Industrial Mathematics at Bell Labs: Past and Present

Debasis Mitra
Vice President, Mathematical and Algorithmic Sciences
Bell Labs, Lucent Technologies

Lucent Technologies
Bell Labs Innovations

Today's Bell Labs

Major Research Areas

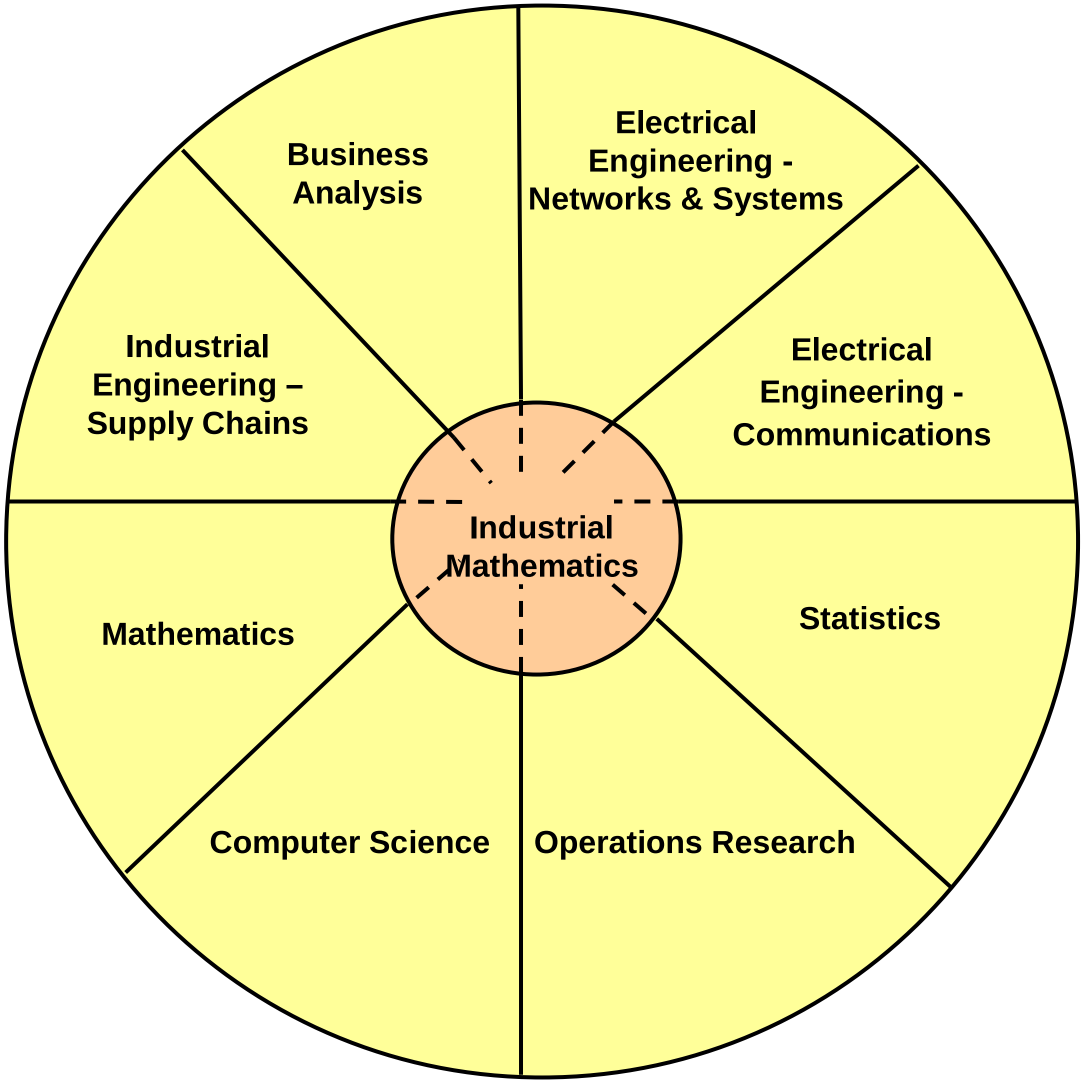
- Mathematical and Algorithmic Sciences
- Network Planning, Performance and Economic Analysis
- Software and Computer Science
- Networking and Network Management
- Wireless and Broadband
- Optical Transport Networks
- Physical Science and Nanotechnology
- Government Communications
- Security

Major Recognition

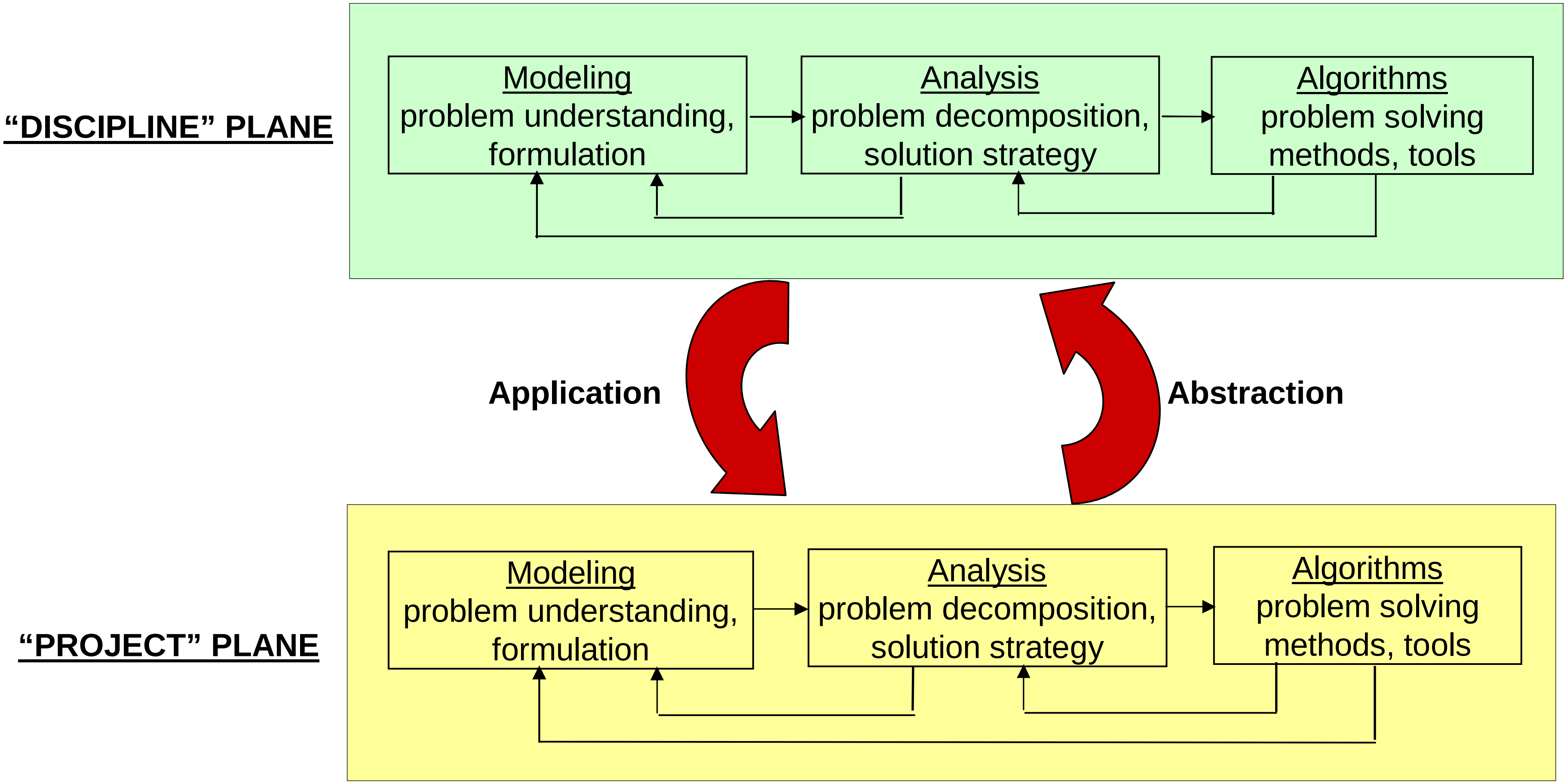
- 11 Nobel Laureates
- 9 National Medals of Science
- 8 U.S. Medals of Technology
- 19 IEEE Medals of Honor

Vision: To be the undisputed industry leader in innovative R&D, giving Lucent a decisive advantage in the marketplace

The Disciplines in Bell Labs' Industrial Mathematics



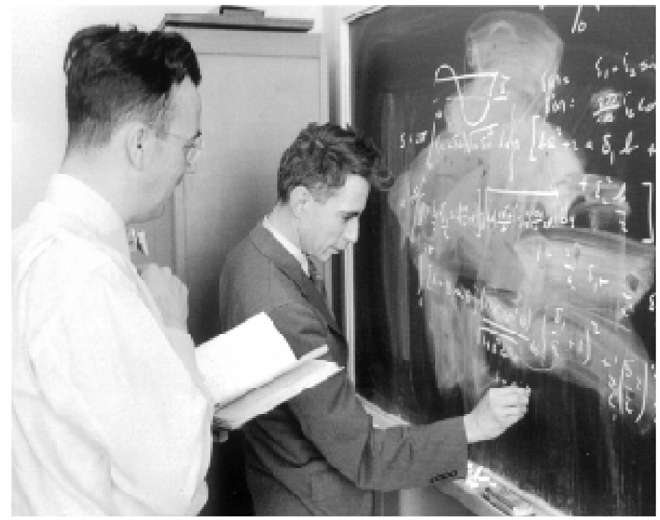
Industrial Mathematics: Modes of Operation



Luminary Bell Labs Mathematicians

Bell Labs Career:
1916 - 1956

Bell Labs Career:
1945 - 1985



Bell Labs Career:
1937 - 1972

<u>Thornton Fry</u>

ü Queuing and Inventory Theories

ü Applied Probability

<u>John Tukey</u>

ü Statistics

ü Data Analysis

<u>Claude Shannon</u>

ü Information Theory

ü Communication Theory

Thornton Fry: ***Queuing Theory, Applied Probability***

- Joined Western Electric in 1916 as a member of the Engineering Research Department
- Earned a doctorate in 1920 from University of Wisconsin based on work done at Western Electric on usefulness of nonconvergent asymptotic series
- In 1925 parts of Western Electric and AT&T were combined to form Bell Telephone Laboratories. Fry headed small separate mathematics department
- Published “Probability and Its Engineering Uses” in 1928. Original treatment of blocking, queuing, congestion, switching machine models and quality control in manufacturing. AT&T colleagues acknowledged: E.C. Molina, Harry Nyquist, W.A. Shewhart. Fry taught MIT course based on book
- Helped to establish *Mathematical Reviews* in 1939
- Wrote paper on “Industrial Mathematics” in 1940

Thornton Fry: ***Queuing Theory, Applied Probability***

- Contributed to industrial research on war problems during WWII

  - Co-organizer, Director of Office of Scientific Research & Development

  - Key role in development of M-9 Gun Director for anti-aircraft fire control

  - Shepherded proposal by 2 young Bell Telephone Labs engineers (D.B.Parkinson and C.A.Lovell)

  - M-9 greatly helped defense of Britain from V1 rockets

- Received Presidential Certificate of Merit for his war efforts in 1948
- Served as Director of Switching Research after WWII; retired 1956

M-9 Gun Director demonstrated at Murray Hill, 1945

The tracking unit with its two operators is shown behind the gun crew.

Fry’s Legacy: **Optimizing Today’s Networks**

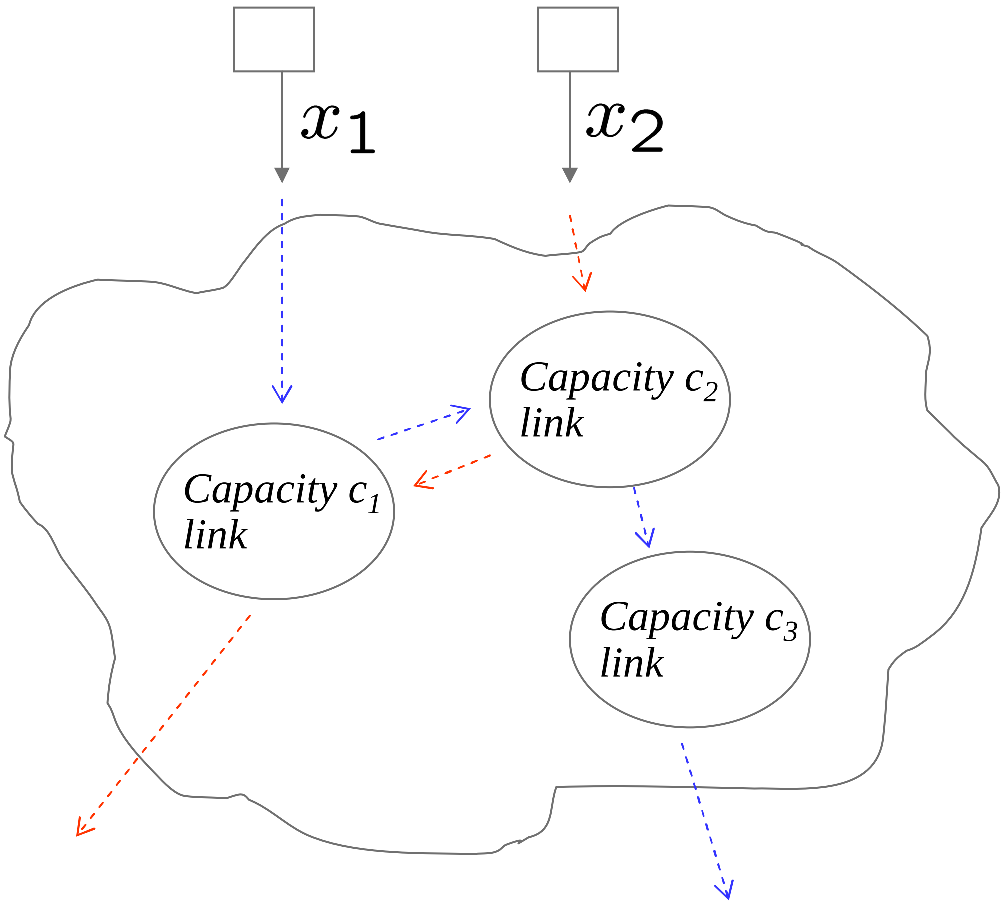
Fry’s Work Then …

- Original treatments of blocking, queuing, congestion, switching machine behavior and quality control in manufacturing

Connected to Breakthroughs Today

- Bell Labs’ Sasha Stolyar’s “Greedy Primal-Dual” method for optimal congestion control in queuing networks
- Theory grounded in deep and extended participation by Math Center team in design of Lucent’s 3G cellular wireless networks- providing QoS over wireless channels
- Algorithm controls network congestion by “scheduling” queues of time-varying “switches” in a dynamic system (network)

Network congestion control



Previous work:

$$\max_x \sum_n U_n(x_n)$$

subject to

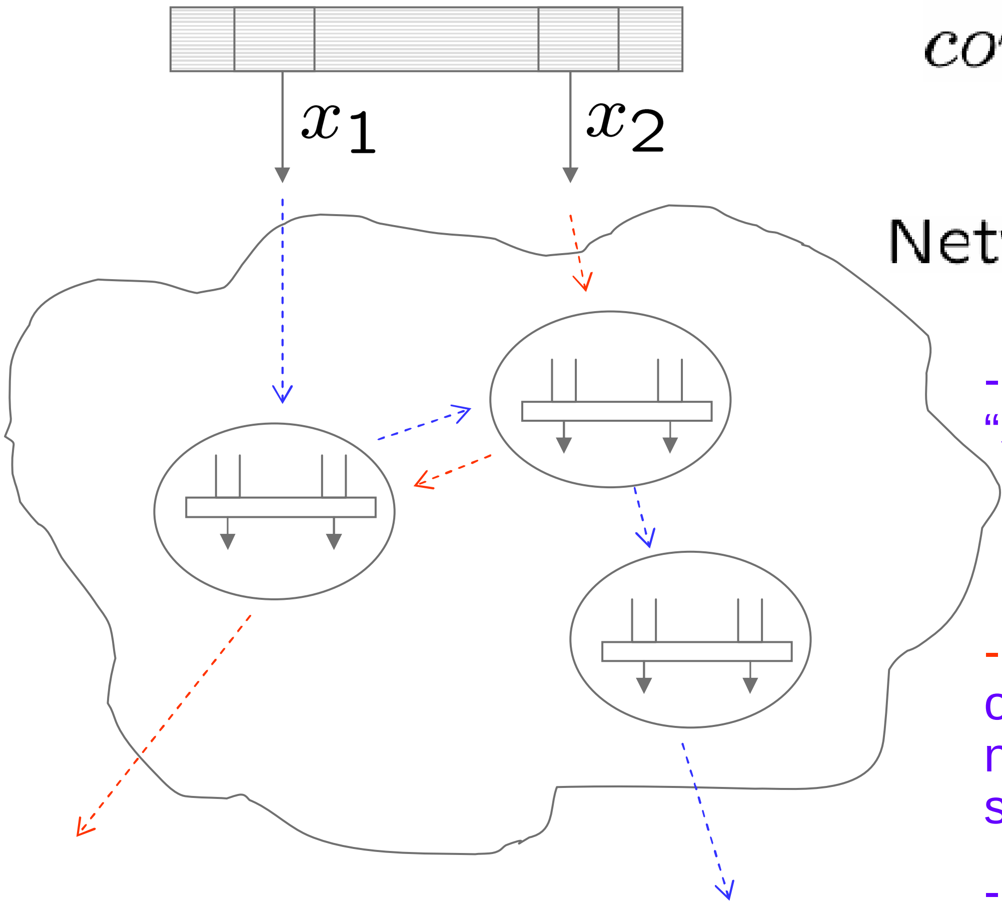
Each link ℓ is not overloaded:

$$\sum_{n \in F(\ell)} x_n \leq c_\ell$$

-TCP congestion control implicitly tries to solve this problem

-Very large and very active field.
F.Kelly et al., S.Low et al., …

Consider more complex networks



$$\max_{\textit{control strategies}} U(x)$$

subject to

Network queues are stable

-Network nodes are time-varying "switches", which need to be scheduled

-Sources may be dependent and need to be scheduled jointly

-Maintaining "desired" average injection or service rates may not be practical – need control strategy which uses "current state" only.

-Network nodes may have power usage constraints

-Optimization problem is defined implicitly

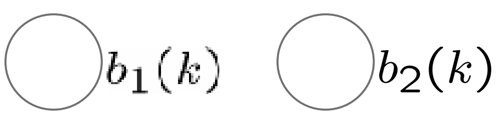
General framework: Controlled queueing network

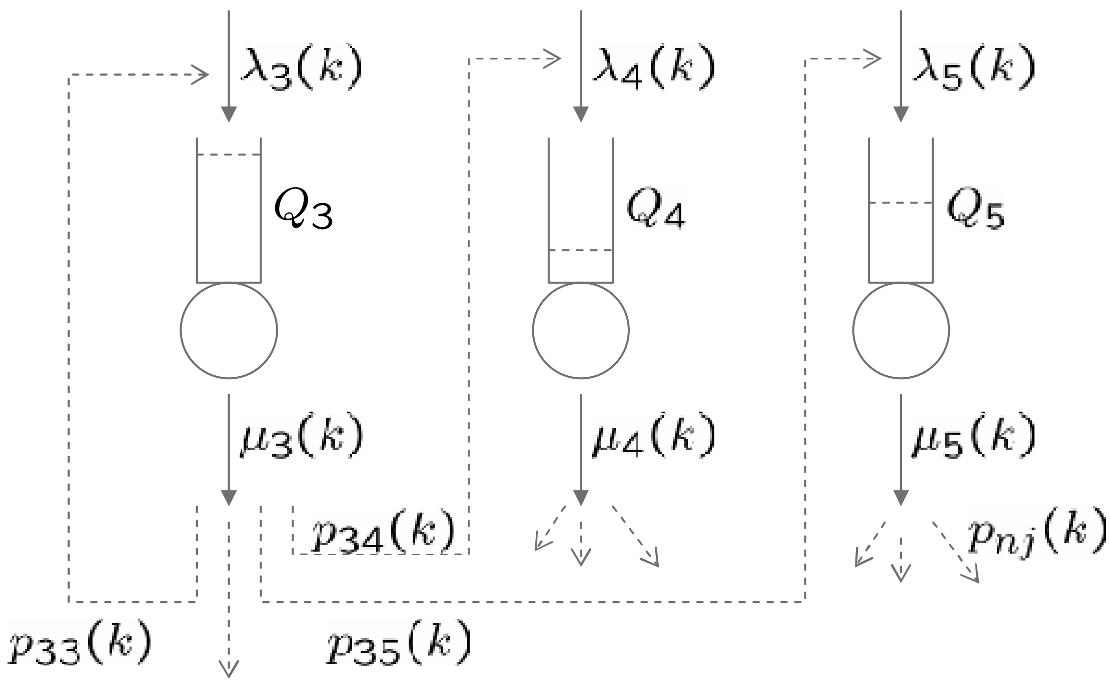
"Utility generating" nodes

e.g., traffic generating

$\mathcal{N}^u\ [=\{1,2\}]$

"Processing" nodes $\mathcal{N}^p\ [=\{3,4,5\}]$





Discrete time $t=0,1,2,\ldots$

Control ***k*** at ***t*** is chosen from a finite set ***K(m(t))***, which depends on the underlying random "network mode" ***m(t)***.

Network control problem

$$b(k) = (b_n(k), n \in \mathcal{N}^u)$$

$x = E[b(k(t))]$ Average commodity rate vector, ***under a given control strategy***

$$Q(t) = (Q_n(t), n \in \mathcal{N}^p)$$

$$\max U(x)$$

$$\text{s.t. } Q(t) \text{ is stable}$$

Utility function *U* is continuously differentiable concave (possibly non-strictly)

Asymptotically optimal network control: Greedy Primal-Dual (GPD) algorithm

$$k(t) \in \arg\max_{k} \nabla U(X(t)) \cdot b(k) - \beta Q(t) \cdot \overline{\Delta Q}(k)$$

$$\beta > 0 \quad \text{small parameter}$$

$$X(t+1) = \beta b(k(t)) + (1-\beta) X(t)$$

$$(\overline{\Delta Q})_n(k) \doteq \lambda_n(k) - \mu_n(k) + \sum_{j \in \mathcal{N}^p} \mu_j(k) p_{jn}(k), \; n \in \mathcal{N}^p$$

$\overline{\Delta Q}(k) =$ expected ***nominal*** queue drift vector

MAIN RESULT (informally):

GPD algorithm is close to optimal when β is small.

GPD algorithm: Discussion

$$k(t) \in \arg\max_k \nabla U(X(t)) \cdot b(k) - \beta Q(t) \cdot \overline{\Delta Q}(k)$$

$$X(t+1) - X(t) = \beta[b(k(t)) - X(t)]$$

$$\overline{\Delta Q}(k) = \text{expected queue drift vector}$$

GPD rule interpretation: *"Greedily" maximize expected drift of*

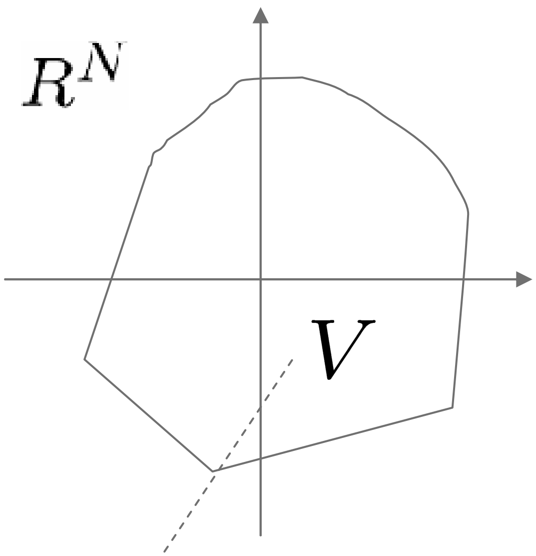
$$H(X(t), Q(t)) = U(X(t)) - \frac{1}{2}\beta \sum_{n \in \mathcal{N}^p} Q_n(t)^2$$

If NO processing nodes: GPD **=>*"Gradient"*** alg., *U(X(t))* is "almost" Lyapunov function

If NO utility nodes: GPD => ***"MaxWeight"*** alg., $\sum Q_n(t)^2$ is Lyapunov function

GPD may be viewed is a "naïve" combination of Gradient and MaxWeight. Optimality is non-trivial, because for this general model ***H(X(t),Q(t))*** is NOT a Lyapunov function

Underlying (implicit) convex optimization problem; System dynamics



Rate region *V* is convex compact

–Rate region *V* is NOT given explicitly

–Standard convex optimization methods cannot be applied

Rate region *V* = All feasible vectors representing joint long-term average commodity rates and long-term drifts of the processing queue lengths

$$\max_{x \in V} U(x)$$

subject to

$$x_n \le 0, \;\; n \in \mathcal{N}^p$$

V^* optimal set

Q^* optimal set for the dual

Under GPD algorithm, for any initial state *X(0)* and *Q(0):*

$$X(t) \to V^* \text{ and } \beta Q(t) \to Q^* \;\text{ as } t \to \infty.$$

John Tukey: ***Founder of Modern Statistics, Data Analysis***

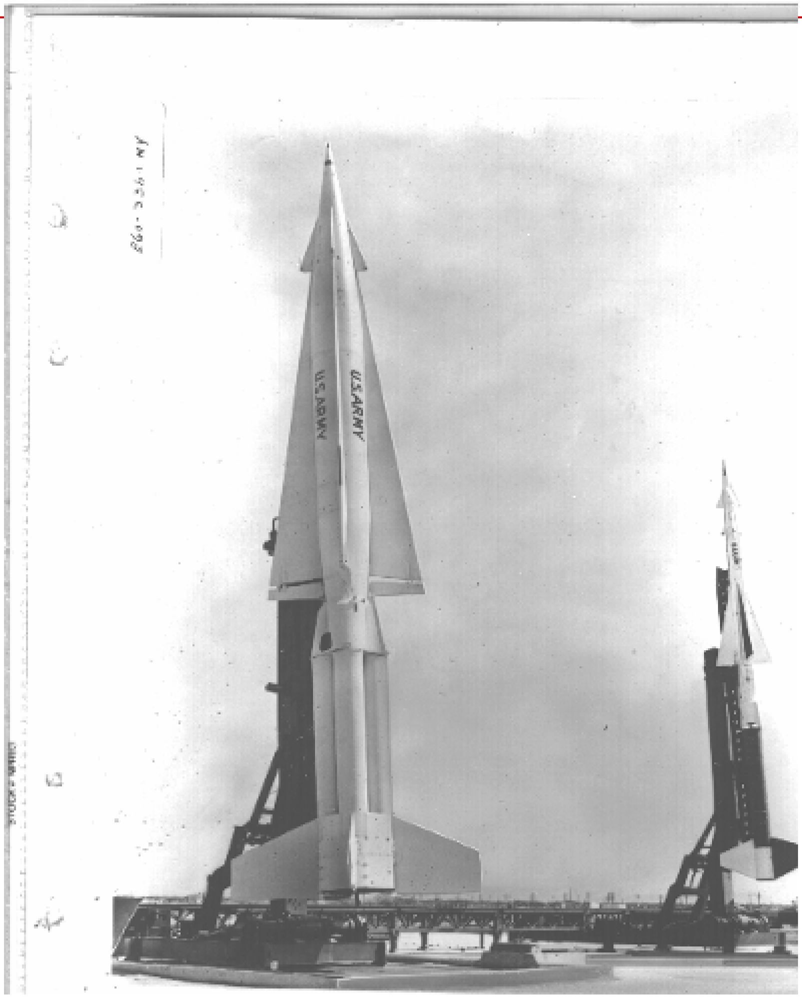
- Statistician, mathematician, data analyst, information scientist, researcher, educator, Presidential advisor and Bell Labs executive
- Joined Bell Labs in early1945, while war was still on. Worked on Nike anti-missile program on trajectory, aerodynamics and warhead design, as well as radar research. After war, remained interested in radar research
- **Time Series and Spectral Analysis** In late 40s a Bell Labs engineer working on radar talked to Tukey and Hamming on estimating power spectrum from sample autocorrelation function.

  The spectrum, obtained from standard Fourier techniques, had a ripple. They tried smoothing with weights 1/4, 1/2, ¼, with striking success.

  "So, Dick and I took off a considerable amount of time to try and understand why this would be…"

  "The origin of the later time series work probably comes from a number of practical problems, …., which was interesting to the boys in Whippany"

Nike Missile



John Tukey: ***Founder of Modern Statistics***

- **Fourier Analysis** In 1963 Tukey showed that if the number of samples N=4**K, then a straightforward computational procedure gives the empirical Fourier transform, which requires fewer than 2N + Log N multiplications.

  In 1965 this became formalized into the Tukey-Cooley algorithm; considered one of the top 10 most important algorithms of the 20th century

- **Data Analysis** Tukey's 1962 landmark paper, "The future of data analysis" changed the paradigm and language of statistics. Bell Labs became a test bed for Tukey's ideas: boxplots, stem-and-leaf diagrams were tried and popularized.

- **Advisor to Presidents** "We have watched at least four Presidents of the United States listen to John and heed his counsel": William O. Baker, Chairman, Bell Labs.

  Tukey participated in the nuclear test ban treaty negotiations in 1959. He became interested in the problem of distinguishing earthquakes from nuclear explosions.

  The model is of a signal laid on top of itself with a delay; the superposition of signals leads to a rippling spectrum. This model was also familiar in the Labs since it arises with echo cancellation, pitch determination and radar tracking. With Bell Labs researchers, Bruce Bogert and Mike Healey, Tukey introduced cepstra – spectra of log spectra, and a complete theory.

Tukey's Legacy: **Analyzing Internet Traffic**

Tukey's Work Then …

- Recognized two types of data analysis – Exploratory Data Analysis (EDA) and conformatory data analysis (CDA).
- Principal aim of EDA is to look for unexpected patterns and let the data point to a model
- Bell Labs' style of data analysis, which is strongly data-driven, is largely Tukey's creation

Leading to Breakthroughs Today

- Bell Labs' Jin Cao uses Tukey's processes to collect data on Internet packet data and create models and analysis; discovered increased smoothness further from the network edge.
- Simple exploratory analysis indicated that the statistical properties of packet traffic changes with changing traffic load; unexpected patterns drove model
- 2 terabytes of compressed packet headers in database –statistical analysis techniques challenged to process such large amounts of data

Internet Packet Traffic Modeling

Goals

- Simple, valid models for network packet traffic generation
- Apply models in traffic engineering problems such as bandwidth estimation

Environment for Empirical Study

- A large Packet header database
- S-Net: analysis software environment

Results

- Changing statistical properties of traffic as load increases
- Packet and application level model for Internet packet traffic
- Minimum bandwidth allocation for Internet best effort traffic on a single link

Environment for Empirical Study

Packet Header Database

- 2 terabyte of compressed packet headers
- 22 Internet interfaces with linkspeed from 10mbps to 2.5Gbps
- Trace durations: 90sec (NLANR) ~ 4 years (Bell Labs)
- Trace sources: NLANR, U. of Waikato DAG group, NCNI/NCREN, MFN, Bell Labs

S-Net

- An open source UNIX software system for comprehensive statistical analysis of IP packet headers
- Addresses full set of tasks from raw traces to final results of data analyses
- Based on S system for graphics and data analysis

Strategy for Analysis

Simple exploratory analysis indicates that the statistical properties of packet traffic changes with changing traffic load

Strategy for comprehensive and yet manageable analysis (consider many days of a packet trace from a single link)

- Time blocking: break the trace into time blocks
- Load sampling: sample blocks according to load if number of blocks is too large
- Repetitive analysis: apply the same analysis to blocks
- Summary analysis of all blocks

Exploratory data analysis and visualization is key for developing insights

Visualization and Fitting for Exploratory Analysis

About 30 visualization and fitting tools lie at the core of our exploration

- Time series plot
- Quantile plots
- Power spectrum plots

Very Large Displays for comprehensive visualization

- One plot per time block
- Many pages, possibly many panels per page

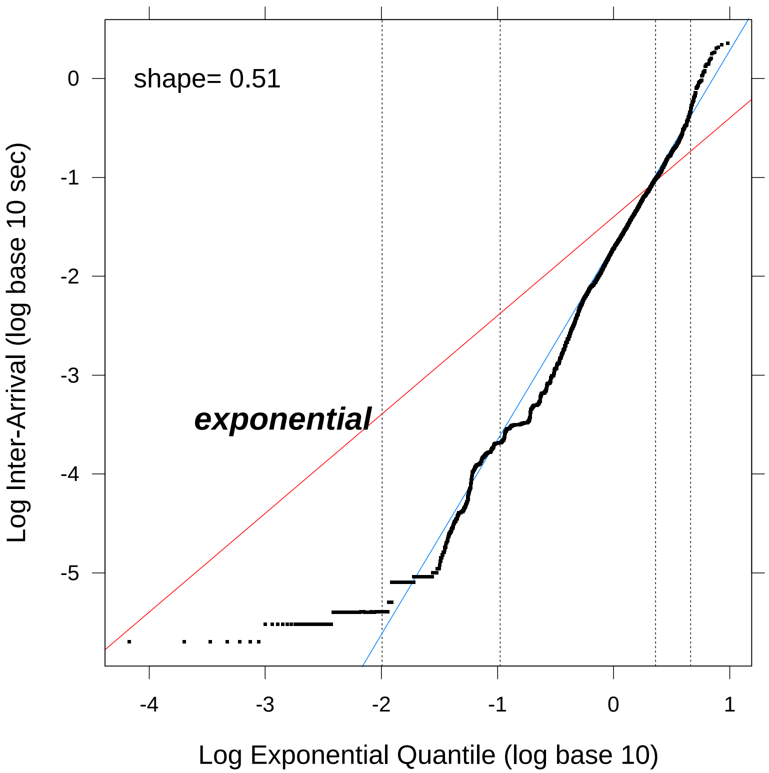
Summary displays

- One point per time block
- A single plot for all time blocks

Example: Distribution of Packet Inter-arrival

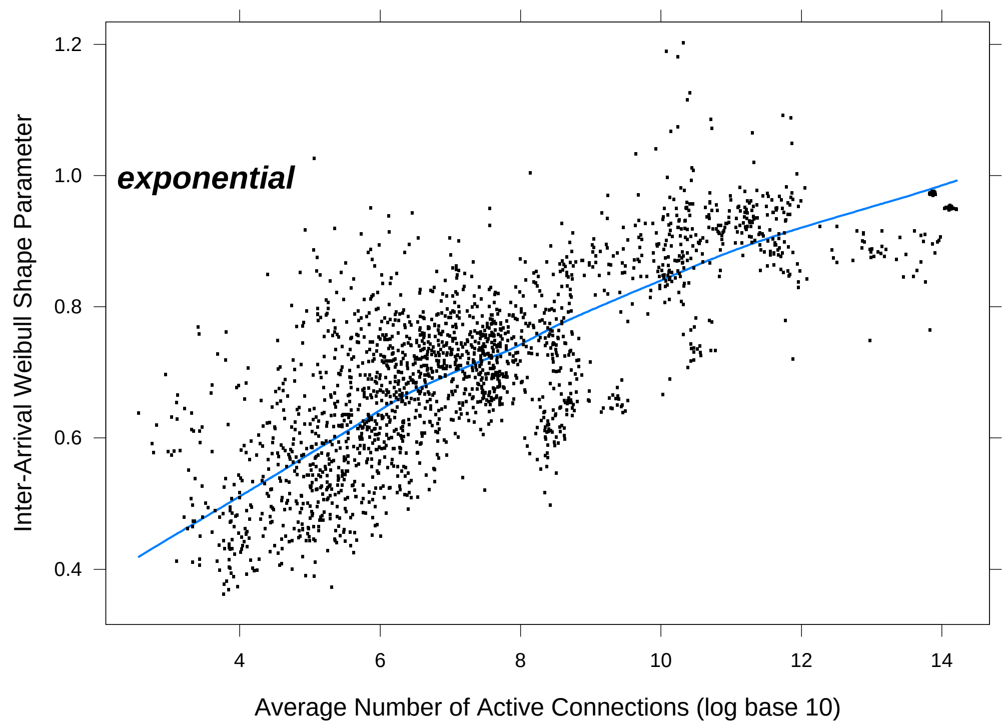
Weibull Quantile Plot

- *X*~ Weibull if $(X/scale)^{shape}$~Exponential
- One Bell Labs 5min trace
- Quantiles of log inter-arrivals ~ Quantiles of log Exponential



Weibull Shape Parameter vs. Traffic Load

- 2072 traces after time blocking and sampling
- Shape parameters tend toward 1 implies packet inter-arrivals tend toward exponential as load increases



Results

Discovery about changing statistical properties as absolute load increases (load increases with distance from edge to core)

- Packet arrivals tend to Poisson
- Packet sizes tend to independence
- Long range dependency in arrivals and sizes dissipates
- Change is critical to engineering the Internet

Simple and valid models for Internet traffic

- Packet level traffic model
  - Hierarchical Fractional-Sum-Difference model to account for both long-range dependency and multiplexing effects with load
  - Empirical and validated using 2072 packet traces
- Application level traffic model for web traffic
  - Validated both at the application and packet level using Network Simulator (NS)

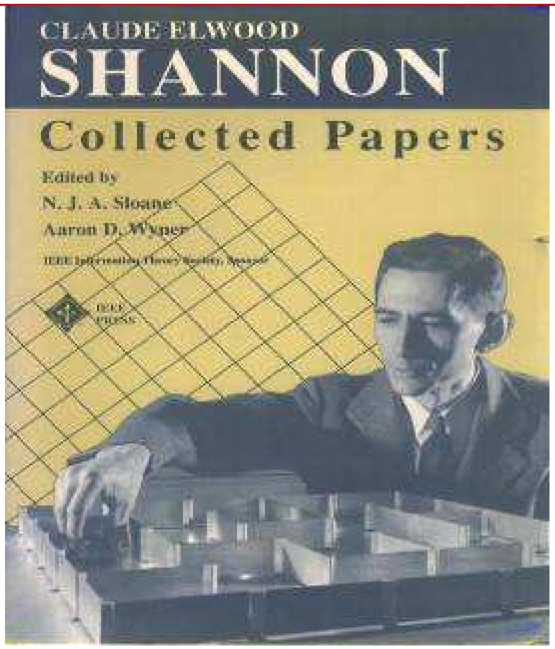
Bandwidth allocation

- Developed a simple solution using queuing simulation and statistical analysis of live and synthetic traces based on traffic models

Claude Shannon: *Information Theory, Communication Theory*

- Having received his B.Sc. In EE and Math, Shannon spent the summer of 1936 in Bell Labs working on relay and switching circuits.
- In 1941 Shannon was in Princeton when Thornton Fry invited Shannon to come to Bell Labs and join the war effort on anti-aircraft fire control.

  Shannon, Blackman and Bode's theory of filtering and prediction complemented Wiener and Kolmogorov's. It was used in the M-9 Gun Director.
- Shannon also worked on the SIGSALY speech scrambling system that allowed Churchill and Roosevelt to talk over an unbreakable coding system .



Claude Shannon: ***Information Theory, Communication Theory***

- Shannon's masterwork, "A Mathematical Theory of Communications" was published in the BSTJ in 1948. "Few other works in this century have had greater impact on science and engineering."
- In 1956 Shannon and Moore showed that switching circuits can be built which are arbitrarily reliable, regardless of how unreliable the original relays are.
- Shannon also contributed to the fields of telephone exchange design, cryptography, chess-playing machines and theoretical genetics.

Shannon’s Legacy: **Space-Time Communications**

Shannon’s Work Then …

- Prior to Shannon, Error-Free Transmission of Information was inconceivable; Information was regarded as a continuous (analog) entity
- Shannon proved that information is inherently discrete (measured in *bits*) and that every noisy channel has a capacity (*bits/second*); further for all transmission rates less than capacity, error-free transmission is possible through channel-coding

Leading to Breakthroughs Today

- Recent innovations in space-time communications – such as multiple wireless antenna arrays – are driven by Shannon’s theories and the work of Bell Labs researchers Gerald Foschini, Emre Telatar, Tom Marzetta and Bert Hochwald
- Shannon’s theory tells us the structure of the capacity-attaining signals, and suggests practical space-time modulation
- Using multiple antennas, as theorized by Bell Labs in 1995, results in huge throughput improvements – with no extra power or spectrum use

Prior to Shannon, Error-Free Transmission of Information was Inconceivable

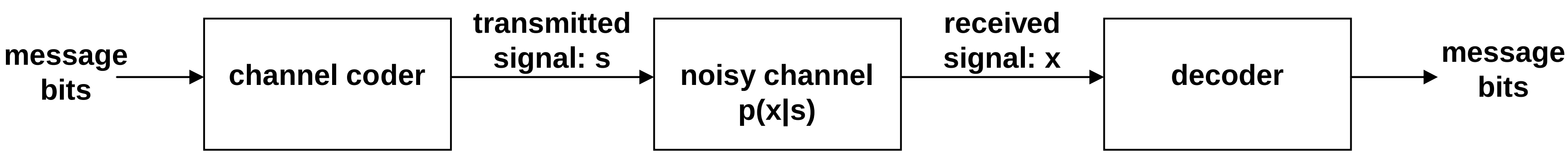
- Information was regarded as a continuous (analog) entity
- Linear modulation (AM; single-sideband)
  - increased transmission power helps
  - no advantage derived from increased channel bandwidth
- Narrow-band nonlinear modulation
  - no advantage over linear modulation

"In fact, as more and more schemes are analyzed and tested, and as the essential nature of the problem is more clearly perceived, we are unavoidably forced to the conclusion that static, like the poor, will always be with us." [J. R. Carson, Proc. IRE, 1928]

- Wideband FM (Armstrong, 1933)
  - bandwidth expansion confers noise immunity

Shannon (1948): Error-Free Transmission is Possible

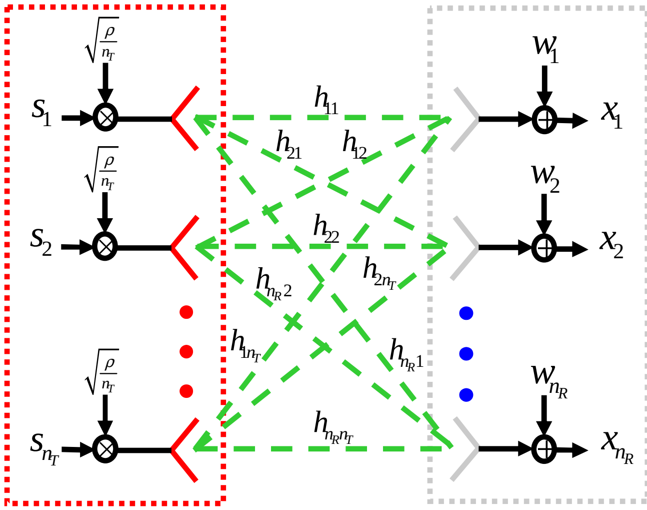
- Information is inherently discrete (measured in *bits*)
- Every noisy channel has a capacity (*bits/second*)
  - for all transmission rates less than capacity, error-free transmission is possible through channel-coding



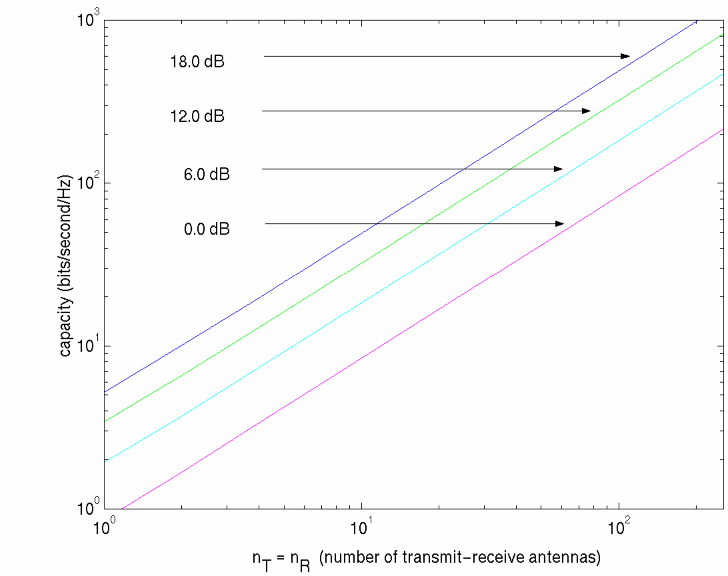
- Computation of capacity
  - use a random codebook, characterized by $p(s)$
  - choose probability density to maximize mutual information

$$C = \sup_{p(s)} E\left\{\log_2\left(\frac{p(x|s)}{p(x)}\right)\right\}$$

Multiple-Antenna Wireless (1995: Foschini; Telatar)



$$x = \sqrt{\frac{\rho}{n_T}} sH + w$$



- Huge throughput improvements: no extra power or spectrum
  - Rayleigh fading conditions
  - receiver knows channel matrix, H

$$\{s(1), \cdots, s(n_T)\} \text{iid CN}(0,1) \longrightarrow C = \log_2 \det\left(I + \tfrac{\rho}{n_T} HH^*\right)$$

Multiple-Antenna Wireless: Nobody Knows the Channel (Marzetta & Hochwald, 1998)

$$\underset{T \times n_R}{X} = \sqrt{\frac{\rho}{n_T}}\underset{T \times n_T}{S}\,\overset{n_T \times n_R}{H} + W$$

- Block fading: unknown channel remains constant for T symbols
- Additive and multiplicative noise: all components iid CN(0,1)
- Shannon theory tells us the structure of the capacity-attaining signals, and suggests practical space-time modulation!
  - there is no point in using more than T transmit antennas
  - $S = \Phi D$
    - $\Phi: T \times n_T$ isotropically random unitary (generalization of phase modulation to space-time)
    - $D: n_T \times n_T$ nonnegative, real, diagonal, random (amplitude modulation)
  - for high SNR, use $S = \sqrt{T}\,\Phi$ (unitary space-time modulation): the subspace spanned by the columns of Φ survives multiplication by unknown H

Standing on the Shoulders of Giants

Since 1925, Bell Labs has been the intellectual inspiration and incubator for generations of scientists and engineers.

Bridging Past and Present, the Invariant Themes:

- Real-world problems that motivate research
- Intimacy between the “discipline” and “project” worlds
  - applications and abstractions
- Commitment to fundamental research
  - time to abstract and replenish the disciplines
- Multi-disciplinary team of high-caliber minds
- Active participation in the scientific community
- Academic collaborations and partnerships

Acknowledgements

Many thanks to Bob Abbott, Jin Cao, Tom Marzetta, Henry Pollak, Devon Prutzman and Sasha Stolyar for their help with this presentation.