Demand of workshop from labs

技术合作部

www.huawei.com

HUAWEI TECHNOLOGIES CO., LTD.

Contents:

(Network Lab) Network Modeling and Network Calculus

Background:
Currently we just try our best to optimize the network, and to improve some QoS item as much as possible. No one could tell what the final goal we could reach. What is the theoretical limit of each QoS item? Is there any theory could support and express different network scenarios?

Topics:
What about the mathematical modeling for a chain of nodes, where multi-priority queues locate on every single node?
How to abstract the network if we know its topology, and to represent its diverse QoS items in a mathematical method?
How to represent the efficient bandwidth of the network in a mathematical method, assuming the data flows injected coincident with some definite distribution, such as Gauss, Poisson and etc. ?

Expectation:
To get the corresponding mathematical modeling for different network scenarios.
To express diverse QoS factors via mathematical method.
To explore the theoretical limit of each QoS item

(WT Lab)Distributed Multi-agent Scheduling Problem for Multi-Cell Scenario

1. Business scenarios:

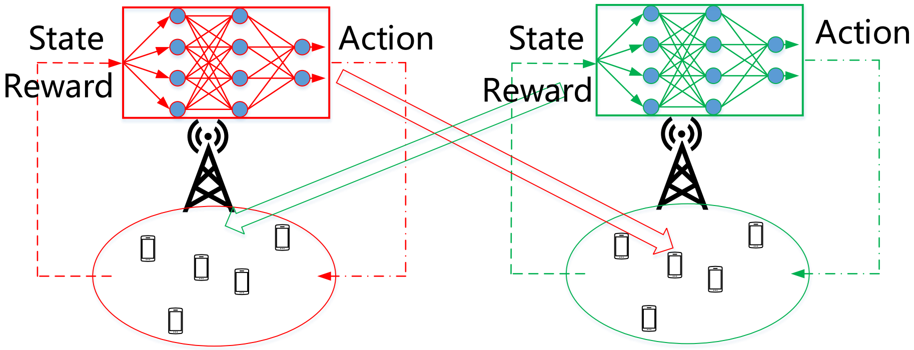
Scheduling is the key to improve system capacity, while for a multi-cell network, distributed but cooperative scheduling algorithm is needed.

2. Problem and benefit:

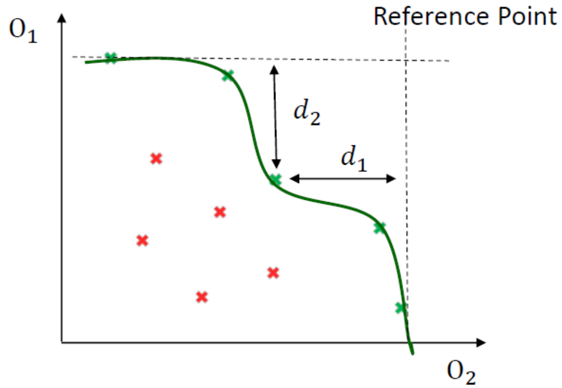
We expect a distributed multi-agent cooperative scheduling framework which can improve system capacity in multi-cell scenario.

Problem: The environment is influenced by all the agents, how to coordinate them? It is a multi-objective (throughput, fairness, packet loss, delay) problem, how to find the Pareto frontier and choose the most suitable solution?

Benefit: With this algorithm, it is possible to make good tradeoff between the above mentioned multiple objectives in a distributed and cooperative manner.



Interaction between two cell



Pareto Frontier (throughput and Fairness)

3. Expectations:

Inputs:

The system state: system settings, channel condition, buffer condition, HARQ condition, historic statistics

Output:

Scheduling framework (how to allocate the radio resource), which may use neural networks.

Analysis & optimization of Polar successive-cancellation-list (SCL) decoder

Further decoder optimizations

Existing optimizations in literature

1) Pruning on the SC decoding tree [Yazdi,2011] (parallelizing constituent code blocks)

- Rate-0 (R0), repetition (Rep) nodes [Heshemi,2016].
- General (Gen) nodes comprised of consecutive bits [Sarkis,2014], [Yuan,2015].

2) Reduce the number of path splitting

- Rate-1 (R1), single parity-check (SPC) nodes [Heshemi,2017].
- Do not split upon the most reliable bits [Li, 2015].

3) Reduce the latency of list pruning

- Adopt bitonic sort [Lin,2013] for efficient pruning.
- Quick list pruning by double-thresholding [Fan,2015].

Analyze the performance of SCL decoder:

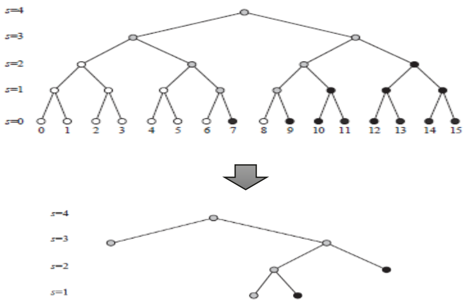
1) SC decoder can be well analyzed:

- Gaussian approximation [Trifonov 2012] can well predict performance, and enable sub-channel reliability estimation.

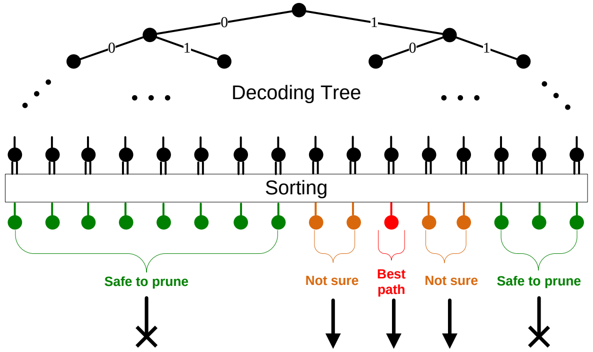
2) SCL decoder is an open problem:

- PM-based final path selection (code distance spectrum analysis)
- Genie-aided final path selection (list extension & pruning analysis)

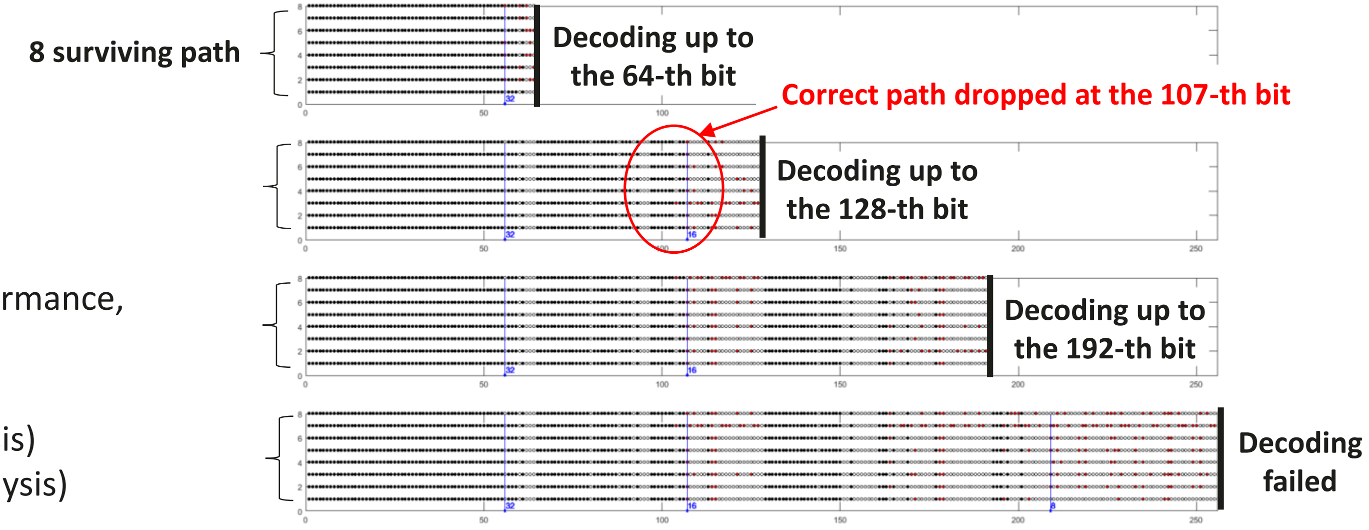
Pruning on the SC tree



Pruning on list decoding tree



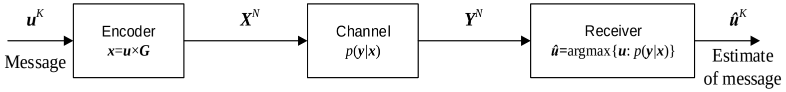
A "decoding error" event in an SCL-8 decoding attempt



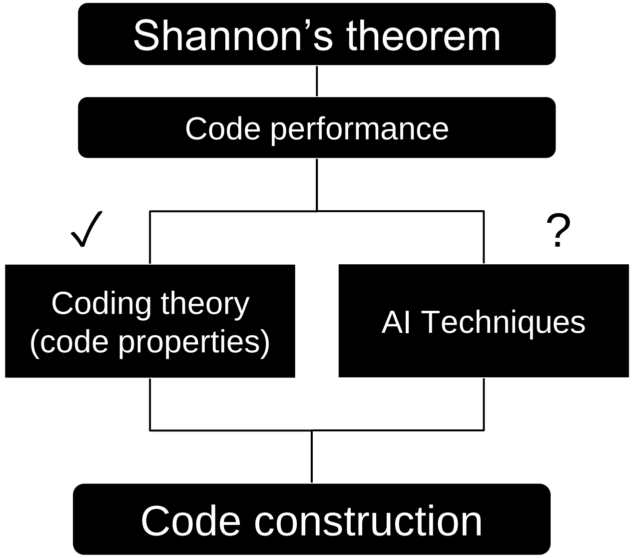
Question 1: can we further improve SCL performance or further reduce complexity?

Question 2: can we theoretically or numerically analyze the SCL decoder, in order to better guide code construction.

AI-driven source coding and channel code construction, as an alternative approach to Shannon limit



A channel coding system


Shannon's theorem

Code performance

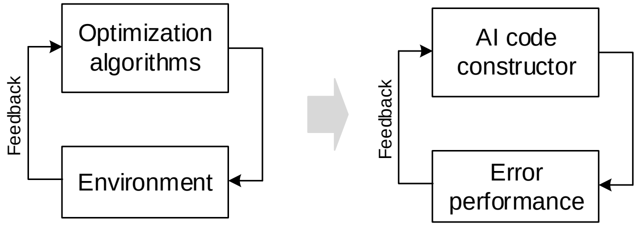
✓ Coding theory (code properties)

? AI Techniques

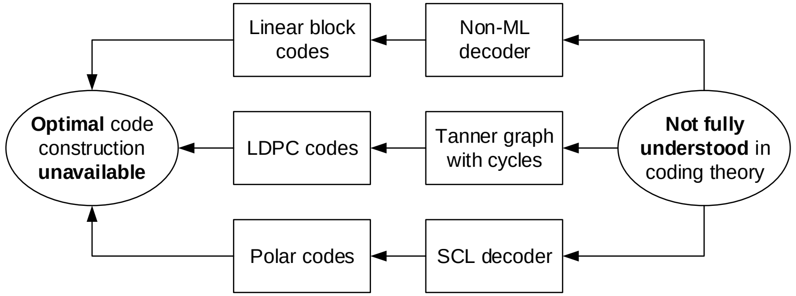
Code construction

Two fundamental questions:

Whether AI algorithms can learn to construct codes without expert knowledge?



Whether the learned codes can have comparable or better error performance with respect to those derived in classic coding theory?



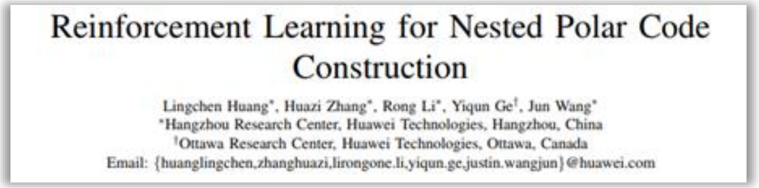
Preliminary works:


AI Coding: Learning to Construct Error Correction Codes

Lingchen Huang[*], Huazi Zhang[*], Rong Li[*], Yiqun Ge[†], Jun Wang[*]
[*]Hangzhou Reserach Center, Huawei Technologies, Hangzhou, China
[†]Ottawa Reserach Center, Huawei Technologies, Ottawa, Canada
Email: {huanglingchen,zhanghuazi,lirongone.li,justin.wangjun,yiqun.ge}@huawei.com

[2019 ArXiv] AI coding: three AI algorithms for three types of code representations:

- Binary matrix;
- Binary vector;
- Nested codes


Reinforcement Learning for Nested Polar Code Construction

Lingchen Huang[*], Huazi Zhang[*], Rong Li[*], Yiqun Ge[†], Jun Wang[*]
[*]Hangzhou Research Center, Huawei Technologies, Hangzhou, China
[†]Ottawa Research Center, Huawei Technologies, Ottawa, Canada
Email: {huanglingchen,zhanghuazi,lirongone.li,yiqun.ge,justin.wangjun}@huawei.com

[2019 Globecom] Reinforcement learning (RL) for nested polar code construction.

- RL algorithms: A2C → PPO
- Integrated learning for higher efficiency

Problems to be solved:

1) From short codes to long codes (N≥1024)
    - Current AI algorithms are low-efficient. Significant improvement of AI algorithm efficiency is required;
2) Machine learning for discrete/graph representations
    - Channel codes are typically represented by binary matrix/vectors or graphs, which is very different from existing popular AI tasks such as image/speech recognition…
    - New AI algorithms are required for channel coding applications, including new algorithms and new neural networks.

Seeking the massive MIMO capacity in closed-form

1. Business scenarios:

Massive MIMO is a key technique for next-generation wireless networks. Thus, it is very important to find the exact massive MIMO capacity in closed-form.

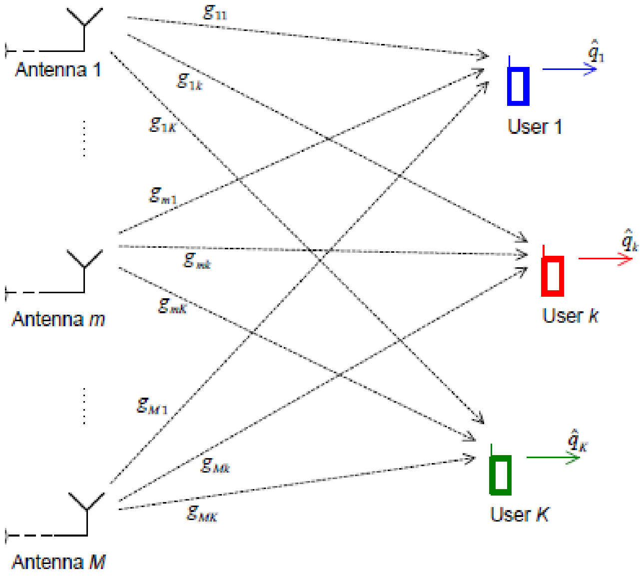
2. Problem and benefit:

Problem:

1) The channel model is very complicated
    a. Path loss, shadowing
    b. Small-scale fading
    c. Line-of-sight
    d. Antenna correlation
    e. Interference
    f. Imperfect channel state information

$$\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K] = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1K} \\ g_{21} & g_{22} & \cdots & g_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ g_{M1} & g_{M1} & \cdots & g_{MK} \end{bmatrix}$$

$$\boldsymbol{h}_k = \sqrt{d_k^{-\gamma} \varsigma_k} \left(\boldsymbol{R} \sqrt{\frac{1}{1+\kappa_k}} \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_M) + \sqrt{\frac{1}{1+\kappa_k}} \boldsymbol{a}_k \right) + \boldsymbol{e}_k$$



2) Wireless communication systems today are very large

3) What is the exact ergodic sum-capacity of the massive MIMO system?

$$\boldsymbol{C}_{ul} = \mathbb{E}\{\log_2 \det(I_k + p_u \boldsymbol{H}^H \boldsymbol{H})\} = \mathbb{E}\{\sum_{k=1}^{K} \log_2(1 + p_u \lambda_k)\} = \int_0^\infty \log_2(1 + p_u \lambda) p(\lambda)\, d\lambda,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0$ are ordered eigenvalues of $\boldsymbol{H}^H \boldsymbol{H}$.

Benefit: The exact massive MIMO capacity can facilitate understanding of the underlying relationships within the respective research problems.

3. Expectations:

Two possible ways:

- Directly derive the distribution of the eigenvalues of $H^H H$, and then obtain the exact capacity
- Use the connection between the Riemann zeta function and the random Hermitian matrix, and then obtain the exact capacity

The interference management in full-duplex multiple BS networks

1. Business scenarios:

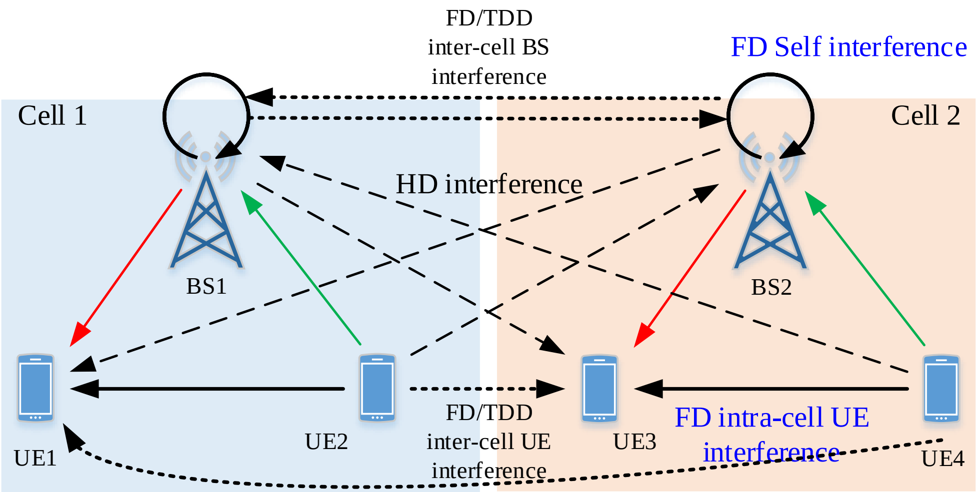
Full-duplex has the potential to double the spectral efficiency at the physical layer and provide many other benefits at higher layers, e.g., significant reduction in end-to-end delay, network security enhancement, and eliminating hidden node issue.

2. Problem and benefit:

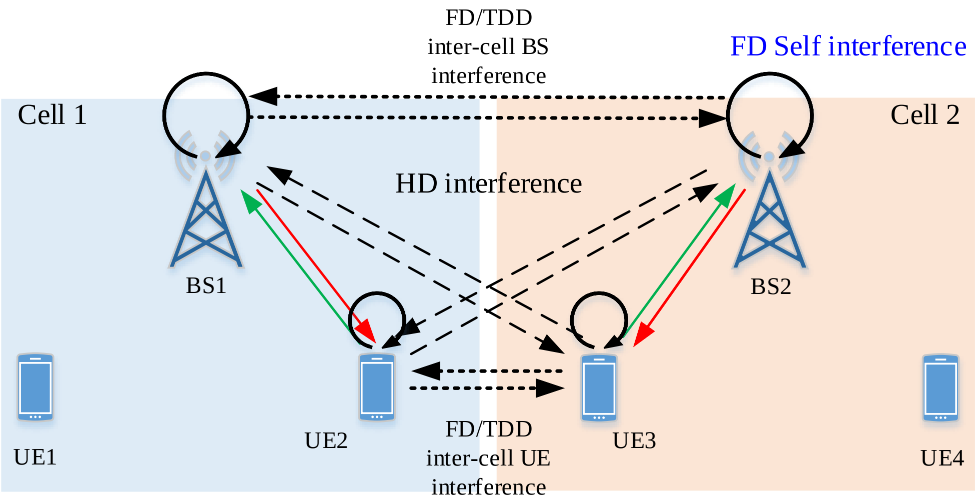
Problem:

1) Strong Self-interference cancellation at the full-duplex node → efficient self-interference cancellation technique is needed (for high transmit power, wide bandwidth, multiple-antenna configuration)

2) Interference from other cells, and also from other wireless links → proper interference management is required



Full-Duplex @BS,   Half-Duplex @UE



Full-Duplex @BS&UE

3. Expectations:

1) New network architecture is urgent to eliminate the strong interference in the full-duplex system
2) Efficient strategies should be designed to coordinate the BS and UE to suppress the interference
3) New signalling which has forward compatibility should be proposed.

(Theory Lab) P1: Mathematical Modeling in Future IoT Network Traffics

1. Business Scenario:

Traffic model is the foundation of communication network/system design and analysis. Different traffic model leads to

2. Problem Statement:

Human to Human:
0.5~1.5B connections

- Basic assumption： i.i.d., low probability
- 1838, Poisson distribution found
- Math base: 1738-De Moivre, 1809-Gauss, 1810-Laplace (probability density, L.L.N., C.L.T.)
- Voice service: stationary intensity, Poisson-distributed (light-tailed)

Human to Machine:
5~15B connections

- Basic assumption: inter-connection between things
- 1988, Barabási found the Power-law on Internet and proposed the BA model.
- Math base: 1895-Pareto, Social fortune 2-8 rule; 1925-Yule power law, math evolution model;
- Data service: intensity fluctuates significantly, traffic self-similarity (burstiness), power-law distributed (heavy-tailed)

Machine to Machine (IoT):
50~500B connections

- Basic assumption： inter-connection between AI things
- 2018, arguments occurs in Nature about the effectiveness of Power-law on Internet

3. Expected Output:

Mathematical models for the space-time network traffic:

- Describe the traffic characteristics in space and time for different services
- Investigate the correlation property of network traffics, e.g. short-correlation (Poisson) or long correlation (self-similarity)

P2: Information Theory of General Interference Channel

1. Business Scenario:

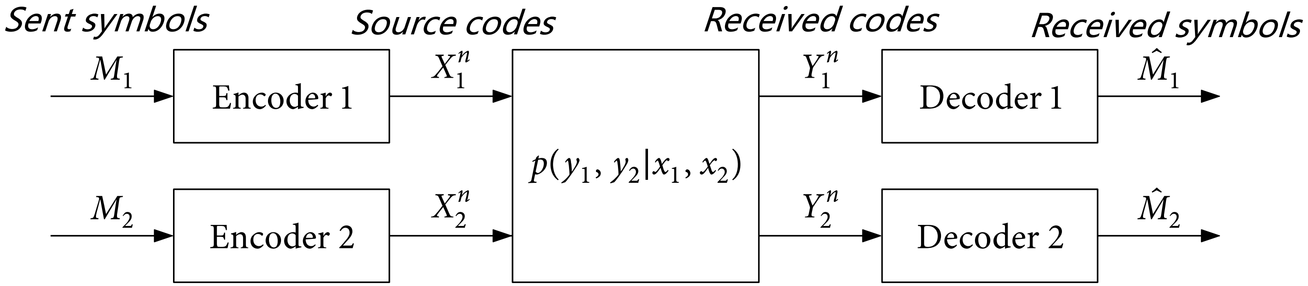
Allowing multiple users communicate through the same channel, although causing interference, increase total channel capacity.

2. Problem Statement:

R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Probab.*, vol. 2, no. 5, Oct. 1974.

Rudolf Ahlswede

- ❑ Germany mathematician R. Ahlswede proposed and find the pre. result.
- ❑ Two senders transmit info. to two receivers on the same channel, a model of communication interference.



Sent symbols *Source codes* *Received codes* *Received symbols*

M_1 → Encoder 1 → X_1^n → $p(y_1, y_2|x_1, x_2)$ → Y_1^n → Decoder 1 → $\hat{M}_1$

M_2 → Encoder 2 → X_2^n → $p(y_1, y_2|x_1, x_2)$ → Y_2^n → Decoder 2 → $\hat{M}_2$

Gauss interference channel capacity region

$$y_1 = x_1 + \sqrt{a_{12}}\, x_2 + n_1,$$

$$y_2 = \sqrt{a_{21}}\, x_1 + x_2 + n_2$$

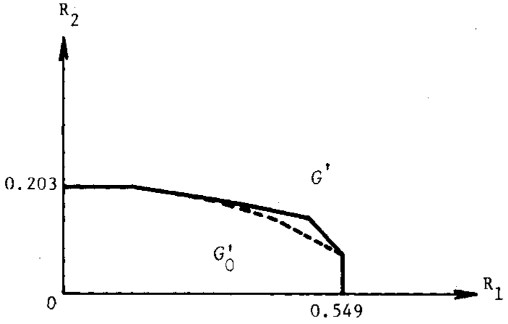


Fig. 9. $P_1 = 2.0$, $P_2 = 0.5$, $a_{12} = 1.0$, $a_{21} = 0.25$.

$$R_1 \le \frac{1}{2}\log(1 + S_1), \qquad R_2 \le \frac{1}{2}\log(1 + S_2)$$

$$R_1 + R_2 \le \min\left\{\frac{1}{2}\log(1 + S_1 + I_1), \frac{1}{2}\log(1 + S_1 + I_2)\right\}$$

A classic open problem in information theory field, a dark cloud in the IT sky

- ❑ 1961, Shannon studied the two-way comm. problem, a prototype of interf. chnanel problem
- ❑ 1974, Ahlswede explicitly proposed the problem and the achievable rate region
- ❑ 1981, Han and Kobayashi proposed the SOTA inner bound in the Gauss interference case
- ❑ 1987, Costa and El Gamal generalized Han-Kobayashi's bound to general discrete memory-less case
- ❑ 2008, Bresler and Tse: deterministic info. theory, approached the bound within 1 bit

Academic research focuses on capacity analysis: no significant progress for a long time

3. Expected Output:

- › Find the capacity region (a set of inequalities) of the general interference channel

P3: Information Capacity of Network Space

1. Business Scenario:

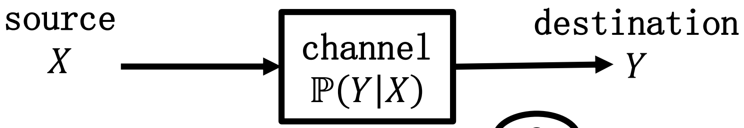
Network information theory provides guidelines for networked system design and optimization (planning, channel coding)

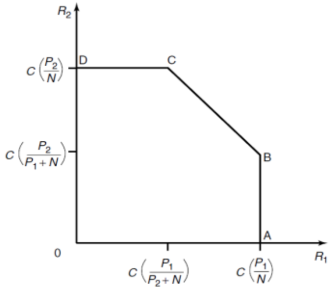
2. Problem Statement:

- **Shannon Information Theory**
  - Point-to-point, channel modeled as transition probability
  - Characterize capacity (achievable rate)
- **Network Information Theory**
  - Multi-access/broadcast channel: capacity-> capacity region
  - Network coding: best paper in IEEE TIT
    - Involve computation in forwarding
    - Single source-multicast: Max Flow-Min Cut Theorem

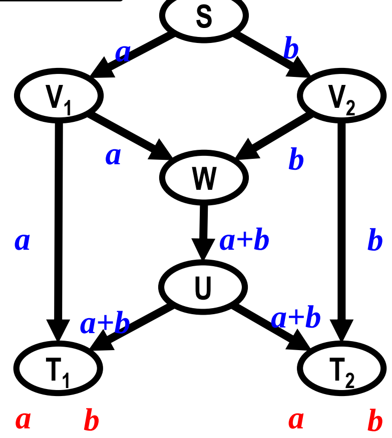
Channel capacity

$$C = \max_{\mathbb{P}(X)} I(X;Y)$$

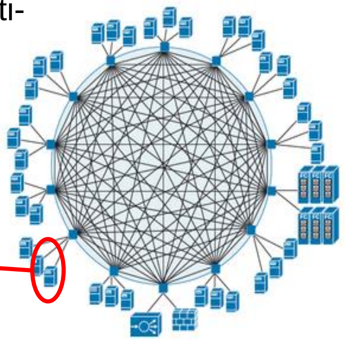




Capacity region of Gauss Multi-access channel



Network coding:
Involve computing in forwarding



Each node can forward, compute and store

Information capacity of network space

3. Expected Output:

Given a network (topology and link capacities) and source-destination pairs, where information can be forwarded, computed and stored

- Find the capacity region of the network

P4: Optimal Information Transport Network

1. Business Scenario:

Communication problems can be viewed as transporting information through the network. Many problems in the communication may be solved efficiently with the Optimal Transport Theory.

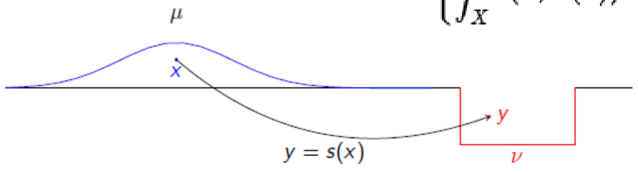
2. Problem Statement:

Optimal Transport (OT) Theory

- ❑ French Mathematician Gaspard **Monge** （1746-1818） proposed OT: how to transport mass with minimum cost
- ❑ 1942, Nobel Laureate, Soviet-Union mathematician **Kantorovich** generalized Monge’s problem
- ❑ 2009, Fields Laureate, French mathematician **Villani** summarized OT problem systematically

Monge problem

$$\inf\left\{\int_X c(x,T(x))\,\mathrm{d}\mu(x)\;\middle|\;T_*(\mu)=\nu\right\}$$



$$\inf\left\{\int_{X\times Y} c(x,y)\,\mathrm{d}\gamma(x,y)\;\middle|\;\gamma\in\Gamma(\mu,\nu)\right\}$$

Kantorovich’s

Sub P1: Resource allocation problem

› Compute resource allocation with OT

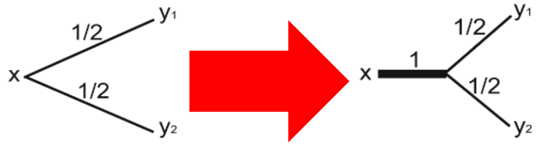
$$\begin{aligned}\min_{\mathbf{x}}\quad & -\sum_{k=1}^{K}\log\left(\|\mathbf{x}_k\|_1\right)+\alpha\max_{l}\frac{\mathbf{R}[l]\mathbf{x}}{c_l}\\ s.t.\quad & \mathbf{R}\mathbf{x}\le\mathbf{c}\\ & \|\mathbf{x}_k\|_0\le 1,\quad\forall k\\ & \mathbf{x}\ge 0\end{aligned}$$

Bandwidth variable $\boldsymbol{x}$
Routing matrix $\boldsymbol{R}$
Link capacity $\boldsymbol{c}$

Sub P2: Topology generation problem

› Given info. production and consumption, compute the optimal topology and bandwidth

How to build the channel/network?



3. Expected Output:

- › Efficient resource allocation solver
- › Solution of optimal topology given info. generation and demand

P5: Mechanism of Network Dynamic Control

1. Business Scenario:

Data networks rely on Transport Control Protocol (TCP) for reliable data communication. TCP-type protocols utilize sliding window and Ack signals to control transmission. However, TCPs meet many problem in practice, e.g. low network utilization, severe congestion etc. We are seeking more effective dynamic control mechanisms and the corresponding mathematical theories.

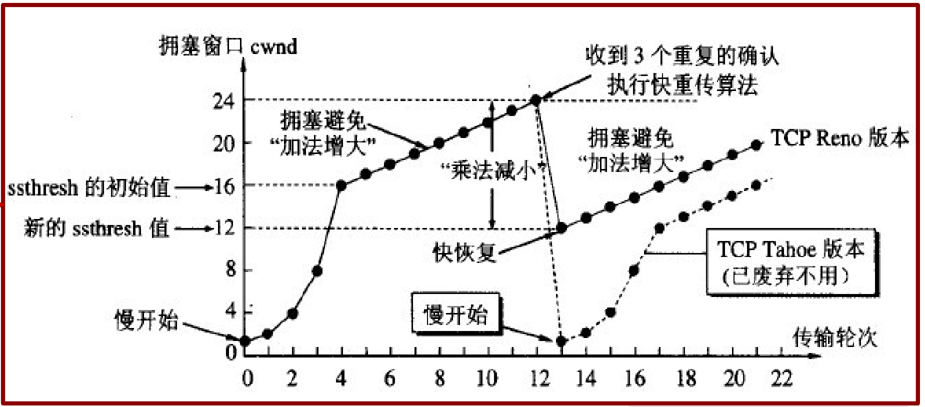
2. Problem Statement:

- Limit network bandwidth: congestion affects network performance and QoS
- As network scale increases, traffic burstiness worsens congestion
- Current network control enhances traffic burstiness
- Classic theory: Network Utility Maximization (NUM)
  - ✓ Application: network resource allocation and congestion control, e.g. TCP-AQM

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathcal{U}(\mathbf{x}) = \sum_{n=1}^{N} u_n(x_n) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\ & \mathbf{x} \in \mathcal{X} \end{aligned}$$

- Existing TCP-AQM: a TCP congestion control mechanism
  - ✓ TCP: adjust congestion window based on AIMD method, e.g. TCP-Tahoe, TCP-Reno
  - ✓ AQM: proactively drop packets, e.g. RED, ECN

- Analysis and design of TCP-AQM:
  - ✓ Based on Fluid model [Misra, et al., 2000]
  - ✓ Based on Stochastic approach [Eun & Wang, 2005]
  - ✓ Math base: PDE, stochastic process (queueing theory)



$$\frac{dx(t)}{dt} = \kappa[w - x(t)p(x(t))]$$

⟺

$$\max_{x_i \ge 0} \sum_i U_i(x_i)$$
$$\text{s.t.} \sum_{i \in l} x_i \le c_l$$

3. Expected Output:

Base on the feedback control in dynamic systems, design new network control methods: achieve high network utility without worsening burstiness.

P6: Complexity Reduction of Large Scale Computation

1. Business Scenario:

Complicated problems occur in many scenarios, e.g. large Cloud Radio Access Network (CRAN). Solving such problems causes huge computational overhead and large delay. Decomposing the large problem into many small problems and solving them individually seems to be a good approach. In this topic, we seek a systematic approach framework and theory to solve this type of problems.

2. Problem Statement:

Complicated Problem

$$\max f(x)$$

Decomposed to low-complexity sub-problems

$$\max g_1(u_1), \ldots, \max g_M(u_M)$$

Solve original problem with solutions of the sub-problems

$$x^* = \phi(u_1^*, \ldots, u_M^*)$$

Lower complexity to ease engineering solution

Decomposition may result in approximate solution

E.G. wireless network decomposition

Overhead of wireless signal processing becomes huge when network scale is large

$$\max SumRate(x_1, x_2, \ldots, x_N)$$

- Decomposing a large network into non-overlapping sub-networks (independent signal processing and much smaller sizes) to lower the whole complexity

$$\max SumRate(x_i | x_i \in cluster1), \ldots, \max SumRate(x_i | x_i \in clusterM)$$

Transmissions in diff. sub-networks interfere with each other; need to contrain the interferences

$$\theta(\{x_i | x_i \in cluster1\}, \ldots, \{x_i | x_i \in clusterM\}) \le I$$

- Then the original problem is transformed into a decomposition problem of Bipartite Graph, can be solved by the Spectral Graph Theory

3. Expected Output:

An algorithm framework and theory for large/complicated problem decomposition and efficient solution.

(ISC Lab) Solving Large Scale MIP Problem Fast

- Problem 1: hierarchical optimization can be very slow when handling multi-tier priority requirements. Should we use parameterized penalty for priority differentiation? Is there a systematic way?
- **min c1 x s.t. Ax >=b, x>=0, where c1 refers to some measurement of demand-supply imbalance;**
- **min c2 x s.t. Ax >=b, c1 x <= (1+e) c1 x*, x>=0, where c2 refers to total backlog;**
- **min c3 x s.t. Ax >=b, c1 x <= (1+e) c1 x*, c2 x <= (1+e2) c2 x*, x>=0, where c3 refers to total cost; …**

✓ Earlier objective becomes constraint for next optimization, but this constraint is very dense, dramatically increases solution time

- Problem 2: very large scale matrix computation – how to maintain and take advantage of matrix sparsity and try to tweak predictor corrector method to solve our types of problem fast? Can we simplify the steps and/or find better meta parameters? – network flow structure; block triangular matrix structure; smoothing constraints; inventory balance constraints

Interior Point/Barrier Method

Predictor step - Affine scaling direction

$$\begin{aligned} \min_x \quad & q(x) = c^T x, \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned}$$

The Karush-Kuhn-Tucker (KKT) conditions for the problem are

$$\begin{aligned} A^T\lambda + s &= c, && \text{(Lagrange gradient condition)} \\ Ax &= b, && \text{(Feasibility condition)} \\ XSe &= 0, && \text{(Complementarity condition)} \\ (x, s) &\geq 0, \end{aligned}$$

These conditions can be reformulated as

$$F(x, \lambda, s) = \begin{bmatrix} A^T\lambda + s - c \\ Ax - b \\ XSe \end{bmatrix} = 0$$

$$(x, s) \geq 0$$

The predictor-corrector method then works by using Newton's method to obtain the affine scaling direction. This is achieved by solving the following system of linear equations

$$J(x, \lambda, s) \begin{bmatrix} \Delta x^{\text{aff}} \\ \Delta \lambda^{\text{aff}} \\ \Delta s^{\text{aff}} \end{bmatrix} = -F(x, \lambda, s)$$

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^{\text{aff}} \\ \Delta \lambda^{\text{aff}} \\ \Delta s^{\text{aff}} \end{bmatrix} = \begin{bmatrix} -r_c \\ -r_b \\ -XSe \end{bmatrix}, \quad r_c = A^T\lambda + s - c, \quad r_b = Ax - b$$

Centering step

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^{\text{cen}} \\ \Delta \lambda^{\text{cen}} \\ \Delta s^{\text{cen}} \end{bmatrix} = \begin{bmatrix} -r_c \\ -r_b \\ -XSe + \sigma \mu e \end{bmatrix}$$

Corrector step

$$\begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^{\text{cor}} \\ \Delta \lambda^{\text{cor}} \\ \Delta s^{\text{cor}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\Delta X^{\text{aff}} \Delta S^{\text{aff}} e \end{bmatrix}$$

$$\sigma = \left(\frac{\mu_{\text{aff}}}{\mu} \right)^3, \qquad \mu = \frac{1}{n} \sum_{i=1}^{n} x_i s_i = \frac{x^T s}{n}.$$

$$\mu_{\text{aff}} = (x + \alpha_{\text{aff}}^{\text{pri}} \Delta x^{\text{aff}})^T (s + \alpha_{\text{aff}}^{\text{dual}} \Delta s^{\text{aff}})^T / n,$$

$$\alpha_{\text{aff}}^{\text{pri}} = \min \left(1, \min_{i : \Delta x_i^{\text{aff}} < 0} - \frac{x_i}{\Delta x_i^{\text{aff}}} \right),$$

$$\alpha_{\text{aff}}^{\text{dual}} = \min \left(1, \min_{i : \Delta s_i^{\text{aff}} < 0} - \frac{s_i}{\Delta s_i^{\text{aff}}} \right),$$

Open Source, Edge Computing and Distributed Algorithms

Many important and large-scale problems (e.g. telecommunications, transport and traffic networks, supply chains, smart cities) adopt an expensive centralized system architecture today, limiting applications. This situation is due to numerous constraints, including the reliance on good-enough Linear / Mixed-Integer Programming (LP/MIP) solvers (e.g. Gurobi, CPLEX and FICO Xpress) that are costly and hardware-demanding, and relatively fewer R&D efforts in distributed algorithms and solvers.

However, the environment is changing. Continuous improvement of open-source solvers and increasing prevalence of cheaper and more capable edge-computing devices may result in a possible future where an otherwise challenging problem can be better managed using highly distributed computing in everyday electronics such as smartphones and devices in cars. Carriers may need split-second dispatch service to direct calls to cell towers, etc.

Can we have some joint research and development in the following diverse topics?

- Partially distributed decision support (heavy centralized + lightweight distributed)
- Practical applications demonstrating benefits of distributed decision support
- Applications using embedded electronics and IoT devices
- Distributed algorithm development
- Joint LP/MIP and heuristics algorithm deployment
- Open-source solver development, and benchmarking to best-in-class

Decision interpretation of complex MIP model

Compared to traditional heuristic algorithm, math optimization has been proved more efficient and effective (in term of global optimum) in solving complex real-life planning problem (e.g. supply chain network optimization, resource allocation, production planning & scheduling). However, one drawback of the optimization approach is the lack of effective mechanism to make results (decision solutions) self-explanatory. Thus most business end-users (planners/decision makers) treat optimization as "black box" without clearly understanding the solving principle. In real life implementation, large amount of time and effort have to be spent on customized post-processing (normally heuristic) in order to help users interpret the result. So, it would be of great necessity and practically useful to have a systematic approach or generic framework to guide the implementation.

In practice, heuristic-based approach starts with analyzing characteristic of various constraints. For example, hard constraints with sparse matrix representation is much easier to be reasoned out than those constraints with dense matrix. Soft constraints involve the component of objectives and thus present a bigger challenge. Rather than handling them directly, deriving rules from model behavior seems a reasonable strategy. Meanwhile, collecting quantitative status of binding and non-binding constraints is also useful to determine the interpretation sequence (which constraints/rules come first) when efficiency is concerned.

Problems in Real World Predictive Analytics

1. The life cycle length of a product is defined to start from the beginning of the sales to the end. For many electronic and high tech products, the cycle lengths get shorter and shorter from one generation to the next. For example, for one particular series of mobile phones, the cycle length changed from 18 months, to 15 months, to 12 months, and is still experiencing shortening. Given the changing life cycle lengths and the uncertain determination for a cycle length due to the data variations caused by other factors and unknown future cycle length, how do we use historical sales data to build models for predict the sales for the next generation of products? Traditional models (ARIMA and Holt-Winters) are based on the inexplicit assumption that the periodicity and cycle length is the same over the time.
2. Given the weekly sales data for a product from its kick-off to the current time, how do we determine its EOF (End of Life, i.e., when to end the sales)? We know the competitive strength of a product can be very important, but the competitive landscape in the market place is always clouded. Given that, what factors should we consider and what models should we build for practical and usefulness purposes? There are ample implications to multiple aspects for deciding the EOF properly and as early as possible: inventory, sales opportunity, the timely introduction of the next generation of products and etc.
3. Different predictive models would perform differently at different aggregation levels. For example, at a very highly aggregated level such as world-wide level for product demand, traditional time series models such as ARIMA and Holt-Winters algorithms would perform well. One the other hand, when predicting the demand for a single short-life product over its lifecycle, neither of them would well, as the derived or estimated trend and seasonality effects from historical data would be too much an extraction for the future in the product life cycle. Isolated and sporadic mappings between models and aggregation levels exist as exemplified above, but a comprehensive mapping between all major analytic models and the application aggregation levels in the high tech industry (based on large scale and systematic study) would be very useful in guiding the practical model applications.

How to measure the performance of supply chain

- Supply chain performance measures are critical for maximizing value in the supply chain and maintaining oversight. What kind of measures companies should use, their advantages and drawbacks? We need useful information on implementing a supply chain performance measurement system. Generally, two characteristics can be used to measure: one is to evaluate the performance of product delivery in terms of quality, cost, service, reliability and lead time; the other is how to deal with uncertain demand and how to manage supply risk.

- How to measure supply chain performance against our competitors

Other Research Topics

1) **building complex mathematical models dealing with various scenarios**
   - **Specially Controlled Material**: Given material C, designate its usage proportion for each finished/semi-finished product. Considering production lead time, deep **BOM** structure and complexity to follow the proportions because of inventory of semi-finished products, it is not easy to express this requirement in mathematical formula
   - **Order time priority**: Guarantee earlier orders are satisfied before later ones
2) **How to utilize specific problem structures to solve large and complex problems fast?**
3) **How to reasonably model uncertainty of demand and supply and ensure some robustness of the solution ?**
4) **How to do what-if analysis effectively? How to scope it?**
5) **Develop an easy to use modeling environment for simple usage (AMPL?)**

(B&P Lab) Solving the nonlinear Schrodinger equation (求解非线性薛定谔方程)

1. Business scenarios:

Numerical solutions of nonlinear Schrodinger equation (NLSE) are widely used in fiber channel simulation platforms and fiber distortion compensation algorithms

2. Problem and benefit:

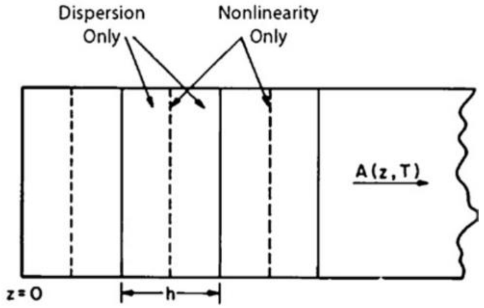
We expect better numerical methods for solving NLSE in terms of robustness, accuracy, speed, degree of parallelism, etc.

Problem: The accuracy and speed for performing the split-step Fourier method (SSFM) to solve NLSE has room to improve; new methods are also desirable.

Benefit: With improved algorithms, simulation platforms for fiber channel model performs faster; efficient fiber nonlinearity compensation algorithms become possible.

$$\frac{\partial A}{\partial z} = \left(\hat{D} + \hat{N}\right) A$$

$$\boxed{\begin{aligned}\hat{D} &= -\frac{1}{2}\beta_2\frac{\partial^2}{\partial T^2} + \frac{1}{6}\beta_3\frac{\partial^3}{\partial T^3} - \frac{\alpha}{2}\\ \hat{N} &= i\gamma\left(|A|^2 + \frac{i}{\omega_0 A}\frac{\partial}{\partial T}\left(|A|^2 A\right) - T_R\frac{\partial |A|^2}{\partial T}\right)\end{aligned}}$$



$$A(z+h,T) \cong \exp\left(\frac{h}{2}\hat{D}\right)\exp\left(\int_z^{z+h}\hat{N}(z')dz'\right)\exp\left(\frac{h}{2}\hat{D}\right)A(z,T)$$

3. Expectations:

- An order of magnitude speed improvements compared with the SSFM.
- Implementable in fiber channel simulation platforms, fiber nonlinearity compensation algorithms.

Optimizing decoding algorithm of Chanel coding method LDPC (优化LDPC信道编码方法的解码算法)

1. Business scenarios:

Chanel coding method LDPC is by far the best performing code. It is widely used in not only in different fiber communication scenarios but also in satellite communications, deep space communication etc.

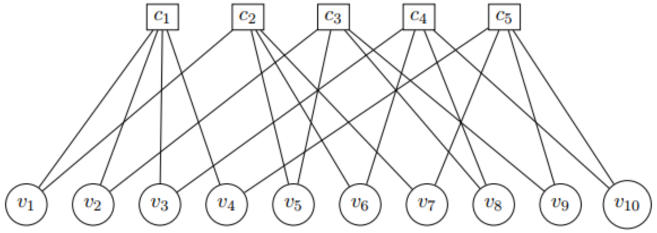
2. Problem and benefit:

We expect LDPC decoder with high performance and low energy consumption.

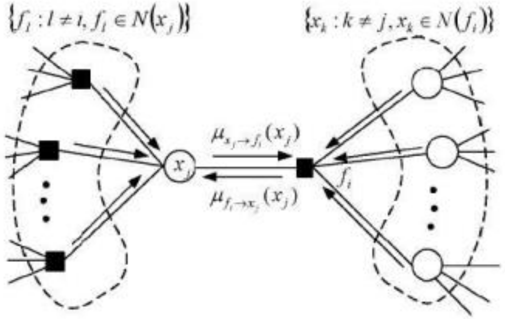
Problem: Now Belief Propagation Algorithm (BPA) is generally utilized in decoding a LDPC code, which is basically a subclass of the iterative message-passing algorithms. We need optimizations of BPA including speedup of convergence, alleviation of implementation complexity; new methods are also desirable.

Benefit: With improved algorithms, energy efficiency of decoding LDPC would be improved. So more complicated code could be used in high performance fiber communication system, and LDPC could be used in power efficiency priority fiber communication system.

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$



A LDPC code can be described by a matrix or a Tanner graph.



System diagram of Belief Propagation

3. Expectations:

- An alternate approximation of BPA
- Optimal compensating parameters of BPA
- Optimal message passing order of BPA

Fiber communication DSP algorithms with other implementation architecture (非传统实现架构的DSP算法)

1. Business scenarios:

Many general optical fiber communication DSP algorithms are widely used in different fiber communication scenarios.

2. Problem and benefit:

Fiber communication algorithms have been researched for many years and are difficult to make a qualitative improvement now. Moore’s Law is also close to the limit, and is difficult to compensate the increasing complexity of the algorithms through the improvement of process. The algorithm implementation architecture change is likely to be the next asic efficiency improvement point. For that we need to re-write our fiber communication DSP algorithms.

Commonly used algorithm in fiber communication DSP: FFT, MLSE(maximum-likelihood sequence estimation), BCJR, Viterbi.

3. Expectations:

- Provide some general fiber communication DSP algorithms with matrix times vector form.
- Evaluate changes in algorithm accuracy before and after rewriting.

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.

HUAWEI