概率图模型理论及应用

Theory and Applications of Probabilistic Graphical Models

(Lesson 7)

欧智坚

清华大学电子工程系

Addr: 罗姆楼 6-104

Tel: 62796193

Email: ozj@tsinghua.edu.cn

课程章节

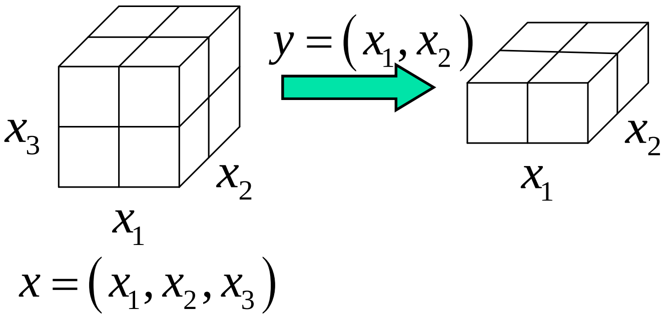
- 第一章 引言（**1**）
- 第二章 图模型的表示理论（**3**）
  - **DGM-UGM**
  - **Semantics**
  - **HMM-CRF**
- 第三章 图模型的推理理论（**6**）
  - 精确推理：**variable-elimination**，**cluster-tree**，triangulate
  - 连续变量：Kalman
  - 采样近似：sampling
  - 变分近似：variational
- 第四章 图模型的学习理论（**3**）
  - 参数学习：**maxlikelihoodEstimate**，BayesEstimate
  - 结构学习：StructureLearning
- 第五章 一个综合例子（1）

Basic concepts – Marginal

- 假设 x 和 y 为变量集，且 $y \subseteq x$。$\phi_X(x)$为定义在x上的函数，则$\phi_X(x)$在y上的边缘化为

$$\Downarrow_y \phi_X(x) = \sum_{x \setminus y} \phi_X(x)$$

$$\Downarrow_y \phi_X(x) = \max_{x \setminus y} \phi_X(x)$$



一、Cluster-tree Elimination（树消除算法）

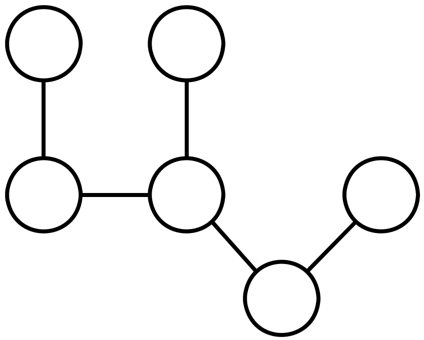
二、Sum-product algorithm on factor graph

三、信道译码应用

Basic concepts – tree, leaf

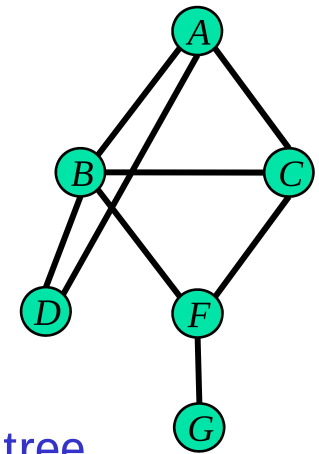
A graph G is connected		图上任两个结点间都有一条迹

tree	T	如果一个图是连通的，并且没有环
Leaf of a tree		树上只有一个邻居的结点



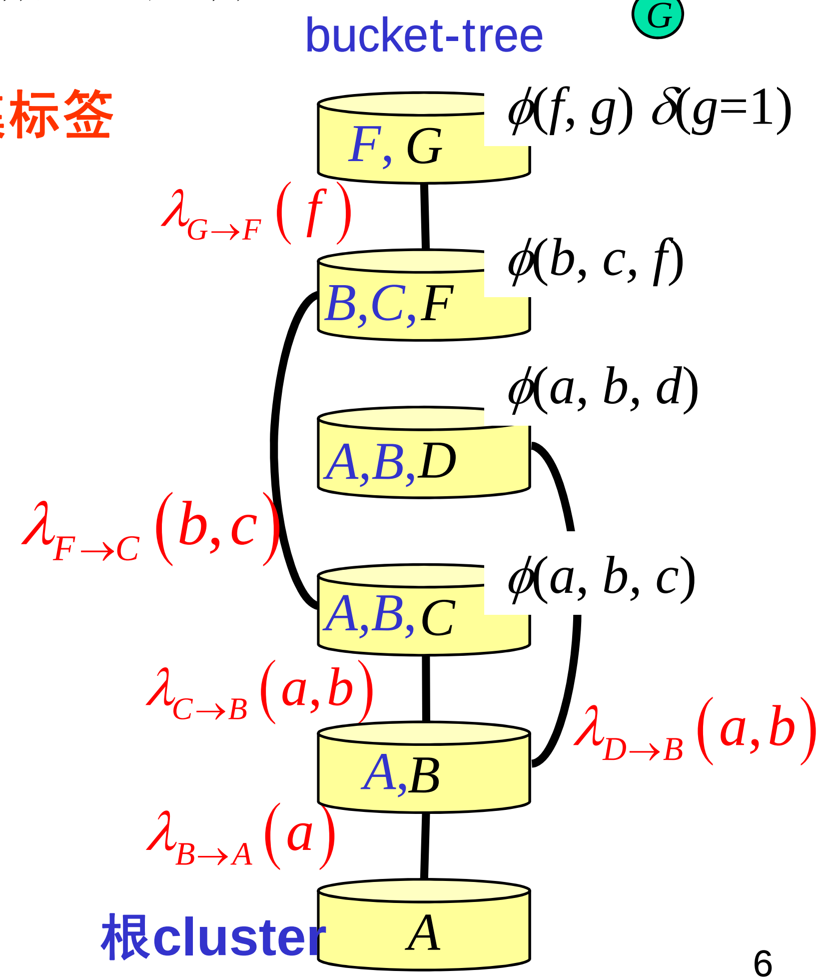
引入 tree-decomposition(树分解)/cluster-tree

$$p(a \mid g=1) \propto \sum_c \sum_b \sum_f \sum_d \sum_g \phi(a,b,c)\phi(b,c,f)\phi(a,b,d)\phi(f,g)\delta(g=1)$$



用一个<u>树型结构</u>来组织**Variable Elimination**的计算。

给定变量消除排序，询问结点 x_Q（这里即 A 结点）居首。



每个桶视为一个cluster，cluster上贴变量集标签

① **初始化：**

将连乘积中的局部函数放入cluster中

② **依次消息计算（从叶子到根）：**

以 x_Q 对应cluster为根；

产生的消息函数与边关联。

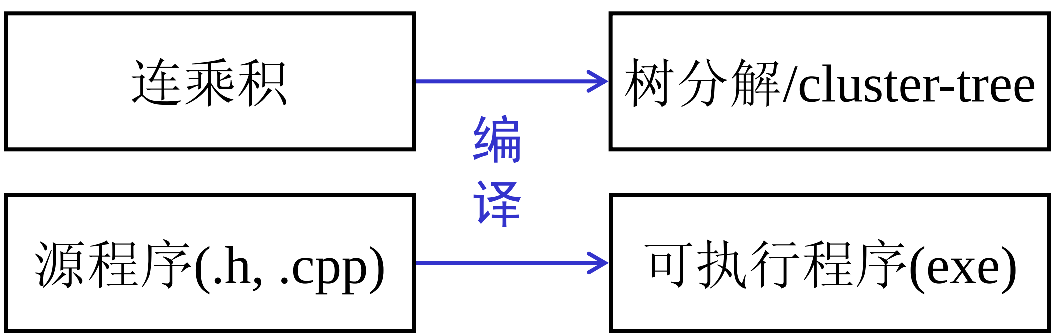
<u>处理一个桶/一次消息计算：</u>

初始放置在桶中的局部函数 与 先后放到桶中的消息函数相乘，再消除变量，得到一个新的消息函数

③ **返回** $p(x_Q \mid e)$：

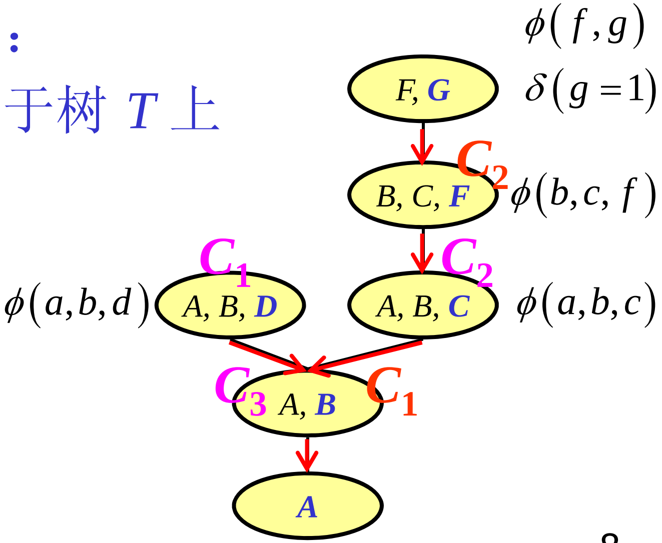
观念的转变

- Variable elimination（变量消除算法）
  - 一边消息计算，一边生成一棵树
- Cluster-tree elimination（树消除算法）
  - 事先编译生成cluster-tree，然后执行计算
  - 把 编译(寻找最优的cluster-tree) 与 计算 分离



tree-decomposition/cluster-tree 定义

- 一个连乘积G: $\prod_i f_i$ 的树分解：一个三元组 <T, χ, ψ>，其中
  - G 中每个乘积项（局部函数）f_i 是一个多变量实值非负函数
  - T 是一棵树，称 树上每个结点C 为 cluster
  - 每个cluster上贴有一个变量集标签 $\chi(C)$ 。
  - 每个cluster内装有一组函数$\psi(C)$。对每个局部函数 f_i，有且仅有一个cluster C 使得 $f_i \in \psi(C)$。所有放置在一个cluster C的局部函数的变量域的并 $\subseteq \chi(C)$
  - 树 T 满足 running intersection property (RIP)：树 T 上任两个cluster的交 $\chi(C_1)\cap\chi(C_2)$ 均包含于树 T 上C_1与C_2之间路径上的每一个cluster的变量集

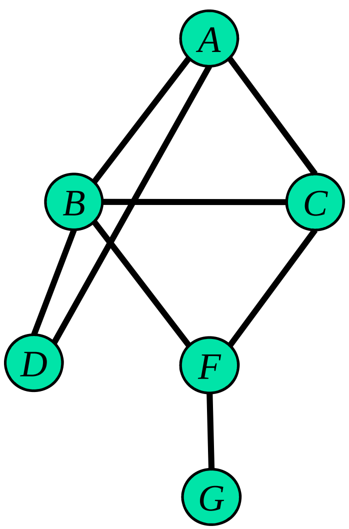


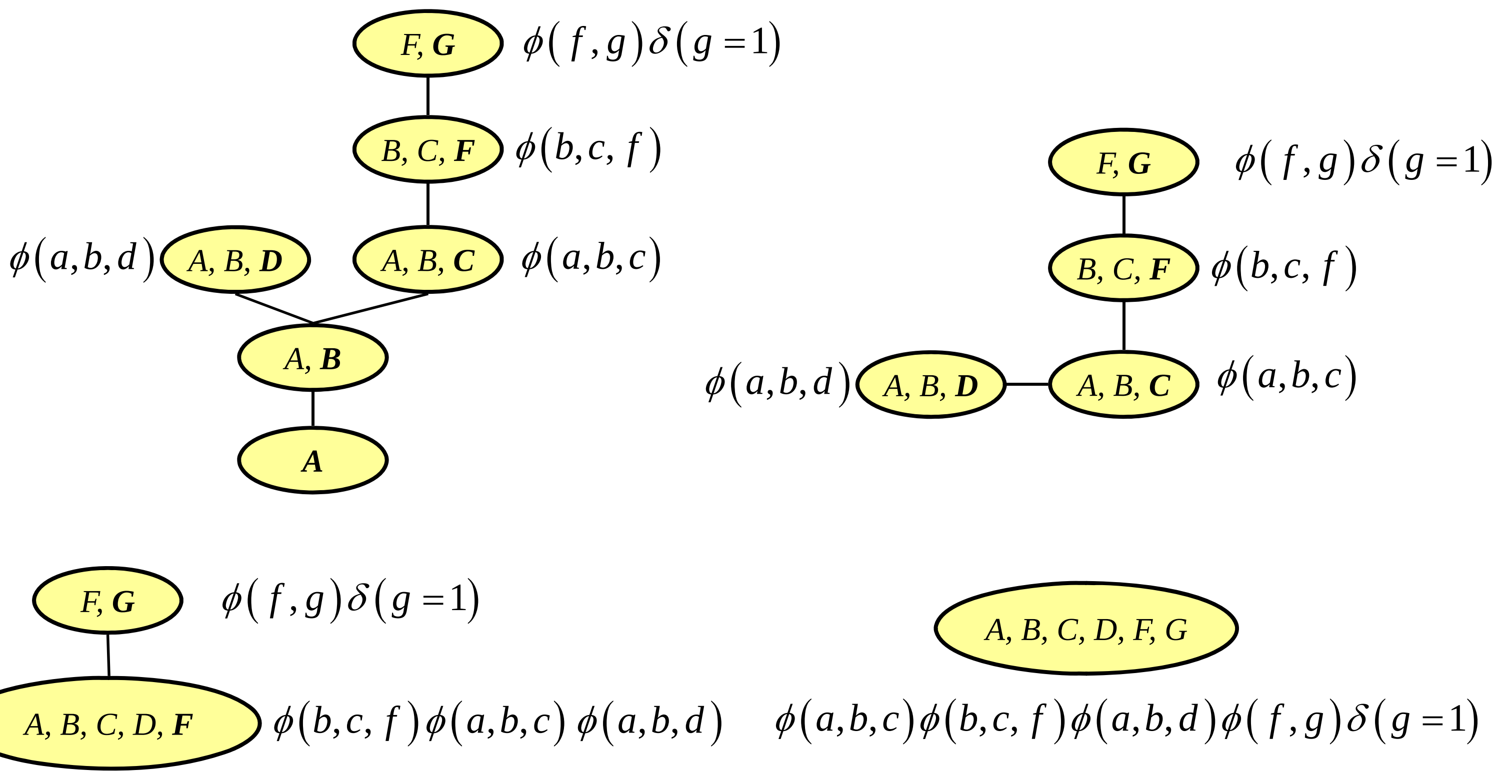
连乘积G：$\phi(a,b,c)\phi(b,c,f)\phi(a,b,d)\phi(f,g)\delta(g=1)$

tree-decomposition—example

一个连乘积的不同树分解

连乘积G：$\phi(a,b,c)\phi(b,c,f)\phi(a,b,d)\phi(f,g)\delta(g=1)$





Cluster-tree elimination

- Consider cluster-tree as a computational data-structure

① 初始化：

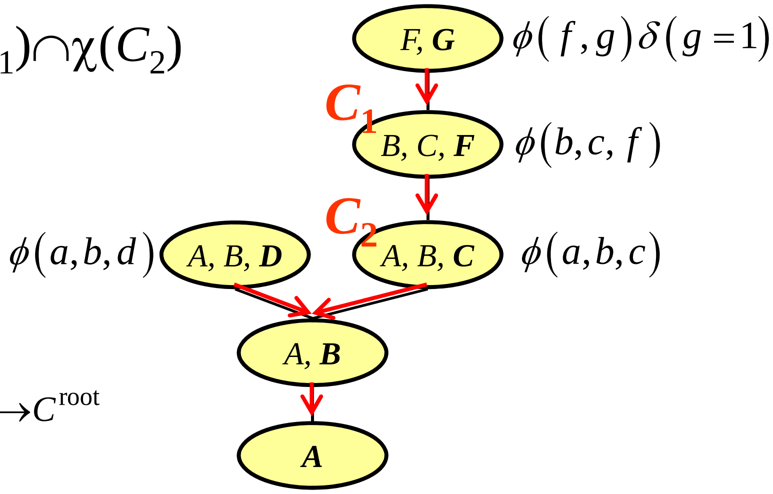
给定连乘积 $\prod_i f_i$ 的一个树分解

② 依次消息计算（从叶子到根）：

任选一个cluster为根C^{root}，进行从叶子至根的消息计算

$$\lambda_{C_1 \to C_2} = \Downarrow_{sep(C_1,C_2)} \left\{ C_1\text{里的函数} \cdot \prod_{Z \in C_1 \text{的邻居} \setminus C_2} \lambda_{Z \to C_1} \right\}$$

树T上两个相邻cluster的交记为$sep(C_1, C_2)=\chi(C_1)\cap\chi(C_2)$



③ 返回：

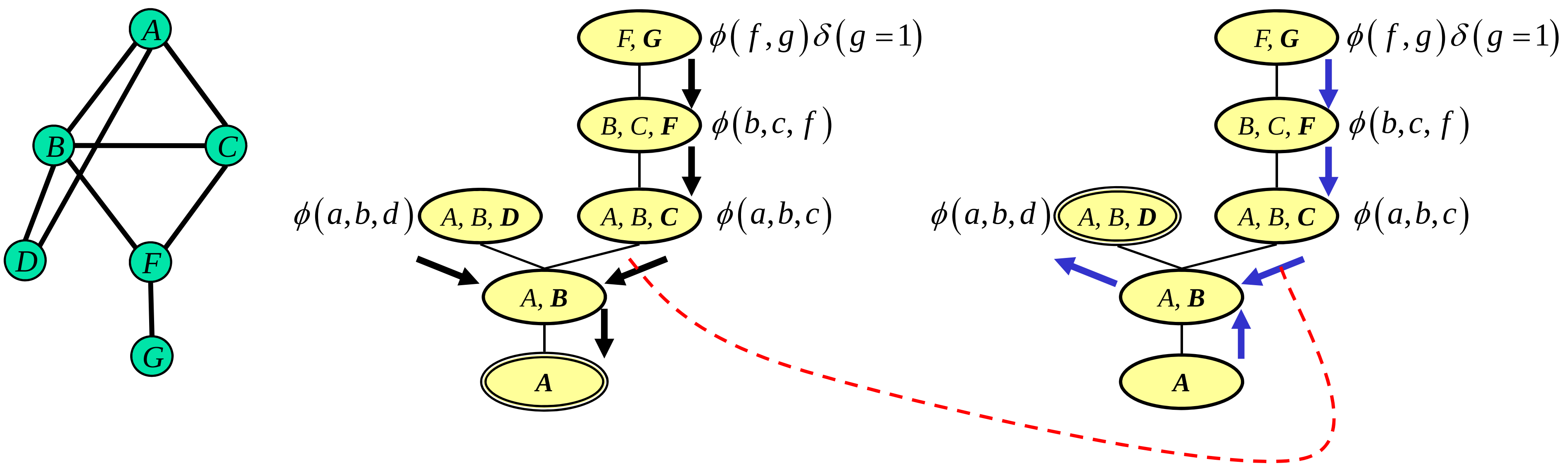
连乘积在根cluster变量集$\chi(C^{\text{root}})$上的边缘化

$$\left(\prod_i f_i\right)\Downarrow_{\chi(C^{\text{root}})} = C^{\text{root}}\text{里的函数} \cdot \prod_{Z \in C^{\text{root}}\text{的邻居}} \lambda_{Z \to C^{\text{root}}}$$

Cluster-tree elimination—all posteriors

- 求多个隐变量的后验分布？ $p(A|evidence), p(D|evidence), \cdots ?$
  - 只需要一棵 cluster-tree.
  - 以不同cluster为根，分别进行叶子至根的消息计算

连乘积：$\phi(a,b,c)\phi(b,c,f)\phi(a,b,d)\phi(f,g)\delta(g=1)$



- **Key insight: 一条边上沿一个方向上的消息总是一样的！**

$= \Downarrow_{隔离集} \{位于边的发送端的所有cluster的函数集中函数连乘积\}$

Cluster-tree elimination—schedule

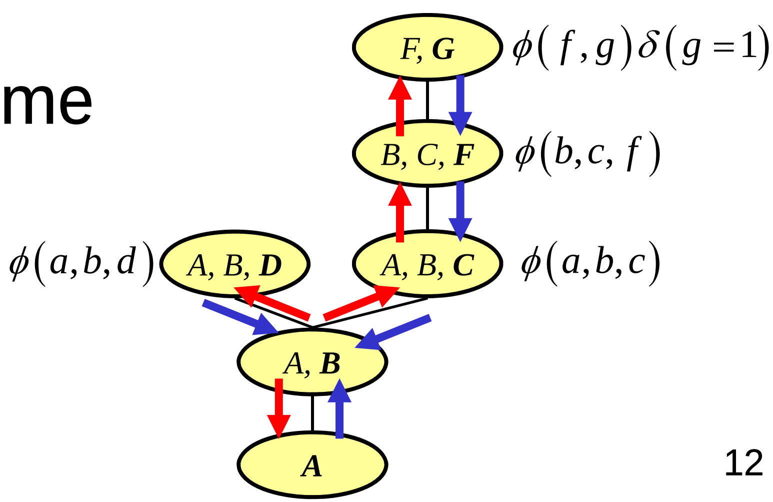
- 求每个cluster的变量集的后验分布？
  - Naïve: 以每个cluster为根，分别进行叶子至根的消息计算
  - 高明：只需计算出每条边 (C_1, C_2) 上的两个消息 $\lambda_{C_1 \to C_2}$ 和 $\lambda_{C_2 \to C_1}$
  - 诸多消息计算的先后顺序？

$$\lambda_{C_1 \to C_2} = \Downarrow_{sep(C_1,C_2)} \left\{ C_1\text{里的函数} \cdot \prod_{Z \in C_1 \text{ 的邻居} \setminus C_2} \lambda_{Z \to C_1} \right\}$$

- Message protocol
  - cluster C_1 可以发消息 $\lambda_{C_1 \to C_2}$ 给相邻cluster C_2，仅当“C_1的除C_2外的所有邻居”发送到C_1的消息都已计算好.

- A simple message passing scheme
  - Leaf-to-root
  - Root-to-leaf

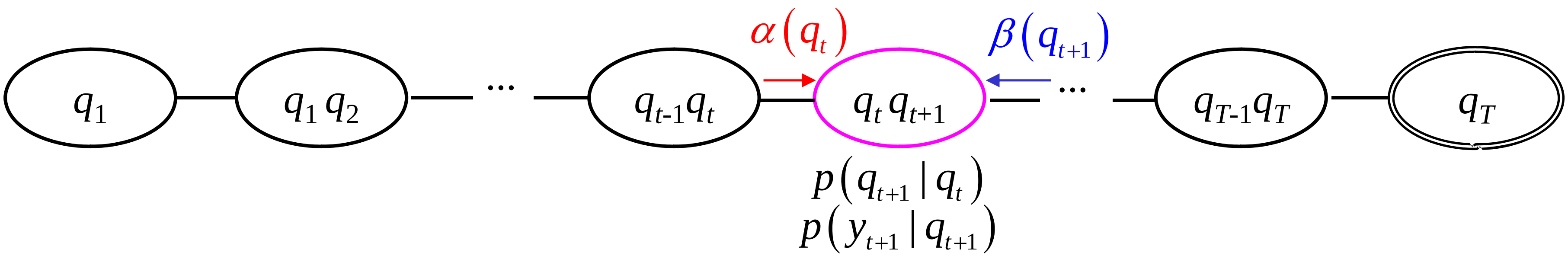
$$\left(\prod_i f_i \right) \Downarrow_{\chi(C)} = C\text{里的函数} \cdot \prod_{Z \in C\text{的邻居}} \lambda_{Z \to C}$$



举例：HMM's Forward-backward algorithm

连乘积：$p(q_1)\cdot\prod_{t=1}^{T-1} p(q_{t+1} \mid q_t)\cdot\prod_{t=1}^{T} p(y_t \mid q_t)$

求：$p(q_t, q_{t+1} \mid y_{1:T})$，$t=1,\cdots,T-1$　　$\xi_t(i,j) \triangleq p(q_t=i, q_{t+1}=j \mid y_{1:T})$



$$\alpha(q_{t+1}) = p(y_{t+1} \mid q_{t+1})\sum_{q_t=1}^{N} p(q_{t+1} \mid q_t)\alpha(q_t)$$

$$\beta(q_t) = \sum_{q_{t+1}=1}^{N} p(y_{t+1} \mid q_{t+1})\beta(q_{t+1})p(q_{t+1} \mid q_t)$$

$$p(q_t, q_{t+1} \mid y_{1:T}) \propto \alpha(q_t)\cdot\beta(q_{t+1})\cdot p(q_{t+1} \mid q_t)p(y_{t+1} \mid q_{t+1})$$

Cluster-tree elimination—history

- Pearl's belief propagation (1988)
  - for single-root query in poly-tree (i.e. DAG forest, e.g. HMM)
- Shafer-Shenoy algorithm (1990)
  - Variable elimination
  - Probability propagation
- Hugin algorithm (1990)
  - F.V. Jensen, S. Lauritzen, and K. Olesen
  - Evolving from clique potentials to clique marginals
- Cluster-tree elimination (Dechter, 2005)
  - Unifying
- Sum-product algorithm (McEliece 1997, Frey 2001)
  - operates in factor graph
  - Loopy sum-product

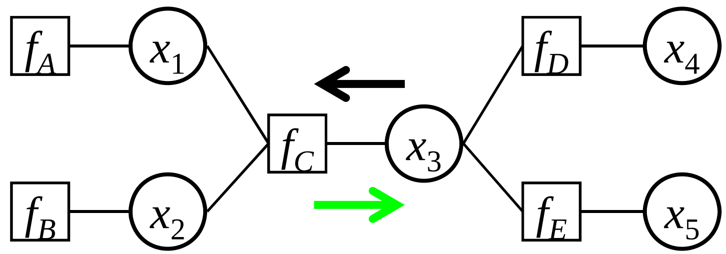
一、Cluster-tree Elimination（树消除算法）

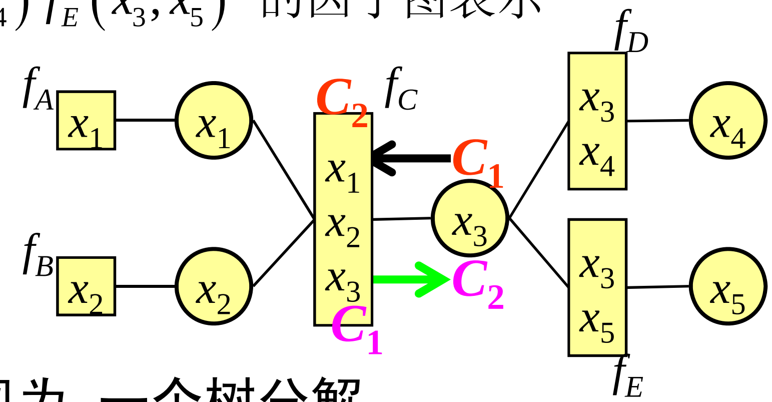
二、Sum-product algorithm on factor graph

三、信道译码应用

Sum-product algorithm on cycle-free FGs

连乘积 $f_A(x_1)\,f_B(x_2)\,f_C(x_1,x_2,x_3)\,f_D(x_3,x_4)\,f_E(x_3,x_5)$ 的因子图表示





无环因子图 自然视为 一个树分解

将变量结点 x 视为单变量cluster： $\chi(C)=\{x\}$, $\psi(C)=\{1\}$

将函数结点 f 视为多变量cluster： $\chi(C)=\arg(f)$, $\psi(C)=\{f\}$

满足 RIP 性质

从变量结点到函数结点的消息：

$$\lambda_{x_3\to f_C}(x_3)=\lambda_{f_D\to x_3}(x_3)\,\lambda_{f_E\to x_3}(x_3)$$

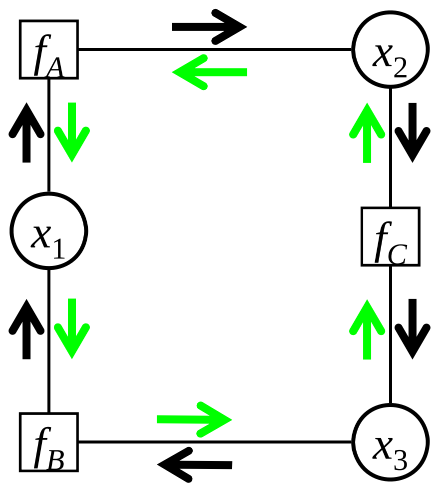
从函数结点到变量结点的消息：

$$\lambda_{f_C\to x_3}(x_3)=\sum_{x_1}\sum_{x_2} f_C(x_1,x_2,x_3)\,\lambda_{x_1\to f_C}(x_1)\,\lambda_{x_2\to f_C}(x_2)$$

均是单变量函数！

Sum-product algorithm on FGs with cycles

连乘积 $f_A(x_1,x_2)\,f_B(x_1,x_3)\,f_C(x_2,x_3)$ 的因子图表示



- 初始假设每条边每个方向都有消息 $\lambda(.) = 1$
- 消息计算循环进行
- 循环结束判断：前后两次后验分布的相对变化小于ε

$$\sum_{i=1}^{3}\left|\frac{posteriori\left(x_i^{(new)}\right)-posteriori\left(x_i^{(old)}\right)}{posteriori\left(x_i^{(old)}\right)}\right|<\varepsilon$$

一、Cluster-tree Elimination（树消除算法）

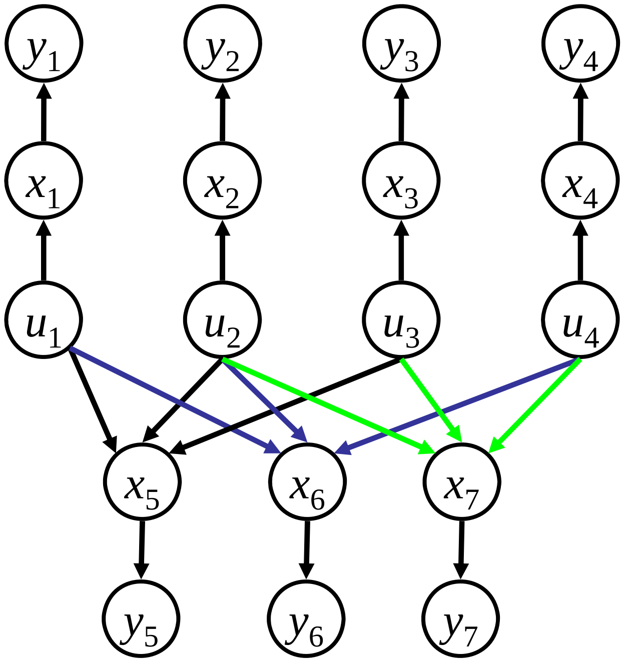
二、Sum-product algorithm on factor graph

三、信道译码应用

Bayes net for a 4/7 Hamming code

- Probabilistic decoding： 基于 $u_{1:4}$ $x_{1:7}$ $y_{1:7}$的联合分布

$$\hat{u}_k = \arg\max_{u_k=0,1} p\left(u_k \mid y_{1:7}\right)$$



- $p(u_k=0)= p(u_k=1)=0.5$ for $k=1,2,3,4$.

- $p(x_1 \mid u_1) = \delta(x_1 , u_1)$
  $p(x_2 \mid u_2) = \delta(x_2 , u_2)$
  $p(x_3 \mid u_3) = \delta(x_3 , u_3)$
  $p(x_4 \mid u_4) = \delta(x_4 , u_4)$
  $p\left(x_5 \mid u_1,u_2,u_3\right) = \delta\left(x_5, u_1+u_2+u_3\right)$
  $p\left(x_6 \mid u_1,u_2,u_4\right) = \delta\left(x_6, u_1+u_2+u_4\right)$
  $p\left(x_7 \mid u_2,u_3,u_4\right) = \delta\left(x_7, u_2+u_3+u_4\right)$

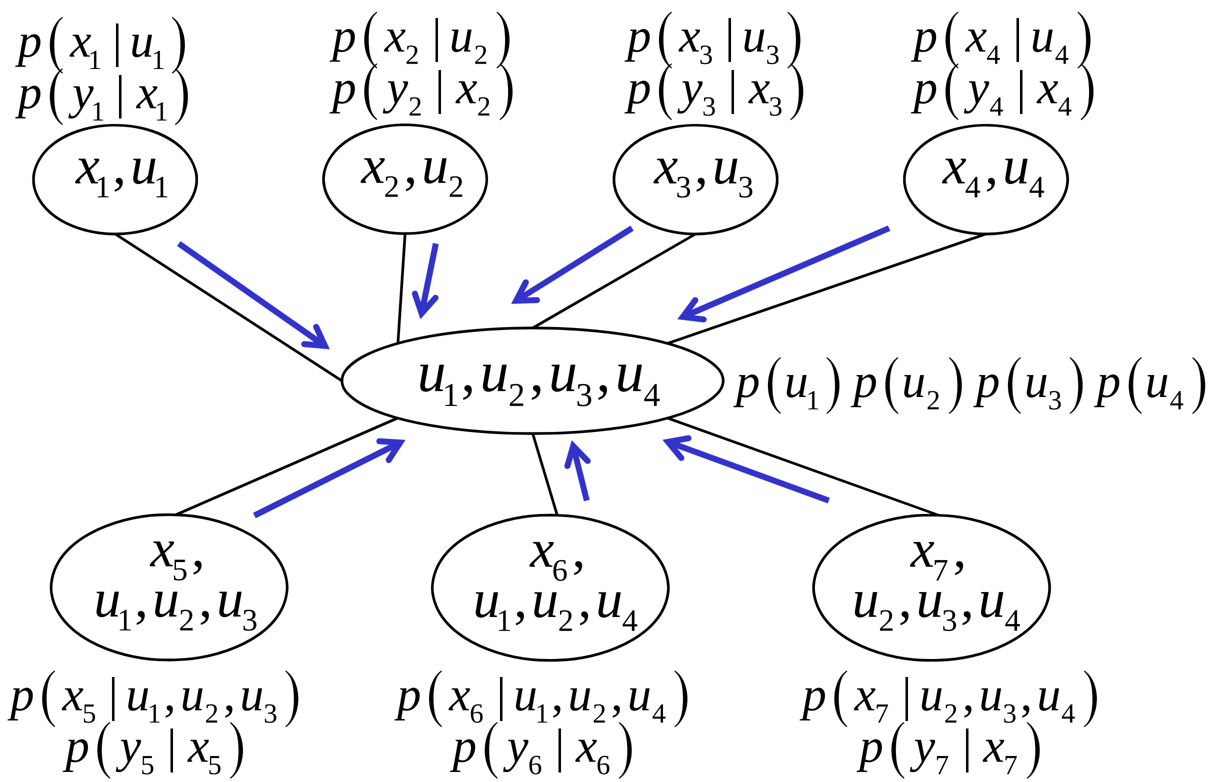
$$\left(x_1,x_2,x_3,x_4,x_5,x_6,x_7\right) = \left(u_1,u_2,u_3,u_4\right)\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

- $p\left(y_n \mid x_n\right) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\dfrac{\left(y_n - x_n\right)^2}{2}\right\}$ for $n=1,\ldots,7$

Bayes net for a 4/7 Hamming code

- Probabilistic decoding： 基于 $u_{1:4}$ $x_{1:7}$ $y_{1:7}$的联合分布

$$\hat{u}_k = \arg\max_{u_k=0,1} p\left(u_k \mid y_{1:7}\right)$$



Factor graph for a 4/7 Hamming code

- Probabilistic decoding： 基于 $u_{1:4}$ $x_{1:7}$ $y_{1:7}$的联合分布

$$\Downarrow_{u_k} \quad p\left(u_{1:4}, x_{1:7}, y_{1:7}\right)$$

- $p(u_k=0)= p(u_k=1)=0.5$ for $k=1,2,3,4$.

- $p(x_1 \mid u_1) = \delta(x_1, u_1)$

  $p(x_2 \mid u_2) = \delta(x_2, u_2)$

  $p(x_3 \mid u_3) = \delta(x_3, u_3)$

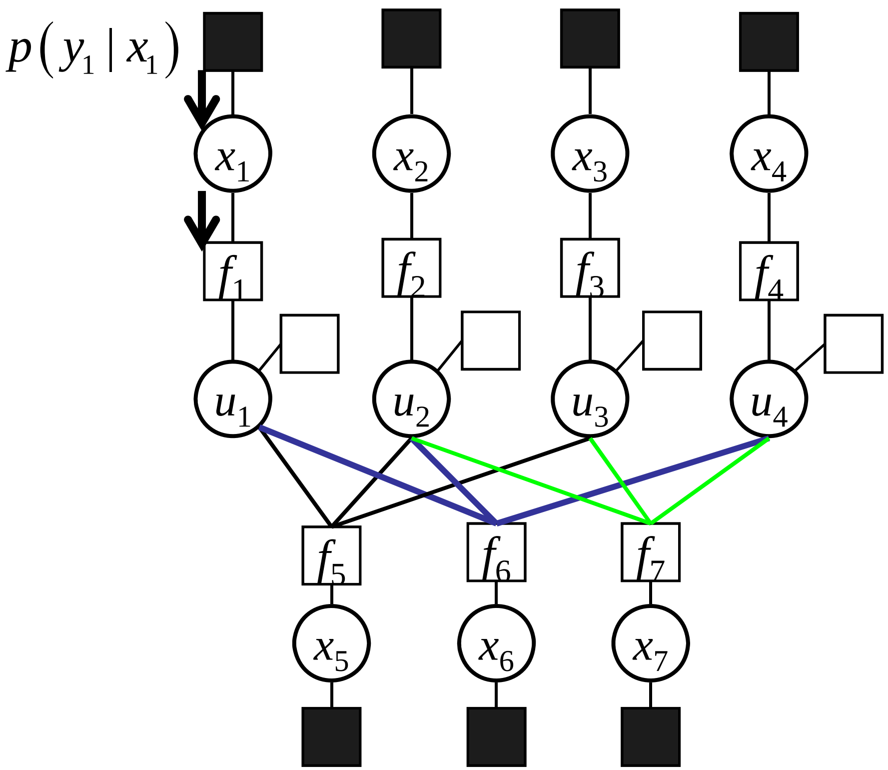
  $p(x_4 \mid u_4) = \delta(x_4, u_4)$

  $p\left(x_5 \mid u_1, u_2, u_3\right) = \delta\left(x_5, u_1 + u_2 + u_3\right)$

  $p\left(x_6 \mid u_1, u_2, u_4\right) = \delta\left(x_6, u_1 + u_2 + u_4\right)$

  $p\left(x_7 \mid u_2, u_3, u_4\right) = \delta\left(x_7, u_2 + u_3 + u_4\right)$

- $p\left(y_n \mid x_n\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(y_n - x_n\right)^2}{2\sigma^2}\right\}$ for $n=1,\ldots,7$



有环因子图

Loopy sum-product for 4/7 Hamming decoding

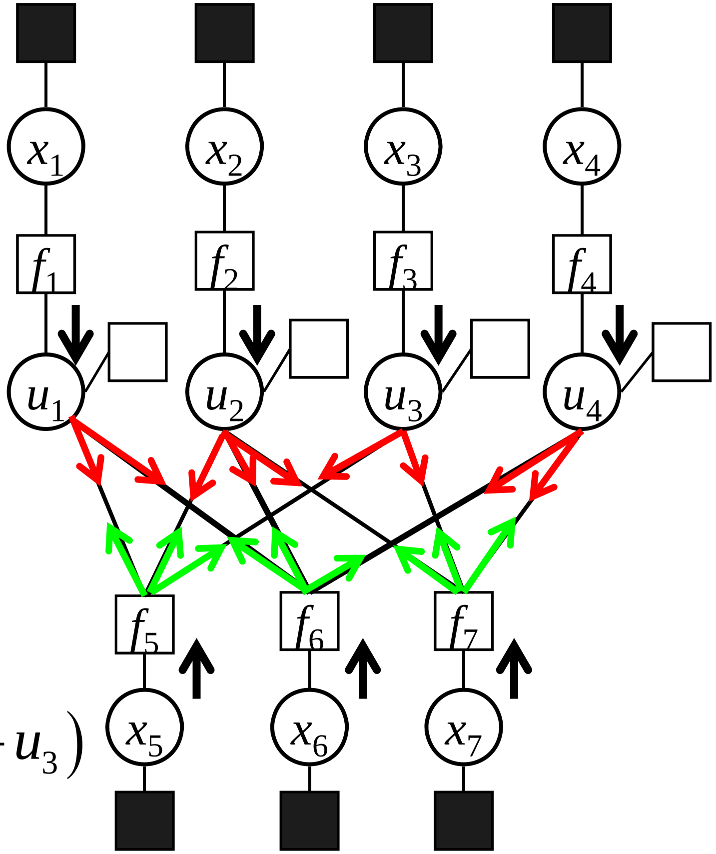
$$\lambda_{f_1 \to u_1}(u_1) = \sum_{x_1} p(y_1 \mid x_1)\delta(x_1, u_1) = p(y_1 \mid u_1)$$

$$\lambda_{u_1 \to f_5}(u_1) = \lambda_{f_1 \to u_1}(u_1)\lambda_{f_6 \to u_1}(u_1)\, p(u_1)$$

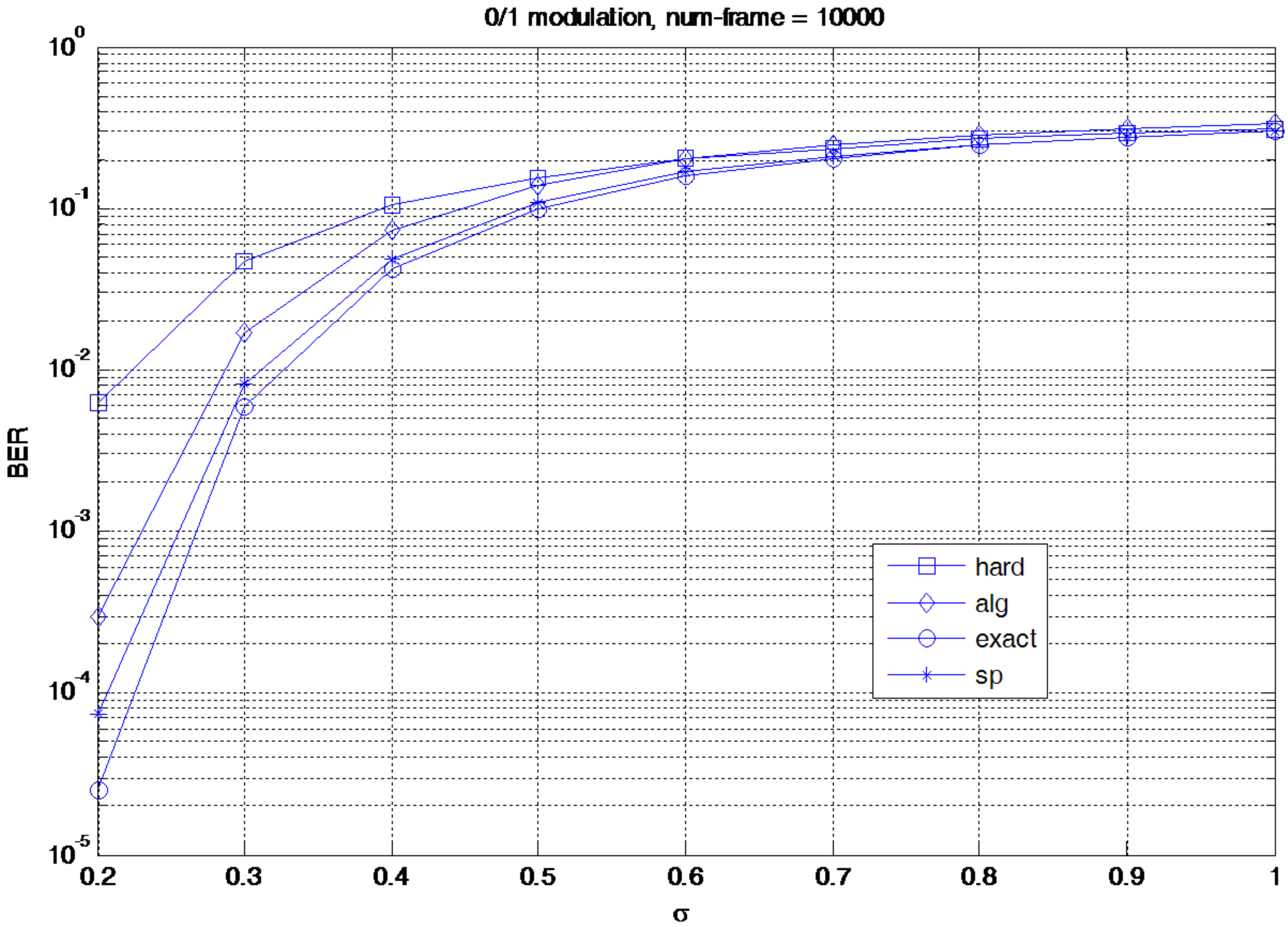
$$\lambda_{f_5 \to u_1}(u_1)$$

$$= \sum_{u_2, u_3, x_5} \lambda_{u_2 \to f_5}(u_2)\lambda_{u_3 \to f_5}(u_3)\lambda_{x_5 \to f_5}(x_5)\delta(x_5, u_1 + u_2 + u_3)$$

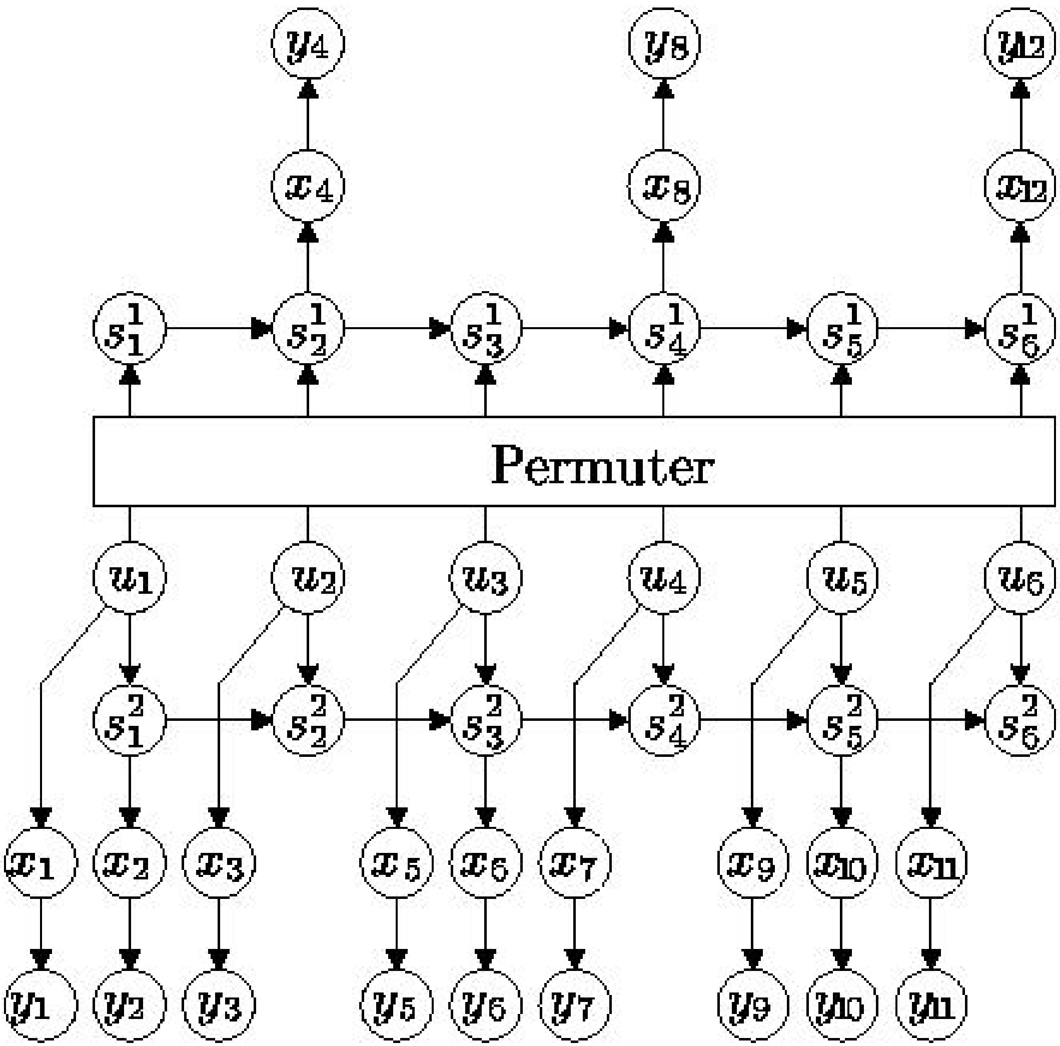
$$\lambda_{x_5 \to f_5}(x_5) = p(y_5 \mid x_5)$$



homework4_hamming



Turbocodes: parallel concatenated convolutional codes (Berrou et al, 1993)



The Bayes net for a rate ½ Turbocode

- Consist of two constituent convolutioal encoder

1) The lower is essentially the same as the systematic convolutional coder described above.

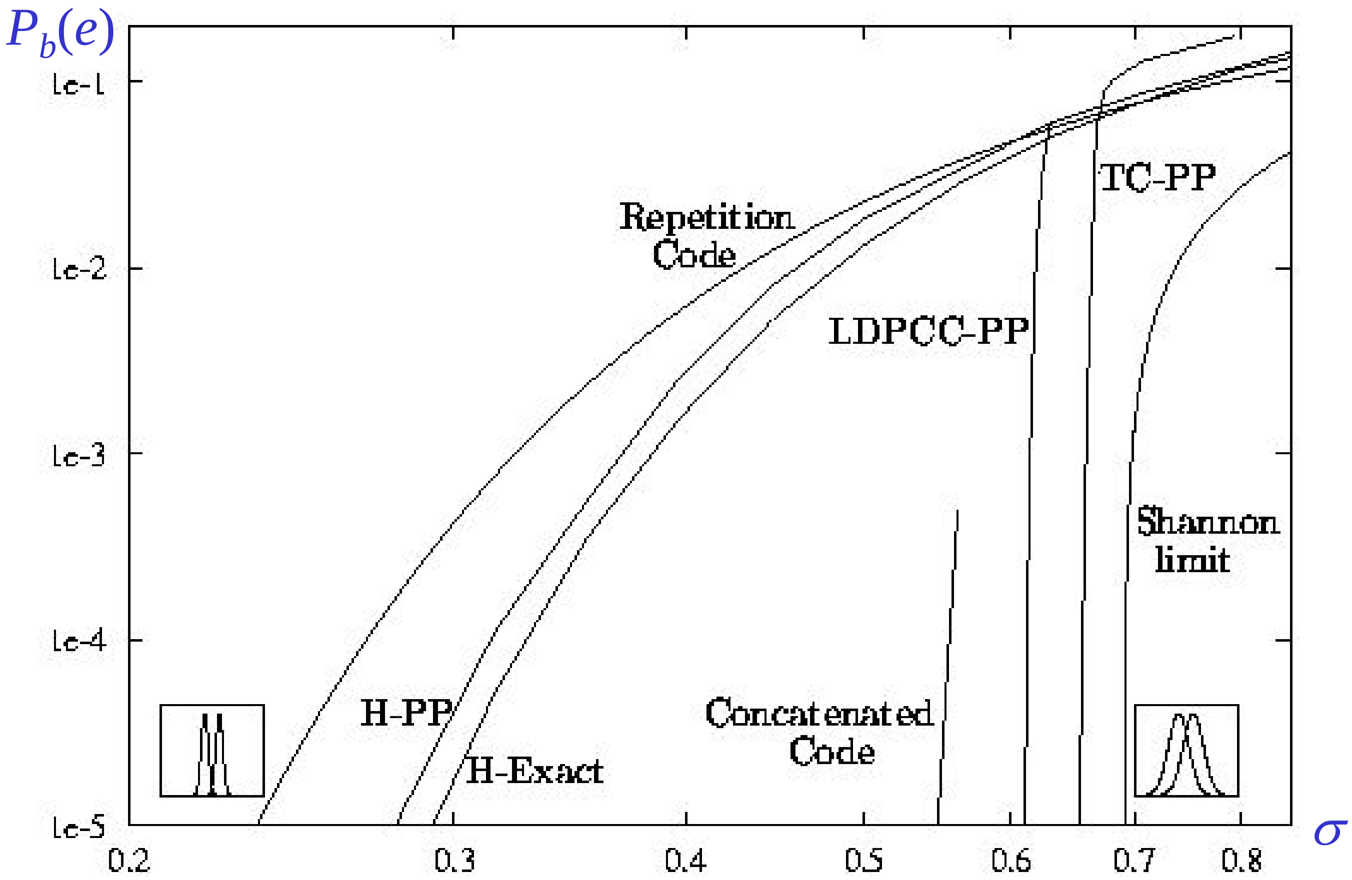
   The only difference is that every second LFSR output is left off.

2) The information bits also fed into the upper coder, but in permuted order.

   Every second LFSR output is also left off.

- So the rate is 1/2.

A plot of bit error $P_b(e)$ versus noise level σ for several codes with rate near ½, using 0/1 signalling.

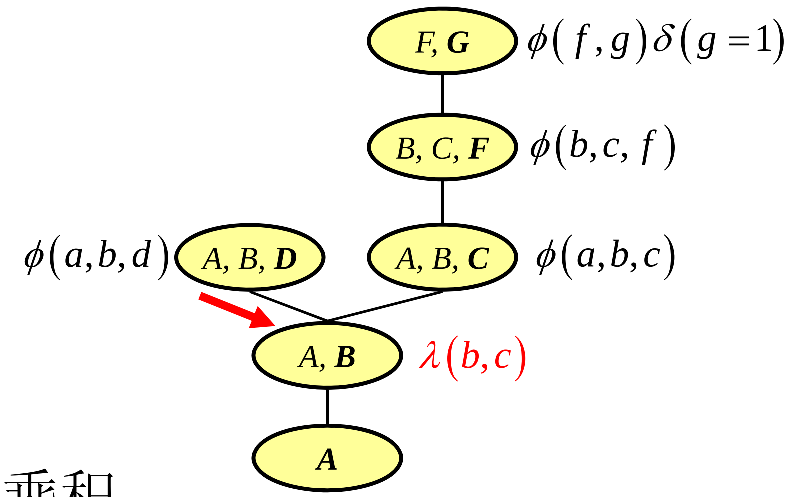


- "TC-PP" = 1/2 Turbocode (K=65536) decoded by loopy sum-product.
- "LDPCC-PP" = 32632/65389 LDPCC decoded by loopy sum-product.

LDPC coding (Gallager, 1963), LDPC decoding (Mackay et al, 1996)

Cluster-tree elimination—'proof'

$$\phi(a,b,c)\phi(b,c,f)\phi(a,b,d)\phi(f,g)\delta(g=1)$$



连乘积 $\prod_i f_i$

= 所有cluster的函数集中函数连乘

进行一步消息传递，

连乘积消除一部分变量仍是一个(新的)连乘积

$$\phi(a,b,c)\phi(b,c,f)\underbrace{\left[\sum_d \phi(a,b,d)\right]}_{\lambda(b,c)}\phi(f,g)\delta(g=1)$$

直至最后，除了根cluster变量集外的所有变量都被消除掉，

得到连乘积在根cluster变量集上的边缘化 $\left(\prod_i f_i\right)\Downarrow_{\chi(C^{\text{root}})}$