one-to-one way. This integral expression for C is more general than $H(x) - H_y(x)$. Properly interpreted (see Appendix 7) it will always exist while $H(x) - H_y(x)$ may assume an indeterminate form $\infty - \infty$ in some cases. This occurs, for example, if x is limited to a surface of fewer dimensions than n in its n dimensional approximation.

If the logarithmic base used in computing $H(x)$ and $H_y(x)$ is two then C is the maximum number of binary digits that can be sent per second over the channel with arbitrarily small equivocation, just as in the discrete case. This can be seen physically by dividing the space of signals into a large number of small cells, sufficiently small so that the probability density $P_x(y)$ of signal x being perturbed to point y is substantially constant over a cell (either of x or y). If the cells are considered as distinct points the situation is essentially the same as a discrete channel and the proofs used there will apply. But it is clear physically that this quantizing of the volume into individual points cannot in any practical situation alter the final answer significantly, provided the regions are sufficiently small. Thus the capacity will be the limit of the capacities for the discrete subdivisions and this is just the continuous capacity defined above.

On the mathematical side it can be shown first (see Appendix 7) that if u is the message, x is the signal, y is the received signal (perturbed by noise) and v is the recovered message then

$$H(x) - H_y(x) \geq H(u) - H_v(u)$$

regardless of what operations are performed on u to obtain x or on y to obtain v. Thus no matter how we encode the binary digits to obtain the signal, or how we decode the received signal to recover the message, the discrete rate for the binary digits does not exceed the channel capacity we have defined. On the other hand, it is possible under very general conditions to find a coding system for transmitting binary digits at the rate C with as small an equivocation or frequency of errors as desired. This is true, for example, if, when we take a finite dimensional approximating space for the signal functions, $P(x,y)$ is continuous in both x and y except at a set of points of probability zero.

An important special case occurs when the noise is added to the signal and is independent of it (in the probability sense). Then $P_x(y)$ is a function only of the difference $n = (y - x)$,

$$P_x(y) = Q(y - x)$$

and we can assign a definite entropy to the noise (independent of the statistics of the signal), namely the entropy of the distribution $Q(n)$. This entropy will be denoted by $H(n)$.

Theorem 16: If the signal and noise are independent and the received signal is the sum of the transmitted signal and the noise then the rate of transmission is

$$R = H(y) - H(n),$$

i.e., the entropy of the received signal less the entropy of the noise. The channel capacity is

$$C = \operatorname*{Max}_{P(x)} H(y) - H(n).$$

We have, since $y = x + n$:

$$H(x,y) = H(x,n).$$

Expanding the left side and using the fact that x and n are independent

$$H(y) + H_y(x) = H(x) + H(n).$$

Hence

$$R = H(x) - H_y(x) = H(y) - H(n).$$

Since $H(n)$ is independent of $P(x)$, maximizing R requires maximizing $H(y)$, the entropy of the received signal. If there are certain constraints on the ensemble of transmitted signals, the entropy of the received signal must be maximized subject to these constraints.

25. Channel Capacity with an Average Power Limitation

A simple application of Theorem 16 is the case when the noise is a white thermal noise and the transmitted signals are limited to a certain average power P. Then the received signals have an average power $P + N$ where N is the average noise power. The maximum entropy for the received signals occurs when they also form a white noise ensemble since this is the greatest possible entropy for a power $P + N$ and can be obtained by a suitable choice of transmitted signals, namely if they form a white noise ensemble of power P. The entropy (per second) of the received ensemble is then

$$H(y) = W \log 2\pi e(P + N),$$

and the noise entropy is

$$H(n) = W \log 2\pi eN.$$

The channel capacity is

$$C = H(y) - H(n) = W \log \frac{P + N}{N}.$$

Summarizing we have the following:

Theorem 17: The capacity of a channel of band W perturbed by white thermal noise power N when the average transmitter power is limited to P is given by

$$C = W \log \frac{P + N}{N}.$$

This means that by sufficiently involved encoding systems we can transmit binary digits at the rate $W \log_2 \frac{P + N}{N}$ bits per second, with arbitrarily small frequency of errors. It is not possible to transmit at a higher rate by any encoding system without a definite positive frequency of errors.

To approximate this limiting rate of transmission the transmitted signals must approximate, in statistical properties, a white noise.[6] A system which approaches the ideal rate may be described as follows: Let

[6]This and other properties of the white noise case are discussed from the geometrical point of view in "Communication in the Presence of Noise," *loc. cit.*

$M = 2^s$ samples of white noise be constructed each of duration T. These are assigned binary numbers from 0 to $M-1$. At the transmitter the message sequences are broken up into groups of s and for each group the corresponding noise sample is transmitted as the signal. At the receiver the M samples are known and the actual received signal (perturbed by noise) is compared with each of them. The sample which has the least R.M.S. discrepancy from the received signal is chosen as the transmitted signal and the corresponding binary number reconstructed. This process amounts to choosing the most probable (*a posteriori*) signal. The number M of noise samples used will depend on the tolerable frequency ϵ of errors, but for almost all selections of samples we have

$$\operatorname*{Lim}_{\epsilon\to 0}\operatorname*{Lim}_{T\to\infty}\frac{\log M(\epsilon,T)}{T} = W\log\frac{P+N}{N},$$

so that no matter how small ϵ is chosen, we can, by taking T sufficiently large, transmit as near as we wish to $TW\log\frac{P+N}{N}$ binary digits in the time T.

Formulas similar to $C = W\log\frac{P+N}{N}$ for the white noise case have been developed independently by several other writers, although with somewhat different interpretations. We may mention the work of N. Wiener,[7] W. G. Tuller,[8] and H. Sullivan in this connection.

In the case of an arbitrary perturbing noise (not necessarily white thermal noise) it does not appear that the maximizing problem involved in determining the channel capacity C can be solved explicitly. However, upper and lower bounds can be set for C in terms of the average noise power N the noise entropy power N_1. These bounds are sufficiently close together in most practical cases to furnish a satisfactory solution to the problem.

Theorem 18: The capacity of a channel of band W perturbed by an arbitrary noise is bounded by the inequalities

$$W\log\frac{P+N_1}{N_1} \le C \le W\log\frac{P+N}{N_1}$$

where

$$\begin{aligned}
P &= \textit{average transmitter power}\\
N &= \textit{average noise power}\\
N_1 &= \textit{entropy power of the noise.}
\end{aligned}$$

Here again the average power of the perturbed signals will be $P+N$. The maximum entropy for this power would occur if the received signal were white noise and would be $W\log 2\pi e(P+N)$. It may not be possible to achieve this; i.e., there may not be any ensemble of transmitted signals which, added to the perturbing noise, produce a white thermal noise at the receiver, but at least this sets an upper bound to $H(y)$. We have, therefore

$$\begin{aligned}
C &= \operatorname{Max} H(y) - H(n)\\
&\le W\log 2\pi e(P+N) - W\log 2\pi e N_1.
\end{aligned}$$

This is the upper limit given in the theorem. The lower limit can be obtained by considering the rate if we make the transmitted signal a white noise, of power P. In this case the entropy power of the received signal must be at least as great as that of a white noise of power $P+N_1$ since we have shown in in a previous theorem that the entropy power of the sum of two ensembles is greater than or equal to the sum of the individual entropy powers. Hence

$$\operatorname{Max} H(y) \ge W\log 2\pi e(P+N_1)$$

[7] *Cybernetics, loc. cit.*

[8] "Theoretical Limitations on the Rate of Transmission of Information," *Proceedings of the Institute of Radio Engineers*, v. 37, No. 5, May, 1949, pp. 468–78.

and

$$\begin{aligned} C &\geq W \log 2\pi e(P+N_1) - W \log 2\pi e N_1 \\ &= W \log \frac{P+N_1}{N_1}. \end{aligned}$$

As P increases, the upper and lower bounds approach each other, so we have as an asymptotic rate

$$W \log \frac{P+N}{N_1}.$$

If the noise is itself white, $N = N_1$ and the result reduces to the formula proved previously:

$$C = W \log\left(1 + \frac{P}{N}\right).$$

If the noise is Gaussian but with a spectrum which is not necessarily flat, N_1 is the geometric mean of the noise power over the various frequencies in the band W. Thus

$$N_1 = \exp \frac{1}{W} \int_W \log N(f)\, df$$

where $N(f)$ is the noise power at frequency f.

Theorem 19: If we set the capacity for a given transmitter power P equal to

$$C = W \log \frac{P+N-\eta}{N_1}$$

then η is monotonic decreasing as P increases and approaches 0 as a limit.

Suppose that for a given power P_1 the channel capacity is

$$W \log \frac{P_1+N-\eta_1}{N_1}.$$

This means that the best signal distribution, say $p(x)$, when added to the noise distribution $q(x)$, gives a received distribution $r(y)$ whose entropy power is $(P_1+N-\eta_1)$. Let us increase the power to $P_1+\Delta P$ by adding a white noise of power ΔP to the signal. The entropy of the received signal is now at least

$$H(y) = W \log 2\pi e(P_1+N-\eta_1+\Delta P)$$

by application of the theorem on the minimum entropy power of a sum. Hence, since we can attain the H indicated, the entropy of the maximizing distribution must be at least as great and η must be monotonic decreasing. To show that $\eta \to 0$ as $P \to \infty$ consider a signal which is white noise with a large P. Whatever the perturbing noise, the received signal will be approximately a white noise, if P is sufficiently large, in the sense of having an entropy power approaching $P+N$.

26. The Channel Capacity with a Peak Power Limitation

In some applications the transmitter is limited not by the average power output but by the peak instantaneous power. The problem of calculating the channel capacity is then that of maximizing (by variation of the ensemble of transmitted symbols)

$$H(y) - H(n)$$

subject to the constraint that all the functions $f(t)$ in the ensemble be less than or equal to $\sqrt{S}$, say, for all t. A constraint of this type does not work out as well mathematically as the average power limitation. The most we have obtained for this case is a lower bound valid for all $\frac{S}{N}$, an "asymptotic" upper bound (valid for large $\frac{S}{N}$) and an asymptotic value of C for $\frac{S}{N}$ small.

Theorem 20: The channel capacity C for a band W perturbed by white thermal noise of power N is bounded by

$$C \geq W \log \frac{2}{\pi e^3} \frac{S}{N},$$

where S is the peak allowed transmitter power. For sufficiently large $\frac{S}{N}$

$$C \leq W \log \frac{\frac{2}{\pi e} S + N}{N} (1 + \epsilon)$$

where ϵ is arbitrarily small. As $\frac{S}{N} \to 0$ (and provided the band W starts at 0)

$$C \Big/ W \log \left(1 + \frac{S}{N} \right) \to 1.$$

We wish to maximize the entropy of the received signal. If $\frac{S}{N}$ is large this will occur very nearly when we maximize the entropy of the transmitted ensemble.

The asymptotic upper bound is obtained by relaxing the conditions on the ensemble. Let us suppose that the power is limited to S not at every instant of time, but only at the sample points. The maximum entropy of the transmitted ensemble under these weakened conditions is certainly greater than or equal to that under the original conditions. This altered problem can be solved easily. The maximum entropy occurs if the different samples are independent and have a distribution function which is constant from $-\sqrt{S}$ to $+\sqrt{S}$. The entropy can be calculated as

$$W \log 4S.$$

The received signal will then have an entropy less than

$$W \log(4S + 2\pi e N)(1 + \epsilon)$$

with $\epsilon \to 0$ as $\frac{S}{N} \to \infty$ and the channel capacity is obtained by subtracting the entropy of the white noise, $W \log 2\pi e N$:

$$W \log(4S + 2\pi e N)(1 + \epsilon) - W \log(2\pi e N) = W \log \frac{\frac{2}{\pi e} S + N}{N} (1 + \epsilon).$$

This is the desired upper bound to the channel capacity.

To obtain a lower bound consider the same ensemble of functions. Let these functions be passed through an ideal filter with a triangular transfer characteristic. The gain is to be unity at frequency 0 and decline linearly down to gain 0 at frequency W. We first show that the output functions of the filter have a peak power limitation S at all times (not just the sample points). First we note that a pulse $\frac{\sin 2\pi W t}{2\pi W t}$ going into the filter produces

$$\frac{1}{2} \frac{\sin^2 \pi W t}{(\pi W t)^2}$$

in the output. This function is never negative. The input function (in the general case) can be thought of as the sum of a series of shifted functions

$$a \frac{\sin 2\pi W t}{2\pi W t}$$

where a, the amplitude of the sample, is not greater than $\sqrt{S}$. Hence the output is the sum of shifted functions of the non-negative form above with the same coefficients. These functions being non-negative, the greatest positive value for any t is obtained when all the coefficients a have their maximum positive values, i.e., $\sqrt{S}$. In this case the input function was a constant of amplitude $\sqrt{S}$ and since the filter has unit gain for D.C., the output is the same. Hence the output ensemble has a peak power S.

The entropy of the output ensemble can be calculated from that of the input ensemble by using the theorem dealing with such a situation. The output entropy is equal to the input entropy plus the geometrical mean gain of the filter:

$$\int_0^W \log G^2\, df = \int_0^W \log\left(\frac{W-f}{W}\right)^2 df = -2W.$$

Hence the output entropy is

$$W\log 4S - 2W = W\log\frac{4S}{e^2}$$

and the channel capacity is greater than

$$W\log\frac{2}{\pi e^3}\frac{S}{N}.$$

We now wish to show that, for small $\frac{S}{N}$ (peak signal power over average white noise power), the channel capacity is approximately

$$C = W\log\left(1+\frac{S}{N}\right).$$

More precisely $C \Big/ W\log\left(1+\frac{S}{N}\right) \to 1$ as $\frac{S}{N}\to 0$. Since the average signal power P is less than or equal to the peak S, it follows that for all $\frac{S}{N}$

$$C \le W\log\left(1+\frac{P}{N}\right) \le W\log\left(1+\frac{S}{N}\right).$$

Therefore, if we can find an ensemble of functions such that they correspond to a rate nearly $W\log\left(1+\frac{S}{N}\right)$ and are limited to band W and peak S the result will be proved. Consider the ensemble of functions of the following type. A series of t samples have the same value, either $+\sqrt{S}$ or $-\sqrt{S}$, then the next t samples have the same value, etc. The value for a series is chosen at random, probability $\frac{1}{2}$ for $+\sqrt{S}$ and $\frac{1}{2}$ for $-\sqrt{S}$. If this ensemble be passed through a filter with triangular gain characteristic (unit gain at D.C.), the output is peak limited to $\pm S$. Furthermore the average power is nearly S and can be made to approach this by taking t sufficiently large. The entropy of the sum of this and the thermal noise can be found by applying the theorem on the sum of a noise and a small signal. This theorem will apply if

$$\sqrt{t}\frac{S}{N}$$

is sufficiently small. This can be ensured by taking $\frac{S}{N}$ small enough (after t is chosen). The entropy power will be $S+N$ to as close an approximation as desired, and hence the rate of transmission as near as we wish to

$$W\log\left(\frac{S+N}{N}\right).$$

PART V: THE RATE FOR A CONTINUOUS SOURCE

27. Fidelity Evaluation Functions

In the case of a discrete source of information we were able to determine a definite rate of generating information, namely the entropy of the underlying stochastic process. With a continuous source the situation is considerably more involved. In the first place a continuously variable quantity can assume an infinite number of values and requires, therefore, an infinite number of binary digits for exact specification. This means that to transmit the output of a continuous source with *exact recovery* at the receiving point requires,

in general, a channel of infinite capacity (in bits per second). Since, ordinarily, channels have a certain amount of noise, and therefore a finite capacity, exact transmission is impossible.

This, however, evades the real issue. Practically, we are not interested in exact transmission when we have a continuous source, but only in transmission to within a certain tolerance. The question is, can we assign a definite rate to a continuous source when we require only a certain fidelity of recovery, measured in a suitable way. Of course, as the fidelity requirements are increased the rate will increase. It will be shown that we can, in very general cases, define such a rate, having the property that it is possible, by properly encoding the information, to transmit it over a channel whose capacity is equal to the rate in question, and satisfy the fidelity requirements. A channel of smaller capacity is insufficient.

It is first necessary to give a general mathematical formulation of the idea of fidelity of transmission. Consider the set of messages of a long duration, say T seconds. The source is described by giving the probability density, in the associated space, that the source will select the message in question $P(x)$. A given communication system is described (from the external point of view) by giving the conditional probability $P_x(y)$ that if message x is produced by the source the recovered message at the receiving point will be y. The system as a whole (including source and transmission system) is described by the probability function $P(x,y)$ of having message x and final output y. If this function is known, the complete characteristics of the system from the point of view of fidelity are known. Any evaluation of fidelity must correspond mathematically to an operation applied to $P(x,y)$. This operation must at least have the properties of a simple ordering of systems; i.e., it must be possible to say of two systems represented by $P_1(x,y)$ and $P_2(x,y)$ that, according to our fidelity criterion, either (1) the first has higher fidelity, (2) the second has higher fidelity, or (3) they have equal fidelity. This means that a criterion of fidelity can be represented by a numerically valued function:

$$v\bigl(P(x,y)\bigr)$$

whose argument ranges over possible probability functions $P(x,y)$.

We will now show that under very general and reasonable assumptions the function $v\bigl(P(x,y)\bigr)$ can be written in a seemingly much more specialized form, namely as an average of a function $\rho(x,y)$ over the set of possible values of x and y:

$$v\bigl(P(x,y)\bigr) = \iint P(x,y)\rho(x,y)\,dx\,dy.$$

To obtain this we need only assume (1) that the source and system are ergodic so that a very long sample will be, with probability nearly 1, typical of the ensemble, and (2) that the evaluation is "reasonable" in the sense that it is possible, by observing a typical input and output x_1 and y_1, to form a tentative evaluation on the basis of these samples; and if these samples are increased in duration the tentative evaluation will, with probability 1, approach the exact evaluation based on a full knowledge of $P(x,y)$. Let the tentative evaluation be $\rho(x,y)$. Then the function $\rho(x,y)$ approaches (as $T \to \infty$) a constant for almost all (x,y) which are in the high probability region corresponding to the system:

$$\rho(x,y) \to v\bigl(P(x,y)\bigr)$$

and we may also write

$$\rho(x,y) \to \iint P(x,y)\rho(x,y)\,dx\,dy$$

since

$$\iint P(x,y)\,dx\,dy = 1.$$

This establishes the desired result.

The function $\rho(x,y)$ has the general nature of a "distance" between x and y.[9] It measures how undesirable it is (according to our fidelity criterion) to receive y when x is transmitted. The general result given above can be restated as follows: Any reasonable evaluation can be represented as an average of a distance function over the set of messages and recovered messages x and y weighted according to the probability $P(x,y)$ of getting the pair in question, provided the duration T of the messages be taken sufficiently large.

The following are simple examples of evaluation functions:

[9]It is not a "metric" in the strict sense, however, since in general it does not satisfy either $\rho(x,y) = \rho(y,x)$ or $\rho(x,y) + \rho(y,z) \ge \rho(x,z)$.

1. R.M.S. criterion.

$$v = \overline{\big(x(t) - y(t)\big)^2}.$$

   In this very commonly used measure of fidelity the distance function $\rho(x,y)$ is (apart from a constant factor) the square of the ordinary Euclidean distance between the points x and y in the associated function space.

$$\rho(x,y) = \frac{1}{T}\int_0^T \big[x(t) - y(t)\big]^2\,dt.$$

2. Frequency weighted R.M.S. criterion. More generally one can apply different weights to the different frequency components before using an R.M.S. measure of fidelity. This is equivalent to passing the difference $x(t) - y(t)$ through a shaping filter and then determining the average power in the output. Thus let

$$e(t) = x(t) - y(t)$$

   and

$$f(t) = \int_{-\infty}^{\infty} e(\tau)k(t-\tau)\,d\tau$$

   then

$$\rho(x,y) = \frac{1}{T}\int_0^T f(t)^2\,dt.$$

3. Absolute error criterion.

$$\rho(x,y) = \frac{1}{T}\int_0^T \big|x(t) - y(t)\big|\,dt.$$

4. The structure of the ear and brain determine implicitly an evaluation, or rather a number of evaluations, appropriate in the case of speech or music transmission. There is, for example, an “intelligibility” criterion in which $\rho(x,y)$ is equal to the relative frequency of incorrectly interpreted words when message $x(t)$ is received as $y(t)$. Although we cannot give an explicit representation of $\rho(x,y)$ in these cases it could, in principle, be determined by sufficient experimentation. Some of its properties follow from well-known experimental results in hearing, e.g., the ear is relatively insensitive to phase and the sensitivity to amplitude and frequency is roughly logarithmic.

5. The discrete case can be considered as a specialization in which we have tacitly assumed an evaluation based on the frequency of errors. The function $\rho(x,y)$ is then defined as the number of symbols in the sequence y differing from the corresponding symbols in x divided by the total number of symbols in x.

28. The Rate for a Source Relative to a Fidelity Evaluation

We are now in a position to define a rate of generating information for a continuous source. We are given $P(x)$ for the source and an evaluation v determined by a distance function $\rho(x,y)$ which will be assumed continuous in both x and y. With a particular system $P(x,y)$ the quality is measured by

$$v = \iint \rho(x,y)P(x,y)\,dx\,dy.$$

Furthermore the rate of flow of binary digits corresponding to $P(x,y)$ is

$$R = \iint P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\,dx\,dy.$$

We define the rate R_1 of generating information for a given quality v_1 of reproduction to be the minimum of R when we keep v fixed at v_1 and vary $P_x(y)$. That is:

$$R_1 = \operatorname*{Min}_{P_x(y)} \iint P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\,dx\,dy$$

subject to the constraint:

$$v_1 = \iint P(x,y)\rho(x,y)\,dx\,dy.$$

This means that we consider, in effect, all the communication systems that might be used and that transmit with the required fidelity. The rate of transmission in bits per second is calculated for each one and we choose that having the least rate. This latter rate is the rate we assign the source for the fidelity in question.

The justification of this definition lies in the following result:

Theorem 21: If a source has a rate R_1 for a valuation v_1 it is possible to encode the output of the source and transmit it over a channel of capacity C with fidelity as near v_1 as desired provided $R_1 \le C$. This is not possible if $R_1 > C$.

The last statement in the theorem follows immediately from the definition of R_1 and previous results. If it were not true we could transmit more than C bits per second over a channel of capacity C. The first part of the theorem is proved by a method analogous to that used for Theorem 11. We may, in the first place, divide the (x,y) space into a large number of small cells and represent the situation as a discrete case. This will not change the evaluation function by more than an arbitrarily small amount (when the cells are very small) because of the continuity assumed for $\rho(x,y)$. Suppose that $P_1(x,y)$ is the particular system which minimizes the rate and gives R_1. We choose from the high probability y's a set at random containing

$$2^{(R_1+\epsilon)T}$$

members where $\epsilon \to 0$ as $T \to \infty$. With large T each chosen point will be connected by a high probability line (as in Fig. 10) to a set of x's. A calculation similar to that used in proving Theorem 11 shows that with large T almost all x's are covered by the fans from the chosen y points for almost all choices of the y's. The communication system to be used operates as follows: The selected points are assigned binary numbers. When a message x is originated it will (with probability approaching 1 as $T \to \infty$) lie within at least one of the fans. The corresponding binary number is transmitted (or one of them chosen arbitrarily if there are several) over the channel by suitable coding means to give a small probability of error. Since $R_1 \le C$ this is possible. At the receiving point the corresponding y is reconstructed and used as the recovered message.

The evaluation v_1' for this system can be made arbitrarily close to v_1 by taking T sufficiently large. This is due to the fact that for each long sample of message $x(t)$ and recovered message $y(t)$ the evaluation approaches v_1 (with probability 1).

It is interesting to note that, in this system, the noise in the recovered message is actually produced by a kind of general quantizing at the transmitter and not produced by the noise in the channel. It is more or less analogous to the quantizing noise in PCM.

29. The Calculation of Rates

The definition of the rate is similar in many respects to the definition of channel capacity. In the former

$$R = \underset{P_x(y)}{\text{Min}} \iint P(x,y) \log \frac{P(x,y)}{P(x)P(y)}\,dx\,dy$$

with $P(x)$ and $v_1 = \iint P(x,y)\rho(x,y)\,dx\,dy$ fixed. In the latter

$$C = \underset{P(x)}{\text{Max}} \iint P(x,y) \log \frac{P(x,y)}{P(x)P(y)}\,dx\,dy$$

with $P_x(y)$ fixed and possibly one or more other constraints (e.g., an average power limitation) of the form $K = \iint P(x,y)\lambda(x,y)\,dx\,dy$.

A partial solution of the general maximizing problem for determining the rate of a source can be given. Using Lagrange's method we consider

$$\iint \left[P(x,y) \log \frac{P(x,y)}{P(x)P(y)} + \mu P(x,y)\rho(x,y) + \nu(x)P(x,y) \right] dx\,dy.$$

The variational equation (when we take the first variation on $P(x,y)$) leads to

$$P_y(x) = B(x)e^{-\lambda\rho(x,y)}$$

where λ is determined to give the required fidelity and $B(x)$ is chosen to satisfy

$$\int B(x)e^{-\lambda\rho(x,y)}\,dx = 1.$$

This shows that, with best encoding, the conditional probability of a certain cause for various received y, $P_y(x)$ will decline exponentially with the distance function $\rho(x,y)$ between the x and y in question.

In the special case where the distance function $\rho(x,y)$ depends only on the (vector) difference between x and y,

$$\rho(x,y) = \rho(x-y)$$

we have

$$\int B(x)e^{-\lambda\rho(x-y)}\,dx = 1.$$

Hence $B(x)$ is constant, say α, and

$$P_y(x) = \alpha e^{-\lambda\rho(x-y)}.$$

Unfortunately these formal solutions are difficult to evaluate in particular cases and seem to be of little value. In fact, the actual calculation of rates has been carried out in only a few very simple cases.

If the distance function $\rho(x,y)$ is the mean square discrepancy between x and y and the message ensemble is white noise, the rate can be determined. In that case we have

$$R = \text{Min}\big[H(x) - H_y(x)\big] = H(x) - \text{Max}\,H_y(x)$$

with $N = \overline{(x-y)^2}$. But the $\text{Max}\,H_y(x)$ occurs when $y-x$ is a white noise, and is equal to $W_1 \log 2\pi e N$ where W_1 is the bandwidth of the message ensemble. Therefore

$$\begin{aligned} R &= W_1 \log 2\pi e Q - W_1 \log 2\pi e N \\ &= W_1 \log \frac{Q}{N} \end{aligned}$$

where Q is the average message power. This proves the following:

Theorem 22: The rate for a white noise source of power Q and band W_1 relative to an R.M.S. measure of fidelity is

$$R = W_1 \log \frac{Q}{N}$$

where N is the allowed mean square error between original and recovered messages.

More generally with any message source we can obtain inequalities bounding the rate relative to a mean square error criterion.

Theorem 23: The rate for any source of band W_1 is bounded by

$$W_1 \log \frac{Q_1}{N} \le R \le W_1 \log \frac{Q}{N}$$

where Q is the average power of the source, Q_1 its entropy power and N the allowed mean square error.

The lower bound follows from the fact that the $\text{Max}\,H_y(x)$ for a given $\overline{(x-y)^2} = N$ occurs in the white noise case. The upper bound results if we place points (used in the proof of Theorem 21) not in the best way but at random in a sphere of radius $\sqrt{Q-N}$.

ACKNOWLEDGMENTS

The writer is indebted to his colleagues at the Laboratories, particularly to Dr. H. W. Bode, Dr. J. R. Pierce, Dr. B. McMillan, and Dr. B. M. Oliver for many helpful suggestions and criticisms during the course of this work. Credit should also be given to Professor N. Wiener, whose elegant solution of the problems of filtering and prediction of stationary ensembles has considerably influenced the writer's thinking in this field.

APPENDIX 5

Let S_1 be any measurable subset of the g ensemble, and S_2 the subset of the f ensemble which gives S_1 under the operation T. Then

$$S_1 = TS_2.$$

Let H^λ be the operator which shifts all functions in a set by the time λ. Then

$$H^\lambda S_1 = H^\lambda TS_2 = TH^\lambda S_2$$

since T is invariant and therefore commutes with H^λ. Hence if $m[S]$ is the probability measure of the set S

$$\begin{aligned} m[H^\lambda S_1] &= m[TH^\lambda S_2] = m[H^\lambda S_2] \\ &= m[S_2] = m[S_1] \end{aligned}$$

where the second equality is by definition of measure in the g space, the third since the f ensemble is stationary, and the last by definition of g measure again.

To prove that the ergodic property is preserved under invariant operations, let S_1 be a subset of the g ensemble which is invariant under H^λ, and let S_2 be the set of all functions f which transform into S_1. Then

$$H^\lambda S_1 = H^\lambda TS_2 = TH^\lambda S_2 = S_1$$

so that $H^\lambda S_2$ is included in S_2 for all λ. Now, since

$$m[H^\lambda S_2] = m[S_1]$$

this implies

$$H^\lambda S_2 = S_2$$

for all λ with $m[S_2] \neq 0, 1$. This contradiction shows that S_1 does not exist.

APPENDIX 6

The upper bound, $\overline{N}_3 \leq N_1 + N_2$, is due to the fact that the maximum possible entropy for a power $N_1 + N_2$ occurs when we have a white noise of this power. In this case the entropy power is $N_1 + N_2$.

To obtain the lower bound, suppose we have two distributions in n dimensions $p(x_i)$ and $q(x_i)$ with entropy powers $\overline{N}_1$ and $\overline{N}_2$. What form should p and q have to minimize the entropy power $\overline{N}_3$ of their convolution $r(x_i)$:

$$r(x_i) = \int p(y_i) q(x_i - y_i)\, dy_i.$$

The entropy H_3 of r is given by

$$H_3 = -\int r(x_i) \log r(x_i)\, dx_i.$$

We wish to minimize this subject to the constraints

$$\begin{aligned} H_1 &= -\int p(x_i) \log p(x_i)\, dx_i \\ H_2 &= -\int q(x_i) \log q(x_i)\, dx_i. \end{aligned}$$

We consider then

$$U = -\int \left[r(x)\log r(x) + \lambda p(x)\log p(x) + \mu q(x)\log q(x)\right] dx$$
$$\delta U = -\int \left[[1+\log r(x)]\delta r(x) + \lambda[1+\log p(x)]\delta p(x) + \mu[1+\log q(x)]\delta q(x)\right] dx.$$

If $p(x)$ is varied at a particular argument $x_i = s_i$, the variation in $r(x)$ is

$$\delta r(x) = q(x_i - s_i)$$

and

$$\delta U = -\int q(x_i - s_i)\log r(x_i)\,dx_i - \lambda \log p(s_i) = 0$$

and similarly when q is varied. Hence the conditions for a minimum are

$$\int q(x_i - s_i)\log r(x_i)\,dx_i = -\lambda \log p(s_i)$$
$$\int p(x_i - s_i)\log r(x_i)\,dx_i = -\mu \log q(s_i).$$

If we multiply the first by $p(s_i)$ and the second by $q(s_i)$ and integrate with respect to s_i we obtain

$$H_3 = -\lambda H_1$$
$$H_3 = -\mu H_2$$

or solving for λ and μ and replacing in the equations

$$H_1\int q(x_i - s_i)\log r(x_i)\,dx_i = -H_3 \log p(s_i)$$
$$H_2\int p(x_i - s_i)\log r(x_i)\,dx_i = -H_3 \log q(s_i).$$

Now suppose $p(x_i)$ and $q(x_i)$ are normal

$$p(x_i) = \frac{|A_{ij}|^{n/2}}{(2\pi)^{n/2}}\exp -\tfrac{1}{2}\sum A_{ij}x_ix_j$$
$$q(x_i) = \frac{|B_{ij}|^{n/2}}{(2\pi)^{n/2}}\exp -\tfrac{1}{2}\sum B_{ij}x_ix_j.$$

Then $r(x_i)$ will also be normal with quadratic form C_{ij}. If the inverses of these forms are a_{ij}, b_{ij}, c_{ij} then

$$c_{ij} = a_{ij} + b_{ij}.$$

We wish to show that these functions satisfy the minimizing conditions if and only if $a_{ij} = Kb_{ij}$ and thus give the minimum H_3 under the constraints. First we have

$$\log r(x_i) = \frac{n}{2}\log\frac{1}{2\pi}|C_{ij}| - \tfrac{1}{2}\sum C_{ij}x_ix_j$$
$$\int q(x_i - s_i)\log r(x_i)\,dx_i = \frac{n}{2}\log\frac{1}{2\pi}|C_{ij}| - \tfrac{1}{2}\sum C_{ij}s_is_j - \tfrac{1}{2}\sum C_{ij}b_{ij}.$$

This should equal

$$\frac{H_3}{H_1}\left[\frac{n}{2}\log\frac{1}{2\pi}|A_{ij}| - \tfrac{1}{2}\sum A_{ij}s_is_j\right]$$

which requires $A_{ij} = \dfrac{H_1}{H_3}C_{ij}$. In this case $A_{ij} = \dfrac{H_1}{H_2}B_{ij}$ and both equations reduce to identities.

APPENDIX 7

The following will indicate a more general and more rigorous approach to the central definitions of communication theory. Consider a probability measure space whose elements are ordered pairs (x,y). The variables x, y are to be identified as the possible transmitted and received signals of some long duration T. Let us call the set of all points whose x belongs to a subset S_1 of x points the strip over S_1, and similarly the set whose y belong to S_2 the strip over S_2. We divide x and y into a collection of non-overlapping measurable subsets X_i and Y_i approximate to the rate of transmission R by

$$R_1 = \frac{1}{T}\sum_i P(X_i,Y_i)\log\frac{P(X_i,Y_i)}{P(X_i)P(Y_i)}$$

where

$P(X_i)$ is the probability measure of the strip over X_i
$P(Y_i)$ is the probability measure of the strip over Y_i
$P(X_i,Y_i)$ is the probability measure of the intersection of the strips.

A further subdivision can never decrease R_1. For let X_1 be divided into $X_1 = X_1' + X_1''$ and let

$$\begin{aligned} P(Y_1) &= a & P(X_1) &= b+c \\ P(X_1') &= b & P(X_1',Y_1) &= d \\ P(X_1'') &= c & P(X_1'',Y_1) &= e \\ & & P(X_1,Y_1) &= d+e. \end{aligned}$$

Then in the sum we have replaced (for the X_1, Y_1 intersection)

$$(d+e)\log\frac{d+e}{a(b+c)} \quad \text{by} \quad d\log\frac{d}{ab} + e\log\frac{e}{ac}.$$

It is easily shown that with the limitation we have on b, c, d, e,

$$\left[\frac{d+e}{b+c}\right]^{d+e} \le \frac{d^d e^e}{b^d c^e}$$

and consequently the sum is increased. Thus the various possible subdivisions form a directed set, with R monotonic increasing with refinement of the subdivision. We may define R unambiguously as the least upper bound for R_1 and write it

$$R = \frac{1}{T}\iint P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\,dx\,dy.$$

This integral, understood in the above sense, includes both the continuous and discrete cases and of course many others which cannot be represented in either form. It is trivial in this formulation that if x and u are in one-to-one correspondence, the rate from u to y is equal to that from x to y. If v is any function of y (not necessarily with an inverse) then the rate from x to y is greater than or equal to that from x to v since, in the calculation of the approximations, the subdivisions of y are essentially a finer subdivision of those for v. More generally if y and v are related not functionally but statistically, i.e., we have a probability measure space (y,v), then $R(x,v) \le R(x,y)$. This means that any operation applied to the received signal, even though it involves statistical elements, does not increase R.

Another notion which should be defined precisely in an abstract formulation of the theory is that of "dimension rate," that is the average number of dimensions required per second to specify a member of an ensemble. In the band limited case $2W$ numbers per second are sufficient. A general definition can be framed as follows. Let $f_\alpha(t)$ be an ensemble of functions and let $\rho_T[f_\alpha(t), f_\beta(t)]$ be a metric measuring

the "distance" from f_α to f_β over the time T (for example the R.M.S. discrepancy over this interval.) Let $N(\epsilon, \delta, T)$ be the least number of elements f which can be chosen such that all elements of the ensemble apart from a set of measure δ are within the distance ϵ of at least one of those chosen. Thus we are covering the space to within ϵ apart from a set of small measure δ. We define the dimension rate λ for the ensemble by the triple limit

$$\lambda = \operatorname*{Lim}_{\delta \to 0} \operatorname*{Lim}_{\epsilon \to 0} \operatorname*{Lim}_{T \to \infty} \frac{\log N(\epsilon, \delta, T)}{T \log \epsilon}.$$

This is a generalization of the measure type definitions of dimension in topology, and agrees with the intuitive dimension rate for simple ensembles where the desired result is obvious.

USA TERRITORIES ONLY

CargoBit is a product of Codex Grandeur LLC, designed for long term mass storage of lossless digital data. More than simple data volume compression, CargoBit provides time compression for customers that are time poor, and can gain performance advantages in terms of centuries into the future. CargoBit compression ratio is 2^20 to 1 or 1,048,576 binary digits are available from every single physical binary digit used.

In the compressed form, data is used with existing hardware, requiring no modification in any way, shape, or form. There is an uncompressed overhead of hash, and meta data to support the compressed file's. The data is NOT usable in compressed format. It will be 100 percent lossless when CargoBit does regenerate the file's, returning their original bit pattern.

The CargoBit can secure the file to specific CargoBit unit's within the customer ownership, or simply secure the file to all CargoBit unit's within the customer ownership. Simply re-run the file through the CargoBit to change file security.

Any copies or intercept over internet transmission, or anywhere after compression, will not provide access to the data, not only because it is a proprietary format, also because 1,048,575 bits are missing for every 1 bit stolen. The compressed files can only be returned with CargoBit unit's, that the customer authorizes. Many different CargoBit units can be on site, or anywhere else inside of United States of America territories's.

CargoBit is compute intensive with very large volumes of the data being compressed, overhead file preparation, and simple transfer of data-in with data-out of CargoBit before compression, and after regeneration does require very expensive raw bit rates to the Data Center.

A **$5.00** 4.7GB M-Disc lasting up to 1,000 year's will store 410 LTO-8 tape cartridges of 12TeraByte raw data costing **$41,000.00** and more. **15 minutes** write-read verify compared to the LTO-8 tape of 9 hours each totaling **3,690 hours** or **153.75 days.**

A single **$100.00** LTO-8 tape cartridge of 12TeraByte raw data will store 12.582,912 ExaByte's with CargoBit coding, eliminating the need for 1,048,576 LTO-8 tape cartridges costing **$104,857,600.00** total. **9 hours** write-read verify compared to **9,437,184 hours** total or **393,276 days** or **1,076.+ years.**

CargoBit compared to Cloud Storage

A single **$5.00** 4.7GB M-Disc lasting up to 1,000 year's will store 410 LTO-8 tape cartridges of 12TeraByte raw data costing **$41,000.00** and more. Using Backblaze B2 Cloud Storage web page as a comparison reference, **1,000 year's** 4.7GB M-Disc **$5.00** CargoBit, compared to every **10 year's** Cloud Storage cost of **$3,406,802.70** B2 Cloud Storage, **$12,284,649.72** Microsoft Azure, **$13,684,810.80** Google Cloud, and **$14,320,091.34** Amazon S3.

Estimate Storage Pricing For Server/ NAS Backup

Using Backblaze B2 Cloud Storage, you will be getting super inexpensive storage that is reliable and instantly accesible.

Initial Upload:	4,928,307	GB
Data over time		
Monthly Upload:	12,000	GB
Monthly Delete:	5	GB
Monthly Download:	12,000	GB
Period of Time:	120	Months

$ USD

STORAGE COSTS	
Storage Cost for Initial Month:	$24,641.54
Data Added Each Month:	$60.00
Data Deleted Each Month:	- $0.03
Net Data:	$59.98
DOWNLOAD COSTS	
Monthly Download Cost:	$120.00
TOTAL	
Total Cost for 120 Months	**$3,406,802.70**

Amazon S3	Microsoft Azure	Google Cloud
$14,320,091.34	$12,284,649.72	$13,684,810.80
+420%	+360%	+401%

* Figures are not exact and do not include the following: Free first 10 GB of storage, free 1 GB of daily downloads, or $.004/10,000 class B Transactions and $.004/1,000 Class C Transactions. All prices subject to applicable local taxes.

CargoBit used in Hospital and Health Clinic

The affordable smart memory card with 1 Megabyte of erasable memory is not enough storage for extremely high detailed medical images, and numerous CAT, MRI, NMR, and X-Ray necessary for emergency rooms, field hospitals, medical clinics all over the planet. A detailed trace history with doctor interview video's, detailed allergic history, blood work test, eliminating the need of the doctor to recreate known past data, when time is of the essence.

While field hospital clinic's will not be able to afford, or even support CargoBit systems, they can afford a simple 1 Megabyte memory card reader, and satellite communications for a link to a Medical CargoBit system somewhere else on planet earth. The patient's 1 Megabyte of memory is sent over the satellite to be regenerated by the remote CargoBit system multiplying the 1 Megabyte of data into 1.048,576 TeraByte's of detailed images, graphs, video's, and medical charts. The entire 1.048,576 TeraByte's can be sent back over satellite, or intelligent remote human selection of graphics displayed by the doctor will speed up critical information delivery.

The patient's smart memory card can then be updated remotely via satellite link, and the remote CargoBit will regenerate a new 1 Megabyte image to update the smart card. Hospital's in the USA can use a similar approach with fiber optical lines, or high speed internet to replace the satellite communication.

Long term Hospital patient record storage can also be reduced, as seen in the Cloud Storage cost jumping into the millions of dollars every 10 year's.

ATMEL AT24C1024 WHITE PVC CARD
Atmel AT24C1024 White PVC Card
SKU: C-A090-PEX
Brand: Atmel
Type: Contact
Material: PVC
Memory Size: 1 MB
For a price please call 1-800-810-4959

HPE

CargoBit is custom configured for each, and every customer site. HPE hardware, and software is used for the Data Center overhead support system. Beginning with the HPE ProLiant DL325 Gen10 Plus Server with AMD EPYC cores, a both physically small, and economically small configuration that is directly compatible with all CargoBit, and FTLDCI systems is possible. Custom hardware is installed inside of the Rack cabinet for the Hatch Secure Processor. The cabinet is considered one station unit, and is physically bounded in 3D location by the sanctioned Hatch Secure Processor, plus or minus a few hundred feet.

The HPE ProLiant DL385 Gen10 Plus Server is used, with two EPYC 64 cores, 8.0 TB RDIMM, communication boards selected to be compatible with existing Data Center data links, SSD drive array, M-DISC drive array as required, MSL3040 Tape Library with 3 Fibre Channel LTO-8 Tape Drives, 8K status monitor, local UPS, and 42U Rack. Custom hardware is installed inside of the Rack cabinet for the Hatch Secure Processor. The cabinet is considered one station unit, and is physically bounded in 3D location by the sanctioned Hatch Secure Processor, plus or minus a few hundred feet.

Multiple optional HPE Apollo 2000 Gen10 Plus Systems are configured, and added when needed for near real-time performance, Data Center interfacing, large scale SSD drive array's with CargoBit 2^20 compression, RAM Disk array's are CargoBit 2^20 compressed that reduce write wear on SSD drive's, with custom customer requirement's. Custom hardware is installed inside of the Rack cabinets for the Hatch Secure Processor. The cabinets are considered one station unit, and are physically bounded in 3D location by the sanctioned Hatch Secure Processor, plus or minus a few hundred feet.

Most hardware / software configurations can function as a CargoBit, or FTLDCI depending on specific hardware, and specific site licensing requirements. Performance is directly linked to RAM, Disk, I/O throughput, and AMD EPYC core availability.

Hatch Secure Processor

The Hatch Secure Processor, called so because it is similar to a submarine hatch that is secured for water tight integrity, the Hatch Secure Processor can not be reverse engineered after site secured first initiation run.

The Hatch Secure Processor is where the compression system resides, and physically secure's the computer to United States of America territory only. The Hatch Secure Processor is currently located inside all of the CargoBit racks used. Also the electrical cabinet racks provide a hardened location system that does not allow movement of any kind after installation, plus or minus a considerable amount, to allow for natural events. The offset can be reset after a possible natural event. If the black boxes are stolen, or tampered with in any way, shape, or form, the programs will cease to exist. The existing HPE hardware is not modified in any way other than programming.

Motion sensors are included with an alarm providing plenty of warning to the customer, and trigger sensitivity can be adjusted for normal access, maintenance, and different environments, including mobile. The motion sensors will only allow a specified distance from original position.

A non-published dynamic changing instruction set, combined with a non-published dynamic changing computer architecture makes it possible to mass produce the trade secret that is CargoBit. The physical hardware is available for reverse engineering, however the number of bits per word, and bit position is always changing from 8 to 800+, with scheduled sequence instruction disorder. Always changing, and never pattern sensitive to applied data, the Hatch Secure Processor is also designed to be evasive.

CargoBit is only installed in the United States of America territories. United States of America customer's only. Compressed files can always be sent anywhere, however they can only be regenerated with the use of a CargoBit physically located in the United States of America territories.

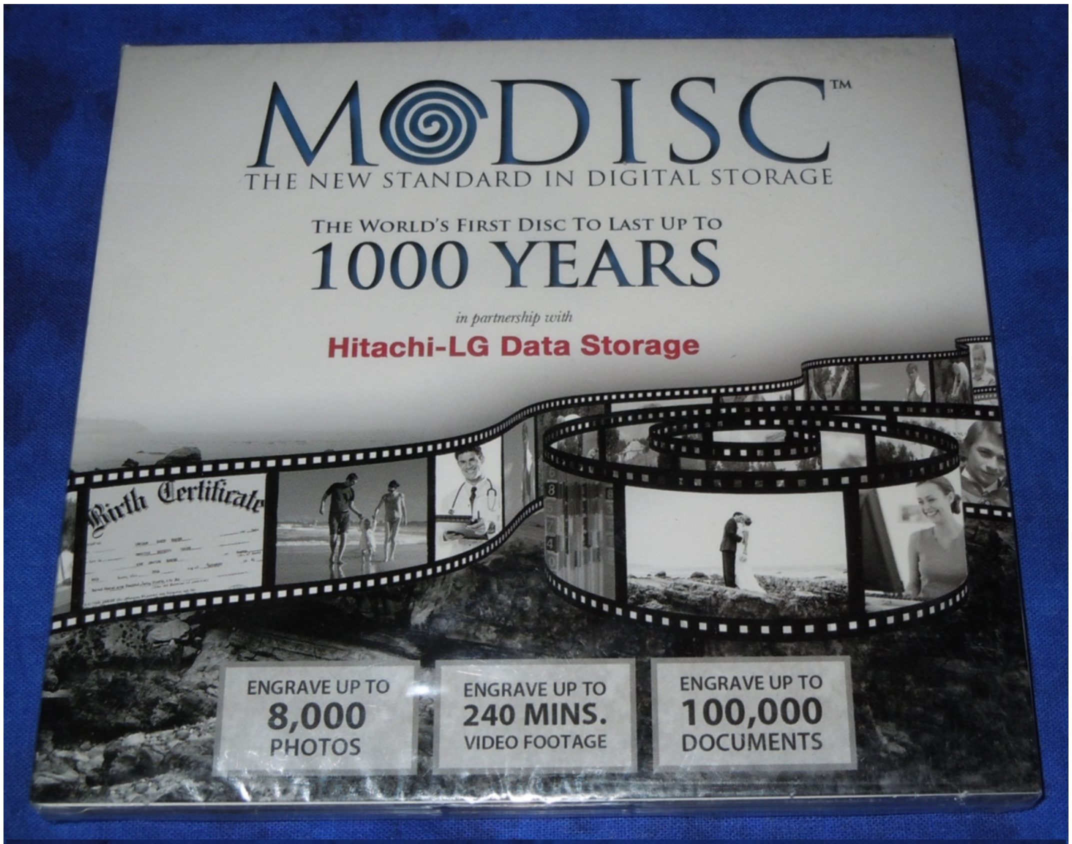
FTLDCI
www.ftldci.com
Faster Than Light Data Center Interconnect

FTLDCI is a Codex Grandeur LLC product based on the same architecture as the CargoBit. Where CargoBit concentrates on High Performance Computing mass storage, and retrieval, the FTLDCI concentrates on Data Communications, specifically for the Data Center Interconnect both internal communications, and Data Center to Data Center.

400Gb Ethernet long haul, and 800Gb Ethernet < long haul, are the practical affordable limits of Fiber Optics data communications at this time. Both 400Gb Ethernet, and 800Gb Ethernet are very expensive, with submarine cables possibly slower, and more expensive. Many submarine cables are going dark, because of decades of usage. FTLDCI will enable the surviving Fiber Optic links to exceed full original cable capacity many, many times over in most cases, with no change in cable. Only that FTLDCI units are installed at both ends.

A single 1Gb Ethernet Fiber Optic link available to most people in large cities now, can carry over 1,310 800Gb Ethernet Fiber Optic data when FTLDCI units are installed at both ends. The same is true inside of a Data Center, allowing one single 1Gb Ethernet Fiber Optic link to carry 1,310 800Gb Ethernet Fiber Optic data when FTLDCI units are installed at both ends.

Please see the FTLDCI pdf at www.ftldci.com.


M-DISC™
THE NEW STANDARD IN DIGITAL STORAGE
THE WORLD'S FIRST DISC TO LAST UP TO
1000 YEARS
in partnership with
Hitachi-LG Data Storage
Birth Certificate
ENGRAVE UP TO 8,000 PHOTOS
ENGRAVE UP TO 240 MINS. VIDEO FOOTAGE
ENGRAVE UP TO 100,000 DOCUMENTS


M•DISC™ The New Standard In Digital Storage For...
Memories • Archival • Documents • Digital • Photos • Records • Medical
Legal • Financial • Music • Video • Personal • Business
M•DISC™
GOLD DVD
ENGRAVE WITH LG DRIVES
Readable on Current DVD Drives Yes Yes
4.7 GB of Storage Yes Yes
Permanent Storage Yes No
Data Carved in Stone Yes No
Endures Exposure to Heat, Humidity & Light Yes No
Eliminates Data Recording in Unstable Dye Yes No
Eliminates Unnecessary Gold Layer Yes No
For more information visit us online
www.m-disc.com
Made in Czech Republic
* SAC35435501 *

= 410X

4.7 GB M-Disc **$5.00** with CargoBit coding is 4.9263072 PetaByte.
10-15 minutes write-read verify.

410 LTO-8, 12TB RAW tape cartridges **$100.00** each.
$41,000.00 total.
9 hours write each = **3,690 hours** total or **153.75 days**.

=8,738X

100 GB M-Disc **$15.00** with CargoBit coding is 104.8576 PetaByte.
192 minutes write-read verify.

8,738 LTO-8, 12TB RAW tape cartridges **$100.00** each.
$873,800.00 total.
9 hours write each = **78,642 hours** total or **3,276.75 days**.

=1,048,576X

Single LTO-8, 12TB RAW **$100.00** with CargoBit coding is 12.582,912 ExaByte.
9 hours write-read verify.

1,048,576 LTO-8 tape cartridges **$100.00** each.
$104,857,600.00 total.
9,437,184 hours total or **393,276 days** or **1,076.+ years.**

Prefix	Symbol	1000^m	10^n	Decimal	Short scale	Long scale	Since[n 1]
yotta	Y	1000^8	10^{24}	1 000 000 000 000 000 000 000 000	Septillion	Quadrillion	1991
zetta	Z	1000^7	10^{21}	1 000 000 000 000 000 000 000	Sextillion	Trilliard	1991
exa	E	1000^6	10^{18}	1 000 000 000 000 000 000	Quintillion	Trillion	1975
peta	P	1000^5	10^{15}	1 000 000 000 000 000	Quadrillion	Billiard	1975
tera	T	1000^4	10^{12}	1 000 000 000 000	Trillion	Billion	1960
giga	G	1000^3	10^9	1 000 000 000	Billion	Milliard	1960
mega	M	1000^2	10^6	1 000 000	Million		1960
kilo	k	1000^1	10^3	1 000	Thousand		1795
hecto	h	$1000^{2/3}$	10^2	100	Hundred		1795
deca	da	$1000^{1/3}$	10^1	10	Ten		1795
		1000^0	10^0	1	One		–
deci	d	$1000^{-1/3}$	10^{-1}	0.1	Tenth		1795
centi	c	$1000^{-2/3}$	10^{-2}	0.01	Hundredth		1795
milli	m	1000^{-1}	10^{-3}	0.001	Thousandth		1795
micro	µ	1000^{-2}	10^{-6}	0.000 001	Millionth		1960
nano	n	1000^{-3}	10^{-9}	0.000 000 001	Billionth	Milliardth	1960
pico	p	1000^{-4}	10^{-12}	0.000 000 000 001	Trillionth	Billionth	1960
femto	f	1000^{-5}	10^{-15}	0.000 000 000 000 001	Quadrillionth	Billiardth	1964
atto	a	1000^{-6}	10^{-18}	0.000 000 000 000 000 001	Quintillionth	Trillionth	1964
zepto	z	1000^{-7}	10^{-21}	0.000 000 000 000 000 000 001	Sextillionth	Trilliardth	1991
yocto	y	1000^{-8}	10^{-24}	0.000 000 000 000 000 000 000 001	Septillionth	Quadrillionth	1991

Codex Grandeur LLC
James Lee McDaniel
Actual

www.codexgrandeur.com
mcdaniel@codexgrandeur.com

www.cargobit.com
mcdaniel@cargobit.com

www.ftldci.com
mcdaniel@ftldci.com